USAARL-TECH-TR--2024-13

UNITED STATES ARMY AEROMEDICAL RESEARCH LABORATORY

Evaluation of a Multisensory Cueing System on Aviators' Performance: Impact of Tactile and Auditory Cueing Sensitivity Levels

Kathryn A. Feltman, Ryan Mackie, Aaron McAtee, Christopher Aura, Jennifer Noetzel, Jordayne Wilkins, Xiaomin Yue, Riley McCormick, Shane Alcock, & Jason A. Gerstner

Notice

Qualified Requesters

Qualified requesters may obtain copies from the Defense Technical Information Center (DTIC), Fort Belvoir, Virginia 22060. Orders will be expedited if placed through the librarian or other person designated to request documents from DTIC.

Change of Address

Organizations receiving reports from the U.S. Army Aeromedical Research Laboratory on automatic mailing lists should confirm correct address when corresponding about laboratory reports.

Disposition

Destroy this document when it is no longer needed. Do not return it to the originator.

Disclaimer

The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation. Citation of trade names in this report does not constitute an official Department of the Army endorsement or approval of the use of such commercial items.

Human Subjects Use Statement

In the conduct of research involving human subjects, the investigator(s) adhered to the policies regarding the protection of human subjects as prescribed by Department of Defense Instruction 3216.02 (Protection of Human Subjects and Adherence to Ethical Standards in DoD-Supported Research) dated 8 November 2011.

IRB Determination and Number

This study, USAARL 2020-007, was approved by the Medical Research and Development Command Institutional Review Board on 9 November 2023.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
14-12-2023	Technical Report	

4. TITLE AND SUBTITLE
Evaluation of a Multisensory Cueing System on Aviators' Performance: Impact of Tactile and Auditory Cueing Sensitivity Levels

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
6.2

6. AUTHOR(S)
Feltman, K. A.[1], Mackie, R.[1,2], McAtee, A.[1,3], Aura, C.[1], Noetzel, J.[1], Wilkins, J.[1,2], Yue, X.[1,3], McCormick, R.[1], Alcock, S.[1], & Gerstner, J. A.[1]

5d. PROJECT NUMBER
MO210066

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Aeromedical Research Laboratory
P.O. Box 620577
Fort Novosel, AL 36362

8. PERFORMING ORGANIZATION REPORT NUMBER
USAARL-TECH-TR--2024-13

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Development Command
Military Operational Medicine Research Program
504 Scott Street
Fort Detrick, MD 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)
USAMRDC MOMRP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution is unlimited.

13. SUPPLEMENTARY NOTES
[1]U.S. Army Aeromedical Research Laboratory; [2]Oak Ridge Institute for Science and Education; [3]Goldbelt Frontier, LLC

14. ABSTRACT
The present study sought to evaluate the utility of a multisensory cueing system on aviators' ability to maintain performance, their experience of workload, and its impact to their situation awareness. Four combinations of cueing sensitivities were evaluated: high auditory/low tactile, high auditory/high tactile, low auditory/low tactile, and low auditory/high tactile. From this study it was concluded that cueing configurations featuring low auditory sensitivity were preferred in terms of performance. Additionally, the majority of participants indicated preference of the configuration featuring low auditory/high tactile cueing, which was also reflected in their workload ratings and performance data.

15. SUBJECT TERMS
aviation, multisensory cueing, flight performance

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UNCLAS	UNCLAS	UNCLAS	SAR	45	Loraine St. Onge, PhD
					19b. TELEPHONE NUMBER *(Include area code)* 334-255-6906

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

This page is intentionally blank.

Acknowledgements

The authors would like to acknowledge the U.S. Army Aeromedical Research Laboratory (USAARL) Flight Systems Branch for their hard work and dedicated participation in this study. The authors would also like to acknowledge Mr. Rolf Beutler for his assistance in programming the USAARL flight simulator. The project would not have been a success without their efforts. Additionally, the authors would like to acknowledge the professionalism and dedication of the research staff in the Warfighter Performance Group.

This research was supported in part by an appointment to the Biomedical Sciences and Engineering Research Participation Program at the U.S. Army Aeromedical Research Laboratory administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Army Medical Research and Development Command.

This page is intentionally blank.

Table of Contents

Page

Acknowledgements ... iii
Background ... 1
Methods ... 3
  Study Design ... 3
  Participants ... 3
  Materials ... 3
    Description of cueing ... 3
    Questionnaires ... 5
  Sensory Capabilities Tasks ... 8
  Equipment ... 9
    Flight simulator ... 9
    Electroencephalogram ... 9
    Eye tracking. ... 9
    Electrocardiogram ... 10
  Procedure ... 10
    Part one - training/screening. ... 10
    Part two - simulated flight tests ... 11
Results ... 12
  Takeoff ... 13
  En-Route ... 13
  Approach ... 14
  Hover ... 15
  NASA Task Load Index ... 16
  Instantaneous Self-Assessment of Workload Ratings ... 16
  Situation Awareness Rating ... 17
  Situation Awareness Rating Technique ... 17
  Physiological Measures ... 18
  Time in VSI Box ... 20
  Time in Speed Cup ... 20
  Tactile Choice Detection Task ... 20
  QuickSIN™ ... 21
  OBVA ... 21
  Usability ... 21
    UMUX-Lite ... 21
Discussion ... 21
Limitations ... 25
Conclusion ... 26
References ... 27
Appendix A. Acronyms and Abbreviations ... 31
Appendix B. Overview of Cueing ... 32
Appendix C. Description of Training ... 44

Table of Contents (continued)

Page

List of Figures

1. Vertical speed indicator 4
2. Vertical speed cup..4
B1. Panel mounted display arrangement 32
B2. Right display: enroute page...32
B3. Flight director path guidance...33
B4. Flight path marker with helicopter terrain avoidance warning system caution/warning 33
B5. Right display: hover approach takeoff page...34
B6. Artificial hover point above landing point. 34
B7. Artificial landing pad dimensions. 35
B8. Torque. 35
B9. HAT page symbols: altitude details. 36
B10. HAT page symbols: vertical speed details. 36
B11. Horizontal velocity vector details 37
B12. HAT page symbols: green markers on guidance symbols. 37
B13. Guidance deviation highlighter. 38
B14. Guidance with horizontal and vertical velocity error highlighters................................38
B15. Bumper radar scan: vertical, real vs. synthetic...39
B16. Obstacle threat assessment. 39
B17. Threat space shape 40
B18. Threat space down inclusion. 41
B19. Cueing multiple obstacles. 41
B20. Power line database obstacle..42
B21. Bumper radar obstacle display: visual. 42
B22. Power line obstacle display: auditory...43

List of Tables

1. Description of Cue Settings 5
2. List of Instruments and Descriptions 5
3. List of Sensory Capabilities Tasks and Descriptions..8
4. Summary of Flight Mission Tasks..11
5. Workload Manipulations 12
6. Simulator Variables Collected for Each Phase of Flight 12
7. Takeoff Descriptive Statistics...13
8. En-Route Descriptive Statistics 13
9. Approach Descriptive Statistics..15
10. Hover Descriptive Statistics...15
11. NASA-TLX Descriptive Statistics...16
12. Instantaneous Workload Ratings Descriptive Statistics 17
13. Situation Awareness Rating Descriptive Statistics..17
14. SART Descriptive Statistics 18

Table of Contents (continued)

Page

List of Tables (continued)

15. Descriptive Statistics for ECG Data 19
16. Descriptive Statistics for Eye Tracking Metrics 19
17. Mixed Effects Model Results....................20
18. Descriptive Statistics for the UMUX-Lite 21
19. Summary of Outcomes for Approach Segment 24
C1. Approximate Time Requirements for Activities....................45

This page is intentionally blank.

Background

Maintaining adequate situational awareness (SA) during flight is an integral aspect of being able to maintain performance. SA has been defined as, "a person's perception of the elements of the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley, 1995, p. 65). During routine flight operations, circumstances can arise that reduce SA. Examples of circumstances that may reduce or alter aviators' SA include, but are not limited to, unanticipated changes in task requirements and operational tempo, equipment failures, erroneous information presented, and the environmental conditions, such as those that cause degraded visual environments (DVE) (e.g., dust, fog, snow, glare, and darkness). The loss of accurate SA can reduce mission effectiveness and is a common cause of accidents.

One method to increase aviator performance and SA is to provide aviators with multisensory cueing systems that deliver cues through multiple senses, such as visual, auditory, and tactile. Most cues that aviators receive during flight are visual. These include visual alerts inside the cockpit (similar to a "Check Engine" light in an automobile), symbology depicting the flight path, and environmental cues out-the-window (e.g., trees, towers). Auditory alerts are also common in the cockpit, such as alerts indicating when malfunctions occur. However, to-date, tactile cueing is not used within Army aircraft. With most of the cueing being delivered visually, the likelihood of an aviator becoming visually overloaded is high. Therefore, the delivery of cues to other senses could help distribute the cognitive resources being used to attend to the cueing. The benefits of multisensory cueing on workload can be understood with Wickens' multiple resource theory (1980; 1984; 2002; 2008) that, in summary, posits that humans have different "pools" of cognitive resources that correspond to different task domains, such as auditory and visual. More than one task requiring resources from the same pool (e.g., all flight information displayed visually) will increase cognitive workload. Alternatively, tasks utilizing resources from different pools (e.g., some information presented visually, some auditorily) will "balance" workload because resources from one pool are not depleted, thus reducing the likelihood of overload. The ability to readily interpret and use information without becoming cognitively overloaded then increases an individual's SA (Wickens, 2002). Auditory and tactile sensory modalities present a viable option for the delivery of cues and have already been established within the literature to significantly reduce workload and improve performance on a variety of tasks (e.g., Baldwin et al., 2012; Hancock et al., 2013; Hancock et al., 2015; Hopkins et al., 2017; Kelley et al., 2012; Oskarsson et al., 2012).

Although distribution of cueing via multiple senses should aid in reducing aviators' workload and increasing SA, there is also the potential to overload the aviator with too many cues. As such, it is critical to identify which tasks to provide cues for and whether to provide the cue using multiple modalities. For example, a low altitude cue could be delivered both visually and auditorily. It is not only important to determine the right types of cues to utilize for delivering important information, but also to pair the cue mechanism with the task. For example, Hopkins et al. (2017) examined the delivery of cues using both auditory and tactile modalities during a dual task of varying workload. The authors found that the benefits of auditory and tactile cues for accuracy on a visual task depended on task type, where the auditory cues impeded performance when the secondary task type was an auditory task. Along similar lines, a meta-analysis of studies examining the use of tactile and visual information displays to improve task performance concluded that how the tactile cues were used (e.g., alerting, providing orientation information) played a significant role in how they impacted behavior, and that tactile cues paired with visual cues often yielded improved performance when compared to visual cues alone (Prewett et al., 2012). In addition to determining how the environment needs to be presented, determining the influence of individual differences in sensory capabilities on how that information is used is key. For example, individuals may be more or less sensitive to tactile cues delivered on the torso depending on body type (Tamé et al., 2019). These differences in sensitivity thresholds can translate to differences in how (and if) these cues are perceived and used during flight. Moreover, the sensitivity level of the cue itself may interact with differences in sensory capabilities, as well as individual preferences, to further impact performance. For example, tactile cues with a higher sensitivity level (tactile cues that are delivered more frequently) may be more readily perceived by someone who is less sensitive to tactile cues.

The objectives of this study were to evaluate how different combinations of auditory and tactile cueing sensitivity levels impacted aviator performance. Here we defined sensitivity as the threshold for which a cue would be triggered. High sensitivity settings would result in a lower threshold required for triggering a cue, such as a cue being triggered at 5-foot (ft) increments versus 10-ft increments. Low sensitivity settings would result in the opposite, where higher increments triggered a cue, such as the 10-ft increments. In addition to effects on performance, we also evaluated how these settings affected workload and SA both during and after using the system, and subjective impressions of the system's usability after its use. Workload and SA were assessed during flight through physiological and subjective measures. Physiological measures to detect changes in aviator workload and likely loss of SA include cortical activity measured through electroencephalogram (EEG), heart rate (HR), and heart rate variability (HRV) measured through electrocardiogram (ECG), and various measures of eye activity, such as gaze patterns and pupil diameter (for a review, see Borghini et al., 2014 and Brookhuis & de Waard, 2010). Additionally, subjective measures of workload and SA were used, to include the Crew Status Survey (CSS) workload scale (Ames & George, 1993) modified into an Instantaneous Workload Rating, the NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988), and the Situation Awareness Rating Technique (SART) (Taylor, 1990). Inclusion of multiple measures increases the likelihood of capturing subtle changes during tasks with performance and psychophysiological measures, and the subject's perspective at the completion of the task (de Waard & Lewis, 2014). The impact of individual differences in sensory capabilities on the use and perception of the different cueing modalities were assessed through various behavioral and clinical tests (e.g., speech in noise assessment, vision assessment).

The following hypotheses were tested:

H_1: By manipulating cueing sensitivities for combinations of auditory and tactile cueing at different levels (e.g., tactile high sensitivity, auditory low sensitivity), the combination of low tactile and low auditory cueing will result in overall improved flight performance.

H_2: By manipulating cueing sensitivities for combinations of auditory and tactile cueing at different levels (e.g., tactile high sensitivity, auditory low sensitivity), the combination of low tactile and low auditory cueing will result in reduced workload and increased situational awareness.

H3: Evaluation of individual performance when cueing sensitivities match personal preference compared to non-preferred cueing combinations will result in better performance at the individual level.

H4: Individual differences in sensory capabilities will affect performance, such that when cueing sensitivity levels match sensory capabilities, performance will be improved.

Methods

Study Design

The study used a within-subjects design to assess the effects of cueing sensitivity configurations on aviators' flight performance, cognitive workload, and situational awareness. There were two within-subjects factors: tactile sensitivity (high vs. low) and auditory sensitivity (high vs. low). Specifically, we evaluated the following combinations of auditory (AUD) and tactile (TAC) cueing: 1) low AUD + low TAC; 2) high AUD + low TAC; 3) low AUD + high TAC; and 4) high AUD + high TAC.

Participants

Sixteen aviators, recruited from the local Fort Novosel, AL area, consented to participate in the study. All participants were current and qualified pilots in the UH-60 helicopter and possessed current DD-2292 up-slips indicating good health. Participants had a range of experience flying the UH-60M, ranging from zero to 2000 hours (M = 687.19, SD = 758.32). UH-60M hours were of interest here as that is currently the most advanced model of the UH-60, and the configuration of the simulator used in this study. Flight hours within the past year, regardless of airframe, ranged from 20 to 400 hours (M = 170.26, SD = 139.14).

