IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 10, 2021, accepted June 19, 2021, date of publication June 22, 2021, date of current version July 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3091528

Integrating Ergonomics Into Safety Management: A Conceptual Risk Assessment Model for Tower Controllers at Multiple Altitudes

QUAN SHAO[1], MIN DONG[1], ZHIYUAN SHEN[1], (Member, IEEE), RUI YANG[2], AND HONG WANG[1]

[1]College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
[2]Department of Engineering Physics, Tsinghua University, Beijing 100084, China

Corresponding author: Quan Shao (shaooquan@nuaa.edu.cn)

This work was supported in part by the National Basic Research Program of China under Grant 2018YFC0809500, in part by the National Natural Science Foundation of China under Grant 71874081, in part by the Qing Lan Project, the Jiangsu Province Natural Science Foundation under Grant BK20201296, in part by the Research on Monitoring and Management of Fatigue State in Air Traffic Controller Project through Air Traffic Management Bureau CAAC, and in part by the Fundamental Research Funds for the Central Universities under Grant kfjj20200720.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Nanjing University of Aeronautics and Astronautics under Approval No. N20201025.

ABSTRACT Plateau and high plateau airports aggravate tower controllers' vulnerability, facing high safety risks, yet the risk evaluation paradigm to manage safety at multiple altitudes still lacks. The study aimed to investigate the effects of altitudes on controllers' mental workload and fatigue to assess the safety risk and introduced voices, mental workload, and fatigue, into a conceptual risk assessment model. Controllers from the Civil Aviation Administration of China (CAAC) conducted the experimental tasks, reporting mental workload and perception fatigue across three altitudes: 0 m, 2243 m, and 3569.7 m. With experimental data: this research (1) quantitatively compared the voice feature differences with feature engineering, and an image quality measure, (2) explored the effects of altitude, sleep, and fatigue, (3) tested the effects of altitude and task complexity on mental workload, and (4) evaluated the airport safety risks under ergonomic factors. Notably, the study revealed that Log-Mel spectrograms outperformed Mel Frequency Cepstral Coefficients (MFCC) in severe fatigue detection. Altitude and task complexity had significant main effects on the mental workload, but altitude had no significant moderator effects on the relationship between sleep and fatigue. The simulation results show that under the low task complexity, the operation risk is low over three airport elevations (with the human error rate $< 10^{-3}$), whereas under the high task complexity, the operation risk increased with altitudes (from 1.73×10^{-3} to 1.02×10^{-1}). Together, these results suggest that ergonomic factors influenced airport safety risk at multiple altitudes and promising real-time fatigue detection with voice features at different altitudes.

INDEX TERMS Voice features, fatigue, mental workload, altitude, safety management.

I. INTRODUCTION

Safety management is a subject with mounting importance in the civil aviation domain due to increased air traffic demand [1], [2]. Statistically, human factors contribute to about 75% of aircraft accidents and incidents [3]. Plateau (airports between 1500 m and 2438 m) and high plateau airports (airports over 2438 m) with hypobaric and hypoxia conditions aggravate operation risk for the immense human vulnerability in these conditions. For instance, the decrease in cognitive performance [4], [5], the impairment of mood [5], and the impairment of working memory [6]. These human vulnerabilities can affect task completion and, in turn, affect operational safety. Applying ergonomics in the operating system is effective and efficient [7] and even proactive to reduce safety occurrences [8]. Thus, analyzing the effects of

The associate editor coordinating the review of this manuscript and approving it for publication was Liang-Bi Chen.

human factors in a risk evaluation system, especially at the plateau and high plateau airports, is a necessity.

At plains, the tower controller is not the most critical safety-related research subject. Since statistically, pilots, flight crew, maintenance personnel, and en route controllers are the most critical safety personnel [2], [3], [9]–[11]. However, the safety data from plain airports occupy a large amount of aviation safety data for the proportion advantage of plain airports. That leads to the failure of safety data in safety management guidance to select key safety personnel at the plateau and high plateau airports. In addition, the hypobaric and hypoxia conditions cast a limited influence on personnel like flight crew and pilots since oxygen supply equipment has isolated these people from the unique natural environment [12].

Moreover, tower controllers undertake more tasks at the plateau and high plateau airports for confronting more complex natural environments, such as unpredictable wind changes and worse equipment conditions caused by extremely low temperatures, than tower controllers at plains. The excessive task load can increase the tower controllers' human error rate [13], which, in turn, increases the operational risk of the plateau and high plateau airports. Therefore, this project examines tower controllers' workload and fatigue in high elevations to manage safety risks.

In practice, applying sensors in controller behavior study in a complex context is a trend for the advances in sensor technologies [14]. Like electrooculogram (EOG), electrocardiogram (ECG), electroencephalogram (EEG), and functional near-infrared spectroscopy (fNIRS), sensors are reliable in measuring human cognitive states [14]. Nevertheless, preceding sensors can disturb the operator's regular duty, for their intrusive resulted in restrictions on the head or hand movements, or uncomfortably sensor wearing experience [15]. Unobtrusively data-collecting technologies are more suitable for exploring the tower controller's behavior patterns than invasive devices. Several studies [16]–[18] have applied voice sensors to learn air traffic controllers successfully. That is also why the authors use voice sensors in the tower controller experiments. Besides, no study has compared physiologic measures of the operators' state collected by sensors in multiple altitudes to our knowledge. So, it remains unclear whether physiologic measures can be informative in human abnormal state detection in diverse altitudes. To fill this gap, the authors compare voice patterns denoted with different fatigue levels at multiple altitudes.

Besides, the present literature focuses on fatigue detection and fails to investigate the effects of fatigue on airport operations. Effects like fatigue can lower the controllers' reaction speed [19] directly associated with airport safety. Despite that fatigue can impair operator performance, tower controllers' fatigue is not the paramount contributor to aviation occurrences [20]. Most literature [20] viewed fatigue as an operational risk element that neglects fatigue's effects on tower controllers' health. Burnout has been a severe problem for air traffic controllers, wherein fatigue is a vital element for emotional exhaustion, a component of burnout. Nevertheless, limited literature has discussed tower controllers' burnout [21]. So, to fill this gap, the authors established a proactive safety management tool including occupational safety and employed fatigue as an indicator.

This paper investigates the effects of altitudes on controllers' mental workload and fatigue to assess the safety risk and discusses the possibility of voice feature fatigue prediction in various altitudes. This study organized the rest of this article as follows: Section II, the context learning and literature part, also the basis for experiment design, illustrates the workflow of tower controllers, radiotelephony communication characteristics between controllers and pilots, the two essential safety risk factors of tower controllers, and experimental hypotheses. Section III is the methodology part, demonstrating data collection approaches, the experiment, factor analysis methods, and a brief introduction to the novel conceptual safety management model establishing process. The detailed experiments, model simulations, and relevant results are in Section IV. Section V is the discussion part, and Section VI is the conclusion part.

II. CONTEXT LEARNING AND LITERATURE REVIEW

A. MENTAL WORKLOAD

Mental workload works as the core of all human factors in controller operation safety for its direct impacts on other mental states [13]. Here the authors defined the mental workload as the present amount of allocated attention. The attention required in the air traffic, a complex situation, can exceed the controller's available attention causing the increasing workload [22]. The increasing mental workload can decrease reaction speed [23]. The excessive mental workload can lengthen the controller's reaction time [24] to increase the operational risk. Thus, it is necessary to analyze mental workload in safety management.

Multi-task and increasing task demands are the sources of mental workload [23]. Although the tower controlling is a multi-task, radiotelephony communication with radio channels is the dominant information exchange route between controllers and pilots during tower controlling operations. Also, radio communication is a vital underlying contributing factor to airport occurrences. Statistically, radio communication errors and problems contributed to 50% of incidents in the airport and were related to 40% of runway incursions [17]. So, the researchers assumed the communication task as the critical task for tower controllers and analyzing radiotelephony communication is the core for mental workload study. That is also the reason why the writers designed our controller simulation tasks based on voices.

Besides, exploring tower controllers' radiotelephony communication data can be a valid method to learn the context of airport control tasks. The communication structures of tower controllers are relatively fixed. Based on the literature [16], [25], [26], this study concluded the communication structures into two types: first calls and other calls and content into six categories: request, readback, courtesy, advisory, report, and others. The first call initiated by a pilot is in the format: control-unit code + aircraft call sign + content. The structure of other calls of the pilot is content + aircraft call sign. After learning the rules of communications between pilots and controllers, this research designed the experimental paradigm for participants to simulate tower controller tasks where the participant would confront similar communication situations and reply with identical rules like the above.

B. VOICE PATTERNS

Another contribution of voice data is the potential to detect controller abnormal cognitive states. There is little literature using voice patterns to explore air traffic controllers' fatigue for limited access to real-world air-ground communication datasets. Voice patterns used to detect human fatigue mainly include non-linear dynamics features like fractal features and phonetic features like formant bandwidth. Krajewski, *et al.* [27] selected 395 non-linear dynamics features and 170 phonetic features. After the correlation filter process, they trained data by AdaBoost and Bagging to obtain the best models for sleepiness detection Shen, *et al.* [16] viewed controllers' fatigue as a binary tuple, fatigue or not fatigue, and employed a support vector machine to detect fatigue via the fractal dimension. Whitmore and Fisher [28] conducted a 36-hour experiment to obtain speech and fatigue data, and via statistical analysis, they proved that fundamental frequency is a valid indicator for fatigue. De Vasconcelos, *et al.* [15] took speech analysis as an accident investigation tool. After comparing voice patterns of non-sleepy data and voice data one hour before the accident, they found pauses during speaking and elocution articulation rates are connected with fatigue. The above literature indicates that voice patterns are practical and useful for fatigue detection.

The Mel Frequency Cepstral Coefficient (MFCC) and Log-Mel images are famous for their robustness in the presence of various noises, which is insensitive to the mild change and the systematic computation technologies utilized in speaker recognition systems [29]. Thus, this study compared the feature diversity of these two caused by the elevation environment. More to the point, this study processed Log-Mel spectrograms comparison as an image discrimination problem. The authors selected an image quality assessment index from the image reconstruction area, the pixel-wise loss function [30], to evaluate the spectrogram differences.

C. HYPOTHESES

Standing on the predicted effects in four leading areas: voice features, workload, fatigue, and safety management, this research proposed the following hypothesis:

Because MFCCs and Log-Mel spectrograms show outstanding robustness in speech recognition [31], thus:

H1. Altitude would not affect MFCCs and Log-Mel spectrograms in fatigue detection.

Further, in machine learning models, Log-Mel spectrograms performed better than MFCCs in speech recognition accuracy [31]. So:

H2. Log-Mel images would be more promising than MFCCs in fatigue detection at various elevations.

The literature suggests that altitude can accelerate fatigue accumulation [32], and fatigue is closely related to sleep-related factors, like sleep duration and circadian rhythm [33]. Interviews with controllers from airports at several altitudes suggest that altitude would influence the relationship between sleep and fatigue. Therefore:

H3. Altitude would moderate the relationship between sleep and perception fatigue.

Apart from impacting fatigue perception, the environment discrepancy caused by the altitude difference may also affect workload perception. Besides, for the high correlation between task load and mental workload [34], hypotheses regarding mental workload are:

H4. The main effect of task complexity on mental workload is significant.

H5. The main effect of altitude on mental workload is significant.

H6. There is a significant interaction effect between task complexity and altitude.

III. EXPERIMENTS AND CALCULATIONS

This project was in a mixed-methods design involving both qualitative and quantitative data. In our design, qualitative data was the voice from airport tower controllers at several altitudes.

A. EXPERIMENT PARADIGM

All the experiments were in the within-subject design. The authors employed a cognitive test battery, paper-and-pencil questionnaires, and subjective scales to investigate cognition performance differences between airport tower controllers at different altitudes.

