Additive Sensory Noise Effects on Operator Performance in a Lunar Landing Simulation

Sage O. Sherman; Young-Young Shen; Daniel Gutierrez-Mendoza; Michael Schlittenhart; Cody Watson; Torin K. Clark; Allison P. Anderson

INTRODUCTION: Adding noise to a system to improve a weak signal's detectability is known as stochastic resonance (SR). SR has been shown to improve sensory perception and cognitive performance in certain individuals, but it is unknown whether this performance improvement can translate to meaningful macrocognitive enhancements in performance for complex, operational tasks.

OBJECTIVE: We investigated human operator performance in a lunar landing simulation while applying auditory white noise and/or noisy galvanic vestibular stimulation.

METHODS: We measured performance ($N = 16$ subjects) while completing simulation trials in our Aerospace Research Simulator. Trials were completed with and without the influence of auditory white noise, noisy galvanic vestibular stimulation, and both simultaneously in a multimodal fashion. Performance was observed holistically and across subdimensions of the task, which included flight skill and perception. Subjective mental workload was collected after completing four trials in each treatment.

RESULTS: We did not find broad operator improvement under the influence of noise, but a significant interaction was identified between subject and noise treatment, indicating that some subjects were impacted by additive noise. We also found significant interactions between subject and noise treatment in performance subdimensions of flight skill and perception. We found no significant main effects on mental workload.

CONCLUSIONS: This study investigated the utility of using additive sensory noise to induce SR for complex tasks. While SR has been shown to improve aspects of performance, our results suggest additive noise does not yield operational performance changes for a broad population, but specific individuals may be affected.

KEYWORDS: stochastic resonance, auditory white noise, noisy galvanic vestibular stimulation, aerospace research simulator.

Sherman SO, Shen Y-Y, Gutierrez-Mendoza D, Schlittenhart M, Watson C, Clark TK, Anderson AP. *Additive sensory noise effects on operator performance in a lunar landing simulation*. Aerosp Med Hum Perform. 2023; 94(10):770–779.

Spaceflight frequently requires astronauts to complete a variety of complex tasks across a mission's duration, necessitating peak human performance. However, living in the spaceflight environment for an extended period of time poses physiological and psychological hazards that impact crew mental health and, thus, human performance.[10] Long-duration deep space missions will also lead to greater morphological and radiative destructive changes in the central nervous system, which may lead to large cognitive and behavioral declines.[14] As such, NASA's Human Factors and Behavioral Performance group identified that on a long duration deep space or planetary mission the risk of adverse cognitive or behavioral conditions on operations requires mitigation.[18] Thus, there stands a need to develop safe, effective, and standalone countermeasures that the crew could use to offset these human performance decrements when performing spaceflight mission tasks. One such countermeasure could leverage the mechanism of stochastic resonance (SR).

From the Ann & H. J. Smead Department of Aerospace Engineering Sciences, The University of Colorado Boulder, Boulder, CO, United States.

This manuscript was received for review in February 2023. It was accepted for publication in July 2023.

Address correspondence to: Sage Sherman, Ann & H. J. Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder, Boulder, CO 80309, United States; sage.sherman@colorado.edu.

DOI: https://doi.org/10.3357/AMHP.6251.2023

SR is a phenomenon where additive noise can improve the detectability of a signal in nonlinear systems.[9] Human experimentation shows that SR can improve perception performance within and across sensory modalities.[2,9,21] While SR has traditionally been explored for perception purposes, Hidaka et al. have shown that noise enhanced sensory information could be used within the central nervous system, suggesting that SR may also affect higher order information processing.[4] This implies that SR could enable a safer, noninvasive form of neuromodulation as it involves stimulating sensory systems instead of directly stimulating the brain, as is the case with alternative neuromodulation techniques, such as transcranial direct current stimulation (tDCS). This notion is supported in the literature through a few human subject experiments that improved cognitive task performance by inducing single modality SR (i.e., applying noise to only one sensory system). Background auditory white noise (AWN) (~78 dB SPL) improved verbal recall, visuo-spatial working memory, and motor response in inattentive school children.[3,19] Noisy galvanic vestibular stimulation (nGVS) was found to reduce visual memory recall speed in healthy adults.[24] In an evaluation of comprehensive cognition, SR cognitive performance enhancement was not found in the broad population, but it was present in subjects who preferred working in noisy environments.[17] These studies, though, focus on specific and separate microcognitive functions, such as working memory, failing to provide insight on the utility of SR in real-world contexts.

Macrocognition refers to cognitive functioning in natural environments.[6] Completing complex mental tasks involves synthesizing many microcognitive skills to execute a relevant task, such as driving a car or landing a spacecraft. To our knowledge, the benefits of SR have not been assessed for macrocognition, as it is potentially difficult to analyze the results of complex tasks with a great degree of sensitivity. Usher and Feingold found SR improved speed of memory retrieval for multiplication.[20] Multiplication may be more analytical and complex than other cognitive domain assessment tasks, but this task is not operationally relevant and a weak indicator for overall performance enhancement. Beyond SR, the literature has explored macrocognitive benefits for traditional forms of neuromodulation. Choe investigated tDCS effects for an nBack task and a flight simulator, finding improved performance and learning in both tasks, implying benefits in the microcognitive and macrocognitive domains.[1] Further, Scheldrup et al. found improvements in multitasking while utilizing tDCS, suggesting improved macrocognitive performance.[16] In addition to direct performance enhancement, tDCS has been shown to reduce perceived temporal workload in surgical simulations.[23] High levels of mental workload can lead to stress and performance decrements in operators[22]; thus, neuromodulation techniques that influence mental workload may indirectly impact operator performance. Given the success of other neuromodulation techniques to improve operationally relevant performance, the absence of research on SR macrocognitive effects presents a substantial literature gap that needs to be addressed.

We aimed to fill this gap in the literature by assessing the potential for enhancing operator performance using sensory noise. We hypothesized that single modality noise (either AWN or nGVS) would enhance performance in human subjects when compared to performance without noise (sham); additionally, we hypothesized that stimulating both modalities simultaneously to induce multimodal SR (MMSR) would have additive benefits and enhance performance to a greater degree than using a single modality alone. To assess this, subjects performed a series of lunar lander simulation tasks under sensory noise and a no noise sham. This task loads several perceptual, cognitive, and motor coordination domains, such that the results provide comprehensive insight into the influence of noise for operational tasks.