Materials

Two separate systems were used to deliver multisensory cues. These included the standard Army-issue communication earplugs (CEPs) (Communication and Ear Protection, Inc) and the Tactile Situational Awareness System (TSAS) (Engineering Acoustics, Inc.). The CEPs delivered the spatial audio sonifications for obstacle detection (further described below) and were worn in combination with the headset used in the simulator. TSAS delivered tactile cues through a waist belt and shoulder harness worn by the participant, and a seat cushion that participants sat on.

Description of cueing.

The multisensory cueing system that was assessed in this study consisted of components of the Integrated Cueing Environment (ICE). ICE was developed by researchers at the Army's Aviation and Missile Command (Godfroy-Cooper et al., 2019). ICE is a cueing system that consists of three cueing components: visual, auditory, and tactile. Visual cueing in this study was displayed on the traditional panel-mounted display. The entirety of visual symbology that makes up ICE are provided in Appendix B. Two visual symbols that are of particular interest are included here. These are the vertical speed indicator (Figure 1 below) and the vertical speed "cup" (Figure 2 below). Both of these symbols provide the aviator with feedback regarding current vertical speed. These are important indicators of performance during the approach phase of flight, which is the phase of flight where a majority of accidents occur (Payan et al., 2017).

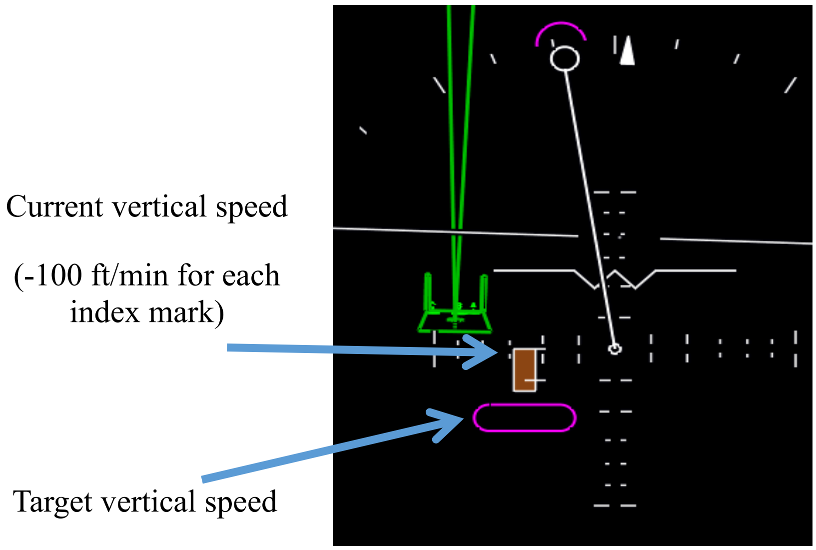


Figure 1. Vertical speed indicator (VSI).

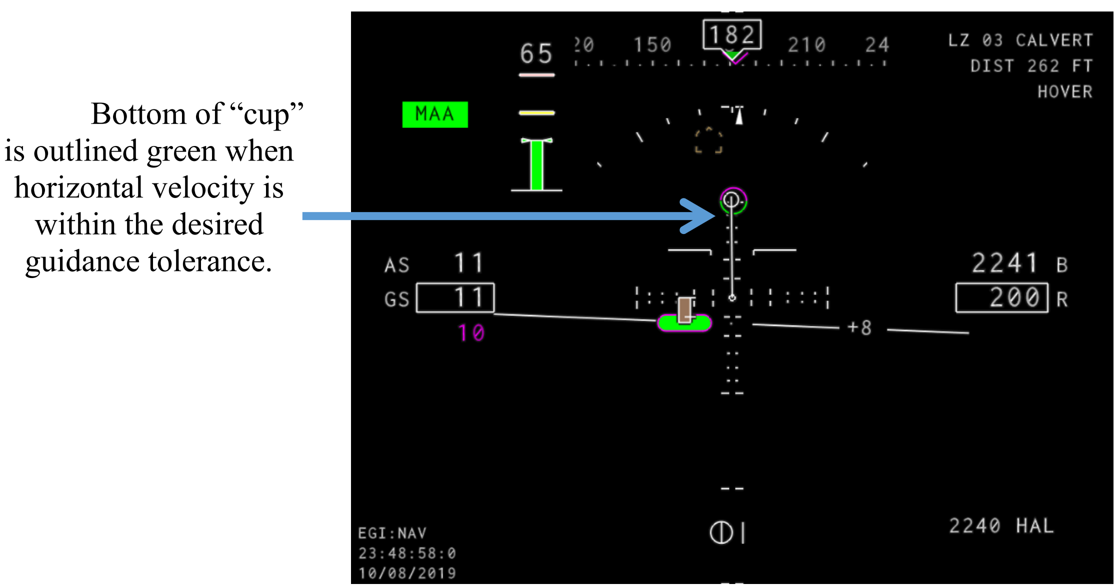


Figure 2. Vertical speed cup.

Auditory cueing includes spatial audio (described in detail in Godfroy-Cooper, 2018; Miller, 2018; Miller, 2019) for obstacle detection and warnings, and auditory alerts for altitude, heading, speed, drift, and power lines (summarized in Appendix B). Tactile cueing was provided using the TSAS, which provides cues for obstacle detection, altitude, and airspeed. Cues regarding the presence of obstacles were delivered using an obstacle threat space assessment, which is described in Appendix B, Figures B14-B22. Cue sensitivities for this study were manipulated for the following tasks: hover position keeping (TSAS), approach to hover/landing (TSAS), and obstacle detecting (see Table 1 below).

Table 1. Description of Cue Settings

Task cued	Modality	High sensitivity settings	Low sensitivity settings
Hover position hold	Tactile	3-6 ft error - seat at 1 Hz	6-9 ft error - seat at 1 Hz
		6-9 ft error - shoulder at 2 Hz	9-12 ft error - shoulder at 2 Hz
		> 9 ft error - shoulder at 4 Hz	> 12 ft error - shoulder at 4 Hz
Approach to landing/hover	Tactile	150-300 FPM - seat at 1 Hz	150-450 FPM - seat at 1 Hz
		300-450 FPM - seat at 2 Hz	> 450 FPM - seat at 4 Hz
		> 450 FPM - seat at 4 Hz	
Obstacle detection	Auditory	Small threat region	Large threat region
	Tactile	Caution region only	Caution region only

Note. ft = feet; Hz = hertz; FPM = feet per minute

Questionnaires.

Various questionnaires and survey instruments were used to evaluate individual differences that may impact response to baseline data and performance outcomes. All questionnaires and survey instruments, along with their descriptions, are presented in Table 2.

Table 2. List of Instruments and Descriptions

Instrument	Description
Sleep Timing Questionnaire (STQ)	The STQ is an 18-item self-report measure of sleep habits and requires 3 minutes for completion. Research shows it to be valid (such that it correlates with sleep diary information) and reliable across repeated administrations (Monk et al., 2003). Key outcome measures included sleep quantity (minutes) and wake after sleep onset (minutes).
State-Trait Anxiety Inventory (STAI)	The STAI is used to measure anxiety (Spielberger et al., 1983). The STAI is a widely used 40-item, self-report anxiety inventory rated on a 4-point Likert-type scale that captures two types of anxiety: state, or event-dependent anxiety, and trait, or persistent demonstrations of anxiety as a personal characteristic. Key outcome measures were total score for state anxiety and trait anxiety.

Beck Depression Inventory (BDI)	Depression symptoms were measured using the Beck Depression Inventory-II (BDI-II; Beck et al., 1996). The BDI-II is a commonly used 21-item, multiple-choice, self-report inventory that captures affect, cognition, and physical symptoms of depression over the most recent two-week period. Higher scores are correlated with greater endorsement of depression symptoms. The key outcome measure was the total score.
Karolinska Sleepiness Scale (KSS)	The KSS is a well-validated single item questionnaire that asks subjects to rate how sleepy they feel at the moment (Kaida et al., 2006). The KSS measures daytime sleepiness with higher scores indicating greater daytime sleepiness. The key outcome measure was subjects' daytime sleepiness score.
Convergence Insufficiency Symptom Survey (CISS)	The CISS is a 15-item questionnaire that asks participants to rate the severity of symptoms consistent with convergence insufficiency (Borsting et al., 1999). Specifically, participants rate 15 statements about how their eyes feel when reading or doing close work on a Likert-type scale from 0 (never occurs) to 4 (always). Key outcome measures were total composite scores, with higher scores indicating greater endorsement of symptoms characteristic of convergence insufficiency and oculomotor functioning.
Shipley Institute of Living Scale (SILS)	The SILS was designed to assess general intellectual functioning in adults and adolescents, and to aid in detecting cognitive impairment in individuals with normal original intelligence. Key outcome measures were vocabulary, abstraction quotient, and combined total score.
Demographic questionnaire	The demographic questionnaire allows subjects to provide basic demographic information that includes rank, ethnicity, basic medical information, career history (e.g., number of deployments), and education history, in addition to flight experience. The key outcome measures were age and flight experience.
Self-Assessment of Training Survey	The Self-Assessment of Training Survey is a five-item survey that was developed in-house. This survey uses a 10-point Likert scale and asks the participant to indicate the extent to which they agree with a statement where 1 = strongly disagree and 10 = strongly agree. Key outcome measures were the participant's level of knowledge regarding various aspects of the advanced pilot cueing system.
NASA Task Load Index (NASA-TLX)	The NASA-TLX is a questionnaire that measures subjective workload (Hart & Staveland, 1988). The subject rates the previous task, in this case flight, on the following categories, using a 100-point scale: mental demand, physical demand, temporal demand, performance, effort, and frustration. Key outcome measures were the weighted total workload score and scores for the six subscales.
Usability Questionnaire	The usability questionnaire is an in-house-developed questionnaire from the U.S. Army Combat Capabilities Development Command (DEVCOM) Data Analysis Center that was used to assess the usability of the different aspects of the advanced pilot cueing system. Subjects were asked to rate how usable each element of the system is

	by selecting from one of the following: not applicable, unsatisfactory, very poor, poor, good, very good, or excellent. The key outcome measure was the ranked order of preferred cueing.
Trust in Automation Questionnaire	The trust in automation questionnaire is a questionnaire developed in-house by the DEVCOM Data Analysis Center. This questionnaire was used to assess subjects' trust in the different aspects of the cueing system. For each aspect of the cueing system, subjects were asked to rate four statements using a 5-point scale ranging from "strongly disagree" to "strongly agree." The key outcome measure was subjects' trust in the cueing system.
Instantaneous Self-Assessment of Workload (ISAW)	Workload was measured in-flight by requesting participants to rate workload using an adaptation of the ISAW Technique (Brennen, 1992; Jordan, 1992). In this adaptation, participants rated workload using the Crew Status Survey Workload Scale (Ames & George, 1993) with ratings defined below. Workload ratings were visually displayed in the out-the-window view and panel-mounted display. Participants provided an initial rating and then were asked to provide updated ratings as they felt their workload changed. A member of the study team would also prompt for updated ratings. Ratings were recorded by a member of the research team. Ratings: 1. Nothing to do; no system demands 2. Light activity; minimal demands 3. Moderate activity; easily managed; considerable spare time 4. Busy; challenging but manageable; adequate time available 5. Very busy; demanding to manage; barely enough time 6. Extremely busy; very difficult; non-essential tasks postponed 7. Overloaded; system unmanageable; essential tasks undone; unsafe
Usability Metric for User Experience (UMUX)-LITE	The UMUX-LITE is a two-item rating scale that measures perceived usability (Lewis et al., 2013). Participants rated each of the items using a 7-point scale ranging from 1 "strongly disagree" to 7 "strongly agree."
Situation Awareness Rating Scale	The Situation Awareness Rating Scale is an in-house-developed scale from the DEVCOM Data Analysis Center and was used to assess participants' perceived internal and external situational awareness. Participants were requested to rate internal and external situational awareness using a 10-point scale that ranges from "high" (1-3), "moderate" (4-6), "low" (7-9), to "none" (10). Internal awareness is defined as the aircraft's state and performance (e.g., heading, speed), while external is defined as the visual scene or sensor image (e.g., terrain slope, obstacles).
Situation Awareness Rating Technique	The SART was used to measure situation awareness post-flight (Taylor, 1990). The SART uses the following ten dimensions to measure SA: familiarity of the situation, focusing of attention,

(SART)	information quantity, information quality, instability of the situation, concentration of attention, complexity of the situation, variability of the situation, arousal, and spare mental capacity. Participants rated each dimension on a seven-point rating scale (1 = low; 7 = high) based on their performance during the flights.

Sensory Capabilities Tasks

Three tasks were used to measure sensory capabilities. Each task represented one of the sensory modalities to which cues were delivered (vision, audition, tactile).

Table 3. List of Sensory Capabilities Tasks and Descriptions

Task	Description
Operational Based Vision Assessment (OBVA)	The OBVA is used to evaluate color vision sensitivity, stereo vision acuity, and fusional range, which are all measures of visual accuracy that might be impacted by macular pigment density. The OBVA is a computer-based battery of assessments interfaced with a remote control for data collection. A high color contrast monitor is used for the color vision test. The stereo vision test uses a standard monitor display to determine fusional range. Outcome measures for this task include color vision, stereo vision acuity, and fusional range.
QuickSIN	The QuickSIN presents participants with sets of six sentences with five key words per sentence in four-talker babble noise. The sentences presented are pre-recorded signal-to-noise ratios (SNRs) that decrease in 5-decibel (dB) steps from 25 (very easy) to 0 (extremely difficult). The SNRs used are 25, 20, 15, 10, 5, and 0, encompassing normal to severely impaired performance in noise. The QuickSIN is scored by marking correct key words repeated from each sentence. The SNR loss is calculated by subtracting the total number correct from 25.5 and averaged from each set of sentences. The key outcome measure for this task is the SNR loss value.
Tactile Choice Detection Task	For this task, participants identified the location of a tactile stimulus, presented in one of eight locations on the torso. To challenge the ability to detect and localize the stimuli, participants wore noise-cancelling headphones with white noise playing. The task required participants to indicate whether the target was present and at which location. A vibration was randomly delivered to one of the eight torso locations. A target stimulus was present in 80% of the trials (20% were "catch trials," where the target is absent to determine a false-alarm rate). When a stimulus occurred, participants were asked to report the location at which they detected the target stimulus. Participants indicated the location by selecting from a diagram of location options displayed on the computer. Key outcome measures for this task included location accuracy conditional on detection (accuracy in localizing stimuli detected vs. missed) and accuracy within one tactor location.