1) A COGNITIVE TEST BATTERY

An approximate 20-min test battery of working memory (WM) and an approximate 30-min controller simulation task comprised the cognitive test battery. A detailed description of these tasks is as follows:

a: DUAL-TARGET N-BACK TASK

The dual-target n-back task, with two stimuli (visual and acoustic stimulus), is a modified version of the single-target task of Townsend and Eidels [35] and the n-back tasks [4, 36]. The dual-target n-back task aimed to investigate the relationships between the response time (RT) and perception load. The character 'n' denotes the exact number of trials from the target trial to the current trial [4]. Participants were required to distinguish the targets and non-targets to respond to the occurrence of the visual target by pressing the letter A or the acoustic target by pressing L. The participants shall react immediately after the perception of the stimuli. The effective response time was from 100 ms, the minimum response time of human-being to a stimulus [37], to 1800 ms. Late and Premature responses were classified as missing. As shown in Fig.1, the visual appeared randomly in a grid of the 9×9 matrix. At the same time, the acoustic stimulus was presented to the participant every 3 seconds (Fig.2). With the n-back task, a classic trial to study working memory, the authors could efficiently manipulate the variable, the perception load (1-, 2-, 3-back) [38]. That is, the task contained three blocks in total, and each block had 45 trials. Before these experimental blocks, three practice blocks with 20 trials familiarized participants with the task rules. Also, to eliminate the practice effect during the experiment, this research used the ABBA balance method to decide the sequences of the blocks (Fig.3). Together with the task starting duration, in practice, the WM battery lasted about 20 mins.

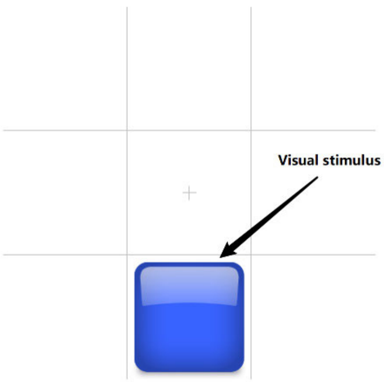


FIGURE 1. Visual stimulus in a 9 × 9 matrix.

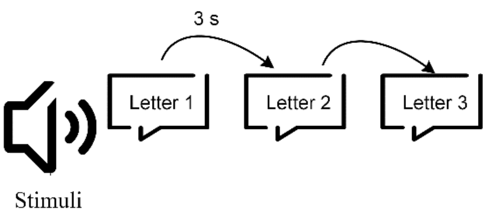


FIGURE 2. The acoustic stimulus of the dual-target n-back task.

Environment Adaptation	Guidance	1	2	3	3	2	1

FIGURE 3. Experimental procedures. The sequence of blocks was generated with the ABBA method.

b: CONTROLLER SIMULATION BATTERY

This study designed task sceneries and rules for the controller simulation battery based on the analysis of radiotelephony communication structures in Section II, included three levels, each containing eighteen PowerPoint slides with six repetitive disparate sceneries presented in Fig.4. Each group lasted about 10 mins, so in total, this battery lasted about 30 mins. The detailed introduction to the controller simulation battery is as follows:

c: BASELINE GROUP

The participants in this experiment would deal with three possible requests from the virtual pilot. Every request corresponded to two situations (as shown in Fig.4a). Combined the situation with a request from the virtual pilot, the participant shall give a proper reply. To elaborate, the participant shall approve the taxi request unless there is another aircraft (marked in yellow) on the taxiway. In the latter condition, they shall stop the aircraft by giving instructions like *hold position*. After that, the virtual pilot would read the instruction back. Then the participant shall judge whether the readback is correct. Addressing the departure and approach requests, the participant shall react following a similar logit. That is approving requests when no obstacle is detected and rejecting requests otherwise.

d: GROUP 2

In the second level, this research increased the task complexity by introducing an addition problem at the end of the readback or in step 3 in Fig.4b. The participant shall complete the additional calculation except for the judgment of the readback.

e: GROUP 3

Unlike the second level, the study adjusted the addition problem's location to the initial request's end (as illustrated in Fig.4c). The participant shall speak out the answer before issuing the instruction to the virtual pilot. That meant increased cognition demands and working memory demands for the participant.

2) SUBJECTIVE SCALES AND QUESTIONNAIRES

This research administered paper-and-pencil questionnaires and scales to collect basic information and assess subjective symptoms of fatigue and subjective workload at various working stages and experimental conditions. Stanford Sleepiness Scale (SSS), the oldest self-report measure to evaluate the human state of sleepiness, was adopted to assess tower controllers' current state [39]. The SSS is a single-item state scale consisting of seven ranked statements. Participants shall select the number that best matches their current state. As for subjective workload, this study adopted the NASA task load index (NASA-TLX) scale, dividing the workload into six dimensions: mental demands, physical demands, temporal demands, performance, effort, and frustration. The authors employed the adapted general task load index (GTLX) calculated from the NASA-TLX scale [24] to express the mental workload:

$$GTLX = (MD + TD + (10 - PE) + EF + FR + PD)/6, \quad (1)$$

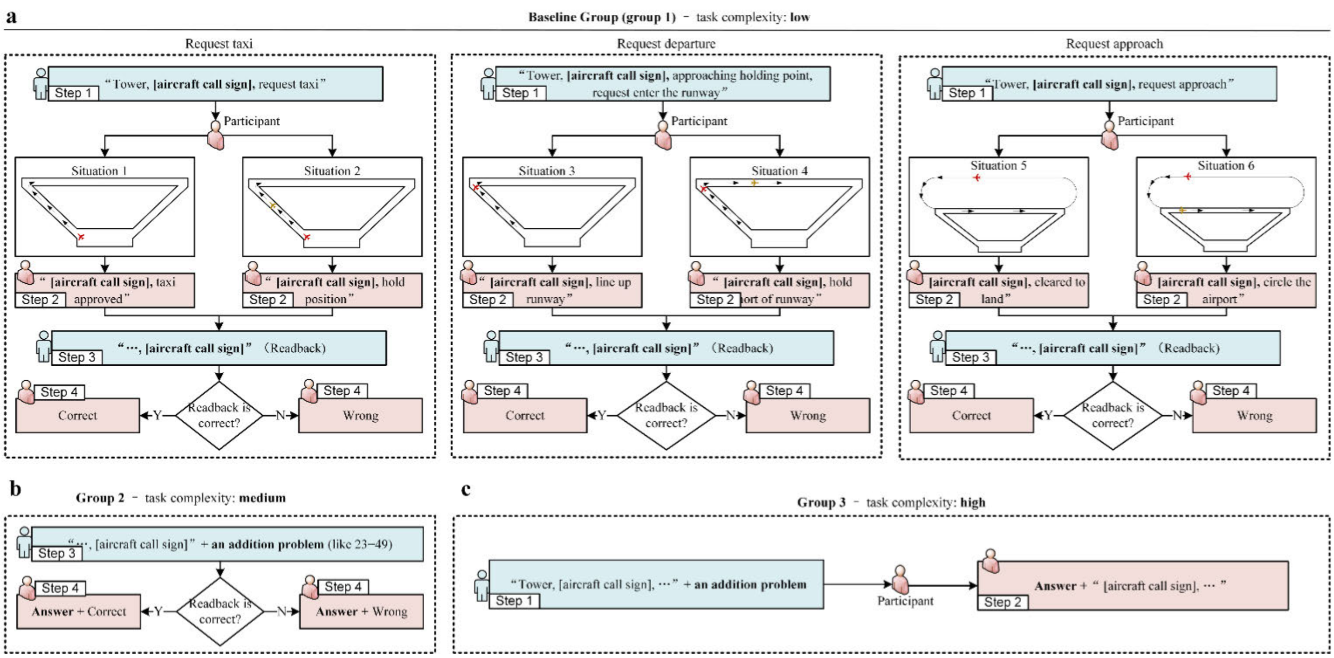


FIGURE 4. Overview of the tower controller simulation tasks and the three experimental groups. The primary tower controller simulation task contained three pairs of controller sceneries, six situations in total. To be noted, in six cases, the red aircraft is the target waiting for the participant's reply. The yellow is the obstacle. b The second level task included an addition problem in step 3. c The most challenging mode included an addition problem in step 1.

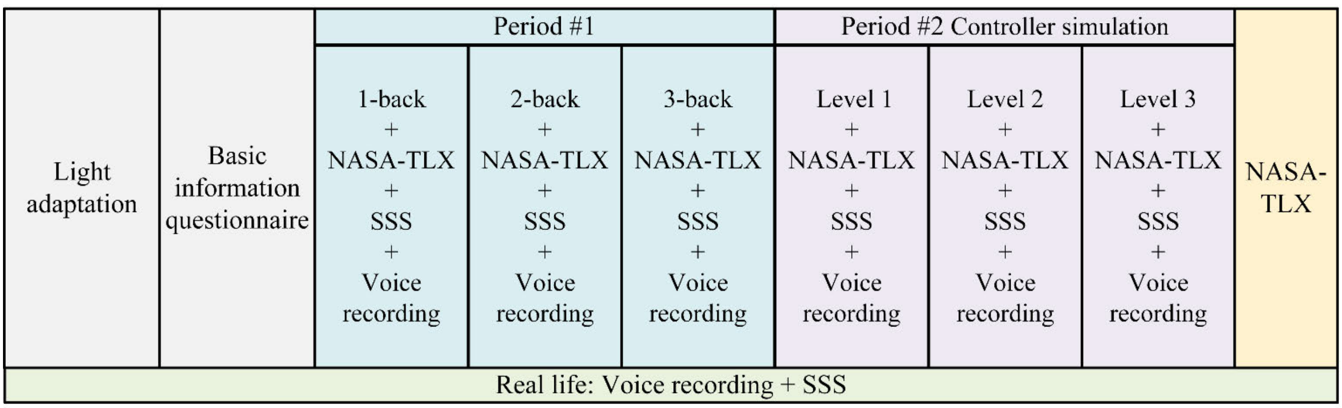


FIGURE 5. Task sequence for each experimental session. (Timeline not to scale).

where *MD, TD, PE, EF, FR, PD* stand for mental demands, temporal demands, physical demands, effort, frustration, and performance.

3) EXPERIMENT PROTOCOL

Prior to the study, the authors obtained written, informed consent from all participants. Participants were seated in front of a screen (resolution: 1024 × 768 pixels). These tasks were taking place indoor, so a 10-min light adaptation was required. Fig.5 illustrates the complete task sequence. After the light adaptation, participants conducted the dual-target n-back tasks, following by the controller simulation tasks. Before starting every task, participants read letters A, B, E, F aloud, and professional devices recorded their voices with an audio sampling rate of 16000 HZ. At the same time, they needed to report their state based on SSS. At the end of every task, participants would assess the workload via NASA-TLX. Finally, they need to rate the workload of real work. Meanwhile, the authors accepted state self-report from participants who were not in the ongoing experiment.

B. RADIOTELEPHONY COMMUNICATION ANALYSIS

This research analyzed the radiotelephony communication interval and voice patterns in dealing with voice data captured in the experiments. The voice patterns analysis mainly included two processes: feature extraction and feature selection.