Additionally, SR perception studies imply that perception performance enhancement is greater for at-threshold perceptual stimuli, but suprathreshold enhancement is possible.[8,15] Thus, we believed that “at-threshold” operational enhancement could be a factor in noise benefit effectiveness. We hypothesized that the extent of SR performance enhancement may vary as a result of task challenge. Our task design allowed us to modulate task difficulty and assess whether improvements are related to task difficulty.

Finally, SR studies have suggested that some individuals are susceptible to SR improvements, while others are not.[2,12] This suggests that performance enhancement may be seen in some subjects as a result of noise but not others. Thus, we also hypothesized that only some subjects would see SR benefits. Building upon a previous study our lab conducted where we saw individual SR sensitivity within cognitive performance,[17] we evaluated whether there was a positive correlation between operator performance and preference enhancement for working in a noisy environment to help identify individuals who receive benefits from noise.

METHODS

Subjects

A total of 16 subjects (9 women/7 men), ages 29 ± 7 yr (range = 20–41 yr) completed testing in the Bioastronautics Lab at the University of Colorado Boulder. An a priori power analysis based on the results of Scheldrup et al. suggested that we needed 16 subjects for our study design to find an effect size greater than 0.3, as Scheldrup et al. found for tDCS.[16] This research was approved by the University of Colorado-Boulder’s Institutional Review Board (protocol #20-0347) and written informed consent was obtained prior to participation. Subjects were prescreened and excluded if they reported a history of health issues that could impact cognitive abilities, such as severe head trauma or disorders associated with thinking impairment. They were also excluded if they reported health issues that could impact auditory or vestibular processing, such as language impairment or vestibular dysfunction. Additionally, subjects underwent auditory screening to verify healthy and unobstructed ear

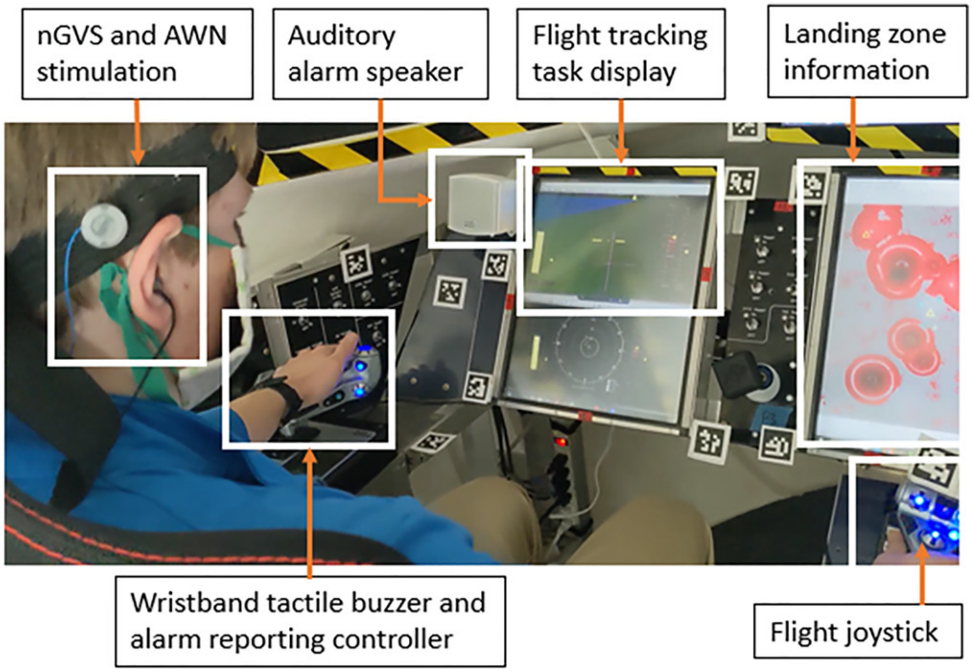

Fig. 1. Over-the-shoulder view of the AReS lunar lander simulation used in this experimental paradigm. Relevant hardware and displays are highlighted in overlaid white boxes. Subjects sit stationary directly in front of the flight tracking task display with their right hand on the flight joystick which they use for the tracking task. To the right of the flight display is a map display indicating lunar topography and possible landing zone locations (i.e., landing zone information). To the left of the flight display is an external auditory speaker that intermittently presents the auditory alarm. A tactile buzzer is attached to the subject's left wrist to present the tactile alarm. The subject rests their left hand on the throttle, which they use to signal when they notice that the auditory or tactile alarm is occurring. nGVS electrodes and AWN earbuds are fixed to the subject's head in all trials, including sham where no noise was administered.

canals (via otoscopy), normal tympanometry, and normal hearing (audiometric thresholds ≤25 dB HL up to 8 kHz).

Equipment and Materials

The task used in this study aimed to be a representative analog for a macrocognitive task that individuals in an operational environment may face. The simulation task was completed using our Aerospace Research Simulator (AReS), shown in **Fig. 1** and **Fig. 2**. AReS is a demonstrated, macrocognitive landing task that incorporates several cognitive processes at once.[11,25] Fig. 1 illustrates the hardware and interface of the AReS fixed-base flight simulation, while the software provides a realistic replication of lunar landing vehicle dynamics, piloting control responses, and fuel consumption. In the AReS lunar lander simulation task, subjects were presented with six landing points scattered across a 2D contour map of the lunar surface. They attempted to choose the optimal landing point, considering its distance with respect to three scientific points of interest (i.e., nearest the centroid) and its potential presence within hazardous areas, such as steep slopes (Fig. 2B). The lander descended at a constant rate, continuously consuming fuel. To navigate to their designated landing zone, subjects were required to complete a tracking task on their primary flight display using a joystick by aligning the spacecraft's pitch and roll attitude (the yellow reticle) to the flight guidance cue (the magenta cue) (Fig. 2A). At a lander altitude of 250 ft (76 m; roughly 40 s into the task), a simulated lidar gave the subject a new topography map which presented additional hazard information not visible initially, such as rock fields where the subject would not be able to land (bottom right panel of Fig. 2B). At this point, the subject could continue to fly toward their original landing zone choice or, by pressing buttons on the joystick, redesignate a new landing site or abort the landing [allowable between 200 and 50 ft (61 and 15 m) of altitude].

Novel to previous work done with AReS,[11,25] we embedded two perception tasks. One was a tactile vibration presented to the wrist and the other was an auditory alarm presented to the cockpit via speakers. Both alarms indicated that a simulated thruster was stuck and consuming more fuel than usual. The magnitude of these perception alarms were initially low, beginning subthreshold and gradually increasing to a suprathreshold level. Subjects pressed a button on the throttle as soon as they identified either alarm to effect a "reset" that solved their fuel leakage problem. The fuel decreased at a faster rate than usual while the alarm was active to incentivize the subjects to attend to the perception task. Each perception task occurred twice during a trial and occurred at random intervals. The timing of the four perception alarms was randomly assigned to four set times (10, 30, 50, and 80 s into the task) with a random time amount (between 1–10 s) added to each of those four set times. Input was only accepted when the alarms were present; unsolicited presses of the button were not registered.

