Journal of Engineering Science and Technology Review 17 (6) (2024) 224-230

Research Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Impact of Audiovisual Dual-Channel Salience on Situational Awareness of Remote Digital Tower Controllers

Yawei Liu[1], Yuzheng Xing[1], Liangru Li[1], Yuansheng Wang[2,*] and Huibin Jin[3]

[1]*School of Civil Aviation, Zhengzhou University of Aeronautics, Zhengzhou 450046, China*
[2]*School of Environmental and Municipal Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450046, China*
[3]*School of Transportation Science and Engineering, Civil Aviation University of China, Tianjin 300300, China*

Received 26 June 2024; Accepted 2 December 2024

Abstract

Remote digital towers are core technologies for improving the operational efficiency of air traffic management systems. Optimizing human–computer interaction design for remote digital towers and ensuring air traffic controllers' situational awareness are focal topics in the next generation of air traffic management systems. To explore the mechanism of audiovisual dual-channel salience on air traffic controllers' situational awareness and its effect on their workload, an experimental platform was developed, integrating a remote digital tower system and an eye tracker. Twenty air traffic controllers were recruited as subjects. Then, eye movement data were collected. The 3D-SART scale and NASA-TLX scale were used to calculate the controllers' situational awareness scores and workload, respectively, which revealed the impact mechanism of audiovisual dual-channel salience on controllers' situational awareness. Results show that compared with the visual-only channel, air traffic controllers' situational awareness in performing remote digital tower tasks is significantly improved under the audiovisual dual-channel condition. Meanwhile, their workload shows an increasing trend but is not significantly different. Additionally, the first fixation duration on dynamic areas of interest decreases significantly, while the proportion of conscious fixation time and scan rate increase significantly. The findings provide a significantly methodological reference for human–computer interaction design in remote digital towers.

Keywords: Remote digital towers; Audiovisual dual-channel; Controllers; Situational awareness; Human factors

1. Introduction

The control tower is the workplace of air traffic controllers, thus serving as the "eyes of the airport." Its core function is to meet controllers' visual requirements for the airport tower control area and provide necessary real-time aircraft dynamic information for controllers to command flight operations. In recent years, remote digital towers (RDTs) have attracted widespread attention from air traffic management researchers because of their significant advantages, including all-weather and all-around monitoring, low construction and labor costs, and a high degree of intelligence [1]. As a study hotspot in the domestic and international air traffic management field, the key to the successful application of RDTs is management transformation, wherein human–computer interface is a crucial transformative element. Situational awareness is not only a critical indicator for evaluating the quality of human–computer interface design [2-3] but also a significant factor affecting controllers' decision-making and operational errors [4]. Poor situational awareness is one of the primary causes of unsafe events and accidents in civil aviation [5]. Therefore, studying the mechanism of human–computer interface design in RDTs on controllers' situational awareness and analyzing its impact on controllers' workload are essential foundations for effectively controlling and preventing human error risks and promoting the application of RDT technology.

However, existing studies on controllers' situational awareness and workload have primarily focused on traditional tower work scenarios [6-8]. As such, scant attention has been paid to work scenarios in RDT, which involve purely visual display terminal operations [9]. Although a few studies have recognized the transformative nature of RDTs, they have mainly focused on technical issues, such as communication networks and image stitching, with a limited exploration of issues related to human factors engineering. The study of RDT operations is a typical human–machine–system domain, where machines and systems are organized around human needs. Therefore, the successful application of RDTs requires optimal human–machine matching, which necessitates extensive human factors engineering tests and human–computer interface evaluations. This process includes evaluating the human factors engineering design of RDTs based on human physiological and psychological characteristics and enhancing controllers' attention allocation and situational awareness through a system-integrated human factors design concept. Existing studies on controllers' situational awareness have largely focused on traditional tower work scenarios. As such, discussions on RDT work scenarios and human–computer interaction have been given insufficient attention. Therefore, this study builds an RDT experimental platform based on an eye tracker and conducts a controlled experiment with visual and auditory dual channels to investigate the impact of different channel stimuli on controllers' situational awareness and workload. Moreover, eye-tracking data are used to analyze the underlying

*E-mail address: wangyuansheng@ncwu.eud.cn
ISSN: 1791-2377
doi:10.25103/jestr.176.25

mechanisms. This study has significant implications for optimizing the human–computer interface design of RDTs.