1) RISK ANALYSIS BASED ON INTERVAL MANAGEMENT

The measure for communication interval analysis was the reaction time of the participants in our paper. To decide whether the captured response time was reasonable and acceptable, the authors employed the human error rate to evaluate. In our experiments, the direct form of human error was that controllers forget or call the wrong aircraft call sign, miss the wrong readback, or give the wrong answer to the addition question. If the direct human error rate were too high, the authors would classify the response time as invalid and give up further analyzing response times in the experimental scenery. To be noted, in this paper, the response time was the human performance measure rather than the mistake or error rate.

a: INTERVAL MODELING

This study employed the lognormal distribution to model the response time, which is believed to fit the reaction time data well [40]. Before modeling, the authors applied the Anderson-Darling test [40] to verify whether the reaction time logarithm was normally distributed. The data, whose logarithms passed the Anderson-Darling test, followed a lognormal distribution. The authors performed the Anderson-Darling test and distributing fitting with MATLAB. Recall that the probability density function of the lognormal distribution is:

$$f(x) = [1/(\sqrt{2\pi} x\xi)] \cdot \exp[-(\ln x - \lambda)^2/(2\xi^2)], \tag{2}$$

for $0 < x < \infty$ and where μ and σ are the sample mean and the variance, respectively; λ and ξ^2 are defined as:

$$\lambda = \ln(\mu/\sqrt{1+\sigma^2/\mu^2}) \tag{3}$$

$$\xi^2 = \ln(1+\sigma^2/\mu^2), \tag{4}$$

b: RISK ASSESSMENT BASED ON THE MENTAL MODEL DISCONNECT

The prolonged reaction time, of course, can contribute to flight delays. Meanwhile, it is also a shared cognitive disconnect between controllers and pilots, viewed as negative operator performance in this article. The shared mental model disconnects between air traffic controllers and pilots may cause accidents or incidents in critical flight stages [41]. In general, if the time left for a pilot to react is less than 20 s, the pilot fails to prevent any accident [42]. According to Yang and Hu [40], the average controllers' reaction time is 7 s. The delayed reaction tends to result in a second-time pilot call, which increases the possibility of accidents or incidents. So, the authors set the adequate reaction time of the controller as 7 s. That is, if the controller's response time is over 7 s, the reaction is a human error. Correspondingly, the human error rate, *HER*, can be calculated with the following equation:

$$HER = 1 - \textstyle\int_{-\infty}^{7000} f(x)dx \tag{5}$$

The influences of human errors on operating safety can be evaluated via the table function (Table 1), according to the research results of Grozdanovic and Bijelić [43].

TABLE 1. The table function between human errors and operational risk.

Risk level	Position	Human error rate
Low	5	10^{-3}
	4	10^{-2}
Medium	3	10^{-1}
High	2	3×10^{-1}
	1	10^{0}

Data Source: Grozdanovic and Bijelić [43].

2) VOICE FEATURE ANALYSIS

a: FEATURE EXTRACTION

The authors extracted the voice features with the adapted MFCC computation technique raised by Sahidullah and Saha [29]. The aim is to obtain twelve MFCCs and twenty-four corresponding dynamic MFCCs. The feature extraction process is as follows: First, the authors manually segmented all collected voices. Then, the authors pre-emphasized the audio to balance the frequency spectrum. After that, the researchers sliced the signal into frames with a 25-ms slice and a 15-ms overlap. Next, the authors applied the Hamming window to each frame. After that, the Fast Fourier Transform (FFT) was performed, and the power spectrum was calculated.

Further, the Discrete Cosine Transform (DCT) was used to compress the filter banks to reduce the correlations among filter bank coefficients. This study only kept the first two to thirteen resulting cepstral coefficients since the discarded ones performed poorly in speech recognition.

As liftering is proved to reduce the signal reverberation of cepstral domain features [44], the following sinusoidal liftering [45] was conducted:

$$w_i = D/2 \cdot \sin(\pi i/D), \tag{6}$$

where, w_i is lifter weight, that conducted on the D cepstral coefficients.

Finally, the first- and second-time deviations of the extracted MFCCs were calculated to obtain the Delta and Delta Delta MFCCs, or in other words, the dynamic MFCCs.

b: FEATURE SELECTION

Due to the small amount of speech data, the machine learning model for feature classification is unsuitable for this project. Therefore, accordingly, the authors employed filter methods to select features.

In this project, for the unique data type and the relationship between different feature classes, the authors adopted the below two procedures to simplify the feature selection process. Initially, this research calculated the mean, standard deviation, maximum, minimum, median, 25th percentile, and 75$^{\text{th}}$ percentile of each feature to neglect the heterogeneity of

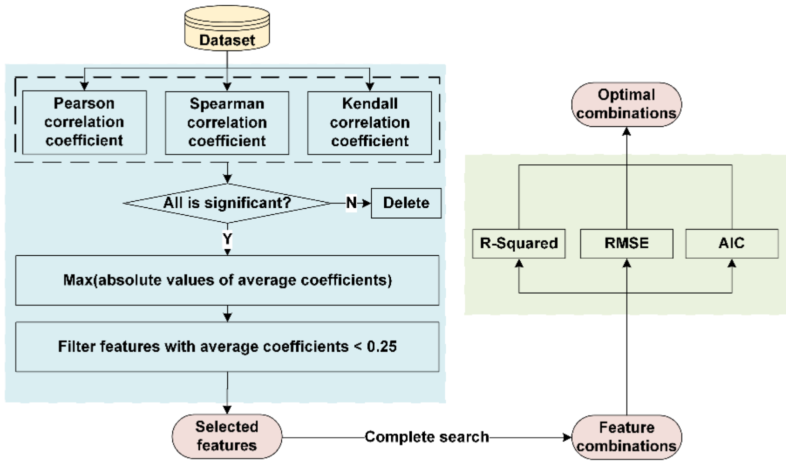


FIGURE 6. **The flowchart of filtering features.**

speech data. After that, the authors used combined filters to reduce features.

As shown in Fig.6, the authors calculated three correlation coefficients: the Pearson coefficient, Spearman correlation coefficient, and the Kendall correlation coefficient. This study would filter this feature if the correlation relationship between the target and the feature were not significant in any three correlation tests. Then, we calculated the average of three correlation coefficients of the left features, and for each voice feature, the authors only picked the statistical feature with the maximum absolute value of the average coefficient. Moreover, from the MFCCs and fatigue correlation analysis result of related documentation [27], the correlation value of those used to predict fatigue tends to be over 0.25. Thus, here the authors only keep features whose average correlation value is above 0.25.

Empirically, the number of selected features is no more than ten. In other words, the number of possible feature combinations is fewer than 1023 (2^{10}-1). Thus, this research conducted a complete search with recursion to find all the feature combinations. Next, the study established multiple linear regression models with all feature combinations and the SSS reference values set. It is important to note that the research focuses mainly on prediction accuracy rather than the relationship between audio features and the SSS reference values. Therefore, the authors ignored the multicollinearity caused by the high intercorrelations among independent variables or, to be precise, the audio features. At last, standing to metrics including R-Squared, RMSE (Root Mean Squared Error), and AIC (Akaike Information Criterion), the optimal feature combination was determined. The above three metrics can be calculated with the following formulas:

$$R^2 = 1 - \sum_i^n (y_i - \hat{y}_i)^2 / \sum_i (y_i - \bar{Y})^2, \tag{7}$$

$$RMSE = \sqrt{1/n \cdot \sum_{i=1}^n (y_i - y'')i}, \tag{8}$$

where, y_i is the observed value, $\hat{y}_i$ is the forecast value, n is the sample size.

$$AIC = 2k - 2\ln(L), \tag{9}$$

where, the log-likelihood estimate is denoted as L; n is the sample size; k is the number of parameters.

Gap analysis based on image processing. Log-Mel spectrograms can be a valuable tool in speech emotion recognition [46] and disease severity judgment [31]. According to Suhas, *et al.* [31], Log-Mel spectrograms outperform MFCCs in voice recognition. Thus, the authors compared the Log-Mel spectrograms of voice at different airports on different fatigue levels (Fig.7).

The pixel-wise loss function is a commonly used quality assessment method in image reconstruction [30]. In this paper, based on the Log-Mel spectrograms, the authors used a pixel-wise discrimination model to compare image appearance differences on differing conditions. The pixel-wise loss consisted of the structural similarity index (SSIM) [30], [47] and the L_1 loss function, also known as the least standard deviation [48]. SSIM is defined as:

$$SSIM(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^\alpha \cdot [c(\mathbf{x}, \mathbf{y})]^\beta \cdot [s(\mathbf{x}, \mathbf{y})]^\gamma, \tag{10}$$
$$l(x, y) = (2\mu_x\mu_y + C_1)/(\mu_s^2 + \mu_y^2 + C_1), \tag{11}$$
$$c(x, y) = (2\sigma_x\sigma_y + C_2)/(\sigma_y^2 + \sigma_y^2 + C_2), \tag{12}$$
$$s(x, y) = (\sigma_n + C_3)/(\sigma_n\sigma_y + C_3), \tag{13}$$

where, *l(x, y), c(x, y),* and *s(x, y)* are used for luminance comparison, contrast comparison, and structure comparison, respectively. x and y are two partial blocks. μ_x is the mean value of x and μ_y is the mean value of y. σ_x is the standard deviation of x. σ_y is the standard deviation of y. Referring to Wang, *et al.* [47], authors set $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$ in this project.

L_1 is less sensitive than *SSIM* to outliers and can be expressed as follows:

$$L_1(\mathbf{x}, \mathbf{y}) = (\sum_{i=1}^N |x_i - y_i|)/N, \tag{14}$$

where x_i and y_i are discrete signals.

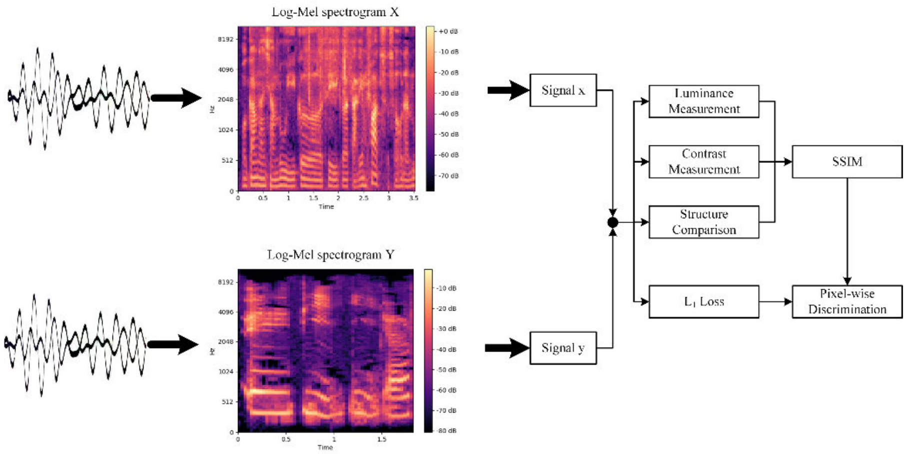


FIGURE 7. Diagram of the pixel-wise discrimination measurement system.

The pixel-wise discrimination function in this article is defined as:

$$C_{pw}(\mathbf{x}, \mathbf{y}) = (1-\lambda) \cdot L_1(\mathbf{x}, \mathbf{y}) + \lambda \cdot (1/N) \cdot \{\sum_{i=1}^{N} [1 - SSIM\,(x_i, y_i)]/2\}, \quad (15)$$

where C_{pw} is the result of pixel-wise discrimination between two images; λ is the hyperparameter working as a weight to balance the L_1 loss and SSIM. Referring to Li [49], this study set λ as 0.85; N is the total number of pixels.

This study compared Log-Mel spectrograms in two situations, as presented in Table 2. In the first situation, on the scale of the same SSS reference values, this research explored whether the image features of Log-Mel spectrograms from participants at the same altitude are stable. Under the second situation, the authors studied whether altitudes affect voice features' stability. The researchers established two indicators to evaluate the above differences, denoted as *SS, SN,* respectively. The exact formulas are as follows:

$$SS(H, S) = \sum_{i,j} C_{pw}(x_i^{H,S}, x_j^{H,S})/N, \quad (16)$$

$$SN(H_a, H_b, S) = \sum_{i,j} C_{pw}(x_{H_a,i}^{S}, x_{H_b,j}^{S})/N, \quad (17)$$

where, H and S are symbols of the altitude and the SSS value; $x_i^{H,S}$ and $x_j^{H,S}$ refer to images i and j with attributes: H and S; $x_{H_a,i}^{S}$ and $x_{H_b,j}^{S}$ are images i and j with attributes: H_a, H_b, and S; N is the comparison combination amount.

C. EFFECTS ANALYSIS OF CONTRIBUTING FACTORS AT DIFFERENT ALTITUDES

The primary purpose of effects analysis is to provide a theoretical ground for the controller schedule arrangement by analyzing perception fatigue and mental workload under the influence of various altitudes and task complexities.

TABLE 2. Gap analysis conditions.