This task was designed to load the operational subdimensions of flight skill, decision-making, and perception. The task's dependent variables were performance metrics that make up the task subdimensions found in **Table I**. Each metric quantified an aspect of performance that we hypothesized may be sensitive to SR

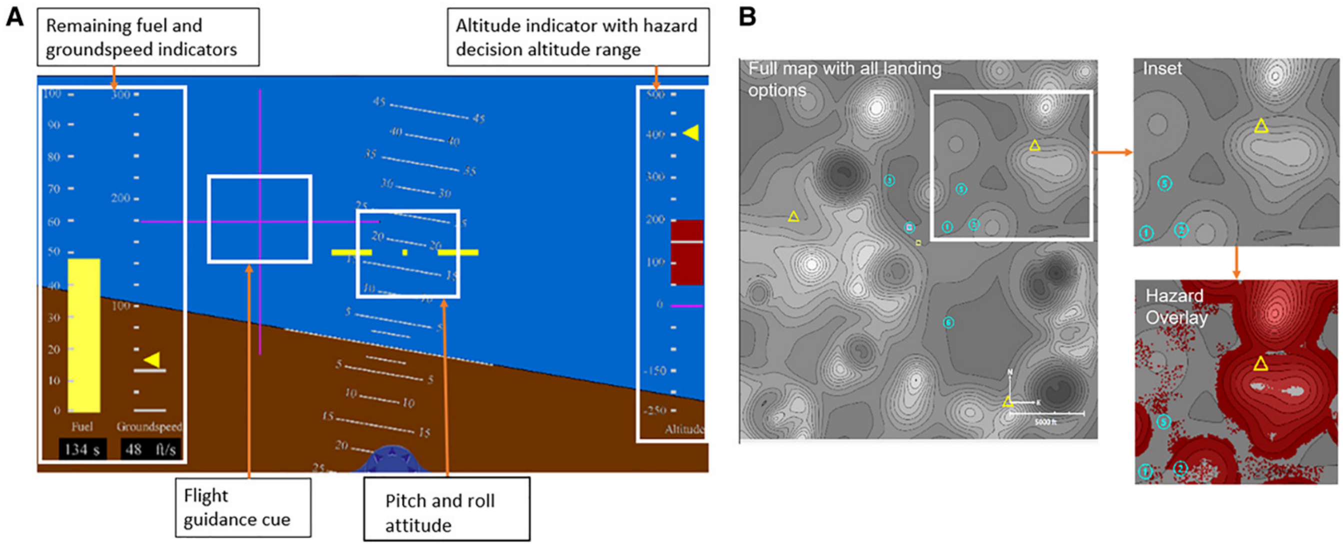

Fig. 2. A) Visual information presented to the subject in the primary flight display. The panel displayed the spacecraft's pitch and roll attitude and altitude (and depiction of the hazard decision range in red), groundspeed, and fuel. The subject tracked the magenta flight guidance cue to align with the spacecraft's pitch and roll attitude as represented by the yellow reticle. B) Topography maps made available to subjects. The left was displayed to the subject at the start of the task. The yellow triangles depicted three scientific points of interest and blue circles were landing zones that subjects chose from. The top right panel is a zoomed-in inset of the map for legibility, the bottom right panel is the same display with hazard data from a simulated lidar sensor overlay that appears once the spacecraft reaches 250 ft (76 m) of altitude.

performance improvement. A description of how each metric relates to performance is also given in Table I. Combining these subdimensions yields a comprehensive performance measure to capture overall operator changes caused by SR.

As hypothesized, we also investigated whether operational enhancement due to SR may be dependent upon task difficulty. For example, performance on an easy task may be insensitive to adding sensory noise. Thus, we tested three levels of task difficulty (easy, medium, or hard) as determined by the layout of hazards, points of scientific interest, and potential landing zones on the landing maps, a description of which is found in **Appendix A** (found online at https://doi.org/10.3357/amhp.6251sd.2023).

The independent variable of this research was the four treatments of sensory noise administered. Broadband AWN (20–20,000 Hz) was administered to subjects through ear buds (Essential Earphones HD; Essential Products, Inc., Palo Alto, CA, USA) and a Samsung Tablet A; the auditory profiles were developed and calibrated by Creare LLC (Hanover, NH, USA).

Table I. A Description of Each Performance Metric in the Lunar Lander Simulation.

SUBDIMENSION & PERFORMANCE METRIC	METRIC DESCRIPTION	METRIC JUSTIFICATION
Flight		
Root mean square distance (RMS; degrees)	The RMS distance error of the yellow reticle from the magenta cue over the simulation duration.	The ability of subjects to track the lander's attitude with the guidance system. Better performance corresponds to a reduction in RMS error.
Joystick input (stick)	The percentage of time the subject spent giving an input to (i.e., deflecting) the joystick during the simulation.	A measure of efficiency, the simulated lander has an attitude hold. If subjects overuse the joystick when it is not necessary, they are spending more fuel.
Smooth flying (smooth)	The number of times the subject crosses over the magenta cue in pitch and roll as they track with the reticle.	A measure of excessive control. A flyer who overshoots the magenta cue spends more fuel correcting for their mistake; better flying results in less overcorrecting.
Decision-Making		
Landing zone (LZ)	A ranked score based on the combination of initial and posthazard display landing zone choices.	Some landing zones are better choices than others in terms of their distance to scientific points and the presence of hazards. Reselecting a better landing zone based upon lidar-updated hazard information was rewarded.
Crash, abort, or land (CAL)	A ranked score based on whether the subject landed, crashed, or aborted when it was or was not possible to land.	Based on their landing zone selections and flight performance, subjects may need to make trade-offs for safety or landing success.
Perception Identification		
Tactile (seconds)	The time it takes for subjects to detect and report each of the two tactile alarms.	A quicker reaction time to press the alarm button results in less fuel loss, suggesting enhanced perceptual performance.
Auditory (seconds)	The time it takes for subjects to detect and report each of the two auditory alarms.	A quicker reaction time to press the alarm button results in less fuel loss, suggesting enhanced perceptual performance.