Equipment

Flight simulator.

Data were collected using the U.S. Army Aeromedical Research Laboratory's (USAARL) NUH-60 research flight simulator. It consists of a simulator compartment containing a cockpit, instructor/operator station, an observer station, and a six-degree-of-freedom motion system. It is equipped with an Rsi CV10R dome and eight Barco FS40 projectors which simulate natural helicopter environment surroundings for day, dusk, or night, and with blowing sand or snow. A Dell Precision laptop receives information concerning changes in the aircraft/simulator state parameters at a 60 hertz (times per second) capture rate. The spatial resolution is 1/256 of a foot, and data files are reported to two decimal places.

Electroencephalogram (EEG).

EEG data were collected using the B-Alert X-24 wireless wet electrode system with 20 channels corresponding to scalp locations according to the International 10-20 system (frontal channels: Fp1, Fp2, F7, F3, Fz, F4, F8; central channels: C3, Cz, C4, T3, T4; parietal and occipital channels: P3, POz, Pz, P4, T5, T6, O1, O2). Power spectral density (PSD) values were computed using the automated algorithms provided through the B-Alert Live Software (B-Alert Live, 2009). Prior to computing PSD values, artifacts were identified and removed using the Advanced Brain Monitoring algorithms for artifacts associated with electromyography (EMG), eye blinks, excursions, saturations, and spikes (B-Alert Live, 2009). PSD values evaluated in this study were from the following frontal channels: Fp1, Fp2, F7, F3, Fz, F4, and F8, with values averaged across each flight scenario as well as averages for phases of flight of interest to be used in data analyses. The PSD frequency ranges included in this study were alpha (9-13 Hz), beta (14-30 Hz), and theta (4-8 Hz).

EEG data were segmented by flight phase using the approach segment start and stop times extracted from the simulator data files. These times were generated from the simulator's internal clock time, which does not synchronize with other clocks (such as the EEG computer's clock) as the simulator computer remains fully off-line, not connected to the internet. As such, the research team entered markers into the EEG data when the simulator started and stopped recording data. Next, we compared the total time (in seconds) of the simulator recording to the time between EEG markers. This time difference ranged from 0.14 seconds to over 400 seconds. Approximately 30% of the data had a time difference less than 5 seconds. Approximately 50% of the data had a time difference less than 10 seconds. We opted to use the end (stop) marker to align the data, with the assumption that the end marker for the EEG data collection occurred at the same time the simulator stopped collecting data.

Eye tracking.

Data were collected using the Pupil-Core binocular headset (Pupil Labs, Berlin, Germany). The Pupil-Core camera system utilizes small (0.5 centimeter [cm] x 1.5 cm) cameras mounted on a lensless frame similar to eyeglasses. The cameras were positioned to a location 2-4 cm off each cheek, outside of the forward visual field so that they did not obstruct vision or interfere with task performance. This camera system collected pupil-size, and gaze-position information using an infrared, pupil-based, eye-tracking approach (Kassner et al., 2014). Prior to data collection, a calibration, consisting of a series of dots or landmarks located in the primary field of view of interest that appear to the participant one at a time was performed. The participants' responsibility is to fixate on each dot until it disappears and continue this process through the matrix of fixation points. The visual landscape was collected using a forward-facing camera, which was then matched to a reconstructed cockpit model, upon which the gaze-position information was registered for fixation and gaze-pattern analysis.

Electrocardiogram (ECG).

ECG data were collected using the BioPac MP150 base unit with electrocardiogram amplifier module (ECG100). Single-lead electrodes were placed on each of the participant's clavicles and one below the right pectoral area. Data were sampled at a rate of 1000 Hz. The raw ECG data was imported from Biopac files and pre-processed using a 0.5 Hz high-pass Butterworth filter and a 60 Hz powerline filter. The signal was then divided into 60-second segments for feature extraction. R-peaks, representing ventricular depolarization, were identified within each segment. Heart rate variability (HRV) was calculated by measuring the time difference between consecutive R-peaks. HRV indices, such as the mean and standard deviation, were then computed using these time differences.

Procedure

Participation in this study consisted of two parts. The first part included consenting, screening, measuring of sensory capabilities, and training. The training portion of the study took place over multiple visits, ensuring participants were familiar with the advanced cueing system prior to beginning testing. The second part included completion of a series of simulated flights where the flight environment was manipulated.

Participants were required to meet the following guidelines prior to data collection activities: 1) avoid taking medications that cause drowsiness; 2) not consume alcohol or sedatives 24 hours prior; 3) not consume nicotine 2 hours prior; and 4) not consume caffeine 16 hours prior. Participants were screened for compliance upon arrival at the laboratory using self-report. There were no participants rescheduled due to noncompliance.

Part one - training/screening.

During part one, participants completed informed consent procedures. After written consent was obtained, participants completed a series of electronic questionnaires consisting of the demographics questionnaire, Sleep Timing Questionnaire, Beck Depression Inventory, State-Trait Anxiety Inventory, Convergence Insufficiency Symptom Survey, followed by a hardcopy Shipley Institute of Living Scale. Sensory capabilities were assessed using the Operational Based Vision Assessment, QuickSIN, and the Tactile Choice Detection Task. Following this, participants engaged in their initial training activities, followed by completion of the Self-Assessment Training Survey regarding their comfort with the training received. Once proficiency was reached, participants completed one final flight to practice ISAW ratings prior to the testing session. All training was completed within the same week over the course of two days. Training activities are described in Appendix C.

Part two - simulated flight tests.

The second part of the study consisted of a series of tasks over the course of one day. Upon arrival, participants completed the KSS to assess current level of sleepiness. Next, a familiarization training in the simulator was conducted with a research pilot for a maximum of two hours. No participants exceeded the time limit for familiarization training. Participants also completed practice on the ISAW ratings. Following familiarization training, the physiological recording devices were placed on the participants and baseline data was collected. Test flights consisted of two routes with each including three flight scenarios featuring different missions (air assault, resupply, medical evacuation [MEDEVAC]). These missions are listed below along with the maneuvers performed during each mission (Table 4). Participants were instructed to maintain airspeeds between 80 and 90 knots indicated air speed (KIAS) and to maintain altitude at or below 300 feet above ground level (AGL). The research pilot provided verbal indications to the participant in the event that the parameters were exceeded or were trending to become exceeded.

Table 4. Summary of Flight Mission Tasks

Mission	**Maneuvers**
Air Assault	VMC takeoff
	VMC approach
	Terrain flight
	Shipboard operations
Resupply	VMC takeoff
	VMC approach
	Terrain flight
	Landing
MEDEVAC	VMC takeoff
	VMC approach
	Rescue hoist operations
	Landing

Note. VMC = visual meteorlogical conditions

Presentation of routes were varied across participants, while the order of missions remained the same (i.e., air assault to resupply to MEDEVAC). Each route included one of four workload manipulations (Table 5 below). The manipulations were presented in a counterbalanced order across the cueing configurations to ensure participants did not become complacent with repeated flights. The mission types were selected to reflect anticipated operations of the Future Attack Reconnaissance Aircraft and Future Long-Range Assault Aircraft, as described by Northrop et al. (2020).

This space is intentionally blank.

Table 5. Workload Manipulations

Option number	Manipulation	Description
Option 1	Emergency malfunction/ procedure	Boost servo malfunction
Option 2	Weather	Clouds and decreased visibility
Option 3	Co-pilot experience level	Co-pilot low experience in UH-60M
Option 4	Airspeed	Increased airspeed requirements

Note. The co-pilot was always the Research Pilot.

Each mission scenario lasted approximately 10 minutes, resulting in approximately 30-minute routes. During each scenario, flight performance data (Table 6) and physiological data were continuously collected. The ISAW ratings were collected throughout the duration of each flight scenario. Additional subjective data (NASA-TLX, Usability Questionnaires, Situation Awareness Rating Technique, UMUX-Lite, and the Trust in Automation Questionnaire) was collected at the end of each scenario.

Table 6. Simulator Variables Collected for Each Phase of Flight

Outcome variable	Units	Definition	Phases of flight captured
RMSD heading deviation	Degrees	Higher deviations = worse performance	Takeoff, approach, hover, landing
RMSD lateral deviation	Feet (ft)	Higher deviations = worse performance	En-route
RMSD speed Deviation	Knots (kts)	Higher deviations = worse performance	En-route, approach
Time in VSI box	Seconds (s)	Increased amount of time = better performance	Approach
Time in speed cup	Seconds (s)	Increased amount of time = better performance	Approach
RMSD altitude deviations from 50 ft	Feet (ft)	Higher deviations = worse performance	Hover
Time in target altitude box	Seconds (s)	Increased amount of time = better performance	Hover

Note. RMSD = root mean squared deviations

Results

Analyses were conducted using R version 4.2.1 (R Core Team, 2022; RStudio Team, 2022). The following packages were used for analyses: lmerTest (Kuznetsova et al., 2017), rstatix (Kassambra, 2023), and tidyverse (Wickham et al., 2019). The identification and removal of outliers are discussed within the descriptions of findings below. Unless otherwise stated in the text, analyses were done using mixed-effects linear regression models, consisting of a fixed effect of condition with a random intercept for each participant, to assess the effects of cueing conditions. Total UH-60M flight hours were explored as a covariate in several of the models.

However, none found a significant relationship between condition and flight hours, so flight hours were removed from the final models.

H1: By manipulating cueing sensitivities for combinations of auditory and tactile cueing at different levels (e.g., tactile high sensitivity, auditory low sensitivity), the combination of low tactile (TAC) and low auditory (AUD) cueing will result in overall improved flight performance.

To evaluate this hypothesis, primary outcome measures from each of the following four phases of flight, takeoff, en-route, approach, and hover (MEDEVAC missions only), were evaluated separately.

Takeoff

The heading deviations dataset had a right-tailed distribution with overall mean and median values of 3.39 and 2.91, respectively. To account for the skewed distribution, a square root transformation of the dependent variable was completed. Cueing condition did not have a statistically significant effect on heading deviations during takeoff, $F(3, 173) = 1.29$, $p = 0.281$. Descriptive statistics for RMSD heading deviations are reported in Table 7 below.

Table 7. Takeoff Descriptive Statistics

Condition	*N*	*Mean (deg)*	*SD*
High AUD/High TAC	46	3.70	2.93
High AUD/Low TAC	45	3.12	2.13
Low AUD/High TAC	49	3.10	2.36
Low AUD/Low TAC	48	3.63	1.91

En-Route

Separate models were used to assess the effects of cueing conditions on the two outcome measures (lateral deviations and speed deviations). Regarding lateral deviations, there was right-tailed distribution with overall mean and median values of 106 and 90, respectively. After three outliers with values > 500 were removed, the overall mean and median values were 100 and 89, respectively. To account for the skewed distribution, a log transformation was completed. Regarding speed deviations, there was a right-tailed distribution with overall mean and median values of 21.4 and 18.9, respectively. To account for the skewed distribution, a log transformation was also completed. Cueing condition did not have a statistically significant effect on lateral deviations, $F(3, 297) = 0.28$, $p = 0.843$ nor speed deviations, $F(3, 300) = 1.01$, $p = 0.391$. Descriptive statistics for each are reported in Table 8 below.

Table 8. En-Route Descriptive Statistics

	Lateral deviations			Speed deviations		
Condition	*N*	*Mean (ft)*	*SD*	*N*	*Mean (kts)*	*SD*
High AUD/High TAC	77	103	55.40	78	22.6	8.38
High AUD/Low TAC	76	102	59.10	76	21.5	8.80
Low AUD/High TAC	81	98.7	50.50	82	21	8.58
Low AUD/Low TAC	78	96	48.60	79	20.7	8.61

Approach

Separate models were used to assess the effects of cueing conditions on the outcome measures (heading deviations, speed deviations, time in VSI box, and time in speed cup).

Heading deviations had a right-tailed distribution with overall mean and median values of 7.5 and 4.8, respectively. After five outliers, with values > 50, were removed, the overall mean and median values were 6.2 and 4.8, respectively. To account for the skewed distribution, a log transformation of the dependent variable was completed. Cueing condition did not have a statistically significant effect on heading deviations, $F(3, 293) = 1.49$, $p = 0.217$.

Speed deviations had a very slightly skewed distribution with overall mean and median values of 18.9 and 18.2, respectively. There were no outliers. To account for the skewed distribution, a log transformation of the dependent variable was completed. Cueing condition did not have a statistically significant effect on speed deviations, $F(3, 299) = 1.42$, $p = 0.238$.

Time in the VSI box had a right-tailed distribution with overall mean and median values of 11.5 and 8.3, respectively. After one outlier, with a value > 60 was removed, the overall mean and median values were still 11.5 and 8.3, respectively. To account for the skewed distribution, a square root transformation was completed. Cueing condition had a statistically significant effect on time in the VSI box, $F(3, 297) = 3.76$, $p = 0.011$. A post hoc analysis, consisting of pairwise comparisons of the cueing conditions, revealed the following effects of cueing condition:

- Time in VSI box was significantly longer in the Low AUD/Low TAC condition than in the High AUD/High TAC condition, $F(1, 141) = 4.19$, $p = 0.043$.
- Time in VSI box was significantly longer in the Low AUD/High TAC condition than in the High AUD/Low TAC condition, $F(1, 142) = 4.59$, $p = 0.034$.
- Time in VSI box was significantly longer in the Low AUD/High TAC condition than in the High AUD/High TAC condition, $F(1, 144) = 11.1$, $p = 0.001$.

Time in the speed cup had a right-tailed distribution with overall mean and median values of 17.3 and 13.4, respectively. There were no outliers. To account for the skewed distribution, a square root transformation of the dependent variable was completed. Using data from both segments, cueing condition did not have a statistically significant effect on time in the speed cup, $F(3, 298) = 0.31$, $p = 0.820$.

This space is intentionally blank.