Type	SSS reference values (same or not)	Altitudes (same or not)
1	same	same
2	same	not

1) MODERATOR EFFECTS OF ALTITUDE ON PERCEPTION FATIGUE

This article took the task itself as a contributing factor to cognitive performance instead of a direct contributing factor to fatigue to simplify the performance risk assessment model. According to Kelly and Efthymiou [50] and Raslear, *et al.* [51], the authors included sleep duration, circadian rhythms, and sleep inertia as fatigue-inducing factors. This study furthered to investigate whether altitude would affect perception fatigue. To quantitively express sleep duration, circadian rhythms, and sleep inertia reasonably, the authors modeled them based on the bio-mathematical fatigue model [52].

$$R_t = \begin{cases} R_{t-1} - Kt, \text{ during wake} \\ 2400\,(a_s+1)\left(\begin{array}{l} 1-\exp\left(-t/\tau_d\right) + \\ R_{t-1}\exp\left(-t/\tau_d\right), \text{ during sleep,} \end{array}\right. \end{cases} \quad (18)$$

$$C_t = \cos(2\pi(T-18)/24) + 0.5 \cdot \cos\left(4\pi\,(T-21)/24\right), \quad (19)$$

$$I_t = -0.05 \cdot e^{-0.04t/(-a_s C_t + f(R_c - R_t))}, \quad (20)$$

$$S_t = 1 - (100\,(R_t/R_c) + (a_1 + a_2\,(R_c - R_t)/R_c)\,C_t + I_t), \quad (21)$$

where, R_t is the homeostatic sleep reservoir, C_t is the circadian rhythm, and I_t is sleep inertia. K is the depletion rate, and this study set it as 30 units/hour. t is the awakening or sleep time. a_S and f are weighting factors, set as 0.235 unit and 0.0026564 min, respectively. τ_d is the recovery time with a value of 4.2 h. T is the time of the day. R_c represents the reservoir capacity, set as 2880 units. Specifically, sleep inertia

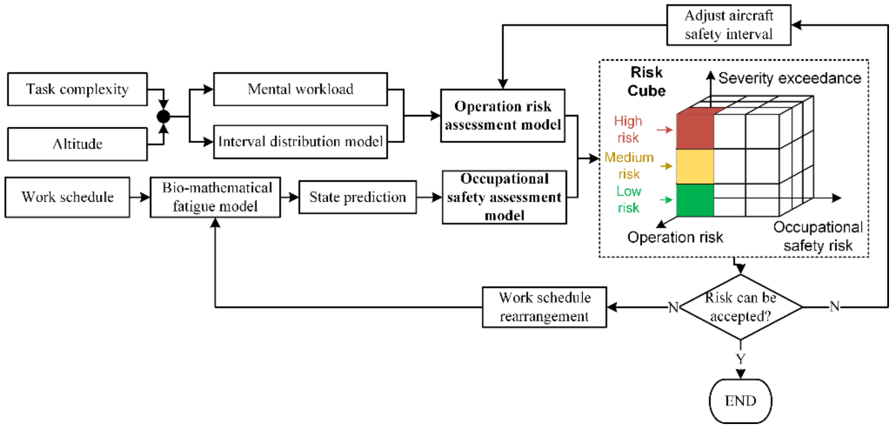


FIGURE 8. **The mechanism of the conceptual safety management model.**

I_t works only within the two hours after waking. S_t is the combined effect of the factors mentioned above.

This research tested whether altitude moderates the relationship between sleep-related contributing factors and perception fatigue with the linear regression hierarchical analysis [53]. The authors hypothesized a moderate effect of altitude on the relationship between sleep factors and the SSS reference values. If the p-value of the interaction term of the combined effect S_t and the altitude dummy variable were more than 0.05, this research would reject the null hypothesis. Otherwise, this study would accept the null hypothesis. The authors also utilized zero-centered independent variables to avoid the possible multicollinearity brought by interaction terms of independent and moderator variables.

2) MAIN EFFECTS AND INTERACTION EFFECTS ANALYSIS ON THE MENTAL WORKLOAD

To test hypotheses H4-6, the authors introduced a MANOVA test. Also, the authors utilized the ordinal regression model to explore the main effects of these two factors. Initially, this study reordered these six tasks according to the mean values of their mental workload. Afterward, the authors validated whether the ordinal logit models' slope coefficients are the same across response categories via the test of parallel lines. In other words, if the statistics pass the test of parallel lines, the statistics have proportional odds.

Meanwhile, the authors utilized the variance inflation factor (VIF) to test the severity of collinearity among factors. If VIF is more than ten, there is multicollinearity among variables. At last, the authors established the ordinal logit model for one to analyze altitude and task complexity's main effects on mental workload, for the other one to predict mental workload.

D. A CONCEPTUAL RISK ASSESSMENT MODEL

The conceptual risk assessment model contained four primary dimensions: operation risk assessment, controller occupational safety assessment, risk elimination, and safety reassessment (Fig.8). This study adopted a two-dimension risk assessment model with an embedded risk cube to manage safety dynamically. The rule to operate the risk cube is that the authors mapped the two types of risks to the airport safety risk axis, and the combined risk depended on the below rule:

$$risk = \max\{operational\ risk, occupational\ risk\} \quad (22)$$

1) MENTAL WORKLOAD BASED RISK ASSESSMENT

Although much relevant literature [25], [40] has calculated aircraft conflict possibility via modeling controller communication interval, they all built a relative macro risk assessment model. Most scholars neglected that the opportunity of human error is not equal under different levels of mental workload, although they have established other interval fitting models under different working conditions (traffic volume). Thus, besides altitude and task complexity, this study took mental workload into account in our operation risk assessment model. The elaborate process to establish the mental workload prediction model based on altitude and task complexity is in Section III-C. The revised human error rate is:

$$RHER_{ij} = MW_{ij}/MW_{0m,j} \cdot HER, \quad (23)$$

where, $RHER_{ij}$ is the revised human error rate of whose mental workload is MW_{ij}. i is the symbol of altitude, and j represents task complexity. $\sum_j MW_j$ is the sum of all possible mental workload. HER is the human error rate without the effect of mental workload.

2) OCCUPATIONAL SAFETY ASSESSMENT

Participants could not select the ensured value on the SSS scale during the experiment but would like to give an interval. Thus, the authors modified the SSS scale and divided the human state into three categories: awake, mild fatigue, and severe fatigue. They corresponded to three risk levels: low, medium, and high risk, as shown in Table 3. After that, the authors trained an ordinal regression classifier based on the Modified SSS scale. The thorough process is as follows.

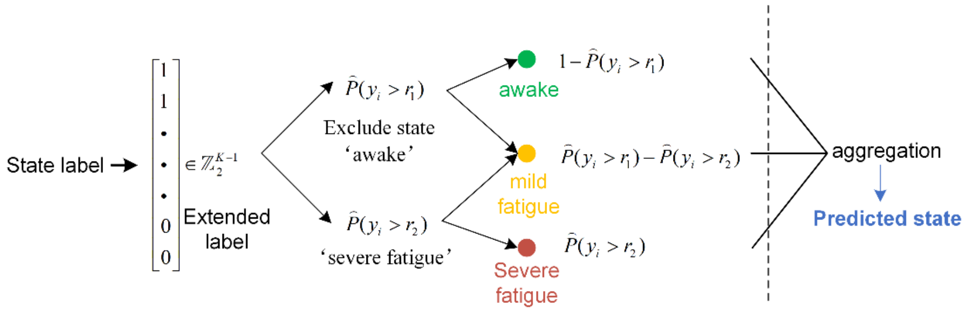


FIGURE 9. The diagram of identifying controllers' state with the ordinal regression model.

TABLE 3. The table function between human state and occupational safety risk.

Symbol	Quantity
Awake	Low
Mild fatigue	Medium
Severe fatigue	High

Initially, this research regrouped the SSS reference values into three levels based on their responses and used the dummy variable *state* as a replacement:

$$state = \begin{cases} 0(\text{awake}), & \text{if } SSS = 1 \text{ or } 2 \\ 1(\text{mild fatigue}), & \text{if } SSS = 3 \text{ or } 4 \\ 2(\text{severe fatigue}), & \text{otherwise,} \end{cases} \tag{24}$$

Then, this research checked whether the classify is reasonable with the aid of the combined sleep-related effect S_t: Firstly, this research employed the Kendall and the Spearman coefficient to explore the relationship between the combined sleep-related effect S_t and the SSS values; Secondly, if they were statistically significantly correlated with each other, this study would conduct the Anova test to investigate whether the variance was quite different among state groups. The authors believed that if the variance of S_t was significantly different among state groups, the state classification was reasonable.

Fig.9 demonstrates the mechanism to build an ordinal regression classifier. For the target variable, there were three state classes with an order. In order not to lose the information about the ordered relationship, before implementing the model, this research carried out the following steps:

1. The authors extended the state labels with three dummy variables. Firstly, this research set the state, *awake*, as 0, the remaining states as 1. Next, this study assigned 0 to state, *awake*, and state, *mild fatigue*, and 1 to the rest. Then, this study set the state, *severe fatigue*, to be 1, and others as 0.

2. For each dummy target variable, this research established a logit model. These models can predict input data as class 0 or 1. The relevant formula is shown in Fig.10.

3. At last, aggregating three predicted probabilities, and the authors can obtain the predicted state.

Furthermore, for the small data sample, the authors used K-Fold cross-validation to evaluate the established model.

FIGURE 10. The pie chart of selected features at different airport altitudes.

The authors split the data into five groups. Every subset would work as a test set, while the left four would work as the training sets to get five models. The accuracy was the mean of the correct rates of the five models.

IV. MODELING AND RESULTS

A. NUMERICAL EXAMPLES

1) PARTICIPANTS

The authors conducted the experiments mentioned in Section III(A) at 0 m, 2243 m, and 3569.7 m. Eight controllers from CAAC between the ages of 22 to 34 volunteered for this study. The authors did not find noticeable individual differences in basic information processing abilities in all the experiments. Data collection lasted from 2018 to 2021.

2) SIMULATION PARAMETERS SETTING

The research only retained the first two-level controller simulation task data due to the extreme human error rate in the third-level controller simulation task. The parameters set for task-related risk analysis are in Table 4. Based on the interviews with local airport controllers at three altitudes, the authors set the simulation situations and parameters for the human state analysis (Table 4) conforming to the actual working arrangement. In every case, this study appointed several agents based on the actual tower controller work schedule and repeated every situation seven times to obtain multiple results in the regular duty.

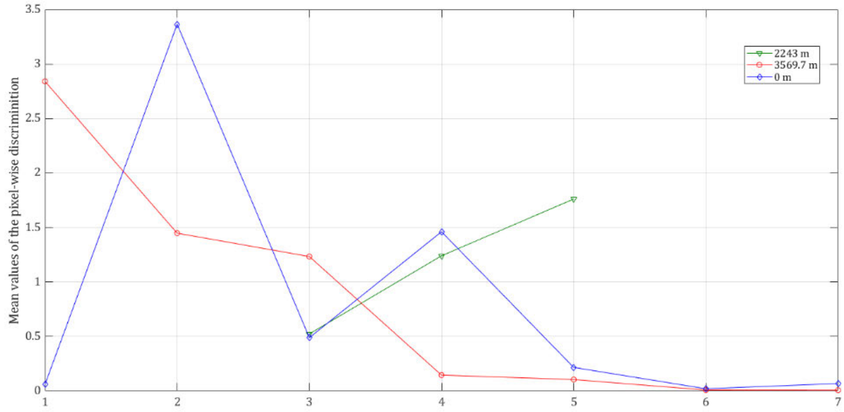


FIGURE 11. The pixel-wise discrimination among images with the same SSS reference values at the same altitudes.

TABLE 4. Simulation situations and associated parameters for the operation risk analysis.