2. State of the Art

For multichannel human–computer interaction, existing literature primarily relies on controlled experiments and statistical methods to analyze the impact of various factors on interaction efficiency. Visual and auditory channels are the primary means for humans to receive information in human–computer interaction systems. Based on visual and auditory dual-channel human–computer interaction, this approach is not only natural and efficient but also provides users with a strong sense of presence and realism [10]. Hu et al. [11] found through visual search tasks under multichannel stimulation that multichannel salience outperforms single-channel salience in high-difficulty visual search tasks, thus laying the foundation for study on the influencing conditions of multichannel salience on search performance. In addition, Esposito et al. [12] found that the amount of information conveyed by auditory and visual channels for emotional information, either individually or combined, is roughly the same: each channel robustly encodes emotional features. Channel redundancy may help restore emotional information when one channel is impaired. Moreover, Dunn et al. [13] conducted flight simulation experiments, which revealed that the combination of onboard cameras and auditory feedback can effectively improve the flight performance (horizontal accuracy, vertical accuracy, and timeliness) of remote pilots. Sun et al. [14] used a literature review and a typical case analysis to summarize the relevant technologies of single-channel sensory stimulation and the specific implementation methods of multichannel sensory stimulation fusion; furthermore, they introduced the application of multichannel sensory stimulation fusion in typical scenarios, such as medical rehabilitation, virtual shopping, and intelligent education. Cun et al. [15] found through experimental studies that the combination of auditory and tactile alerts in aircraft cockpits is effective; they also concluded that pilots have rapid selection responses. Yuan et al. [16] administered face and sound stimulation experiments and determined that dual-channel stimulation can capture attention more effectively than single-channel auditory stimulation. Moreover, as the verbal working memory load increases, the benefits of auditory stimulation from multichannel promotion increase. Meanwhile, the benefits of visual stimulation from multichannel promotion do not change significantly. Conversely, as the visual working memory load increases, the benefits of auditory stimulation from multichannel promotion do not change significantly. Meanwhile, the benefits of visual stimulation from multichannel promotion decrease significantly. The current study on multichannel human–computer interaction focuses on statistical significance analysis and subjective inference, thus lacking objective technical tools for verification. Data collection and analysis based on eye-tracking can effectively reveal human attention and cognitive processing mechanisms, thus providing important supplementary information for the study of multichannel human–computer interactions.

In addition, current domestic and international studies on RDTs focus on operational mode analysis and technical verification. For instance, Romero et al. [17] deduced air traffic controllers' taskload driven impact factors under various weather phenomena and developed a mixed integer programming model for controllers' staff scheduling. Xu et al. [18] elaborated on the definition and basic components of RDTs as well as the application of augmented reality technology in remote tower control. In addition, Zhang et al. [19] systematically introduced the operational concepts and technical requirements of European RDTs and provided reference opinions for the construction of RDTs in China. Furthermore, Tong et al. [20] proposed a detection algorithm suitable for airport moving target detection and ADS-B target matching, thus solving the occlusion problem in moving target detection for RDTs. Li et al. [21] recruited 15 qualified air traffic controllers to participate an experiment bases on pupil labs eye tracker, the result showed that a 43-inch panoramic display provided a better human–computer interaction experience for target recognition than a 55-inch panoramic display. Larger screens present larger stimuli, thus making them easier for air traffic controllers to detect. However, they also lead to larger head and eye movements, more severe distortion at the edges of the screen, and longer target recognition reaction times. Moruzzi et al. [22] designed a specific eye-tracking system using Microsoft Kinect V2 to enable the design of virtual RDT applications. Ohneiser et al. [23] thought that the application of speech recognition systems in RDTs can significantly reduce the workload of controllers. Moreover, Li et al. [24] collected eye movement data of controllers in different RDT control scenarios and revealed that the visual attention and monitoring capabilities of controllers are affected by the information presentation method, the complexity of information presentation, and the operating environment of the RDT. Josefsson et al. [25] conducted a simulated experiment on remote tower control at a Swedish airport to analyze the performance of controllers in resolving potential conflicts in different scenarios; they subsequently proposed optimizing the controller duty time to address controller workload issues reasonably. Kearney et al. [26] spotted that enhanced visual RDTs provide sufficient technical support for a single controller to complete tasks that would otherwise be performed by four air traffic controllers; however, this method also significantly increases controllers' workload. Additionally, Heintz et al. [27] conducted interviews with six air traffic controllers at the Leipzig RDT center and found that RDTs do not have a significant impact on controllers' psychological and social capabilities apart from significant changes in perception due to the inability to observe maneuvering areas and airspace directly. Peter et al. [28] conducted 50 field tests where multiairport operating modes did not reduce the safety level of air traffic control operations nor have a negative impact on airport capacity and personnel performance. Although automatic target tracking technology in RDTs can improve controllers' situational awareness, it can also increase the workload of controllers. Maxime et al. [29] studied the impact of sound focusing and vibrotactile feedback on controller performance in RDT scenarios; they found that sound focusing significantly improved subjects' accuracy in aircraft recognition under low visibility conditions compared with pure 3D spatial sound. When sound focusing was combined with vibrotactile feedback, subjects demonstrated improved performance. Data from subjective scales showed that the interactive mode combining sound focusing and vibrotactile feedback had high usability and a low subject fatigue index. Meanwhile, Zhang et al. [30] used a combined model of K-means clustering and support vector machines to classify and identify the situational awareness level of RDT controllers; their findings verify the effectiveness of analyzing controller situational awareness using eye movement data. RDTs still focus on software and hardware equipment. They have insufficient consideration of human–computer interface design. Moreover, a mechanism of the impact of different channel stimuli on controller situational awareness is severely lacking. In summary, visual and auditory dual-channel highlights can enhance the effectiveness of operators. However, the applicability of this system in human–computer interaction in RDTs needs to be verified. The existing research on controller situational awareness mainly focuses on traditional tower work scenarios, thus paying little attention to RDT scenarios. The measurement of situational awareness still relies mainly on subjective scales, which have little use of objective technical tools.