Broadband, unipolar, zero-mean white noise (0–100,000 Hz) was bilaterally administered to subject mastoids through the Galvanic Vestibular Oscillating Stimulator (model 0810, Soterix Medical, Woodbridge, NJ, USA) using electrodes with a contact area of 2 square cm.[21] The third sensory noise treatment consisted of using both AWN and nGVS administered simultaneously in a multimodal fashion. A sham treatment where no sensory noise was administered, but with electrodes and earbuds applied, served as the baseline.

Procedure

A within-subject experimental design was implemented. After enrollment, subjects watched a 15-min tutorial video to orient them to the lunar landing task. They then completed a minimum of nine practice trials of the task, or until they felt comfortable with the controls, displays, and goals. This was done to ensure they had fully learned how to operate the simulation and understood all dimensions of the task. Further, a test operator assessed the subject's basic competency level with the task before proceeding.

On a separate test day (within 1 wk of their initial visit), subjects completed 34 trials of the task. Each trial contained a unique map with differing terrain (and thus hazards) and landing points from the other trials. There were two phases to the experimental trials on the test day. The first phase identified the subject-specific optimal noise levels in AWN and nGVS for testing in the second phase, as will be described in the next paragraph. In the second phase, we investigated our main hypotheses for task performance and subjective workload using our four sensory noise treatments (AWN, nGVS, MMSR, and sham).

There is an optimal level of noise to induce SR that depends on subject, task, and sensory system.[9] This has been demonstrated in studies evaluating noise enhancement of sensory perception within and across modalities.[2,12,21] We believed this would be the case for cognitive performance enhancement; thus, an initial suite of three nGVS levels (0.2, 0.5, and 0.8 mA) and three AWN levels (40, 55, and 70 dB SPL) were tested in a randomized order, as has been done in our prior work.[17] Subjects completed three trials for each level, resulting in 18 total trials in this first phase to identify the subject-specific best noise level. Raw performance in each metric was fractionally ranked across the 18 trials and assessed. In order to identify each subject's best noise level, broad task performance was quantified (Eq. 1) from this initial set of trials. This metric is the sum of each individual metric captured and equally weighted among the three subdimensions of flight performance, decision-making, and perception.

$$P_{Total} = \frac{1}{3} * \left(P_{RMS} + P_{Stick} + P_{Smooth}\right) + \frac{1}{2} * \left(P_{LZ} + P_{CAL}\right) + \frac{1}{2} * \left(P_{Tactile} + P_{Auditory}\right) \qquad \text{Eq. 1}$$

The SR noise level that yielded the best performance described by Eq. 1 was selected as the subject-specific best (experimentally close to optimal) AWN and nGVS level. The performance value calculated in Eq. 1 was not used for any further analysis beyond identifying these best noise levels.

Once the subject-specific best SR levels were obtained in the first phase, subject-specific best level of AWN, nGVS, and MMSR were tested across 16 additional unique trials (4 trials per treatment) in the second phase. Within each treatment, four trials were administered based on the map difficulty (one easy, two medium, and one hard map) in a randomized order. After each treatment was tested, mental workload was captured using a modified Bedford workload scale.[5,13] This allowed us to assess average subjective workload independent of map difficulty. All sensory noise treatments were presented in a randomized order. Data from these 16 trials were retained for analysis.

After completing all trials, subjects completed a subjective five-point Likert scale questionnaire that asked how well they could maintain focus in quiet and noisy environments. Their noisy environment preference score was defined as the difference in their ranking between quiet and noisy environments (i.e., a negative score means the subject prefers working in quiet places and a positive score means they prefer working in noisy places).[17] This survey can be found in **Appendix B** (found online at https://doi.org/10.3357/amhp.6251sd.2023).

Statistical Analysis

A within-subjects analysis was completed to evaluate operator performance differences due to sensory noise treatments. Each of the performance metrics described in Table I have different measurement units, making them difficult to combine into a composite performance score. Thus, ranking was used. For each metric, the raw performance values in each of the 16 trials were fractionally ranked for each subject [e.g., when assessing performance for root mean square (RMS) distance, each of the subject's 16 trials were ordered and ranked from best to worst]. This allowed us to compile ranked data across metrics to assess overall operator performance and per subdimension by isolating the subdimension metrics of flight, decision-making, and perception.

Upon visualization, the subdimension of decision-making yielded substantial violations of normality assumptions for residuals. This is due to the skewed nature of the nominal and ordinal data collected for the decision-making metrics. Specifically, across all subjects, 89% of trials were landed successfully and in 66% of trials subjects identified the optimal landing zone (LZ) in their first selection (this increased to 70% by their second selection). Therefore, we determined that this subdimension was not sensitive enough to observe deviations in performance based upon nonparametric rankings, so data related to this subdimension was removed from our overall operator performance analysis. Thus, we conducted a separate Chi-squared goodness of fit analysis to observe differences between treatments in each separate decision-making subdimension metric, including crash, abort, or land (CAL) and LZ selection. Subdimensions of flight and perception were analyzed as no observable assumption violations were present. This suggests that a parametric statistical analysis for this data was appropriate and retained for overall operator performance analysis.

For overall operator performance analysis, and the subdimensions of flight and perception, a repeated measures analysis of variance (RM ANOVA) was conducted between noise treatments on the fractional ranked values. Fixed main effects included in the model were noise treatment, from which our main hypothesis was investigated, and map difficulty. Additionally, the interaction of noise treatment with subject was included since only some subjects may exhibit performance changes with SR.[17] An interaction between noise treatment and map difficulty was included to test whether the effect of sensory noise on performance was influenced by task difficulty. Assumptions for homogeneity and residual normality were tested to ensure that parametric statistics were appropriate. If the *F*-test results from the RM ANOVAs were significant, Tukey HSD multiple pairwise comparisons were used to identify which treatments were different from one another.

A nonparametric Friedman test was used to assess mental workload from our Bedford scale data, as the data is ordinal in nature. We also applied an RM ANOVA to the Bedford scale data for completeness, using the same factors as described for the fractionally ranked performance data.

Additionally, we aimed to see whether a subject's noisy environment preference indicated operator performance sensitivity to the noise treatments. Within individual metrics, performance was averaged across the four maps, resulting in one value per noise treatment per subject. From there, performance in the sham treatment was subtracted from their performance in each sensory noise treatment. This data resulted in 240 outcomes (16 subjects × 3 baseline-adjusted sensory noise conditions × 5 metrics in Table I = 240 outcomes). Linear regression models were fit to this entire performance dataset against subjects' noisy environment preference scores to identify if subjects with a preference for noisy environments benefited more from SR.