Table 9. Approach Descriptive Statistics

Condition	Heading deviations (deg)			Speed deviations (kts)			Time in VSI box (s)			Time in speed cup (s)		
	N	***Mean***	***SD***	***N***	***Mean***	***SD***	***N***	***Mean***	***SD***	***N***	***Mean***	***SD***
High AUD/ High TAC	74	6.32	4.98	77	20.10	7.02	77	9.00	7.69	77	16.70	15.20
High AUD/ Low TAC	75	5.75	4.83	76	17.70	7.02	75	10.60	8.29	76	17.70	15.30
Low AUD/ High TAC	82	5.80	4.32	82	18.60	7.44	82	13.00	11.60	82	15.70	13.90
Low AUD/ Low TAC	78	6.86	5.50	79	19.10	7.62	79	12.60	12.50	79	19.30	18.10

Hover

Separate models were used to assess the effects of cueing conditions on the outcome measures (altitude deviations, time in target altitude box, and heading deviations).

Altitude deviations had a right-tailed distribution with overall mean and median values of 12.1 and 7.2, respectively. After two outliers, with values > 50 were removed, the overall mean and median values were 9.6 and 7.1, respectively. To account for the skewed distribution, a log transformation was completed. Cueing condition did not have a statistically significant effect on altitude deviations, $F(3, 35) = 0.79$, $p = 0.508$.

Time in the target altitude box had a right-tailed distribution with overall mean and median values of 4.5 and 3.3, respectively. There were no outliers. Cueing condition did not have a statistically significant effect on time in target altitude box, $F(3, 37) = 0.49$, $p = 0.690$.

Heading deviations had a right-tailed distribution with overall mean and median values of 16.0 and 12.4, respectively. After three outliers, with values > 45 were removed, the overall mean and median values were 12.8 and 11.8, respectively. Cueing condition did not have a statistically significant effect on heading deviations, $F(3, 39) = 2.27$, $p = 0.095$.

Table 10. Hover Descriptive Statistics

Condition	Altitude deviations (ft)			Time in target altitude box (s)			Heading deviations (deg)		
	N	***Mean***	***SD***	***N***	***Mean***	***SD***	***N***	***Mean***	***SD***
High AUD/High TAC	10	7.36	4.11	11	3.29	3.50	11	13.10	9.47
High AUD/Low TAC	11	10.50	7.47	11	4.87	5.69	11	11.70	6.26
Low AUD/High TAC	13	8.49	5.27	14	3.64	3.25	11	9.05	5.72
Low AUD/Low TAC	14	11.50	11.40	14	6.02	6.36	14	16.40	8.23

H2: By manipulating cueing sensitivities for combinations of auditory and tactile cueing at different levels (e.g., tactile high sensitivity, auditory low sensitivity), the combination of low tactile and low auditory cueing will result in reduced workload and increased situational awareness.

To evaluate workload and situational awareness, the NASA-TLX, ISAW ratings, Situation Awareness Rating Scale, and Situation Awareness Rating Technique were examined using a repeated-measures analysis of covariance (ANCOVA). Flight experience (number of UH-60M hours) was used as the covariate.

NASA Task Load Index

One outlier was removed prior to analysis. Results of the ANCOVA found no effect of cueing condition on workload ratings, $F(3, 39) = 0.77$, $p = 0.52$.

Table 11. NASA-TLX Descriptive Statistics

Condition	*Mean*	*SD*
High AUD/High TAC	46.78	15.53
High AUD/Low TAC	44.50	13.18
Low AUD/High TAC	40.28	14.88
Low AUD/Low TAC	43.61	14.70

Instantaneous Self-Assessment of Workload Ratings

These ratings were aggregated across phases of flight. The flights that included two en-routes, approaches, and landings were combined into a singular metric. For the en-route ratings, an effect of condition was found, $F(3, 299.2) = 4.19$, $p = 0.006$. Pairwise comparisons showed ratings were significantly higher in High AUD/High TAC compared to Low AUD/High TAC, and higher in each Low AUD/Low TAC and High AUD/Low TAC conditions compared to the Low AUD/High TAC condition.

For the approach ratings, an effect of condition was found, $F\ (3, 298.2) = 7.45$, $p < 0.001$. Pairwise comparisons found similar results to those during en-route where ratings were significantly *higher* in High AUD/High TAC compared to Low AUD/High TAC, and *higher* in each Low AUD/Low TAC and High AUD/Low TAC conditions compared to the Low AUD/High TAC condition.

No significant effects of condition were found for the hover phase, nor a significant interaction with flight hours, $F(3, 34.3) = 2.31$, $p = 0.09$.

For landing, an effect of condition was found, $F(3, 234.2) = 6.92$, $p < 0.001$. Pairwise comparisons showed the same pattern of results as those of the en-route and approach results, such that ratings were significantly higher in High AUD/High TAC compared to Low AUD/High TAC, and higher in each Low AUD/Low TAC and High AUD/Low TAC conditions compared to the Low AUD/High TAC condition.

Table 12. Instantaneous Workload Ratings Descriptive Statistics

	High AUD/ High TAC		High AUD/ Low TAC		Low AUD/ Low TAC		Low AUD/ High TAC	
Flight phase	***Mean***	***SD***	***Mean***	***SD***	***Mean***	***SD***	***Mean***	***SD***
En-route	2.72	1.27	2.55	1.00	2.42	1.00	2.29	0.94
Approach	3.39	1.38	3.12	1.07	3.20	1.09	2.83	0.98
Hover	4.09	1.81	3.66	1.20	3.61	1.36	3.50	1.35
Landing	3.73	1.64	3.52	1.23	3.73	1.26	3.05	1.18

Note. Scale ranged from 1 = nothing to do; no system demands to 7 = overloaded; system unmanageable; essential tasks undone; unsafe.

Situation Awareness Rating

An ANCOVA was used to examine these ratings, which did not support an effect of cue condition on internal situation awareness ratings, $F(3,42) = 0.17$, $p = 0.91$ nor for external situation awareness ratings, $F(3,42) = 2.03$, $p = 0.13$. Overall, participants rated internal and external situation awareness within the high to moderate ranges (see Table 13).

Table 13. Situation Awareness Rating Descriptive Statistics

	Internal		***External***	
Condition	***Mean***	***SD***	***Mean***	***SD***
High AUD/High TAC	3.13	1.09	3.75	1.65
High AUD/Low TAC	3.56	1.03	4.56	1.55
Low AUD/High TAC	3.75	1.65	4.38	1.59
Low AUD/Low TAC	3.94	1.34	5.06	1.88

Note. The ranges for categorizing the ratings are as follows: 1-3 = high; 4-6 = moderate; 7-9 = low; 10 = none.

Situation Awareness Rating Technique

ANCOVAs were used to examine these ratings, and an effect of condition was found for the complexity subscale ratings, $F(3, 42) = 3.11$, $p = 0.04$. Post hoc analyses did not yield any significant differences between the conditions.

This space is intentionally blank.

Table 14. SART Descriptive Statistics

Subscale	High AUD/ High TAC *Mean*	High AUD/ High TAC *SD*	High AUD/ Low TAC *Mean*	High AUD/ Low TAC *SD*	Low AUD/ High TAC *Mean*	Low AUD/ High TAC *SD*	Low AUD/ Low TAC *Mean*	Low AUD/ Low TAC *SD*
Instability	3.38	1.45	3.50	1.41	3.31	1.40	3.06	1.57
Complexity	3.69	1.66	3.63	1.50	3.25	1.44	3.13	1.54
Variability	3.91	1.47	3.94	1.77	3.63	1.41	3.19	1.22
Arousal	4.75	1.39	4.69	1.45	4.94	1.53	5.38	1.15
Concentration	4.81	1.17	4.69	1.35	4.94	1.53	5.38	1.09
Division	5.00	1.21	4.31	1.14	4.63	1.31	4.63	1.20
Spare mental	4.25	1.29	4.63	1.20	5.00	1.10	4.88	1.09
Information quantity	5.31	0.95	4.81	0.98	5.19	1.28	5.44	0.89
Familiarity	5.03	1.40	5.13	1.09	5.56	0.96	5.38	0.96

Note. Scale ranged 1 (high) to 7 (low).

Physiological Measures

In addition to evaluating the subjective measures of workload, physiological data were evaluated as measures of workload. Only data collected during the approach segment were evaluated as this segment of flight introduces the highest workload (Payan et al., 2017) and was the only phase of flight where condition had an effect on performance. EEG measures examined included: frontal theta, beta, and alpha values, as well as the beta-ratio.

Regarding frontal theta values, cueing condition did not have an effect, $F(3, 172) = 1.31$, $p = 0.272$. Frontal alpha values also were not affected by cueing condition, $F(3, 172) = 2.32$, $p = 0.078$. There was no effect of cueing condition on frontal beta values, $F(3, 172) = 1.69$, $p = 0.171$. Finally, the combination of these values into the beta-ratio also did not yield a significant effect of cueing condition, $F(3, 172) = 1.44$, $p = 0.233$.

An effect of cueing was found with the ECG metrics. Two outlier data points were removed from the analysis. Mean heart rate (beats per minute) was statistically significant, $F(3, 118) = 8.02$, $p < 0.0001$. Pairwise comparisons found that mean heart rate was significantly *higher* during the Low AUD/Low TAC compared to each of the remaining three conditions (see Table 15 for summary statistics). Additionally, heart rate during the High AUD/Low TAC condition was significantly *lower* than both the Low AUD/High TAC and High AUD/High TAC conditions.

There was also an effect of condition on HRV means, $F(3, 118) = 9.39$, $p < 0.0001$. HRV values in the Low AUD/Low TAC condition were significantly *lower* compared to each of the remaining three conditions. Additionally, HRV values in the High AUD/Low TAC condition were significantly *higher* than both the Low AUD/High TAC and High AUD/High TAC conditions.

Table 15. Descriptive Statistics for ECG Data

	High AUD/ High TAC		**High AUD/ Low TAC**		**Low AUD/ High TAC**		**Low AUD/ Low TAC**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Heart rate	66.53	10.57	64.85	10.53	66.13	11.31	68.53	11.14
HRV mean	924.97	150.88	950.52	161.57	932.19	149.28	897.98	143.05

From the eye tracking data, a significant effect of condition was found for two of the three outcome variables. There was a significant effect of condition on fixation counts, $F(3, 172) = 3.30$, $p = 0.02$. For fixation counts, there were more fixations in the High AUD/High TAC condition compared to the Low AUD/Low TAC condition (see Table 16 for descriptive statistics). However, the average duration of individual fixations was not significantly different between conditions. There were also more fixations in both High AUD/Low TAC and High AUD/High TAC compared to Low AUD/High TAC. There was also an effect of condition on pupil diameter, $F(3, 170) = 5.61$, $p = 0.001$. Diameter was larger in the Low AUD/Low TAC condition compared to the High AUD/Low TAC condition. Additionally, pupil diameter in both Low AUD/High TAC and High AUD/High TAC were larger compared to the High AUD/Low TAC condition.

Table 16. Descriptive Statistics for Eye Tracking Metrics

	High AUD/ High TAC		**High AUD/ Low TAC**		**Low AUD/ High TAC**		**Low AUD/ Low TAC**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Fixation counts	23.64	10.96	23.21	10.88	20.28	10.58	20.98	10.49
Fixation duration (ms)	285.92	64.94	286.81	61.14	295.73	58.12	291.86	58.26
Pupil diameter mean	39.63	7.08	36.81	7.81	40.25	7.95	39.36	7.01

Note. ms = milliseconds

H$_3$: Evaluation of individual performance when cueing sensitivities match personal preference compared to non-preferred cueing combinations will result in better performance at the individual level.

To evaluate whether individual preference for cueing combinations resulted in improved performance when flying under that combination, first we evaluated the frequencies of preferred combinations. Of the 16 participants, 13 preferred the combination of Low AUD/High TAC; the remaining 3 participants each had different preferences (1 = High AUD/High TAC; 1 = High AUD/Low TAC; 1 = AUD only).

Using only the data from the 13 participants whose preferred cueing method was Low AUD/High TAC, the effect of condition on each of the outcome measures was assessed. The goal was to see if using only participants whose preferred cueing method was Low AUD/High TAC would change any of the results previously found when using all of the participants. To complete a comparable analysis with the other cueing preferences, more participants would be needed, thus no analyses were completed to look at those participants with regard to personal preference. Additionally, only approach phase data were examined given this was the only phase that found significant results, with participants maintaining better performance related to vertical speed in the Low AUD/Low TAC and Low AUD/High TAC conditions. The results of the mixed effects regression model with condition and run as predictor variables are listed in Table 17 below. Those with a significant effect of condition are indicated with post hoc analyses reported below.

Table 17. Mixed Effects Model Results

Outcome measure	***F*(3, 235)**	***p***
RMSD lateral dev.	0.51	0.68
RMSD heading dev.	1.95	0.12
RMSD speed dev.	1.75	0.16
Time in VSI box (s)	3.58	0.01*
Time in speed cup (s)	4.29	0.01*

Note. *Post hoc analyses and descriptive statistics reported below.

Time in VSI Box

The post hoc analysis of the effect of condition showed that mean time in the VSI box was significantly longer in the Low AUD/Low TAC condition compared to the High AUD/High TAC condition, $t(255) = 2.93$, $p = 0.02$. These findings agree with the findings of using all participants.

Time in Speed Cup

The post hoc analysis of the effect of condition showed that mean time in the speed cup was significantly longer in the Low AUD/Low TAC condition compared to the High AUD/High TAC condition, $t(254) = 3.49$, $p < 0.01$, and compared to the Low AUD/High TAC condition, $t(254) = 3.14$, $p = 0.01$. This finding differed from that when all participants were used, which resulted in no significant effects.

H_4: Individual differences in sensory capabilities will affect performance, such that when cueing sensitivity levels match sensory capabilities, performance will be improved.

To evaluate whether differences in sensory capabilities affected performance, first, correlations between the outcome flight performance measures and the outcome sensory measures were conducted. Following this, a series of linear regression models were completed. Each model included a single sensory metric as a covariate to examine the main effects of condition and run.

Tactile Choice Detection Task

Two measures were calculated for this task: 1) percent correct responses, and 2) percent correct responses within one tactor from the target tactor. Participants performed well on this task, with response ranges for percent correct of 83% to 99% ($M = 0.94$, $SD = 0.04$) and even higher for within one tactor, ranging from 98% to 100% ($M = 0.99$, $SD = 0.006$). Given the lack of variability in the task, this measure was not used to evaluate the impact of sensory capabilities.