This study addresses the deficiencies of the existing studies by constructing an experimental platform for RDT control tasks. Moreover, it utilizes eye-tracking devices to collect eye movement data from subjects. It also combines the 3D-SART scale and NASA-TLX scale to analyze the impact of visual and auditory dual-channel highlighting on controllers' situational awareness and workload. This study optimizes the human–computer interface design of RDTs to reduce human errors by controllers.

The remainder of this study is organized as follows. Section 3 describes the construction of the eye-tracking experimental platform, the selection of subjects, the experimental tasks and procedures, and the selection and processing of eye movement indicators. Section 4 presents the result analysis. Finally, Section 5 summarizes this study and provides relevant conclusions.

3. Methodology

3.1 Experimental platform

An experimental platform based on an RDT simulation system and an eye tracker was established. The RDT simulation system can fully implement the task flow of tower control and can be customized in terms of flight traffic, audiovisual presentation, and other modules. The eye tracker has a simultaneous error deviation of less than 30 ms and can achieve binocular tracking with an accuracy of 0.5°.

3.2 Experimental subjects

Twenty male air traffic controllers from the Henan Air Traffic Control Bureau and the Zhengzhou Shangjie General Aviation Airport participated in the study. The average age of the participants was 29.7 years with a standard deviation of 4.578. Their work experience ranged from 2 to 10 years. All participants were familiar with the RDT simulation system used in this study and could independently complete the assigned control tasks.

3.3 Experimental tasks and procedures

In this experiment, the Zhengzhou Shangjie General Aviation Airport was selected as the scenario for the RDT control. Flight data packets were captured from the airport's actual operation data. Subjects were required to clear the aircraft for takeoff and guide the aircraft to land on the runway while ensuring safe aircraft spacing and minimizing conflicts. The experiment lasted 20 minutes and involved five takeoffs and three landings. During the experiment, the pilots, ramp controllers, and approach controllers who interacted with the tower controllers were all ideal individuals set by the computer. Their responses were quick and accurate.

The control group included participants who performed control tasks based solely on the visual display and air-ground communication presented on the RDT. In the experimental group, the participants were also exposed to the sounds of aircraft taking off and landing in addition to the monitoring images and air-ground communication presented on the RDT. The sounds began when the aircraft appeared on the monitoring screen and ceased when the aircraft disappeared from the screen. The intensity and direction of the sounds varied depending on the aircraft's position and its status, such as taxiing, waiting, taking off, and landing. Other sound elements, such as background noise and special warning sounds, were not included in this experiment.

In the specific experimental scenario, the airport's surface wind direction was 122° with a wind speed of 4 m/s, visibility of 8000 m, and a corrected sea level pressure of 100.8 kPa. The sky was clear and cloudless with no weather phenomena significantly affecting flight operations. In addition, no flight conflicts occurred. All aircraft were operating normally.

The control group and the experimental group followed the same experimental procedure as shown in Fig. 1:

1) The experimenter initiated the RDT simulation system and set the parameters for the simulated tower control task.

2) The participant sat in front of the experimental system interface. Under the guidance of the experimenter, the participant completed the wearing, calibration, and alignment of the eye tracker.

3) The RDT control task began, and the eye tracker simultaneously recorded the participant's eye movement data.

4) Upon completion of the experiment, the participant completed the 3D-SART situational awareness scale and the NASA-TLX workload scale.