Altitude	Task complexity
0 m	1
	3
2243 m	1
	3
3569.7 m	1
	3

B. VOICE FEATURE COMPARISONS AT THREE AIRPORT ALTITUDE

1) FEATURE SELECTION RESULTS

Tables 6-8 provide information about the selected features at three airport altitudes. There were ten filtered features at 0 m, which consisted of 1023 effective combinations. Among these combinations, the combination with the maximum R-Squared value and the minimum RMSE value (denoted as combination 01) was not the combination with the minimum AIC (denoted as combination 02). For the differences of the R-Squared and RMSE values between the above two combines were narrow (combination 01: R-Squared = 0.85, RMSE = 1.51; combination 02: R-Squared = 0.82, RMSE = 1.51), the authors thought that combination 02 was the optimal combination to avoid overfitting brought by redundant features. Combination 02 contained the *Standard deviation of delta MFCC 12* and the *Maximum of delta-delta MFCC 8*. On the condition of 2243 m, nine selected features comprised 511 possible combinations. Following the same optimal selection rules, the optimal feature combination at 2243 m included the *median of MFCC 5, minimum of MFCC 11, 25th percentile of MFCC 10, and the 25th percentile of MFCC 4* (R-Squared = 0.96, RMSE = 0.79, and AIC = 85.85). Moreover, at Lhasa airport (3569.7 m), based on the 31 combinations of five features, the authors selected the optimal combination with the *median of MFCC 2, minimum of delta MFCC 5, minimum of MFCC 5,* and *mean of MFCC13* (R-Squared = 0.86, RMSE = 1.22, and AIC = 214.96).

2) HYPOTHESES 1 & 2: EFFECTS OF ALTITUDES ON MFCCs AND LOG-MEL SPECTROGRAMS

The results of this paper partially rejected H1 but supported H2. The pie chart (Fig.10) compares the selected feature amount and the feature nature difference (dynamic or static) of MFCCs at three airport elevations. The research shows that the features of MFCCs left at 2243 m and 0 m were more than those at 3569.7 m. Besides, at sea level, dynamic features constituted over 50% of the total features. By contrast, at the other two altitudes, static features exceeded dynamic features. The graph indicates that dynamic features tended to conduct a favorable at lower levels in fatigue prediction, while at higher levels, they lost strength in fatigue prediction. This result suggests that altitude did affect MFCCs in fatigue detection. Fig.11 compares the *SS* values of seven states at three altitudes. In general, the *SS* value was no more than 3.5. Besides, the *SS* value was lower than 0.1 under the condition with the SSS value of 5, 6, or 7. It indicates that Log-Mel images had considerable minor differences at high SSS values at the same elevation. Together, altitudes would affect MFCCs in fatigue detection but slightly affect Log-Mel spectrograms when the SSS values were tremendous.

Fig.12 shows the mean values of the pixel-wise discriminations among images with the same SSS reference values at different altitudes. It is important to note that the higher the pixel-wise discrimination value was, the more significant the gap between two Log-Mel spectrograms. The graph shows that the pixel-wise discrimination values of compared groups with SSS values 1, 6, and 7 were much lower than groups with other SSS references. To be more precise, the *SN* values of compared groups with SSS values 1, 6, and 7 were lower than 0.15. That suggests that the Log-Mel image can be an excellent choice to predict fatigue without being disturbed by altitudes. Together, Log-Mel images would be more promising than MFCCs in fatigue detection at multiple elevations.

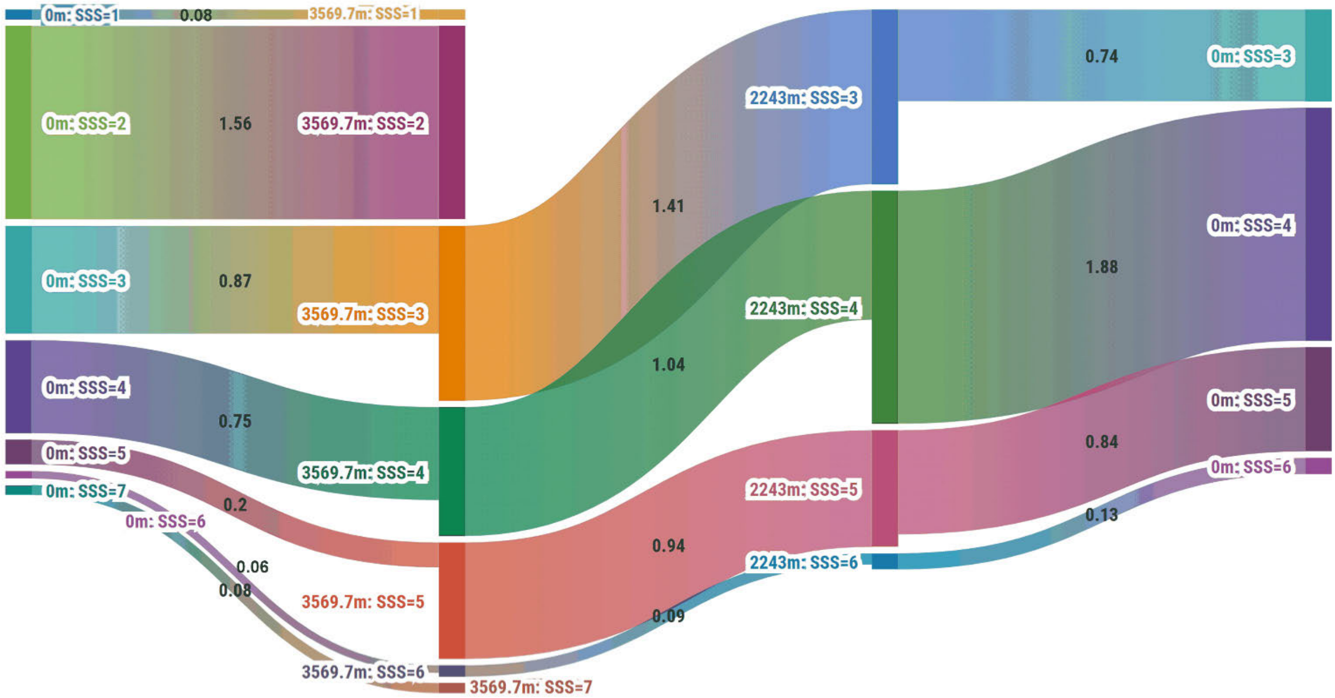


FIGURE 12. The pixel-wise discrimination among images with the same SSS reference values at different altitudes. The label naming rule is altitude + SSS reference value. The number in black is the pixel-wise discrimination value.

C. EFFECTS ANALYSIS RESULTS

1) HYPOTHESIS 3: ALTITUDE EFFECTS ON PERCEPTION FATIGUE

H3 was not supported in this article. The linear regression hierarchical analysis result indicated no statistically significant moderator effects of altitude on perception fatigue (the p-value of the interaction term of centered S_t and altitude is more than 0.05). Although the moderator effects were not statistically significant, Fig.13 illustrates the possible influence of altitude on perception fatigue. People at 2243 m had higher perception fatigue at the same S_t level. Also, the slope of data collected at 3569.7 m was more extensive than that at 0 m. That indicates that human fatigue was more sensitive to sleep factors at 3569.7 m than at 0 m.

2) HYPOTHESES 4-6: EFFECTS ANALYSIS RESULTS OF TASK COMPLEXITY AND ALTITUDE ON MENTAL WORKLOAD

Table 9, the MANOVA test results for mental workload, shows:

1. H4 was supported: the main effect of task complexity on mental workload was significant ($p < 0.001$).

2. H5 was supported: the main effect of altitude on mental workload was significant ($p < 0.001$).

3. H6 was not supported: there was no significant interaction effect of task complexity and altitude ($p > 0.5$).

The parallel lines test (Table 10) shows that the null hypothesis could be accepted ($\chi^2 = 282.017$, $p = 0.834$). That is, the regression equations were parallel to each other, and the authors could use ordinal logit regression to analyze statistics. Since the VIFs of variables were all less than 10, there was no multicollinearity among variables. Table 11 illustrates that the ordinal logit regression model fitted well with the data, as p-values of the Person test and Deviation test were over 0.05.

The authors drew the graphs about the main effects of the two factors, altitude, and task complexity, based on the established ordinal logit model, as shown in Fig.14 and Fig.15. These two graphs show that:

1. When the task complexity was fixed, the higher the altitude, the higher the mental workload.

2. When the altitude was fixed, in general, as the task complexity increasing, the mental workload went higher. However, when the complexity values were close to each other, the mental workload sometimes would not strictly follow the rule.

D. MODEL IMPLEMENTATION RESULTS

1) VALIDATION FOR MODIFIED STATE CLASSIFICATION

Constructed on the statistically significant correlation between the combined sleep-related effect S_t and the SSS values (the Kendall coefficient is 0.323 with a p-value less than 0.01, and the Spearman coefficient is 0.424 with a p-value less than 0.01), the authors believed that if the variance of S_t was significantly different among state groups, the state classification was reasonable. Table 12 displays Levene's test results on SSS groups, and the Levene statistic based on the mean was 2.139 with the corresponding p-value of 0.121. Because the p-value was over 0.05, the study failed to reject the null hypothesis. In other words, the research could not prove that the variance of the combined sleep-related effect S_t between the different SSS value groups

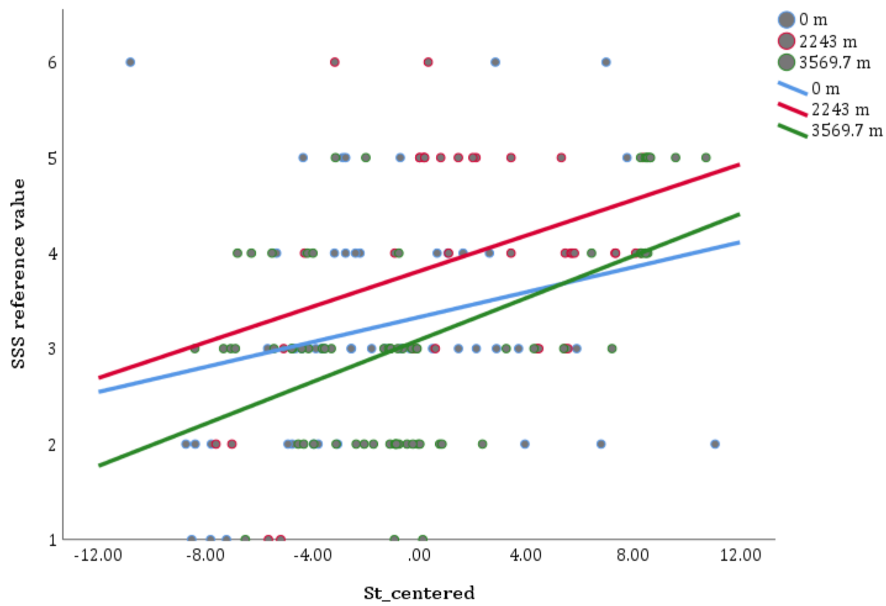


FIGURE 13. The graph of moderator effects of altitude.

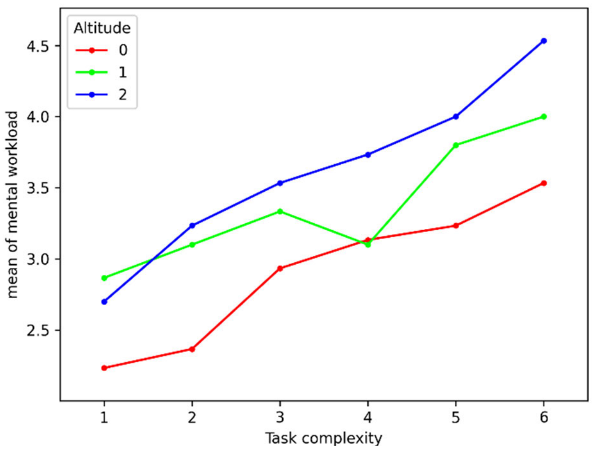


FIGURE 14. Main effects of altitude. Zero stands for sea level. One stands for 2243 m, and two stands for 3569.7 m.