RESULTS

The AWN and nGVS noise levels presented in these results are the subject best noise levels which were derived from pretrial performance evaluation across all metrics (Eq. 1). Visualizations of this pretrial performance data for each subject are given in **Appendix C** (found online at https://doi.org/10.3357/amhp.6251sd.2023). **Table II** displays the RM ANOVA results that correspond to overall operator performance (flight and perception subdimensions combined). Contrary to our hypothesis, no significant differences were found for our noise treatment alone. However, consistent with our hypothesis, a significant interaction between subject and noise treatment was identified. A main effect of map difficulty was also identified for this compiled dataset. A multiple comparisons analysis for the main effects of map difficulty on overall performance found that performance in "easy" maps [mean (M) = 0.10 ± SD 0.26] was significantly better than "medium" maps (M = −0.01 ± 0.23) and "hard" maps (M = −0.05 ± 0.24) and performance in "medium" maps was significantly better than "hard" maps. Contrary to our hypothesis, no significant interaction effects were identified between noise treatment and map difficulty. These results are visualized in **Fig. 3A**.

Table II. RM ANOVA Results for Overall Operator Performance.

FACTOR	*F* (DOF)	*P*-VALUE	η_p^2
Noise Treatment	0.069 (3, 1223)	0.61	0.003
Map Difficulty	35.28 (2, 1223)	<0.005*	0.055
Noise Treatment × Subject	2.13 (45, 1223)	<0.005*	0.073
Noise Treatment × Map Difficulty	1.29 (6, 1223)	0.26	0.006

*Factors that met a statistical significance below 0.05.

Table III displays the RM ANOVA results that correspond to the subdimensions of flight and perception. For the flight subdimension, we found a significant main effect of map difficulty, as well as a significant interaction between noise treatment and subject. A multiple comparisons analysis for the main effects of map difficulty on flight data found that performance in "easy" maps (M = 0.17 ± 0.3) was significantly better than "medium" maps (M = −0.02 ± 0.25) and "hard" maps (M = −0.09 ± 0.25) and performance in "medium" maps was significantly better than "hard" maps. Contrary to our hypothesis, no significant effects were identified for the noise treatments or the interaction of noise treatment and map difficulty. These results are visualized in **Fig. 3B**.

For the perception subdimension, we found a significant main effect of noise treatment and a significant interaction between treatment and subject. A multiple comparisons analysis for the main effects of treatment on the perception data found that performance in the AWN treatment (M = −0.04 ± 0.22) was significantly lower (i.e., worse) than in the sham treatment (M = 0.04 ± 0.17). No other significantly different comparisons were identified. As might be expected, map difficulty had no effect on the perception task. These results are visualized in **Fig. 3C**. Note that no significant interactions between noise treatment and map difficulty were identified for the subdimension (Table III) performance evaluations.

For the subdimension of decision-making, a separate analytical approach was applied. The frequency of the nominal outcomes is presented in **Table IV**. A Chi-squared goodness of fit test was applied to each decision-making metric presented in Table IV. When assessing the CAL metric, due to the low frequency of aborts or crashes, these outcomes had to be combined to meet the assumption for sufficiently sized expected frequencies. Thus, the statistical test was applied to the outcomes of "land" and "not-land". For the CAL metric, the resulting test statistic was $\chi^2(3) = 4.17$. For the landing zone selection metric, the resulting test statistics were $\chi^2(3) = 0.63$ for the choice before hazards were displayed and $\chi^2(3) = 1.84$ for the choice after hazards were displayed. Thus, contrary to our hypothesis, no significant effects were identified between the noise treatments when it came to our decision-making metrics.

A nonparametric Friedman analysis was used to assess the Bedford workload scale data. Contrary to our hypothesis, our results showed no significant main effects of noise treatment [$\chi^2(3) = 4.49$, $P = 0.21$]. For completeness, an RM ANOVA test was also performed since it may have had more power, but the ordinal data technically violated the model's assumptions; however, it yielded the same conclusion [$F(3,45) = 1.45$, $P = 0.24$].