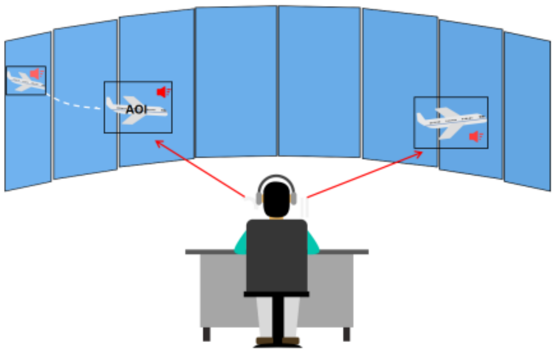

Fig.1. Experimental scenario diagram

The calculation methods for the 3D-SART situational awareness scale and the NASA-TLX workload scale are as follows:

$$Workload = \left(\sum_{i=1}^{6} e_i w_i \right) / 15 \quad (1)$$

In Equation (1), e represents the dimension score in the scale, and w represents the weight of that dimension.

$$SA = U - (D - S) \quad (2)$$

In Equation (2), U represents the level of situation awareness (i.e., understand), D represents the demand for attentional resources, and S represents the supply of attentional resources. All dimensions are scored on a 100-point scale.

3.4 Eye movement index selection and data processing

Based on the interpretation of eye movement indices [31] and the study objectives of this study, three eye movement indices were selected for analysis: first fixation duration on the area of interest (AOI), percentage of conscious fixation time, and saccade velocity. The AOI was dynamically defined as the region of the screen containing aircraft information. The first fixation duration on the AOI was used as the reaction time for participants to perceive the aircraft appearing on the screen. The percentage of conscious fixation time was calculated as the percentage of fixation duration longer than 320 ms in the total fixation duration.

The data processing procedure was as follows. First, data were screened. The subsequent analysis only included participants with a sampling rate of over 80% for eye movement data, which ensured data reliability and reflected the participants' true eye movements during the experiment. Second, the participants were divided into two skill groups based on their years of experience: the novice group (less than five years) and the expert group (five years or more). Finally, the receiver operating characteristic (ROC) curve analysis demonstrated that this classification effectively distinguished between the two groups (AUC = 1, $P < 0.001$), thus confirming the validity of the grouping. In the subsequent analysis, the following levels of statistical significance were used: $P < 0.05$ (*), $P < 0.01$ (**), and $P < 0.001$ (***).

4. Result Analysis and Discussion

4.1 Comparative analysis of situational awareness

A two-way ANOVA revealed no interaction between the skill level and the audiovisual channel on air traffic controllers' situational awareness ($F = 1.456$, $P = 0.222$). However, significant differences were observed in the situational awareness between different skill levels ($F = 41.393$, $P = 0.000$) and between the visual-only and audiovisual channels ($F = 22.124$, $P = 0.000$). Paired t-tests further demonstrated that the situational awareness of all participants was significantly improved when performing RDT control tasks using the audiovisual channel compared with the visual-only channel.

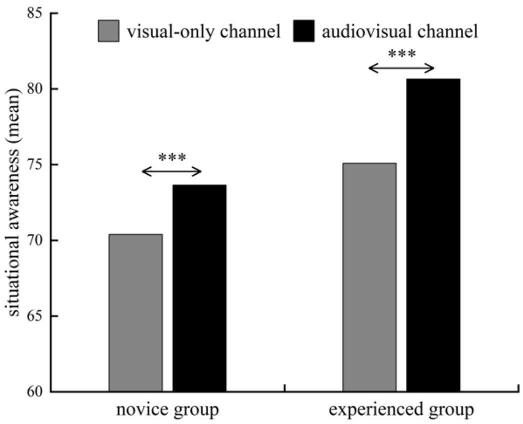

Fig. 2. Comparison of situational awareness of subjects under different perceptual channels

As shown in Figure 2, regardless of whether the visual-only or audiovisual channel was used, novice participants exhibited lower levels of situational awareness than experienced participants. This outcome indicates that work experience and skill level are important factors that influence situational awareness. According to Endsley's three-level model of situation awareness, the perception of information at the first level is a critical foundation. By integrating and utilizing the advantages of information conveyed through individual channels, the audiovisual channel can enhance air traffic controllers' perception of flight dynamic information. Moreover, the audiovisual channel closely resembles the real-world working environment of air traffic controllers in traditional towers, thus facilitating skill transfer and reducing cognitive friction. Therefore, the situational awareness of participants was significantly improved under the audiovisual channel.