QuickSIN™

All 16 participants successfully completed the two QuickSIN™ lists. Average dB SNR loss was -0.4375 (*SD* = 1.93) for all participants, suggesting better than normal speech in noise understanding for the group. Participants performed better on list two when compared to list one as indicated with a paired *t*-test, $t(15) = 4.69$, $p < 0.001$. Linear regression models were performed to evaluate the effect of SNR loss scores as a covariate while examining the main effect of cueing condition on each performance metric. There was some evidence of this measure improving the linear model to predict performance; however, none of these findings differed from what was previously found without its inclusion.

OBVA

The following outcome measures were evaluated: stereo search near test threshold (near test), and stereo fusion range threshold (horizontal break, vertical break, and vertical recovery). The remaining variables for cone contrast (long wavelength cone and medium wavelength cone), and stereo fusion range threshold horizontal recovery were not examined due to missing data. A series of linear regression models were used with each OBVA measure as a covariate and examined the main effects of condition and flight number. None of the outcome measures were predictive of performance.

Usability

UMUX-Lite.

Descriptive statistics for the UMUX-Lite are reported in Table 19 below. Two conditions had higher ratings of agreement, Low AUD/High TAC, and Low AUD/Low TAC.

Table 18. Descriptive Statistics for the UMUX-Lite

	High AUD/ High TAC		**High AUD/ Low TAC**		**Low AUD/ High TAC**		**Low AUD/ Low TAC**	
Survey item	***Mean***	***SD***	***Mean***	***SD***	***Mean***	***SD***	***Mean***	***SD***
Q1. This cueing system's capabilities meet my requirements.	4.88	1.63	4.88	1.36	5.06	1.73	5.19	1.22
Q2. This cueing system is easy to use.	4.94	1.29	4.75	1.48	4.88	1.59	4.81	1.17

Note. Ratings ranged from 1 (strongly disagree) to 7 (strongly agree).

Discussion

The present study sought to evaluate the utility of a multisensory cueing system on aviators' ability to maintain performance, their experience of workload, and its impact on their SA. More specifically, the study sought to determine whether different combinations of auditory and tactile cueing sensitivity levels impacted aviators' performance and how they experienced workload and SA across three flight mission scenarios. In addition, we were interested in whether or not individual differences in sensory capabilities influenced outcomes. The usability of the cueing was also assessed for development purposes. Four hypotheses were evaluated in this study.

H_1: By manipulating cueing sensitivities for combinations of auditory and tactile cueing at different levels (e.g., tactile high sensitivity, auditory low sensitivity), the combination of low tactile and low auditory cueing will result in overall improved flight performance.

In order to understand the impact of the cueing sensitivities across the flight scenarios, we isolated the flight performance data based on phase of flight. This was due to the differences in the types of tasks that aviators are engaged in throughout each phase. This resulted in the following four phases of flight, which are also characteristic of most rotary-wing missions: takeoff, en-route, approach, and hover (MEDEVAC missions only, with hover occurring in an urban environment). From these data, only one flight performance metric was significant, and that occurred during the approach portion of the flight. During the approach phase, time in the VSI box, which aids in maintaining vertical speed during the approach phase of flight, showed differences between cueing conditions. Here it was found that participants spent significantly more time within these parameters for the Low AUD/Low TAC condition compared to the High AUD/High TAC condition, as well as for the Low AUD/High TAC condition compared to each the High AUD/Low TAC and High AUD/High TAC conditions. This suggests that for maintaining approach performance related to vertical speed, the combinations of the Low AUD/Low TAC and Low AUD/High TAC conditions are more effective for maintaining performance parameters as compared to the High/High conditions. Given that the approach phase of flight has frequently been identified as the phase of flight where a majority of accidents occur (Payan et al., 2017), this finding provides insight regarding cueing recommendations to make. However, further research using different flight scenarios would be needed to determine whether these findings are practically significant and generalizable to different flight operations.

That none of the other performance measures were impacted by condition, with only approach performance resulting in significant effects of cueing, is not entirely surprising. The vertical speed information is most relevant during approach to landing and/or approach to hover, when pilots often experience the highest workload. Additionally, there were a number of obstacles present when participants were conducting their approach, and the obstacle cueing is where the two modalities overlapped the most. It is, however, surprising that there were no effects of condition during the hover phase of flight. This finding may be due to participants experiencing a high workload regardless of the cueing configuration, as visual inspection of the reported means show that for three of the four conditions, this phase of flight received the highest ISAW ratings.

H_2: By manipulating cueing sensitivities for combinations of auditory and tactile cueing at different levels (e.g., tactile high sensitivity, auditory low sensitivity), the combination of low tactile and low auditory cueing will result in reduced workload and increased situational awareness.

Workload and SA were evaluated using both subjective evaluations and physiological measures. Regarding the subjective workload ratings, no significant differences were found for the NASA-TLX ratings. However, the aggregated ISAW ratings did yield significant differences between conditions within several phases of flight. For each the en-route, approach, and landing phases, participants rated workload in the High AUD/High TAC condition significantly *higher* than the Low AUD/High TAC condition. Workload ratings for the Low AUD/High TAC condition in these same phases were also rated significantly *lower* than both the Low AUD/Low TAC and High AUD/Low TAC conditions.

Physiological data can be used in conjunction with subjective ratings to evaluate workload. Here, we focused solely on the physiological data recorded during the approach phase, as that was the only phase where significant differences in performance were found. Historically, EEG indices have typically been the most reliable for differentiating workload conditions (Feltman et al., 2020). However, in this study, we found no effect of cueing condition on the EEG measures during the approach phase of flight. Regarding the ECG data, heart rate and HRV values suggest *increased* workload in the Low AUD/Low TAC condition compared to all other conditions, and *lower* workload in the High AUD/Low TAC condition compared to both the Low AUD/High TAC and High AUD/High TAC conditions. The pupil diameter results supported these findings, with *higher* workload in the Low AUD/Low TAC condition compared to the High AUD/Low TAC condition and *lower* workload in the High AUD/Low TAC condition compared to the Low AUD/High TAC and High AUD/Low TAC conditions.

To better interpret these findings, it is useful to examine the subjective, physiological, and performance results all together. In Table 19 below, we have included the pairwise comparisons for each of these measures for just the approach phase of flight. To interpret the table, the results within the cells are the comparison of the *top row* condition to the *first column* condition. For example, in looking at the *top row* condition Low AUD/Low TAC and the *first column* condition of High AUD/Low TAC, we can see that heart rate and pupil diameter were higher in the Low AUD/Low TAC condition compared to the High AUD/Low TAC condition, whereas HRV had the opposite effect.

This space is intentionally blank.

Table 19. Summary of Outcomes for Approach Segment

	Compared to Low AUD/ Low TAC	**Compared to High AUD/ High TAC**	**Compared to High AUD/ Low TAC**	**Compared to Low AUD/ High TAC**
Low AUD/ Low TAC	-	↑time in VSI box ↑ISAW ratings ↑HR ↓HRV ↓fixation count	↑ISAW ratings ↑HR ↓HRV ↑pupil diameter	↑ISAW ratings ↑HR ↓HRV
High AUD/ High TAC	↓time in VSI box ↓HR ↑HRV ↑fixation count	-	↑HR ↓HRV ↑pupil diameter	↓time in VSI box ↑ISAW ratings ↑fixation counts
High AUD/ Low TAC	↓HR ↑HRV ↓pupil diameter	↓HR ↑HRV ↓pupil diameter	-	↓time in VSI box ↑ISAW ratings ↓HR ↑HRV ↓pupil diameter ↑fixation count
Low AUD/ High TAC	↓ISAW ratings ↓HR ↑HRV	↑time in VSI box ↓ISAW ratings ↑fixation count	↑time in VSI box ↓ISAW ratings ↑HR ↓HRV ↑pupil diameter ↑fixation count	-

Note. ↑time in VSI box = better performance; ↑ISAW ratings = workload rated higher; ↑HR is related to higher workload; ↓HRV is related to higher workload; ↑pupil diameter is related to higher workload.

Looking at all significant comparisons, we see there are some inconsistencies between the three measures. For example, in the comparison of the Low AUD/Low TAC condition to the High AUD/ High TAC condition, we see better performance, but also higher ratings of workload and increased heart rate with decreased HRV, which is also consistent with a higher workload. Typically, it would be assumed that a higher workload would lead to worse performance (Howard et al., 2021). However, in this case we may interpret this as increased engagement, particularly taken together with the fact that SA ratings remained in the high to moderate ranges, suggesting participants were aware of their environment. Thus, it can be interpreted that while the aviators may have experienced an increase in workload, this was accompanied by good SA, which ultimately lead to improved performance. Improved performance may result in safer flight profiles, reducing the likelihood of aviation accidents.

In determining which condition resulted in the least workload and best performance, the Low AUD/High TAC column suggests this condition is preferred. The lack of physiological measures supporting a low workload in this condition may simply be due to the participants being engaged throughout the study and exposure to various conditions. Alternatively, this finding may also suggest that the differences in cueing conditions were subtle. Indeed, previous research using an earlier iteration of this cueing system concluded that additional cueing leads to improved performance (Hartnett et al., 2020). However, this finding is also in line with recent suggestions that a singular physiological metric for identifying workload is likely insufficient. Additionally, there are many cases where increased workload is reflective of increased engagement, which promotes improved performance.

Regarding the SA measures, there was no effect of condition on the SA ratings for either internal or external SA ratings. Alternatively, the results of the Situation Awareness Rating Technique found an effect of condition on ratings, but no significant effects were identified after performing pairwise comparisons. This suggests a result similar to the previous paragraph where just having the additional cues improves SA and there is not necessarily a clear difference between the various conditions.

H_3: Evaluation of individual performance when cueing sensitivities match personal preference compared to non-preferred cueing combinations will result in better performance at the individual level, and H_4: Individual differences in sensory capabilities will affect performance, such that when cueing sensitivity levels match sensory capabilities, performance will be improved.

Our evaluation of individual sensitivities did not yield any significant findings in predicting performance under the different cueing combinations. Although SNR loss scores did improve model accuracy for some measures, their addition did not change the overall findings where they were not included. Further evaluation may be needed to determine whether SNR loss or a similar hearing assessment would be beneficial in evaluating performance. Additionally, the lack of findings related to individual sensory capabilities may be due to the screening that already takes place to ensure aviators possess the necessary visual and hearing capabilities for safe flight. In evaluating personal preferences, the majority of participants indicated a preference for the Low AUD/High TAC configuration, which agrees with the subjective workload ratings and performance results for the approach phase.

Limitations

The study is not without limitations. One limitation is how we aligned the physiological data with the simulator data. At the time of this study, we did not have a method to perfectly sync these data by recording them into a singular file. As such, we relied on “start” and “stop” markers entered into the physiological data by a research technician to indicate when the flight began and ended. This, of course, introduces human error. Additionally, we were limited in the amount of time participants were able to train on the cueing system. Further training may have yielded different results. The manipulation of the cueing sensitivities was also limited. We chose to only manipulate one aspect of audio cueing, which was the obstacle avoidance cues, whereas tactile cues were introduced for multiple tasks. However, given that there are already standardized auditory cues for many of those tasks, such as altitude deviations, we chose to reserve the manipulation of auditory cue sensitivity for only its novel application of obstacle avoidance.

Conclusion

From this study we can conclude that cueing configurations featuring low auditory sensitivity are preferred in terms of performance. Moreover, the majority of aviators indicated preference of the Low AUD/High TAC configuration, which was also reflected in the ISAW ratings and in the performance data. More specifically, the approach phase showed the only significant differences for performance. Based on these flight scenarios, approach phase of flight is where we can most effectively reduce aviation workload with the ICE multisensory cueing system. Future studies should design flights to include more approach-like skills/tasks (and focus less on take-off and en-route) to further evaluate and refine the cues to increase their utility.

While we collected physiological data as an additional measure of workload, the physiological outcome measures did not necessarily reflect the same pattern of results. Specifically, performance was found to increase during the approach phase, as measured by the amount of time spent within the VSI box, for the two conditions featuring Low AUD. However, in the Low AUD/High TAC condition, which demonstrated both *improved* performance and *lower* workload ratings, the physiological data would suggest a higher workload. This may just suggest that the aviators in the study were working hard during this phase of flight, but did not reach the point of oversaturation, thus their physiological response may just reflect engagement with the task (more engaged = better performance). The ISAW ratings for this phase of flight fell within the two to four range, where the definition for a rating of four is, "busy; challenging but manageable; adequate time available." Therefore, it could be posited that the physiological measures simply reflected that the aviators were indeed busy and being challenged, whereas the performance measures reflected the ability to manage the challenge.

References

Ames, L. L., & George, E. J. (1993). *Revision and verification of a seven-point workload estimate scale* (AFFTC-TIM Report No. 93-01). Air Force Flight Test Center.

Baldwin, C. L., Spence, C., Bliss, J. P., Brill, J. C., Wogalter, M. S., Mayhorn, C. B., & Ferris, T. K. (2012). Multimodal cueing: The relative benefits of the auditory, visual, and tactile channels in complex environments. In *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*, 1431–1435.

B-Alert Live (2009). [Computer software]. Advanced Brain Monitoring, Inc.

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

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience and Biobehavioral Reviews*, *44*, 58–75.

Borsting, E., Rouse, M. W., & De Land, P. N. (1999). Prospective comparison of convergence insufficiency and normal binocular children on CIRS symptom surveys: Convergence Insufficiency and Reading Study (CIRS) group. *Optometry and Vision Science*, *76*(4), 221–228.

Brennan, S. D. (1992). *An experimental report on rating scale descriptor sets for the instantaneous self-assessment (ISA) recorder* (DRA Technical Memorandum (CAD5) 92017). DRA Maritime Command and Control Division.

Brookhuis, K. A., & de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis and Prevention*, *42*, 898–903.

de Waard, D., & Lewis-Evans, B. (2014). Self-report scales alone cannot capture mental workload. *Cognition, Technology and Work*, *16*, 303–305.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, *37*(1), 65–84.

Feltman, K. A., Bernhardt, K. A., & Kelley, A. M. (2020). Measuring the domain specificity of workload Using EEG: Auditory and visual domains in rotary-wing simulated flight. *Human Factors*, *63*(7), 1271–1283.

Godfroy-Cooper, M. M. (2018). Isomorphic spatial visual-auditory displays for operations in DVE for obstacle avoidance. *Proceedings of the European Rotorcraft Forum*, Delft, The Netherlands.

Godfroy-Cooper, M., Miller, J. D., Szoboslay, Z., & Hartnett, G. (2019). Multimodal pilot cueing for 360° situation awareness. *46th European Robocraft Forum*, Warsaw, Poland.