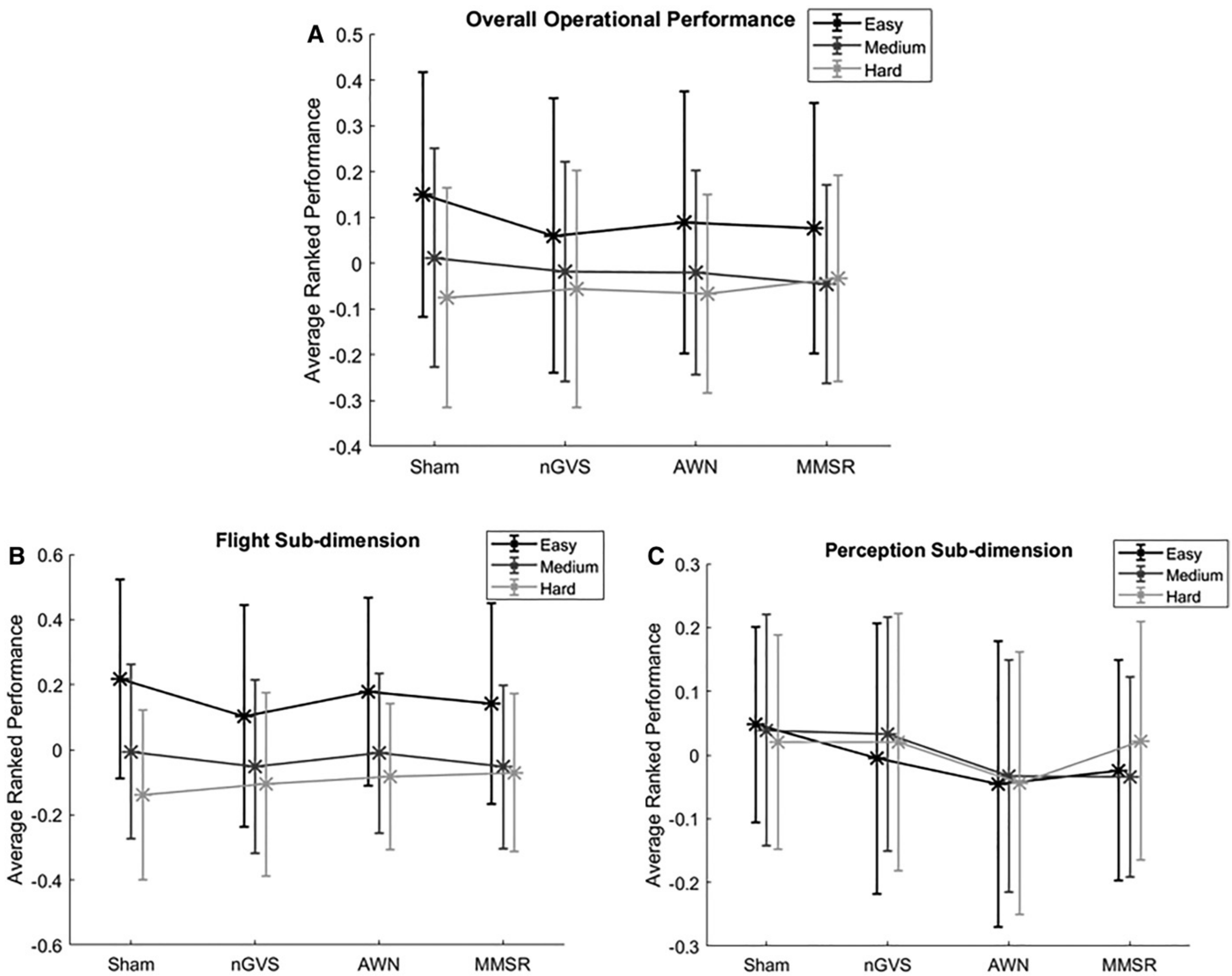

Fig. 3. A) Main effects plot of noise treatments and map difficulty for the overall performance aggregated dataset. The three sensory noise treatments were applied at subject-specific best levels determined in the first phase of testing. Higher ranks correspond to better performance. Error bars represent the standard deviation. B) Main effects plot of noise treatment and map difficulty for the flight subdimension. Error bars represent the standard deviation. C) Main effects plot of noise treatment and map difficulty for the perception subdimension. Error bars represent the standard deviation.

Additionally, a linear regression was fit between subject performance difference for the aggregated dataset across all noise treatments and the subject's noisy environment preference. Contrary to our hypothesis, we did not find a significant correlation between noisy environment preference and operator performance relative to sham (slope = −0.46, $P = 0.57$).

DISCUSSION

This research aimed to understand the utility of using additive sensory noise to improve operator performance. To our knowledge, this is the first assessment of SR for macrocognitive tasks. This was done by having subjects complete a complex lunar landing task, requiring participants to make decisions, actively

Table III. RM ANOVA Results for the Flight and Perception Subdimensions.

	FLIGHT			PERCEPTION		
FACTOR	*F* (DOF)	*P*-VALUE	η_p^2	F (DOF)	*P*-VALUE	η_p^2
Noise Treatment	0.53 (3, 711)	0.67	0.005	2.75 (3455)	0.049*	0.026
Map Difficulty	43.3 (2, 711)	<0.005*	0.130	0.14 (2455)	0.87	0.001
Noise Treatment x Subject	2.49 (45, 711)	<0.005*	0.136	1.62 (45,455)	0.008*	0.138
Noise Treatment x Map Difficulty	0.97 (6, 711)	0.44	0.008	0.58 (6455)	0.75	0.008

*Factors that met a statistical significance below 0.05.
DOF: degrees of freedom.

Table IV. Outcome Frequency Table for the Decision-Making Metrics.

OUTCOME	SHAM	nGVS	AWN	MMSR
Crash – Abort – Land Metric				
Land	53	57	60	58
Abort	5	5	3	6
Crash	6	2	1	0
Optimal landing zone selection (prior to hazard appearance)				
Selects Optimal LZ	43	43	44	40
Fails to select Optimal LZ	21	21	20	24
Optimal landing zone selection (after hazard appearance)				
Selects Optimal LZ	41	47	47	44
Fails to select Optimal LZ	23	17	17	20

nGVS: noisy galvanic vestibular stimulation; AWN: auditory white noise; MMSR: multimodal stochastic resonance; LZ: landing zone.

track moving stimuli, and vigilantly identify perceptual alarms under sensory noise aimed to induce SR.

By observing performance across subdimensions of flight and perception, we intended to identify what attributes of operations that sensory noise may influence. We found no main effect of noise treatment on performance in the flight task, but noise had a significant main effect in the perceptual task; however, based upon our pairwise comparison, we found this significant difference results from AWN masking the auditory alarm, reducing auditory detection relative to the sham treatment. While certain levels of additive AWN are shown to reduce auditory thresholds and enhance perception,[9] these intensity levels are often low, in contrast to higher levels inducing masking behavior.[12] However, the auditory noise levels needed to induce SR across sensory modalities (e.g., in visual perception) and enhance cognitive functions are sufficiently suprathreshold.[7,19] This is a relevant concern when it comes to implementing auditory white noise treatments in operational environments (i.e., the high level of AWN necessary to induce cross-modal SR may produce decrements in auditory perception via masking).

Interestingly though, the interaction results reflect findings in other perceptual and cognitive SR literature. Previous perception studies found that only some individuals exhibit SR benefits where noise can lower perception thresholds.[2,12] This has also been shown in microcognitive task performance. Söderlund et al. reported that AWN improves cognitive performance in inattentive school children, whereas attentive school children did not exhibit benefits from AWN.[19] Previous work that we conducted found that applying AWN or nGVS had no effect on overall cognition for the broad population. However, subjects who self-reported preferring to work in noisy environments received cognitive enhancement from additive sensory noise.[17] Building upon this work, the noisy environment preference questionnaire was included in this study to further investigate this notion. While we found a significant interaction between subject and noise treatment, no correlations were identified between noise preference and performance changes. Our results suggest that individual differences may be a dominant factor in whether SR improves operator performance, but our noise preference questionnaire may not be a useful indicator in this context.

Map difficulty was identified as a main effect in influencing flight skill, but not perception. This may be expected, as map difficulty modifies the simulation's optimal flight pattern and trajectory without changing aspects of the perception task (auditory and tactile detection response times). We hypothesized that there may be an interaction between sensory noise treatment and map difficulty, as SR effects might be more pronounced at certain levels of task difficulty. Our results did not find a significant interaction between treatment and difficulty