4.2 Comparative analysis of workload

A two-way ANOVA on air traffic controllers' workload revealed a significant main effect of skill level (novice vs. expert) ($F = 44.510$, $P = 0.000$). However, it did not demonstrate any significant main effect of the visual or audiovisual channel ($F = 1.230$, $P = 0.275$). In addition, no significant interaction was observed between skill level and channel ($F = 0.801$, $P = 0.377$).

Paired t-tests showed that, compared with the visual-only channel, the audiovisual channel significantly increased the workload of novice participants ($P = 0.035$) but not of expert participants ($P = 0.746$), as shown in Figure 3. Overall, the workload was higher with the audiovisual channel than with the visual-only channel. However, the difference was not statistically significant ($P = 0.055$).

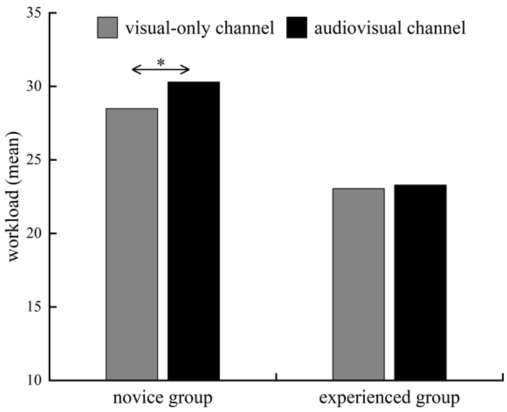

Fig. 3. Comparison of workload of subjects under different perceptual channels

As shown in Figure 3, regardless of whether the visual-only or audiovisual channel was used, novice participants consistently exhibited higher outcomes in specific metrics (e.g., workload, stress) than experienced participants. Experienced participants have more work experience, are more capable of monitoring and controlling multiple aircraft simultaneously, and have a higher tolerance for workload. Additionally, experienced participants have better situational awareness and employ more effective control strategies, thus allowing them to break down complex control tasks proactively. Therefore, experienced participants generally reported low levels of workload. The significant increase in workload for novice participants but not for experienced participants under the audiovisual condition may be explained by the increased information load caused by the addition of auditory information. Some studies have suggested that multimodal human–computer interaction can increase the bandwidth of human–computer information interaction and reduce cognitive load. However, Gu Jiyou's [32] analysis on attention and audiovisual integration has revealed that the redundancy gain effect, where multimodal perceptual integration enables individuals to make fast and accurate cognitive judgments about information, only occurs when semantically consistent audiovisual stimuli are at the peak of attention and individuals attend to visual and auditory conditions simultaneously. Based on these findings and the postexperiment interviews, another possible reason for the increased workload among novice participants is that the appearance of aircraft with sound in the visual field attracts the attention of controllers. However, when novice controllers complete tasks for one flight and then turn their attention to other flights, the sound of the previously completed flight becomes background noise. Missing the auditory information at the attention peak increases the workload of the novice group. By contrast, experienced participants are accustomed to this background noise. Some participants even find it more immersive than silent visual scenes, thus allowing them to maintain alertness without feeling fatigued. Therefore, the experienced group did not experience a significant change in perceived workload.

4.3 Comparative analysis of eye movement data

The results of the two-factor ANOVA for eye movement indicators are shown in Table 1 below.

Table. 1 Results of two-way ANOVA for eye movement indicators

Two-way ANOVA (Analysis of Variance)	**First fixation duration on AOI (s)**		**Proportion of conscious fixation time (%)**		**Scan rate (px/ms)**	
	F	**P**	**F**	**P**	**F**	**P**
Audiovisual	202.999	0.000	37.837	0.000	1.920	0.174
Novice and expert	0.073	0.909	4.151	0.049	2.333	0.135
Audiovisual * Novice and expert	1.399	0.245	1.118	0.297	0.143	0.707

As shown in Table 1, no interaction effect exists between the audiovisual channel and the skill level on the three eye movement metrics (P > 0.05). The main effect for the first fixation duration on the AOI was the visual-only channel and the audiovisual channel. The proportion of conscious fixation time for participants under different sensory channels and skill levels was statistically significant. However, no statistically significant difference was observed in the saccade velocity among participants under different sensory channels and skill levels. Paired t-tests revealed significant differences in the first fixation duration on the AOI, the proportion of conscious fixation time, and saccade velocity between participants using the audiovisual channel and those using the visual-only channel. These outcomes are shown in Figure 4.