Hancock, P. A., Mercado, J. E., Merlo, J., & Van Erp, J. B. F. (2013). Improving target detection in visual search through the augmenting multi-sensory cues. *Ergonomics, 56*(5), 729–738.

Hancock, P. A., Lawson, B., Cholewiak, R., Elliot, L. R., Van Erp, J. B. F., Mortimer, B. J. P., Rupert, A., & Redden, E. S. (2015). Tactile cuing to augment multisensory human-machine interaction. *Ergonomics in Design*, 1–9.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183.

Hartnett, G., Hicks, J., Durbin, D., Feltman, K. A., St. Onge, P., Aura, C., Stewart, J., Godfroy-Cooper, M., & Miller, J. (2020). *Pilot cueing for 360◦ obstacle awareness during DVE missions* (CCDC DAC-TR-2020-030). CCDC DAC.

Hopkins, K., Kass, S. J., Blalock, L. D., & Brill, J. C. (2017). Effectiveness of auditory and tactile crossmodal cues in a dual-task visual and auditory scenario. *Ergonomics, 60*(5), 692–700.

Howard, Z. L., Innes, R., Eidels, A., & Loft, S. (2021). Using past and present indicators of human workload to explain variance in human performance. *Psychonomic Bulletin & Review*, 1923–1932.

Jordan, C. (1992). *Experimental study of the effect of an Instantaneous Self Assessment Workload recorder on task performance* (Technical Report DRA Technical Memorandum [CAD5] 92011). DRA Maritime Command Control Division.

Kaida M., Takahashi T., Åkerstedt A., Nakata Y., Otsuka T., Haratani K., & Fukasawa. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology, 117*, 1574–81.

Kassambra, A. (2023). *Rstatix: Pipe-friendly framework for basic statistical tests.*

Kassner, M., Patera, W., & Bulling, A. (2014). Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. *UbiComp '14 Adjunct: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 1151–1160.

Kelley, A. M., Cheung, B., Lawson, B. D., Rath, E., Chiasson, J., Ramiccio, J. G., & Rupert, A. H. (2012). Tactile cues for orienting pilots during hover over moving targets. *Aviation, Space, & Environmental Medicine, 84*(12), 1255–1261.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). ImerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26.

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013). UMUX-LITE – When there's no time for the SUS. *CHI* April 27 – May 2, 2013; Paris France pp. 2099-2102.

Miller, J. D. (2019). Augmented-reality cueing for obstacle avoidance: Towards a new topology for threat-level presentation. *Proceedings of the 75th American Helicopter Society Annual Forum & Technology Display.*

Miller, J. G.-C. (2018). ARSAD: An argumentated-reality spatial auditory display for obstacle avoidance for all phases for flight. *Proceedings of the 74th of the American Helicopter Society Annual Forum & Technology Display.*

Monk, T. H., Buysse, D. J., Kennedy, K. S., Potts, J. M., DeGrazia, J. M., & Miewald, J. M. (2003). Measuring sleep habits without using a diary: The sleep timing questionnaire. *Sleep*, *26*(2), 208–212.

Oskarsson, P. A., Eriksson, L., & Carlander, O. (2012). Enhanced perception and performance by multimodal threat cueing in simulated combat vehicle. *Human Factors*, *54*(1).

Payan, A. P., Lin, P., Johnson, C., & Marvis, D. N. (2017). Helicopter approach stability analysis using flight data records. *American Institute of Aeronautics and Astronautics, Inc.*

Prewett, M. S., Elliot, L. R., Walvoord, A. G., & Coovert, M. D. (2012). A meta-analysis of vibrotactile and visual information displays for improving task performance. *IEEE Transactions on Systems, Man, and Cybernetics: Applications and Reviews, 42*(1), 123–132.

R Core Team. (2022). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

Spielberger, C. D., & Gorsuch, R. L. (1983). *State-trait anxiety inventory for adults: Manual, instrument, and scoring guide*. Mind Garden, Incorporated.

Tamé, L., Azañón, E., & Longo, M. R. (2019). A conceptual model of tactile processing across body features of size, shape, side, and spatial location. *Frontiers in Psychology*, *10*(291). https://doi.org/10.3389/fpsyg.2019.00291

Taylor, R. M. (1990). Situation awareness rating technique (SART): The development of a tool for aircrew systems design. *In Situational Awareness in Aerospace Operations* (Chapter 3). NATO-AGARD-CP-478.

Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., Fancois, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Muller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4(43),* 1686.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 239–257). Erlbaum.

Wickens. C. D. (1984). Processing resource in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63-101). Academic Press.

Wickens, C. D. (2002). Situation awareness and workload. *Current Directions in Psychological Science, 11*(4), 128–133.

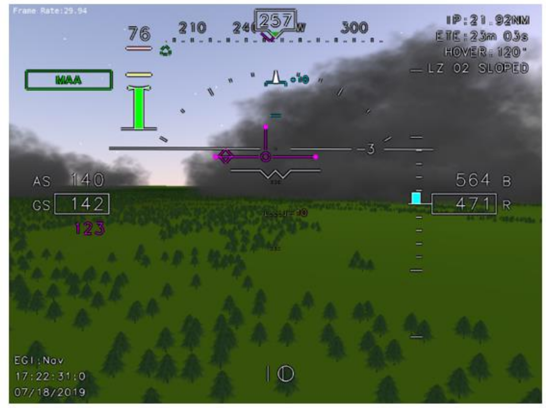
Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455.

Appendix A. Acronyms and Abbreviations

ANCOVA	Analysis of covariance
AUD	Auditory
CEPs	Communication Earplugs
CSS	Crew Status Survey
DVE	Degraded visual environment
ECG	Electrocardiography
EEG	Electroencephalography
FPM	Feet per minute
HR	Heart rate
HRV	Heart rate variability
Hz	Hertz
ICE	Integrated cueing environment
MEDEVAC	Medical evacuation
SA	Situational awareness
TAC	Tactile
TSAS	Tactical Situational Awareness System
USAARL	U.S. Army Aeromedical Research Laboratory
VMC	Visual meteorological conditions
VSI	Vertical speed indicator

Appendix B. Overview of Cueing

<u>Hover/Approach/Takeoff (HAT) Symbology Page</u>
- Used at low speeds (typically below 40 knots)
- Terrain image in background



<u>Enroute (Cruise) Symbology Page</u>
- Used at high speeds
- Terrain image in background

Figure B1. Panel mounted display arrangement.

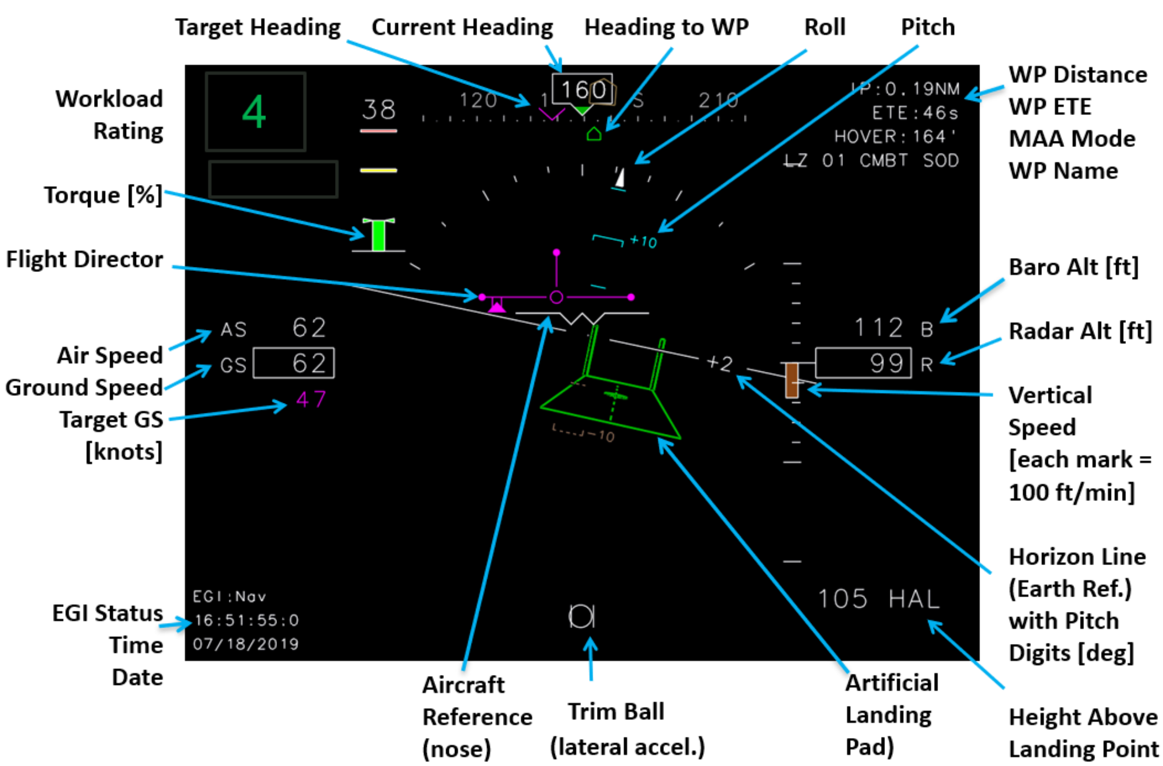


Figure B2. Right display: enroute page.

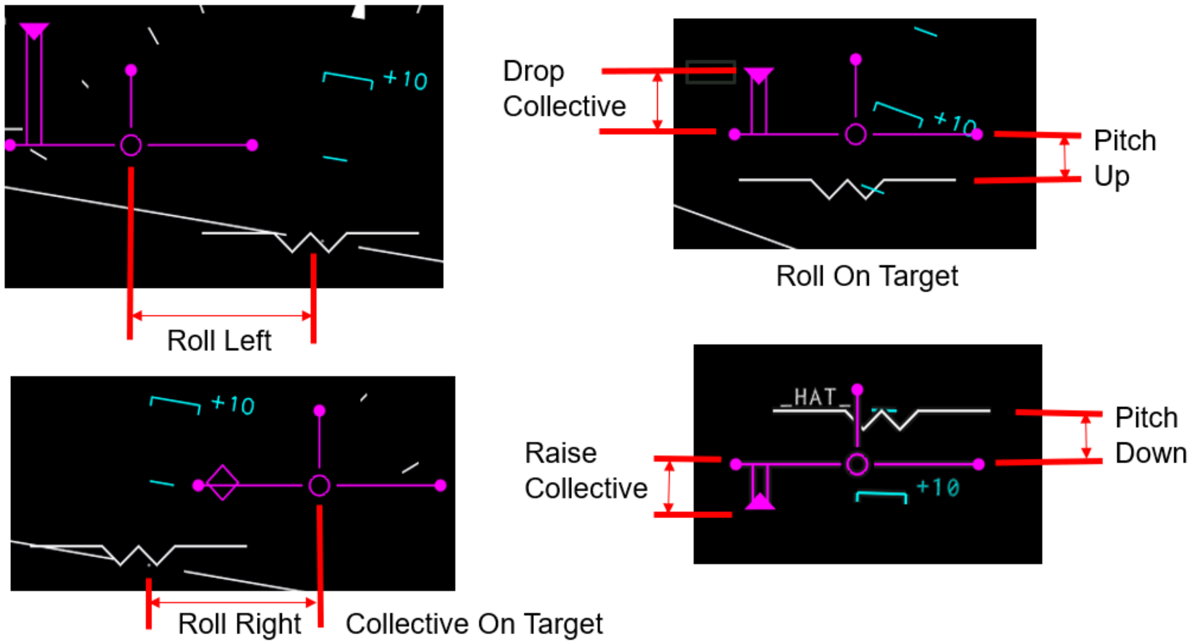


Figure B3. Flight director path guidance.

Note. Flight director strategy for pitch and roll:

- Think of aircraft reference symbol (W) as ownship longitudinal centerline.
- Think of flight director as the leader to be followed.
- Move cyclic to fly the aircraft reference symbol (W) to the leader.

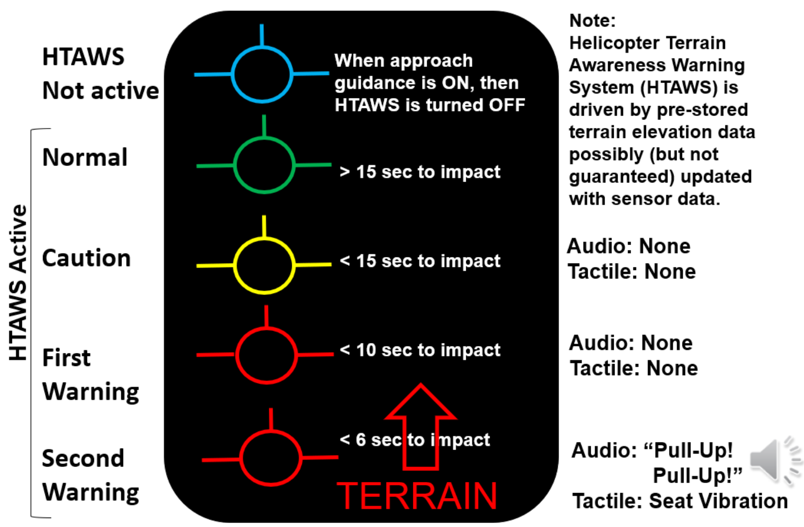


Figure B4. Flight path marker with helicopter terrain avoidance warning system (HTAWS) caution/warning.