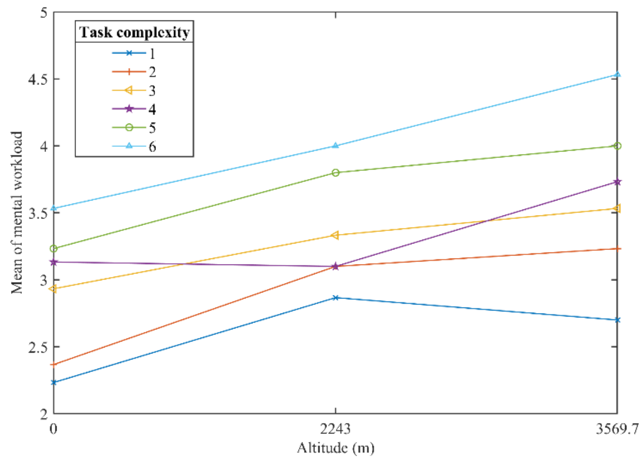


FIGURE 15. Main effects of task complexity.

was statistically significantly different. That means this study validated the assumption that each SSS value group had an equal variance via Levene's test, and the statistics met the requirement to conduct ANOVA (one-way analysis of variance) analysis. Table 13 indicates a statistically significant difference between the previously-mentioned groups regarding the combined sleep-related effect S_t for the significance value was 0.000, which was below 0.01. Therefore, it can be concluded that the modified state classification was reasonable and acceptable.

2) OPERATION RISK WITHOUT THE INTERFERENCE OF MENTAL WORKLOAD

Because when the task complexity set as one, the reaction times at three altitudes were not significantly different; therefore, the authors used one lognormal distribution to fit these data. The study utilized three more lognormal distributions to fit the reaction times for task complexity of three at three altitudes, respectively. The logarithms of the four groups of data all passed the AD test. In other words, these four groups of data followed the lognormal distribution. The four lognormal equations are shown in (25), (26), (27), and (28).

$$f_{complexity=1}(x) = \begin{cases} 1/(0.451x \cdot \sqrt{2\pi}) \\ \cdot \exp\left\{-(\ln x - 6.374)^2/0.4068\right\}, & x > 0 \\ 0, \quad \text{otherwise} \end{cases} \tag{25}$$

$$f_{complexity=3,altitude=0}(x) = \begin{cases} 1/(0.480x \cdot \sqrt{2\pi}) \\ \cdot \exp\left\{-(\ln x - 7.643)^2/0.4608\right\}, & x > 0 \\ 0, \quad \text{otherwise} \end{cases} \tag{26}$$

$$f_{complexity=3,altitude=2243}(x) = \begin{cases} 1/(0.491x \cdot \sqrt{2\pi}) \\ \cdot \exp\left\{-(\ln x - 7.875)^2/0.4822\right\}, & x > 0 \\ 0, \quad \text{otherwise} \end{cases} \tag{27}$$

$$f_{complexity=3,altitude=3569.7}(x) = \begin{cases} 1/(0.555x \cdot \sqrt{2\pi}) \\ \cdot \exp\left\{-(\ln x - 8.103)^2/0.6161\right\}, & x > 0 \\ 0, \quad \text{otherwise} \end{cases} \tag{28}$$

Fig.16 shows the four fitting curves, from which the authors can see that the higher the complexity was, the longer the reaction time would be. Besides, under the same task complexity, the reaction time tended to be longer with altitude. The calculated HER and corresponding risks are in Table 14.

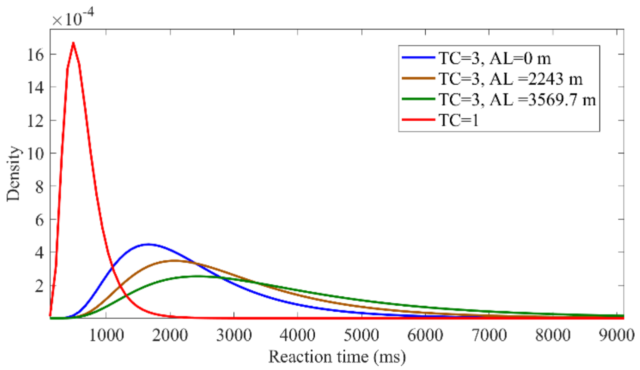


FIGURE 16. The lognormal fitting curves of four groups of data. TC stands for task complexity. AL stands for altitude.

3) NUMERICAL RESULTS OF RISK CUBE

With (23), the authors obtained the adjusted human error rates, and the detailed results are in Table 14. The cross-validation results of the ordinal logit model to predict the human state is in Table 15. The accuracy of the model was calculated as follows:

$$ACC = (54.545 + 66.667 + 76.190 + 71.428 + 80.952)/5 = 69.96.$$

The predicted states based on differing simulation situations and corresponding risk levels are in 17. Based on Tables 15 and 16, the study computed the combined risks or, in other words, the severity exceedance in the risk cube, as in Table 18.

V. DISCUSSION

A. APPLYING VOICE FEATURES IN FATIGUE DETECTION

Tables 6-8 are the results of features of MFCCs that were selected for their relatively high correlation with the human states, the SSS values. The filtered features were slightly different among the three altitudes. Previous studies found that MFCCs have strength in speaker recognition for their most sensitivity to fatigue [29], [45]. Compared to the analysis results of the correlation between MFCCs and KSS reference values in Krajewski *et al.*'s work, in the environment without the effects of hypobaric hypoxia, MFCC 8 performed well in fatigue prediction [27], which is consistent with our correlation analysis results of MFCCs. Fig.10 shows the most apparent difference among selected MFCCs features of the three elevations. That is, dynamic features had a higher correlation to fatigue in lower altitudes. While in higher altitudes, static MFCCs performed better than dynamic features. The results indicate that when choosing MFCCs to predict human fatigue, the prediction model shall vary with the airport elevations.

As for the Log-Mel spectrograms, Fig.11 and Fig.12 depict the pixel-wise discrimination comparison results among images with the same SSS reference values at the same and different altitudes. The results suggest that altitudes did not affect Log-Mel spectrograms in predicting severe fatigue, differing from MFCCs. Similar results are noted in Suhas *et al.*'s work [31], which indicates that Log-Mel images outperformed MFCCs in speaker recognition. Therefore, in conclusion, audio features can detect human fatigue, but the feature, Log-Mel spectrogram containing more information, is better than MFCCs when carrying out fatigue prediction in various elevations.

B. THE INTERACTIONS AMONG ALTITUDE, SLEEP, AND FATIGUE

The linear regression hierarchical analysis result of altitude, sleep, and fatigue shows airport altitude did not significantly affect the relationship between sleep and human fatigue in this article. Fig.13 demonstrates that at 2234 m, tower controllers precepted higher fatigue on the analogous sleeping condition than tower controllers at plain airports. Next, sleep seems to conduct an enormous impact on tower controllers' fatigue at 3569.7 compared with sea level.

Some literature has found that altitude affects cognitive fatigue. For instance, Duan, *et al.* [32] reported that altitude could significantly shorten the driver's fatigue duration Bouak, *et al.* [4] recovered that altitude increased general fatigue. Comparing experiments with experiments in the preceding literature, the authors found that the average experiment site elevation exceeded 4000 m. That may explain the insignificant moderator effect of altitude on sleep and fatigue since this paper's highest research elevation is 3569.7 m. Documentation by some scholars can support this assumption. These scholars recovered that fatigue increased when the environment passed a specific elevation, and a relatively low elevation would not affect fatigue accumulation [5], [54].

In addition, another possible cause of the insignificant moderator effect is that altitude can affect sleep from operator sleep duration and quality [55] but would not moderate the relationship between sleep and fatigue. Simultaneously, our current experiment could not validate this assumption because, in our experiment design, sleep factors were all independent variables. That led to that our experiment data

TABLE 5. Simulation situations and associated parameters for the human state investigation.

0	Agent type	Sleeping duration (hour)	Waking time	Snap interval	Working interval
1	1	[7, 9]	[6.00, 8.00]	-	[9.00, 17.00]
2	1	[5, 6]	[5.00, 5.45]	[14.00, 16.00]	[6.20,12.00] [18:00, 22:00]
	2	[8, 10]	[8.00, 10.00]	-	[12.00, 18.00] [22.00, 24.00]
3	1	[6.5, 8]	[6.30, 7.30]	-	[9.00, 14.30]
	2	[8, 10]	[8.00, 10.00]	- -	[14.30, 19.20] [23.20, 4.00^{+1}]
	3	[8, 10]	[8.00, 10.00]	- [0.00^{+1}, 3.30^{+1}]	[19.20, 23.30] [4.00^{+1}, 9.00^{+1}]

- means that the agent would not take a snap; +1 means that the working time lasts to the next day.

TABLE 6. Correlation analysis results between features and fatigue reference values at 0 m.

No.	Feature name	Pearson		Spearman		Kendall		Average correlation coefficient
		Correlation coefficient	p	Correlation coefficient	p	Correlation coefficient	p	
1	75th percentile of delta-delta MFCC 2	-0.415	0.006	-0.434	0.004	-0.340	0.003	-0.396
2	Maximum of delta-delta MFCC 8	-0.405	0.007	-0.465	0.002	-0.298	0.009	-0.389
3	Standard deviation of delta MFCC 13	-0.428	0.004	-0.386	0.011	-0.342	0.003	-0.385
4	75th percentile of delta-delta MFCC 3	-0.421	0.005	-0.468	0.002	-0.249	0.028	-0.379
5	75th percentile of delta MFCC 2	-0.398	0.008	-0.404	0.007	-0.271	0.017	-0.358
6*	Median of MFCC 12	0.333	0.029	0.394	0.009	0.274	0.016	0.333
7	Minimum of delta-delta MFCC 9	0.368	0.015	0.408	0.007	0.210	0.065	0.329
8	Standard deviation of delta-delta MFCC 8	-0.340	0.025	-0.343	0.024	-0.244	0.032	-0.309
9	Maximum of delta MFCC 9	-0.303	0.048	-0.350	0.022	-0.244	0.032	-0.299
10*	Minimum of MFCC 10	0.318	0.038	0.308	0.044	0.225	0.048	0.284

At the level of 0.05 (two-tailed test), the correlation between the feature and the target is significant; The feature with * means that the feature is static; otherwise, it is dynamic.

cannot reflect the effects of altitude on sleep. Therefore, further studies can further discuss this question.

C. RISK ASSESSMENT RESULTS

The MANOVA test results in Table 9 show significant main effects of altitude and task complexity on mental workload. Regarding task complexity, d'Engelbronner, *et al.* [56] recovered similar results that task complexity positively correlates with the mental workload. More importantly, this paper explored the effects of altitude on mental workload. The results show that airport tower controllers had more mental workload as height increased when the task complexity was the same. Task complexity had a similar positive effect on mental workload as well.

As for the safety risk, when eliminating mental workload disturbance, operation risk increased as the elevation rising with the task complexity was three. While when the task complexity was low, the altitude had no significant effects on the human error rate, and the operation risk kept at a low level. When considering the mental workload diversity at multiple altitudes, the risk level distinction among different altitudes was more prominent, and the operation risk still expanded with altitude when the task complexity was high. Compared with the workload, occupational risk brought by fatigue presented a slighter difference among the three altitudes, almost staying at the same risk level. Looking at the operational safety risk: the operation risk can stay acceptable, except when the task complexity was high at 2243 m and 3569.7 m. Focusing on the occupational safety risk: the occupational risks remained below the high-risk level in the three situations assumed according to the actual working sceneries. It indicates that the current work schedules for tower controllers are acceptable from the occupational safety management scale. Together, the authors shall adjust the aircraft safety operation interval at the airport to 3569.7 m to lower the safety risk and keep a close eye on the airport operations at 2243 m.

TABLE 7. Correlation analysis results between features and fatigue reference values at 2243 m.

No.	Feature name	Pearson correlation coefficient	Pearson p	spearman correlation coefficient	spearman p	Kendall correlation coefficient	Kendall p	Average correlation coefficient
1*	25th percentile of MFCC 4	0.634	0.000	0.687	0.000	0.536	0.000	0.619
2*	25th percentile of MFCC 10	-0.601	0.000	-0.619	0.000	-0.489	0.000	-0.570
3*	Minimum of MFCC 11	-0.604	0.000	-0.622	0.000	-0.474	0.001	-0.567
4*	Standard deviation of MFCC 3	0.625	0.000	0.547	0.002	0.422	0.003	0.531
5*	25th percentile of MFCC 6	-0.550	0.002	-0.570	0.001	-0.427	0.002	-0.516
6	Standard deviation of delta MFCC 6	0.494	0.005	0.515	0.004	0.375	0.007	0.462
7*	Median of MFCC 5	-0.473	0.008	-0.468	0.009	-0.422	0.003	-0.454
8	Standard deviation of delta MFCC 3	0.546	0.002	0.453	0.012	0.360	0.010	0.453
9*	Median of MFCC 9	-0.380	0.038	-0.371	0.043	-0.308	0.028	-0.353

At the level of 0.05 (two-tailed test), the correlation between the feature and the target is significant; The feature with * means that the feature is static; otherwise, it is dynamic.