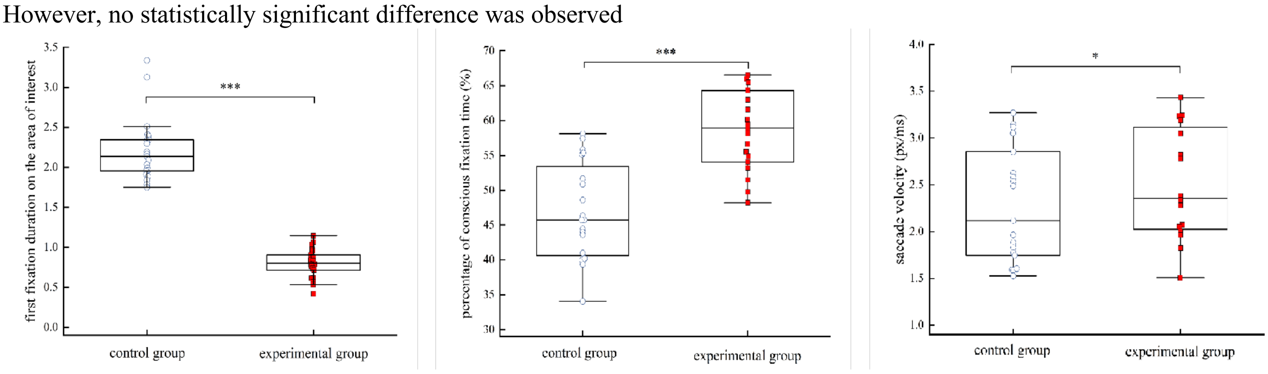

Fig.4 Comparison of eye movement indicators of subjects under different perceptual channels

As shown in Figure 4, the first fixation duration on the AOI for controllers under the audiovisual condition was significantly shorter than that under the visual-only condition. This outcome indicates that the addition of auditory information significantly reduced controllers' reaction time to the aircraft appearing on the screen, thereby enhancing their information perception ability. This observation not only supports the analysis of situational awareness but also aligns with the view that auditory information can guide visual attention. The proportion of conscious fixation time for controllers under the audiovisual condition was significantly higher than that under the visual-only condition. This result suggests that auditory stimuli increased controllers' proactive and conscious gaze behavior, thus reducing their unconscious gaze behavior. Additionally, the increased proportion of conscious fixation time implies that the participants exerted substantial subjective effort, which is one reason for the increased workload. The saccade velocity of controllers under the audiovisual condition was significantly higher than that under the visual-only condition, thereby indicating an improvement in participants' proactivity in actively searching for effective information. This result can enhance participants' ability to grasp global dynamic information, thus preventing them from missing key target information while accurately predicting changing situations and improving their situational awareness.

The audiovisual channel can significantly enhance air traffic controllers' situational awareness. The three eye movement metrics, namely, first fixation duration on the AOI, the proportion of conscious fixation time, and saccade velocity, complement one another to elucidate the cognitive mechanisms underlying the enhancement of situational awareness. However, auditory ergonomics design is a systematic engineering process, and variations in different dimensions and their dynamic combinations can affect individual auditory perception. The cognitive mechanisms behind the auditory channel are still not fully understood. Hence, these factors need to be considered and verified one by one. Similar to the trend of increasing visual information levels in aircraft cockpits and human–computer interfaces in traditional air traffic control towers, the human–computer interface of RDTs will inevitably incorporate additional display information. Under the condition of increasing visual channel load, auditory perception becomes crucial. As such, the integration of audiovisual channels is an inevitable trend in the development of RDT technology.

5. Conclusion

This study addresses the issue of situational awareness among air traffic controllers in RDTs by conducting a controlled experiment with eye-tracking technology. This analysis seeks to analyze the mechanism and intensity of the effects of visual and auditory channels on controllers' situational awareness and explore changes in workload. The following conclusions were drawn:

(1) In the RDT scenario, compared with visual-only channel stimulation, controllers under the audiovisual channel condition exhibited significantly shorter first fixation duration on the AOI. Meanwhile, the proportion of conscious fixation time and saccade velocity increased significantly. Simultaneously, controllers' situational awareness improved significantly. Their workload also increased but not significantly.

(2) Eye-tracking technology can effectively analyze controllers' scanning patterns and attention allocation mechanisms, thus providing a powerful tool for exploring the mechanisms by which visual and auditory channels affect controllers' situational awareness.

(3) The design of RDTs should not only consider the presentation of visual images but also pay attention to the design of auditory stimuli to serve controller operations effectively, improve situational awareness, and ensure civil aviation safety.