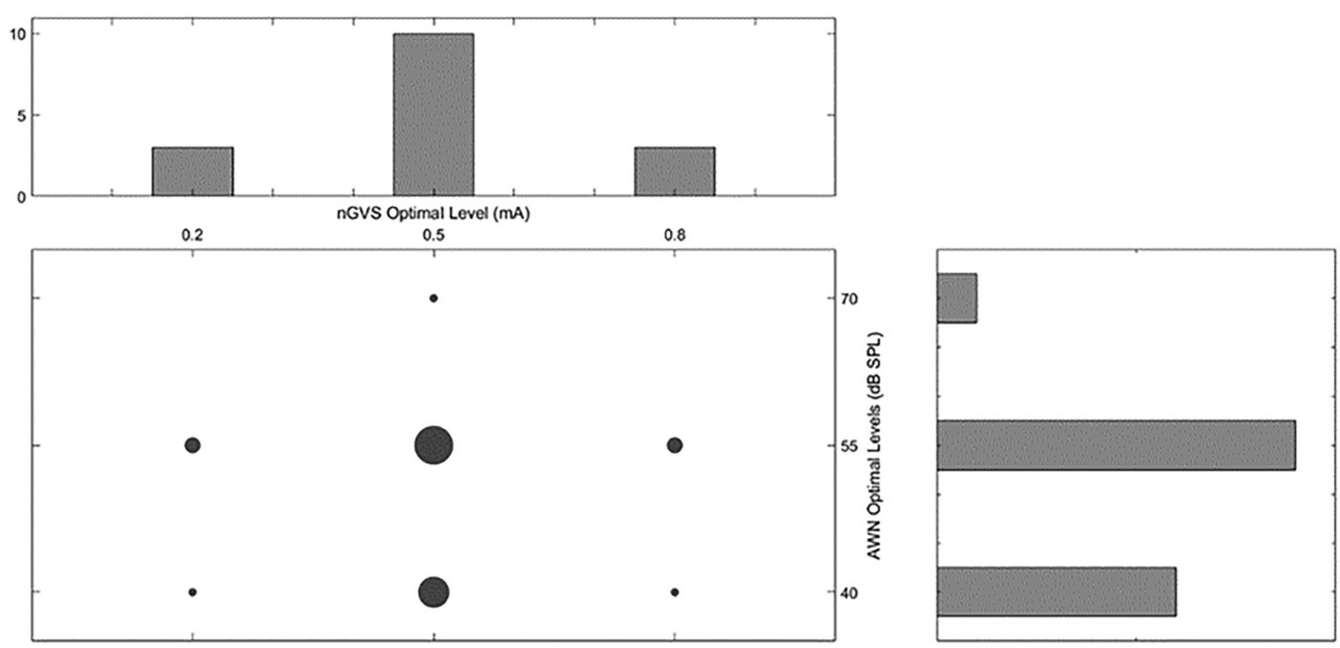

Fig. 4. Scatter bubble plot with marginal histograms showing the frequency of best levels identified for nGVS (x-axis) and AWN (y-axis) across our subject pool. Larger bubbles indicate a higher frequency of that combination. Note that the best levels in each sensory modality were the central levels tested.

in the overall or subdimension performance analyses to support this hypothesis. While SR has been shown to improve suprathreshold performance in sensory systems,[7,15] it is classically believed to modulate threshold, or at-limit, capabilities; therefore, by varying task difficulty we could capture whether improvements are only observed near subject limits. It is possible that our task was not challenging enough for our subjects to achieve this at-limit improvement, as subjects, on average, successfully landed 89% of the trials (95% for hard maps, 86.7% for medium maps, and 87.5% for easy maps). This appears consistent with our average subjectively reported mental workload, as the reported average was 3.2 with a 1.1 standard deviation, which suggests the task was always "satisfactory" or "tolerable."

While a null finding cannot prove there is no effect, this is the first evidence supporting that both nGVS and AWN do not enhance multiple aspects of operational performance. Like any study, it could be that there is an effect and our study was just not sufficiently well powered to identify it. First, this investigation consisted of 16 subjects as guided by our a priori power analysis. While we mention that our task may not have been sensitive enough to find performance differences, it is entirely possible that the effect size is small enough such that a greater number of subjects is needed to increase power and identify significant changes. Small effect sizes can result from large measurement variability. It was noted that older subjects had greater challenges adjusting to pitch inversion, finding the task more challenging than younger subjects, which could result in larger measurement variability. Originally, to avoid this, some subjects were given a longer training session than others. An exploratory analysis found that age had no significant effect on operational performance despite these reported challenges ($P > 0.9$), so age may not be a result of variability. Note that we report effect sizes to enable future meta-analyses. Second, it could be that our specific lunar landing task is not susceptible to SR effects, while other operational tasks may be. This is a first investigation and motivates future work. However, the lack of evidence across multiple subdomains of the complex task does not support benefits in other complex tasks. Third, other SR work has concluded that different levels of sensory noise are optimal for different individuals and tasks, but many cognition-based SR studies investigate a single noise level across all participants. To try to address this we rigorously conducted an initial suite of tests at three different sensory noise levels (0.2, 0.5, and 0.8 mA for nGVS and 40, 55, and 70 dB SPL for AWN) to identify the subject-specific best levels. The frequency of best levels identified are shown in **Fig. 4**; further visualization of subject performance in each metric for each noise level is provided in Appendix C (found online at https://doi.org/10.3357/amhp.6251sd.2023). It is possible that this procedure was inadequate at identifying a level of sensory noise that was beneficial for each individual, either because our suite was not inclusive of the optimal levels for most subjects (e.g., a subject's optimal nGVS level was 1 mA and our suite only extended up to 0.8 mA) or because the suite was not fine enough (e.g., optimal was 0.65 mA and we only tested at 0.5 and 0.8 mA, neither of which yielded much benefit). However, our suite was selected based upon levels at which SR benefits had previously been observed,[2,3] have been used by us previously,[17,21] and reasonably traded off the time required to do the initial suite (and associated subject learning/fatigue/boredom). Fig. 4 shows that levels with increased sensitivity around 0.5 mA nGVS and 55 dB SPL AWN levels should be further explored. Additionally, the results found in Appendix C (found online at https://doi.org/10.3357/amhp.6251sd.2023) show it is possible that some noise levels may be more appropriate for specific subdimensions within subjects (e.g., 40 dB SPL may be best for perception detection, but 70 dB is best for flight). This can pose operational challenges in identifying noise levels that are comprehensive in performance improvement.

While the literature shows that SR may help enhance perception and aspects of cognition, we did not find that it has a substantial influence on operator performance for the broad population. Our work, however, is the most comprehensive assessment of sensory noise effects on operator performance in a complex task to date. As such, for complex aerospace applications like the one investigated in this study, it may not be a critical operationally relevant countermeasure. Nonetheless, since neither nGVS or AWN seemed to affect some individuals, but was not related to an individual's noise preference, future work should explore these individual differences in SR susceptibility for operator performance.