TABLE 8. Correlation analysis results between features and fatigue reference values at 3569.7 m.

No.	Feature Name	Pearson Correlation coefficient	Pearson p	spearman correlation coefficient	spearman p	Kendall correlation coefficient	Kendall p	Average correlation coefficient
1*	Mean of MFCC 13	0.464	0.000	0.378	0.002	0.294	0.002	0.379
2*	Minimum of MFCC 5	-0.471	0.000	-0.376	0.002	-0.284	0.003	-0.377
3	Minimum of delta-delta MFCC 5	-0.404	0.001	-0.341	0.006	-0.260	0.006	-0.335
4	Minimum of delta MFCC 5	-0.410	0.001	-0.269	0.032	-0.190	0.043	-0.290
5*	Median of MFCC 2	0.304	0.015	0.281	0.025	0.216	0.022	0.267

At the level of 0.05 (two-tailed test), the correlation between the feature and the target is significant; The feature with * means that the feature is static; otherwise, it is dynamic.

TABLE 9. The MANOVA test results.

Factors	Sum of squares	F	p
Task complexity	19.339	75.198	0.00
Altitude	7.704	29.956	0.00
Task complexity * Altitude	0.200	0.778	0.38

TABLE 10. Test of parallel lines.

Model	-2Log Likelihood	Chi-Square	p
Null Hypothesis	282.017		
General	0	282.017	0.834

The null hypothesis states that the location parameters are the same across response categories.

D. IMPLICATIONS

The implications of the above findings are multi-faceted. First, present researchers on the influence of hypobaric and hypoxia environment concentrate on military pilots or acute mild hypoxic hypoxia [4], [6]. Whereas the authors stated in the INTRODUCTION section, the hypobaric and hypoxia

TABLE 11. Goodness-of-fit.

	Chi-Square	p
Pearson	348.479	0.05
Deviation	202.229	1.00

TABLE 12. Levene's test results of variance homogeneity.

	Levene Statistic	Significance
Based on the mean	2.139	0.121
Based on the median	1.662	0.193
Based on median and with adjusted data of freedom	1.662	0.193
Based on trimmed mean	2.175	0.117

TABLE 13. ANOVA analysis results.

	Sum of Squares	Mean Square	F	Significance
Between Groups	523.041	261.521	11.863	0.000
Within Groups	3395.046	22.046		
Total	3918.087			

TABLE 14. Operation risk without the interference of mental workload.

Altitude	Task complexity	HER	Risk level	Position
All altitudes	1	$< 10^{-3}$	Low	5
0 m	3	1.73×10^{-3}	Low	4
2243 m	3	2.32×10^{-2}	Medium	3
3569.7 m	3	8.79×10^{-2}	Medium	3

TABLE 15. Operation risk with the interference of mental workload.

Altitude	Task complexity	HER	Risk level	Position
All altitudes	1	$< 10^{-3}$	Low	5
0 m	3	1.73×10^{-3}	Low	4
2243 m	3	4.29×10^{-2}	Medium	3
3569.7 m	3	1.02×10^{-1}	High	2

TABLE 16. K-Fold cross-validation results.

Group	1	2	3	4	5
Accuracy (%)	54.545	66.667	76.190	71.428	80.952

TABLE 17. State simulation results.

Situation	Predicted state	Risk level
1	Awake-mild fatigue	Low-Medium
2	Awake-mild fatigue	Low-Medium
3	Awake-mild fatigue	Low-Medium

environment can impair tower controllers' performance, but no study currently researches this topic. The findings of this study can provide a theoretical basis for future research in the impact of consistent hypobaric and hypoxia environments on operator performance. In practice, these findings can provide a guide for managers at the plateau and high-plateau airports to manage operational and occupational safety.

Second, the findings of the excellent performance of static MFCCs and Log-Mel images in fatigue detection in the plateau and high-plateau environment hint at the extended application range of voice features in fatigue detection. Also, the feature selection results can offer reference to those who want to employ voice features to predict the human state under hypobaric and hypoxia environments.

Another point is that the authors modified the SSS scale, and the results showed that this modification was statistically reasonable. The statistical reasonability implies that this study extended the SSS scale application from the plain to the plateau and high-plateau airport environment.

TABLE 18. The combined risks.

Altitude	Task complexity	Situation	Risk level
All altitudes	1	1	Low-Medium
		2	Low-Medium
		3	Low-Medium
0 m	3	1	Low-Medium
		2	Low-Medium
		3	Low-Medium
2243 m	3	1	Medium
		2	Medium
		3	Medium
3569.7 m	3	1	High
		2	High
		3	High

Finally, we raised a conceptual safety management model involving both operational risk and occupational risk. This combined model incorporates ergonomics in safety management and makes risk management more comprehensive from risk assessment.

VI. CONCLUSION AND FUTURE WORK

In the study, the writers investigated how the two main contributing factors, namely fatigue and mental workload, influence tower controllers' performance in the context of the plateau and high-plateau airports. The authors conducted cognition experiments at three airport elevations: 0 m, 2243 m, 3569.7 m for data collection and safety risk modeling to facilitate operator performance comparisons. The effect test results showed a significant effect on the mental workload related to task complexity, whereas a non-significant effect on the relationship between sleep and fatigue. An increasing airport altitude increased the mental workload under the same task complexity. These findings shed light on the management of the human factor at the plateau and high-plateau airports. Airport managers at higher altitudes shall pay more attention to tower controllers' mental workload to ensure operational and occupational safety.

Besides, the authors employed MFCCs, and an image quality measure based on Log-Mel spectrograms to discuss fatigue detection applications in different environments. Both voice features showed robustness in fatigue detection at different airport elevations. Whereas dynamic MFCCs are more suitable for fatigue detection at low altitudes, while at high elevations, static MFCCs have more apparent strength. In addition, in comparison, Log-Mel spectrograms were relatively stable at high SSS values at the same height. These findings hint at practical application values of fatigue detection in real work. Together, airports at high elevations shall employ Log-Mel spectrograms or static MFCCs to detect operator fatigue.

Furthermore, with a conceptual safety management model, this study found that operation risk and the combined risk gained as the elevation increasing with the high task complexity, and under low task complexity, altitudes had limited influence on risk. These findings strongly suggested more attention to the plateau and high-plateau airport safety management, especially when the airport operations are busy.

In addition, the new modified SSS scale was statistically reasonable, extending the SSS scale application from the plain to the plateau and high-plateau airport environment.

Suggestions for future improvement are multifold. Firstly, the authors used a small sample in our experiments to keep the participants the same at different experimental airports to lower the individual cognition ability differences. That resulted that the authors could not generalize our results to a larger population. However, all participants conducted all the tests at both the plain airports and plateau airports, lowering the disturbance of individual differences. Also, for the bad natural environment and working conditions at the plateau and high-plateau airports, the turnover rate of control staff is high. So, the age range of controllers is relatively low in real life, consistent with the participant age range in the study. Besides, before selecting participants, the authors used sphygmomanometers, blood oxygen instruments, and other equipment to collect basic parameters to ensure no noticeable difference in their healthy level to rule out the impacts from healthy problems. Researchers can employ a hypoxic cabin to simulate the plateau and high-plateau airport environment in future work to hire more participants in the experiment.

The second limitation, or the direction for future work, is about altitude effects on fatigue. For one, the experiment site selection, by far, the airports where the authors conducted our experiments were all below 4000 m. Continuing our experiments in higher altitude airports is informative in recovering more underlying effects of altitude on human fatigue. For the other, the authors did not study the effects of altitude on sleep itself, which may significantly affect the human state in the daytime, and further affects human performance.

Another such area is the interactions between mental workload and fatigue. The authors kept participants at a similar positive cognition state for the limited experiment duration in airport towers to reduce experimental groups. Since that, the current study cannot further explore the effects at multiple altitudes, despite that mental workload impacts human fatigue [13].

APPENDIX

See Table 5–18.

REFERENCES

[1] S. Wilke, A. Majumdar, and W. Y. Ochieng, "Airport surface operations: A holistic framework for operations modeling and risk management," *Saf. Sci.*, vol. 63, pp. 18–33, Mar. 2014.

[2] D. M. Benitez, A. D. C. Valiente, and P. Lanzi, "A novel global operational concept in cockpits under peak workload situations," *Saf. Sci.*, vol. 102, pp. 38–50, Feb. 2018.

[3] H. Kharoufah, J. Murray, G. Baxter, and G. Wild, "A review of human factors causations in commercial air transport accidents and incidents: From to 2000–2016," *Prog. Aerosp. Sci.*, vol. 99, pp. 1–13, May 2018.

[4] F. Bouak, O. Vartanian, K. Hofer, and B. Cheung, "Acute mild hypoxic hypoxia effects on cognitive and simulated aircraft pilot performance," *Aerosp. Med. Hum. Perform.*, vol. 89, no. 6, pp. 526–535, Jun. 2018.

[5] S. J. Legg, A. Gilbey, S. Hill, A. Raman, A. Dubray, G. Iremonger, and T. Mündel, "Effects of mild hypoxia in aviation on mood and complex cognition," *Appl. Ergonom.*, vol. 53, pp. 357–363, Mar. 2016.

[6] C. Malle, P. Quinette, M. Laisney, C. Bourrilhon, J. Boissin, B. Desgranges, F. Eustache, and C. Piérard, "Working memory impairment in pilots exposed to acute hypobaric hypoxia," *Aviation, Space, Environ. Med.*, vol. 84, no. 8, pp. 773–779, Aug. 2013.

[7] W. Zongmin, Z. Damin, W. Xiaoru, W. Hengyang, and Z. Yingwei, "Prediction and analysis of the human errors in the aircraft cockpit," presented at the 5th Int. Conf. Biomed. Eng. Informat. (BMEI), Chongqing, China, Oct. 2012.

[8] P. Dode, M. Greig, S. Zolfaghari, and W. P. Neumann, "Integrating human factors into discrete event simulation: A proactive approach to simultaneously design for system performance and employees' well being," *Int. J. Prod. Res.*, vol. 54, no. 10, pp. 3105–3117, May 2016.

[9] K. M. A. Revell, C. Allison, R. Sears, and N. A. Stanton, "Modelling distributed crewing in commercial aircraft with STAMP for a rapid decompression hazard," *Ergonomics*, vol. 62, no. 2, pp. 156–170, Feb. 2019.

[10] R. Sun, Y. Zhang, Z. Liu, and K. Li, "Task-load evaluation method for maintenance personnel based on the JACK simulation," presented at the 15th Int. Conf. Eng. Psychol. Cogn. Ergonom., vol. 10906, Las Vegas, NV, USA, Jul. 2018.

[11] 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. 67, pp. 135–144, Sep. 2018.

[12] *Guiding Opinions on High Altitude Operations of Civil Aviation in Southwest China (Trial)*, Civil Aviation Admin. China Southwest Regional Admin., China, 2020.

[13] P. Arico, G. Borghini, G. Di Flumeri, S. Bonelli, A. Golfetti, I. Graziani, S. Pozzi, J.-P. Imbert, G. Granger, R. Benhacene, D. Schaefer, and F. Babiloni, "Human factors and neurophysiological metrics in air traffic control: A critical review," *IEEE Rev. Biomed. Eng.*, vol. 10, pp. 250–263, 2017.

[14] N. Pongsakornsathien, Y. Lim, A. Gardi, S. Hilton, L. Planke, R. Sabatini, T. Kistan, and N. Ezer, "Sensor networks for aerospace human-machine systems," *Sensors*, vol. 19, no. 16, p. 3465, Aug. 2019.