Applying eye-tracking technology to evaluate and analyze the human–computer interface of RDTs can clarify the impact of different perceptual channels on air traffic controllers' performance, identify factors that jeopardize their work capacity, and address the risks of human error control. The results presented in this study, which is based on an eye-tracking model for assessing situational awareness in RDTs, can analyze the mechanisms and intensity of the effects of visual and auditory channels on controllers' situational awareness. Hence, these outcomes can provide accurate experimental support for the design of human–computer interfaces in RDTs. However, this study has only conducted preliminary explorations of variable control in the auditory channel. Future research can employ physiological and psychological experimental methods to investigate the impact of different perceptual channels on the performance of remote virtual tower controllers and the underlying physiological and cognitive mechanisms.

Acknowledgments
This study was funded by the Science and Technology Planning Project of Henan Province (No. 232102240016 and 242102321038) and the Youth Talent Support Project of Henan Province (No. 2024HYTP036) and the Postdoctoral Initiation Fund Project of Henan Province (No. 202103101) and the Key Research and Development Special Project of Henan Province (No. 221111321000).

References

[1] W. C. Li, P. Kearney, G. Braithwaite, and J. J. H. Lin. "How much is too much on monitoring tasks? Visual scan patterns of single air traffic controller performing multiple remote tower operations". *Int. J. Ind. Ergon.*, vol. 67, pp. 135-144, Sep. 2018.
[2] M. R. Endsley and D. G. Jones. "Situation awareness oriented design: review and future directions". *Int. J. Hum.-Comput. Interact.*, vol 40, no 7, pp. 1487-1504, Apr. 2024.
[3] Y. Li, M. Ji, J. J. Lan, M. Chang, and X. Q. You. "From aerial driving behavior to industrial design: Transition and expansion of study on situation awareness". *J. Psychol. Sci.*, vol. 40, no. 2, pp. 263-268, Mar. 2017.
[4] Y. Chi, J. Nie, L. Zhong, Y. Wang, and D. Delahaye. "A review of situational awareness in air traffic control". *IEEE Access*, vol. 11, pp. 134040-143057, Nov. 2023.
[5] H. B. Jin, G. L. Zhu, X. M. Mu, and Y. W. Liu. "Research progress of situation awareness in aviation safety". *J. Chongqing Jiaotong Univ. (nat. sci.)*, vol. 38, no. 6. pp. 103-108, Jun. 2019.
[6] S. C. Black, A. D. Bender, S. J. Whitney, S. Loft, and T. A. Visser. "The effect of multi-tasking training on performance, situation awareness, and workload in simulated air traffic control". *Appl. Cogn. Psychol.*, vol. 36, no. 4, pp. 874-890, Jun. 2022.
[7] M. Bongo and R. Seva. "Effect of fatigue in air traffic controllers' workload, situation awareness, and control strategy". *Int. J. Aerosp. Psychol.*, vol. 32, no. 1, pp. 1-23, Jan. 2022.
[8] M. Z. Suárez, et al. "Understanding the research on air traffic controller workload and its implications for safety: A science mapping-based analysis". *Saf. Sci.*, vol. 176, Aug. 2024, Art. no. 106545.
[9] K. Yamazaki, S. Inoue, H. Aoyama, and T. Kimiaki. "Remote digital tower to support air traffic control systems". *Hum. Fact. Transp.*, vol. 148, no. 56, pp. 43-48. May. 2024.
[10] Y. D. Xu, Z. G. Fang, L. H. Zhang, and J. Ding. "Survey of sonification research". *J. Comput. Aided Des. Comput. Graph.*, vol. 16. no. 1, pp. 14-18, Jan. 2004.
[11] F. P. Hu, L. Z. Ge, and Y. J. Xu. "Visual search performance study between single and multi-modality highlighting on high level difficulty". *J. Ergon.*, vol. 11, no. 3, pp. 1-3. Sep. 2005.
[12] A. Esposito. "The perceptual and cognitive role of visual and auditory channels in conveying emotional information". *Cogn. Comput.*, vol. 1, no. 3, pp. 268-278, May. 2009.
[13] M. J. M. Dunn, B. R. C. Molesworth, T. Koo, and G. Lodewijks. "Effects of auditory and visual feedback on remote pilot manual flying performance". *Ergonomics*, vol. 63, no. 11, pp. 1380-1393, Jul. 2020.
[14] L. Y. Sun, K. X. Wang, P. Chen, Z. H. Zhou, C. Y. Yang, and G. Yang. "Review of sensory feedback simulation methods in multi-modal human-computer interaction". *Packag. Eng.*, vol. 44. no. 8, pp. 81-94, Apr. 2023.
[15] W. Cun, S. Yu, J. Chu, Y. Chen, J. Sun, and H. Fan. "Exploration of multimodal alarms for civil aircraft flying task: A laboratory study". *Hum. Factor. Ergon. Man.*, vol. 34, no. 4, pp. 279-291, Feb. 2024.
[16] Y. C. Yuan, H. Yan, X. He, and Z. Z. Yue. "The modulation of working memory load on attention capture by audiovisual stimuli". *J. Psychol. Sci.*, vol. 47, no. 4, pp. 795-802, Jul. 2024.
[17] E. H. Romero, B. Josefsson, A. Lemetti, T. Polishchuk, and C. Schmidt. "Integrating weather impact in air traffic controller shift scheduling in remote and conventional towers". *EURO J. Transp. Logist.*, vol. 11, Apr. 2022, Art. no. 100076.
[18] G. B. Xu, Z. L. Lin, and M. L. Hou. "Remote virtual tower technology", *China Sci. Tech. Inform.*, vol 20. no. 2, pp. 19-21, Oct. 2017.
[19] J. Q. Zhang, Y. Xiang, Q. Zhang, and Q. Q. Wang. "Enlightenment of european remote tower operation concept and technical requirements for China". *Aeronaut. Comput. Tech.*, vol. 48, no. 6, pp. 122-126, Nov. 2018.
[20] F. Tong, X. Y. Zhang, Y. Qiao, and K. Wang. "A fusion method of civil aviation ASTERIX data with video surveillance at remote tower". *Sci. Tech. Eng.,* vol 23. no. 7, pp. 2916-2921, Mar. 2023.