This investigation evaluated the utility of applying sensory noise to improve performance in a macrocognitive task. We conclude that applying additive noise to auditory and vestibular modalities will not result in improved operator performance or reduced perceived mental workload for the broad population. However, similar to other SR investigations, we find that specific individuals may be affected by additive noise. We had subjects report their preference for working in noisy environments to build upon previous work on whether preference is a useful indicator for individual SR performance effects. We found no correlation between noisy environment preference and performance under noise influence.

ACKNOWLEDGMENTS

We acknowledge Anna Jonsen, Alexander (Sasha) Kryuchkov, Abigail Durell, and Maya Greenstein for their testing efforts on this project. This study was funded by the Translational Research Institute for Space Health (TRISH) through NASA Cooperative Agreement NNX16AO69A (award number T0402).

Financial Disclosure Statement: The authors have no competing interests to declare.

Authors and Affiliation: Sage O. Sherman, Ph.D., M.S., Young-Young Shen, Ph.D., M.A.Sc.,Daniel Gutierrez-Mendoza, B.S., Michael Schlittenhart, M.S., B.S., Cody Watson, M.S. Student, B.S., Torin K. Clark, Ph.D., M.S., and Allison P. Anderson, Ph.D., M.S., University of Colorado Boulder. Boulder, CO, United States.

REFERENCES

1. Choe J, Coffman BA, Bergstedt DT, Ziegler MD, Phillips ME. Transcranial direct current stimulation modulates neuronal activity and learning in pilot training. Front Hum Neurosci. 2016; 10:34.

2. Galvan-Garza RC, Clark TK, Mulavara AP, Oman CM. Exhibition of stochastic resonance in vestibular tilt motion perception. Brain Stimul. 2018; 11(4):716–722.
3. Helps SK, Bamford S, Sonuga-Barke EJS, Söderlund GBW. Different effects of adding white noise on cognitive performance of sub-, normal and super-attentive school children. PLoS One. 2014; 9(11): e112768.
4. Hidaka I, Nozaki D, Yamamoto Y. Functional stochastic resonance in the human brain: noise induced sensitization of baroreflex system. Phys Rev Lett. 2000; 85(17):3740–3743.
5. Kintz JR, Banerjee NT, Zhang JY, Anderson AP, Clark TK. Estimation of subjectively reported trust, mental workload, and situation awareness using unobtrusive measures. Hum Factors. 2022; 2022:001872082211293; online ahead of print.
6. Klein G, Ross KG, Moon BM, Klein DE, Hoffman RR, Hollnagel E. Macrocognition. IEEE Intell Syst. 2003; 18(3):81–85.
7. Lugo E, Doti R, Faubert J. Ubiquitous crossmodal stochastic resonance in humans: auditory noise facilitates tactile, visual and proprioceptive sensations. PLoS One. 2008; 3(8):e2860.
8. McDonnell MD, Stocks NG, Pearce CE, Abbott D. Stochastic resonance: From suprathreshold stochastic resonance to stochastic signal quantization. Cambridge (UK): Cambridge University Press; 2008:67–101; 323–344.
9. Moss F, Ward LM, Sannita WG. Stochastic resonance and sensory information processing: a tutorial and review of application. Clin Neurophysiol. 2004; 115(2):267–281.
10. Patel ZS, Brunstetter TJ, Tarver WJ, Whitmire AM, Zwart SR, et al. Red risks for a journey to the red planet: the highest priority human health risks for a mission to Mars. NPJ Microgravity. 2020; 6(1):33.
11. Pinedo C. Design and evaluation of an achievability limit display and control response types for piloted planetary landing. Ann Arbor (MI): ProQuest Dissertations & Theses; 2021.
12. Ries DT. The influence of noise type and level upon stochastic resonance in human audition. Hear Res. 2007; 228(1–2):136–143.
13. Roscoe A, Ellis G. A subjective rating scale for assessing pilot workload in flight: a decade of practical use. Farnborough (United Kingdom): Royal Aerospace Establishment; 1990. [Accessed 3 August 2023]. Available from https://apps.dtic.mil/sti/citations/ADA227864.
14. Roy-O'Reilly M, Mulavara A, Williams T. A review of alterations to the brain during spaceflight and the potential relevance to crew in long-duration space exploration. NPJ Microgravity. 2021; 7(1):5.
15. Sasaki H, Sakane S, Ishida T, Todorokihara M, Kitamura T, Aoki R. Suprathreshold stochastic resonance in visual signal detection. Behav Brain Res. 2008; 193(1):152–155.
16. Scheldrup M, Greenwood PM, McKendrick R, Strohl J, Bikson M, et al. Transcranial direct current stimulation facilitates cognitive multi-task performance differentially depending on anode location and subtask. Front Hum Neurosci. 2014; 8:665.
17. Sherman SO, Greenstein M, Basner M, Clark TK, Anderson AP. Effects of additive sensory noise on cognition. Front Hum Neurosci. 2023; 17:1092154.
18. Slack KJ, Williams TJ, Schneiderman JS, Whitmire AM, Picano JJ. Risk of adverse cognitive or behavioral conditions and psychiatric disorders. 2016; [Accessed 3 August 2023]. Available from https://ntrs.nasa.gov/citations/20160004365.
19. Söderlund GBW, Sikström S, Loftesnes JM, Sonuga-Barke EJ. The effects of background white noise on memory performance in inattentive school children. Behav Brain Funct. 2010; 6:55.
20. Usher M, Feingold M. Stochastic resonance in the speed of memory retrieval. Biol Cybern. 2000; 83(6):L11–L16.
21. Voros JL, Sherman SO, Rise R, Kryuchkov A, Stine P, et al. Galvanic vestibular stimulation produces cross-modal improvements in visual thresholds. Front Neurosci. 2021; 15:640984.
22. Wickens CD, Hollands JG, Banbury S, Parasuraman R. Engineering psychology and human performance. 4th ed. London (UK): Pearson; 2013:308–330.
23. Wilkinson A, Patel R, Darzi A, Singh H, Leff DR. Transcranial direct-current stimulation (tDCS) attenuates perceived temporal demand during simulated laparoscopic tasks. In: Fairclough SH, Zander TO, editors. Current research in neuroadaptive technology, chapter 8. Cambridge (MA): Academic Press; 2022:139–157.
24. Wilkinson D, Nicholls S, Pattenden C, Kildu P, Milberg W. Galvanic vestibular stimulation speeds visual memory recall. Exp Brain Res. 2008; 189(2):243–248.
25. Zuzula EA, Dixon JB, Davis E, Bretl K, Pinedo C, Clark TK. A numerical algorithm to estimate an achievability limit for crewed planetary landing. 2018 IEEE Aerosp Conference. New York: IEEE; 2018:1–7.