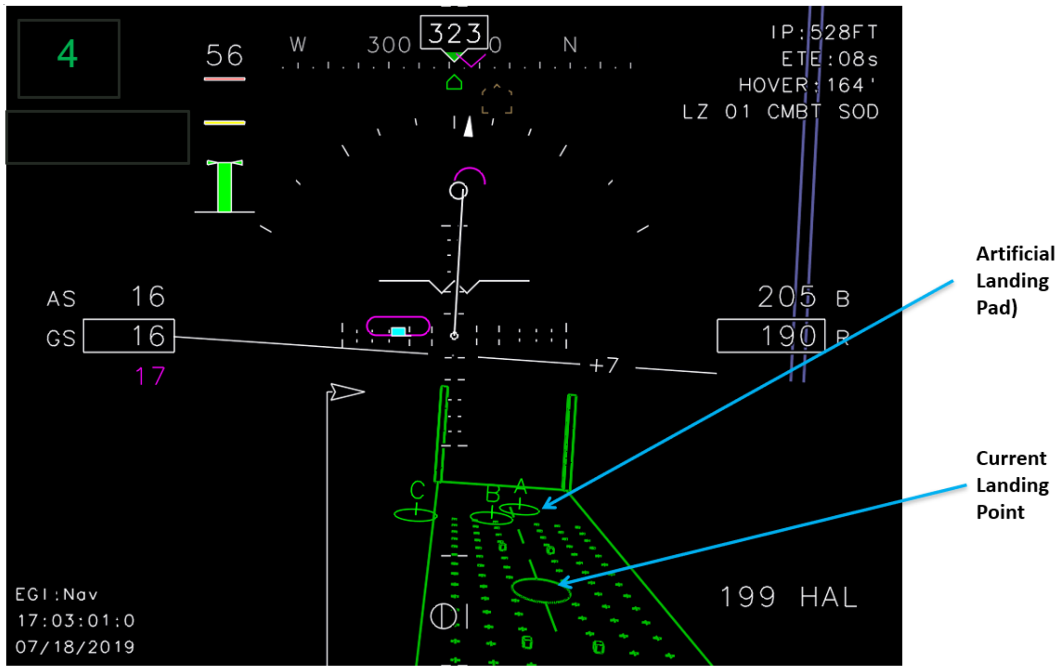


Figure B5. Right display: hover approach takeoff (HAT) page.

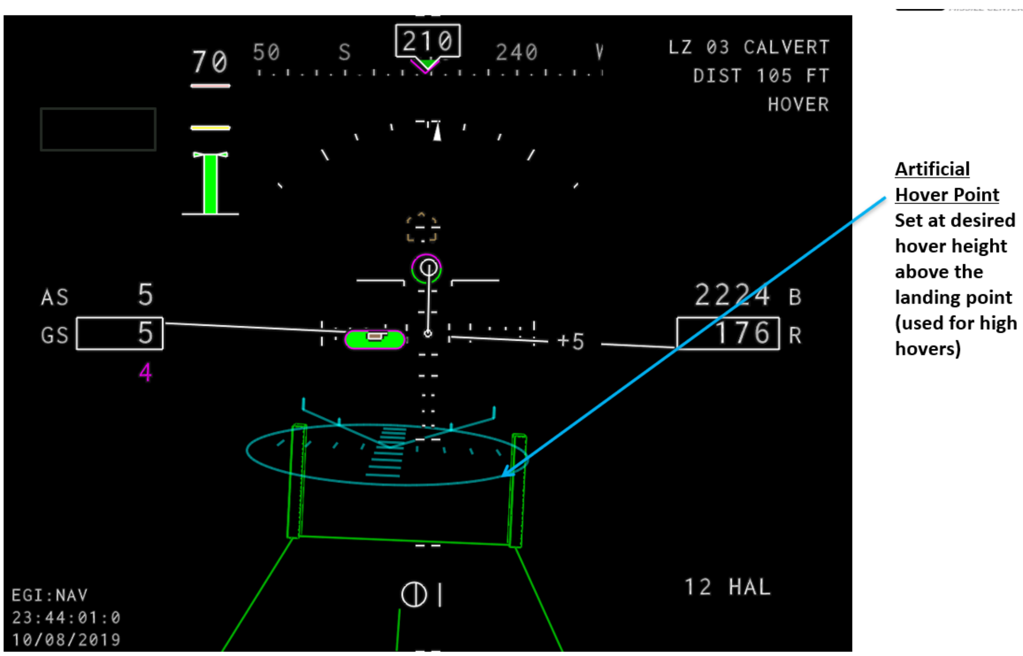


Figure B6. Artificial hover point above landing point.

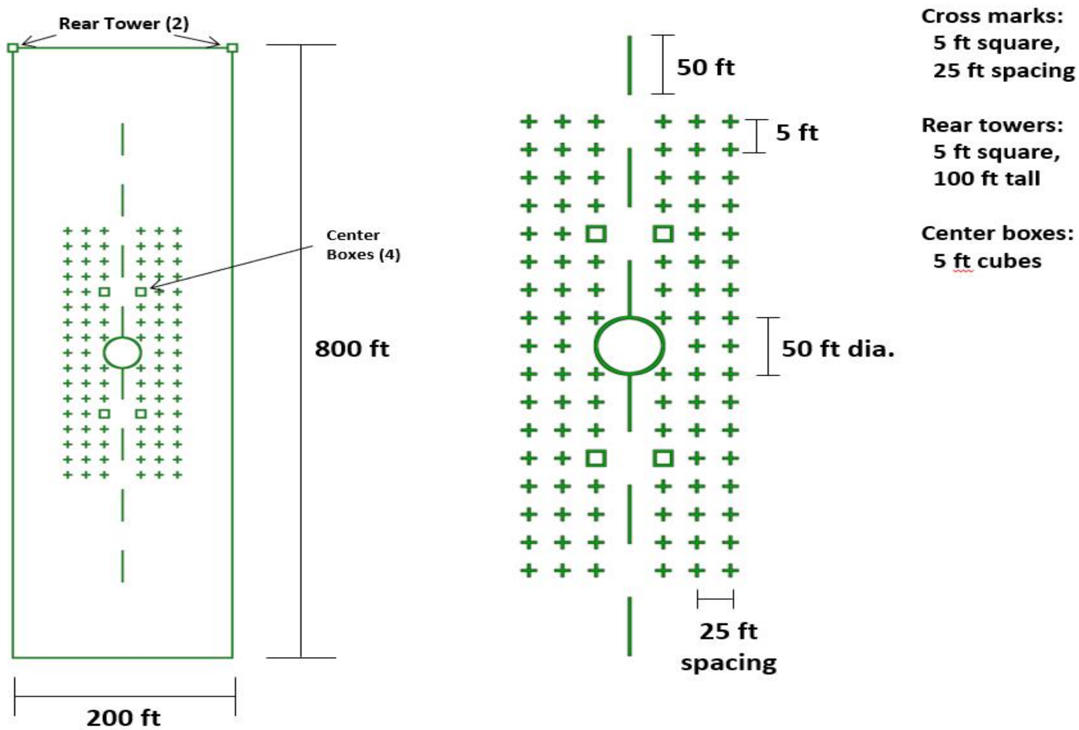


Figure B7. Artificial landing pad dimensions.

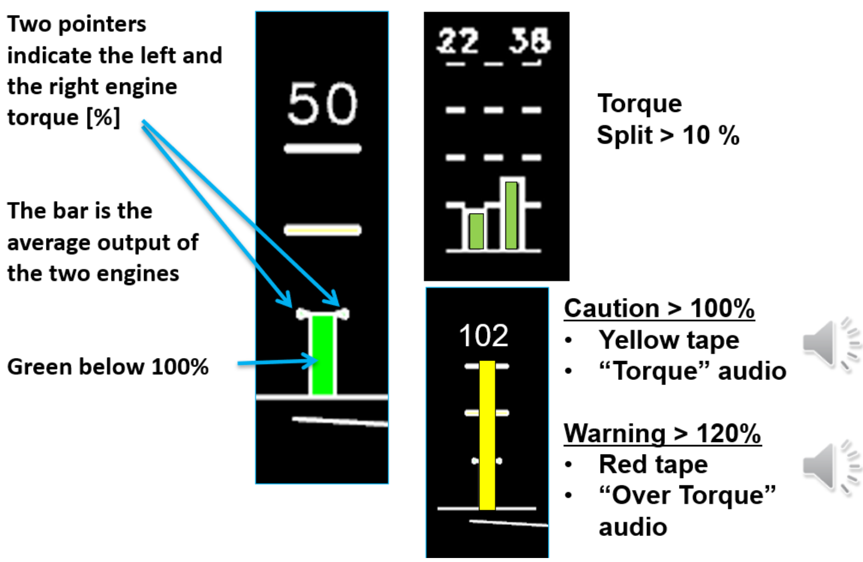


Figure B8. Torque.

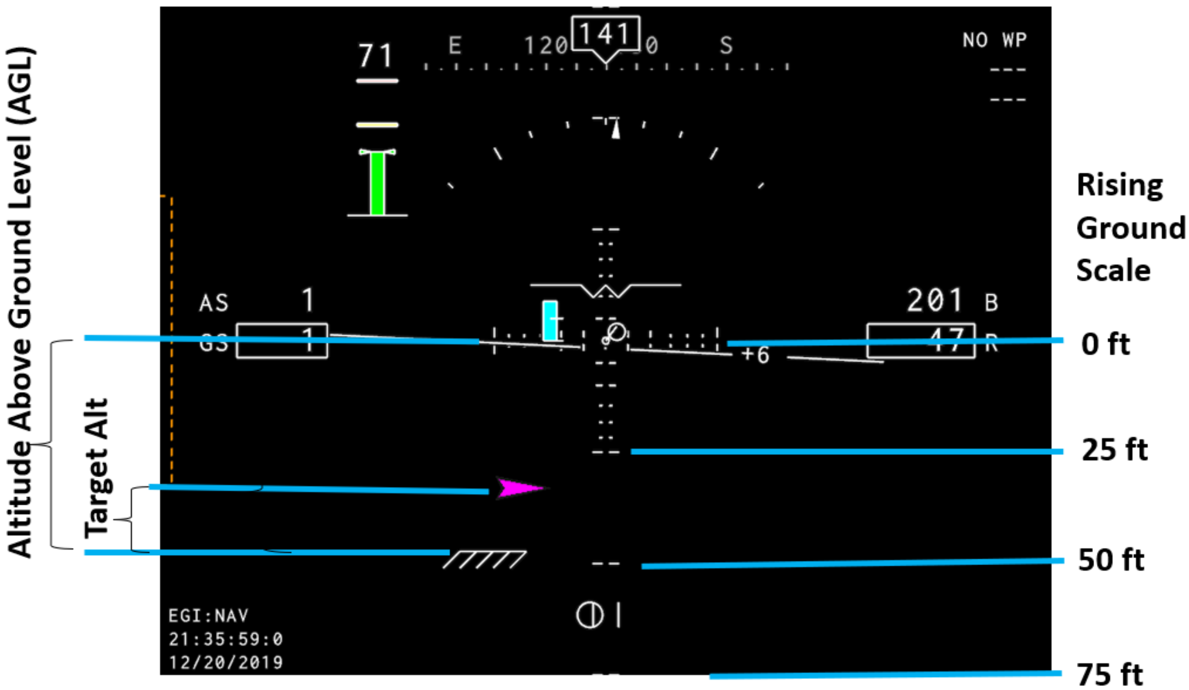


Figure B9. HAT page symbols: altitude details.

Note. Voice synthesizer will say altitude at 100, 50, 40, 30, 20, and 10 feet.

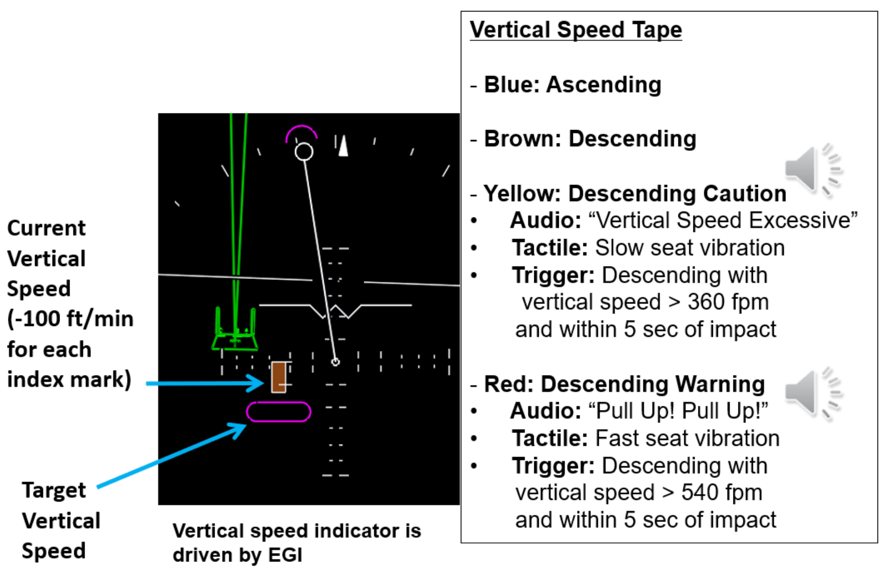


Figure B10. HAT page symbols: vertical speed details.

Note. Tactile cues were modified in this study based on sensitivity settings.

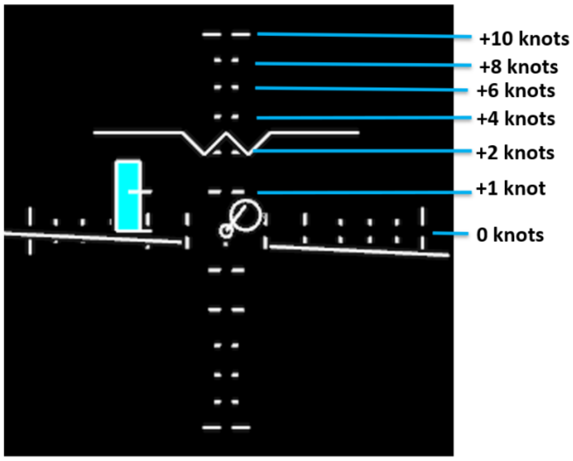


Figure B11. Horizontal velocity vector details.

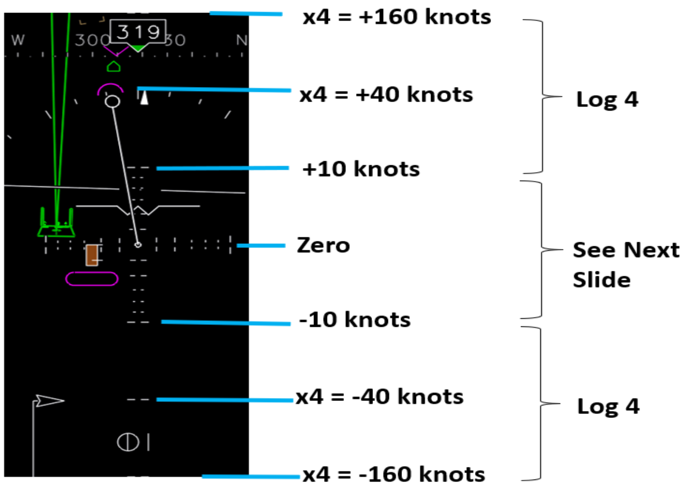


Figure B12. HAT page symbols: green markers on guidance symbols.