[21] W. C. Li, P. Moore, J. Zhang, J. Lin, and P. Kearney. "The impact of out-the-window size on air traffic controllers' visual behaviours and response time on digital tower operations". *Int. J. Hum.-Comput. Stud.*, vol. 166, Oct. 2022, Art. no. 102880.

[22] M. C. Moruzzi, S. Santhosh, M. Corsi, S. Bagassi, and F. D. Crescenzio. "Design and implementation of eye tracking application for generation of augmented reality content on spatial see through display of remote and virtual control tower (RVT)". *Int. J. Interact. Des.*, vol. 17, no. 4, pp. 1859-1867, Mar. 2023.

[23] O. Ohneiser, et al. "Assistant based speech recognition support for air traffic controllers in a multiple remote tower environment". *Aerospace*, vol. 10, no. 6, Jun. 2023, Art. no. 560.

[24] W. C. Li, P. Kearney, G. Braithwaite, and J. J. H. Lin. "How much is too much on monitoring tasks? Visual scan patterns of single air traffic controller performing multiple remote tower operations". *Int. J. Ind. Ergonom*., vol. 2018, no. 67, pp. 135-144, Sep. 2018.

[25] B. Josefsson, T. Polishchuk, V. Polishchuk, and C. Schmidt. "Step toward remote tower center deployment: Optimizing staff schedules". *Air Traffic Contr. Quart.*, vol. 27, no. 3, pp. 121-130, Mar. 2019.

[26] P. Kearney, W. C. Li, J. Zhang, G. Braithwaite, and L. Wang. "Human performance assessment of a single air traffic controller conducting multiple remote tower operations". *Hum. Factors Ergonom. Manuf. Serv. Ind.*, vol. 30, no. 2, pp. 114-123, Nov. 2020.

[27] A. Heintz, and H. Eifeldt. "Ability requirements for remote tower control first insights". *Transp. Res. Procedia*, vol. 66, pp. 148-155, Dec. 2022.

[28] P. Kearney and W. C. Li. "Multiple remote tower for Single European Sky: The evolution from initial operational concept to regulatory approved implementation". *Transp. Res. Pt. A-Policy Pract.*, vol.116, pp. 15-30. Oct. 2018.

[29] M. Reynal, Jean-P. Imbert, P. Aricò, J. Toupillier, G. Borghini, and C. Hurter. "Audio Focus: Interactive spatial sound coupled with haptics to improve sound source location in poor visibility". *Int. J. Hum.-Comput. Stud.,* vol.129, pp. 116-128. Sep. 2019.

[30] Z. N. Zhao and M. Hao. "Remote tower controllers' situational awareness level based on machine learning combinatorial modeling". *Sci. Tech. Eng.,* vol 24. no. 27, pp. 11928-11936, Sep. 2024.

[31] H. B. Jin and Y. W. Liu. "Performance trace experimental study of the eye movement index and its effect on detecting the fatigue of the air traffic controller". *J. Saf. Environ.,* vol. 18, no. 3, pp. 1004-1007, Jun. 2018.

[32] G. L. Kang and X. X. Luo."The integration and conflict control in audiovisual processing". *J. Psychol. Sci.*, vol. 43, no. 5, pp. 1072-1078, Oct. 2020.