[15] C. A. D. Vasconcelos, M. N. Vieira, G. Kecklund, and H. C. Yehia, "Speech analysis for fatigue and sleepiness detection of a pilot," *Aerosp. Med. Hum. Perform.*, vol. 90, no. 4, pp. 415–418, Apr. 2019.

[16] Z. Shen, G. Pan, and Y. Yan, "A high-precision fatigue detecting method for air traffic controllers based on revised fractal dimension feature," *Math. Problems Eng.*, vol. 2020, Feb. 2020, Art. no. 4563962.

[17] Y. Lin, "Spoken instruction understanding in air traffic control: Challenge, technique, and application," *Aerospace*, vol. 8, no. 3, p. 22, Mar. 2021.

[18] A. Srinivasamurthy, P. Motlicek, M. Singh, Y. Oualil, M. Kleinert, H. Ehr, and H. Helmke, "Iterative learning of speech recognition models for air traffic control," in *Proc. 19th Annu. Conf. Int. Speech Commun. Assoc. (Interspeech)*, Sep. 2018, pp. 3519–3523.

[19] X. Zhang, P. Bai, X. Wang, and Y. Zhao, "Effect of fatigue and nervousness of tower controller on the control efficiency," in *Engineering Psychology and Cognitive Ergonomics*. Cham, Switzerland: Springer, 2018, pp. 100–112.

[20] W. Li, P. Kearney, J. Zhang, Y. Hsu, and G. Braithwaite, "The analysis of occurrences associated with air traffic volume and air traffic controllers' alertness for fatigue risk management," *Risk Anal.*, vol. 41, no. 6, pp. 1004–1018, 2021.

[21] C. A. Mas, A. Gallo, A. Ibarra, and J. C. S. García, "The relation between work stress and burnout syndrome in a sample of Chilean air traffic controllers," *Ciencias Psicológicas*, vol. 12, no. 2, pp. 239–248, Nov. 2018.

[22] G. Borghini, V. Ronca, A. Vozzi, P. Aricò, G. D. Flumeri, and F. Babiloni, "Monitoring performance of professional and occupational operators," in *Handbook of Clinical Neurology*, vol. 168. Amsterdam, The Netherlands: Elsevier, 2020, pp. 199–205.

[23] Z. L. Howard, N. J. Evans, R. J. Innes, S. D. Brown, and A. Eidels, "How is multi-tasking different from increased difficulty?" *Psychonomic Bull. Rev.*, vol. 27, no. 5, pp. 937–951, Oct. 2020.

[24] A. Drouin-Picaro, I. Albuquerque, J.-F. Gagnon, D. Lafond, and T. H. Falk, "EEG coupling features: Towards mental workload measurement based on wearables," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Banff, AB, Canada, Oct. 2017, pp. 28–33.

[25] S. R. Wilson, R. D. Leonard, D. J. Edwards, K. A. Swieringa, and J. L. Murdoch, "Model specification and confidence intervals for voice communication," *Qual. Eng.*, vol. 27, no. 4, pp. 402–415, Oct. 2015.

[26] *Radiotelephony Communications for Air Traffic Services*, Standard MH/T 4014-2003, 2003.

[27] J. Krajewski, S. Schnieder, D. Sommer, A. Batliner, and B. Schuller, "Applying multiple classifiers and non-linear dynamics features for detecting sleepiness from speech," *Neurocomputing*, vol. 84, pp. 65–75, May 2012.

[28] J. Whitmore and S. Fisher, "Speech during sustained operations," *Speech Commun.*, vol. 20, nos. 1–2, pp. 55–70, Nov. 1996.

[29] M. Sahidullah and G. Saha, "Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition," *Speech Commun.*, vol. 54, no. 4, pp. 543–565, May 2012.

[30] S. C. Satapathy, N. Sri Madhava Raja, V. Rajinikanth, A. S. Ashour, and N. Dey, "Multi-level image thresholding using Otsu and chaotic bat algorithm," *Neural Comput. Appl.*, vol. 29, no. 12, pp. 1285–1307, Jun. 2018.

[31] B. N. Suhas, J. Mallela, A. Illa, B. K. Yamini, N. Atchayaram, R. Yadav, D. Gope, and P. K. Ghosh, "Speech task based automatic classification of ALS and Parkinson's disease and their severity using log Mel spectrograms," presented at the Int. Conf. Signal Process. Commun., Bangalore, India, Jul. 2020.

[32] Z. Duan, J. Xu, H. Ru, and M. Li, "Classification of driving fatigue in high-altitude areas," *Sustainability*, vol. 11, no. 3, p. 817, Feb. 2019.

[33] P. A. Desmond, M. C. Neubauer, G. Matthews, and P. A. Hancock, "Introduction to operator fatigue," in *The Handbook of Operator Fatigue*, 1st ed., P. A. Desmond, M. C. Neubauer, G. Matthews, and P. A. Hancock, Eds. Farnham, U.K.: Ashgate, 2012, pp. 3–62.

[34] J. Nixon and R. Charles, "Understanding the human performance envelope using electrophysiological measures from wearable technology," *Cogn., Technol. Work*, vol. 19, no. 4, pp. 655–666, Nov. 2017.

[35] J. T. Townsend and A. Eidels, "Workload capacity spaces: A unified methodology for response time measures of efficiency as workload is varied," *Psychonomic Bull. Rev.*, vol. 18, no. 4, pp. 659–681, 2011.

[36] A. Heathcote, J. R. Coleman, A. Eidels, J. M. Watson, J. Houpt, and D. L. Strayer, "Working memory's workload capacity," *Memory Cogn.*, vol. 43, no. 7, pp. 973–989, 2015.

[37] B. J. Baars, "Mind and brain," in *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*, 2nd ed., B. J. Baars and N. M. Gage, Eds. New York, NY, USA: Academic, 2010, pp. 3–32.

[38] B. J. Baars, "A framework," in *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*, 2nd ed., B. J. Baars and N. M. Gage, Eds. New York, NY, USA: Academic, 2010, pp. 33–62.

[39] C. Christodoulou, "Approaches to the measurement of fatigue," in *The Handbook of Operator Fatigue*, P. A. Desmond, M. C. Neubauer, G. Matthews, and P. A. Hancock, Eds. Farnham, U.K.: Ashgate, 2012, pp. 125–138.

[40] Y. Shang-Wen and H. Ming-Hua, "Estimation of air traffic longitudinal conflict probability based on the reaction time of controllers," *Saf. Sci.*, vol. 48, no. 7, pp. 926–930, Aug. 2010.

[41] H.-Y. Lai, C.-H. Chen, P. Zheng, and L. P. Khoo, "Investigating the evolving context of an unstable approach in aviation from mental model disconnects with an agent-based model," *Rel. Eng. Syst. Saf.*, vol. 193, Jan. 2020, Art. no. 106657.

[42] I. Iervolino, D. Accardo, A. E. Tirri, G. Pio, and E. Salzano, "Quantitative risk analysis for the Amerigo Vespucci (Florence, Italy) airport including domino effects," *Saf. Sci.*, vol. 113, pp. 472–489, Mar. 2019.

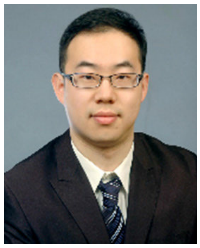
[43] M. Grozdanovic and B. Bijelić, "Impact of human, workplace and indoor environmental risk factors on operator's reliability in control rooms," *Hum. Ecol. Risk Assessment, Int. J.*, vol. 26, no. 1, pp. 177–189, Jan. 2020.

[44] J. L. Sánchez-Bote, J. González-Rodriguez, and J. Ortega-García, "A new approach to dereverberation and noise reduction with microphone arrays," presented at the 10th Eur. Signal Process. Conf., Tampere, Finland, 2000.

[45] I. D. S. Miranda, A. H. Diacon, and T. R. Niesler, "A comparative study of features for acoustic cough detection using deep architectures," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Berlin, Germany, Jul. 2019, pp. 2601–2605.
[46] M. Seo and M. Kim, "Fusing visual attention CNN and bag of visual words for cross-corpus speech emotion recognition," *Sensors*, vol. 20, no. 19, p. 5559, Sep. 2020.
[47] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[48] B. H. Aithal, S. K. Das, and P. P. Subrahmanya, "Urban 3D structure reconstruction through a generative adversarial network model," *Arabian J. Sci. Eng.*, vol. 45, no. 12, pp. 10731–10741, Dec. 2020.
[49] S. Li, "Improvement of deep learning method for single image super-resolution," M.S. thesis, School Aerosp. Eng., Xiamen Univ., Xiamen, China, 2018.
[50] D. Kelly and M. Efthymiou, "An analysis of human factors in fifty controlled flight into terrain aviation accidents from 2007 to 2017," *J. Saf. Res.*, vol. 69, pp. 155–165, Jun. 2019.
[51] T. G. Raslear, S. R. Hursh, and H. P. A. Van Dongen, "Predicting cognitive impairment and accident risk," in *Human Sleep and Cognition* (Progress in Brain Research), vol. 190, H. P. A. Van Dongen and G. A. Kerkhof, Eds. Amsterdam, The Netherlands: Elsevier, 2011, pp. 155–167.
[52] H. T. Peng, F. Bouak, W. Wang, R. Chow, and O. Vartanian, "An improved model to predict performance under mental fatigue," *Ergonomics*, vol. 61, no. 7, pp. 988–1003, Jul. 2018.
[53] Á. Martínez-Sánchez, M.-J. Vela-Jimenez, S. Abella-Garces, and S. Gorgemans, "Flexibility and innovation: Moderator effects of cooperation and dynamism," *Personnel Rev.*, vol. 48, no. 6, pp. 1548–1564, Sep. 2019.
[54] M. A. Kryskow, B. A. Beidleman, C. S. Fulco, and S. R. Muza, "Performance during simple and complex military psychomotor tasks at various altitudes," *Aviation, Space, Environ. Med.*, vol. 84, no. 11, pp. 1147–1152, Nov. 2013.
[55] R. Gupta, J. Ulfberg, R. P. Allen, and D. Goel, "Comparison of subjective sleep quality of long-term residents at low and high altitudes: SARAHA study," *J. Clin. Sleep Med.*, vol. 14, no. 1, pp. 15–21, Jan. 2018.
[56] J. G. d'Engelbronner, C. Borst, J. Ellerbroek, M. M. van Paassen, and M. Mulder, "Solution-space–based analysis of dynamic air traffic controller workload," *J. Aircr.*, vol. 52, no. 4, pp. 1146–1160, Jul. 2015.

QUAN SHAO was born in Ezhou, China, in 1981. He received the B.S. and M.S. degrees from the University of Science and Technology of China and the Ph.D. degree from the Institute of Public Safety, Tsinghua University, Beijing, China, in 2009. He is currently an Associate Professor with the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics. His research interests include civil aviation safety management, risk assessment, and emergency decision-making for emergencies.

MIN DONG was born in Yancheng, China, in 1996. She received the B.S. degree in airport management from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2019, and the B.S. degree in applied science of aviation from the Royal Melbourne Institute of Technology University, Melbourne, VIC, Australia, in 2019. She is currently pursuing the M.S. degree in transportation with the Nanjing University of Aeronautics and Astronautics. Her current research interests include human factors and safety management of civil aviation.



ZHIYUAN SHEN (Member, IEEE) was born in Nanjing, China, in 1985. He received the B.S., M.S., and Ph.D. degrees from the Harbin Institute of Technology University. He is currently a Tutor with the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics. His research interests include airport operations and safety management, human factors in civil aviation, and air traffic control monitoring and information processing.

RUI YANG was born in Qinghai, China, in 1976. He received the B.S and Ph.D. degrees from the University of Science and Technology of China. He is currently an Associate Professor with the Department of Engineering Physics, Tsinghua University. His current research interests include aviation safety and numerical simulation of hazards.

HONG WANG was born in Hangzhou, China, in 1996. She received the B.S. degree in airport management from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2019, where she is currently pursuing the M.S. degree in transportation. Her current research interests include aircraft delay and emotional evolution.