This space is intentionally blank.

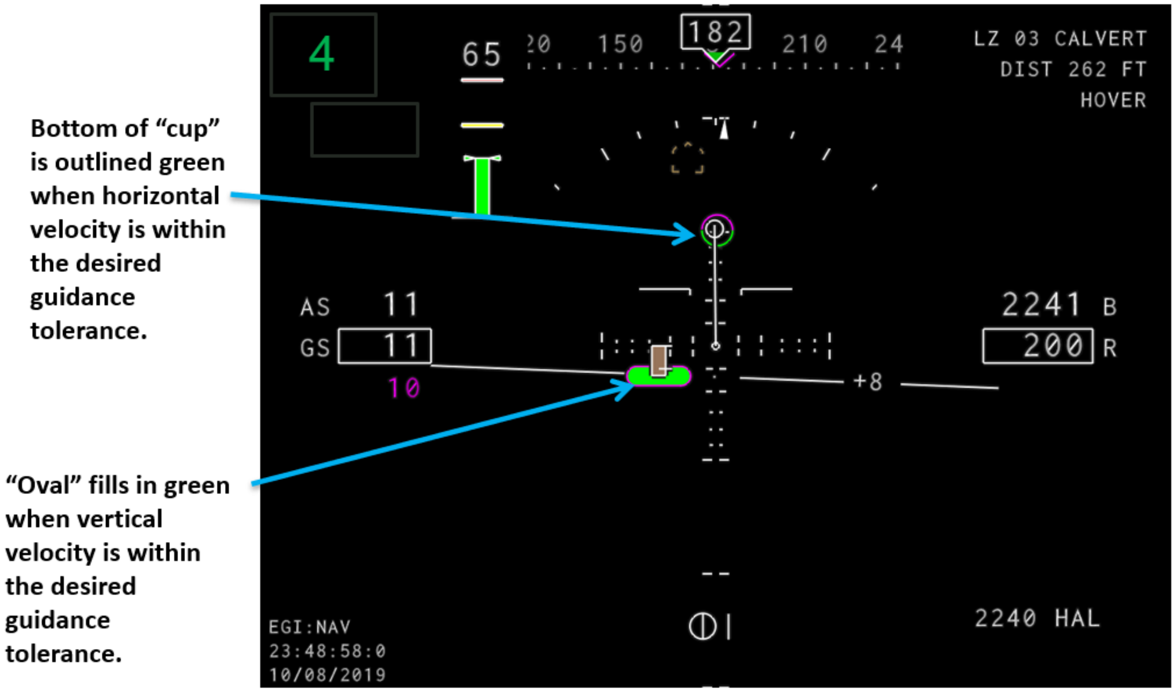


Figure B13. Guidance deviation highlighter.

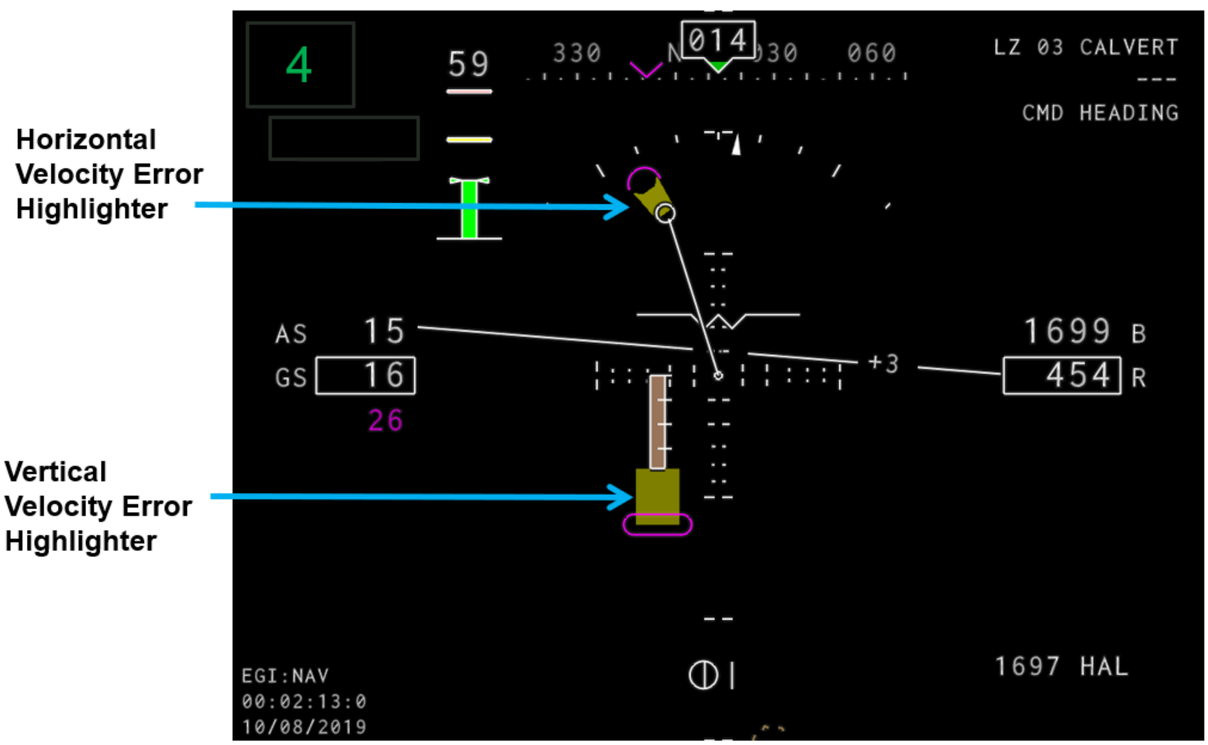


Figure B14. Guidance with horizontal and vertical velocity error highlighters.

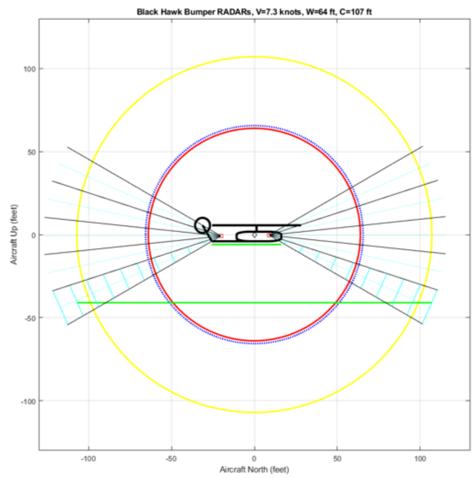


Figure B15. Bumper radar scan: vertical, real vs. synthetic.

Note. The physical bumper radars are rigidly attached to ownship with fixed pointing directions (the synthetic radars are perfectly gimballed).
360-degree (°) azimuth scan at two elevations: 0° and 12° *(*synthetic includes -12°*).*
4° azimuth x 12° elevation beam extent (black lines), 10.65 ft range/bin (cyan lines) (synthetic is laser-like and continuous).
Minimum range during ground testing ≈ 65 ft (dashed blue circle) (synthetic does not have a minimum range).
Default spherical 0-knot Caution (yellow) and Warning (red) cueing regions shown.

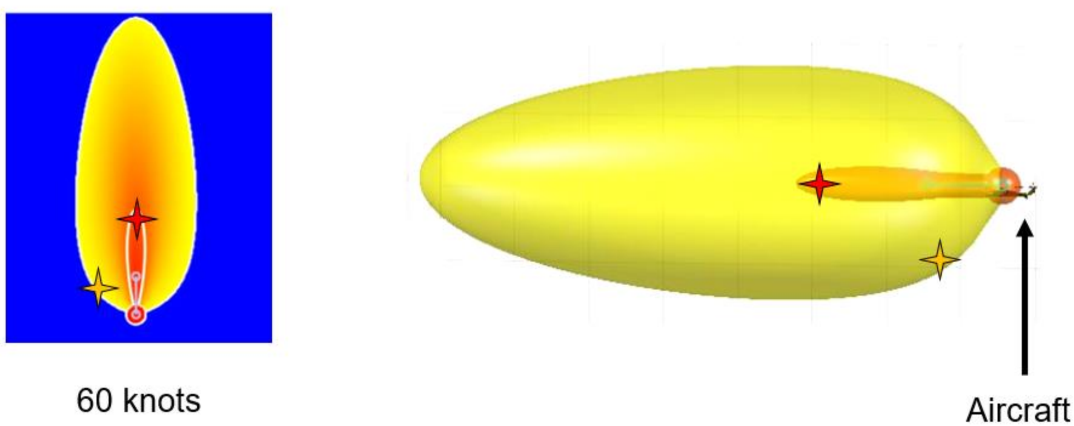


Figure B16. Obstacle threat assessment.

Note. A static-obstacle threat assessment ranks radar hits in urgency. The obstacle *threat space* cueing regions are fixed and spherical at low speeds and extend in the direction of the velocity vector as speed increases. The two most urgent obstacles (Urgent1 and Urgent2) within the threat space are presented.

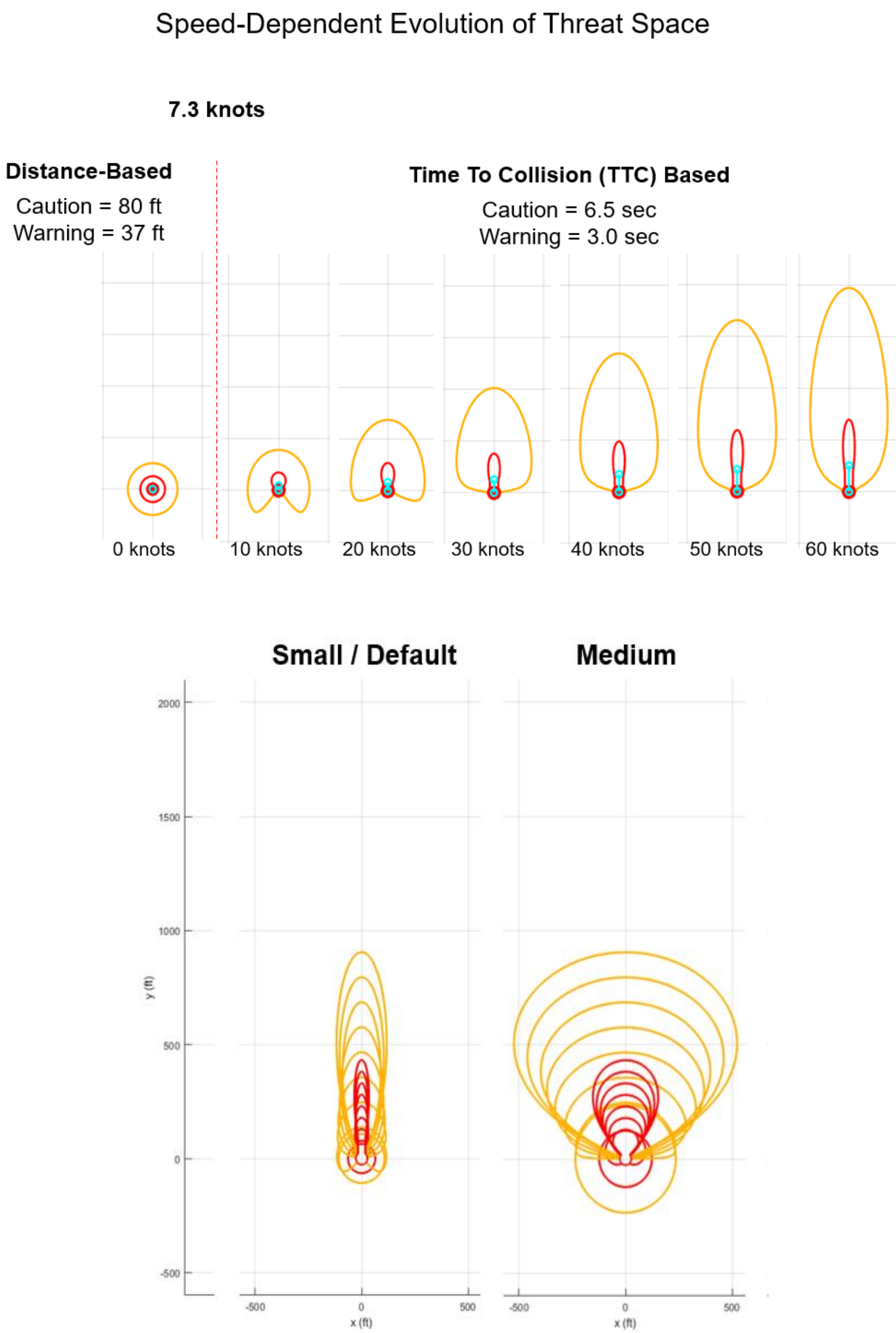


Figure B17. Threat space shape.

Note. Two threat spaces based on cueing sensitivity (0 to 80 every 10 knots).

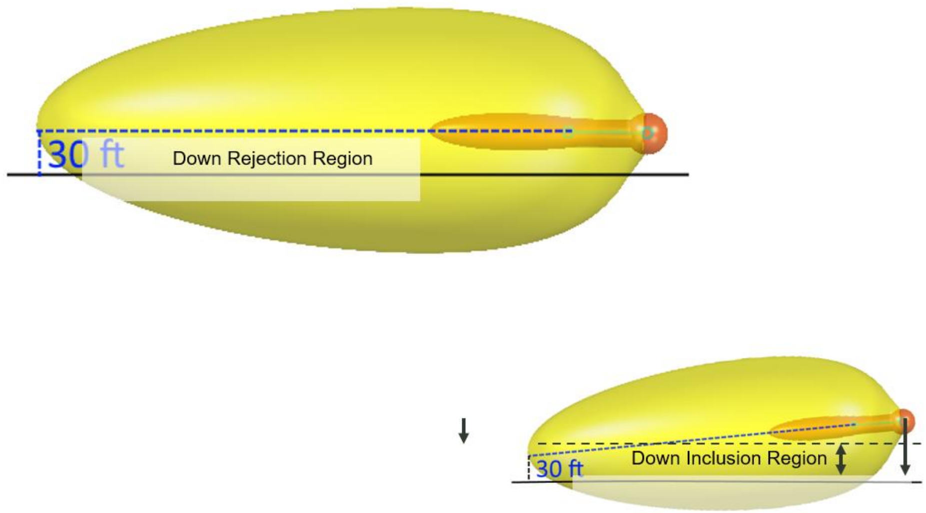


Figure B18. Threat space down inclusion.

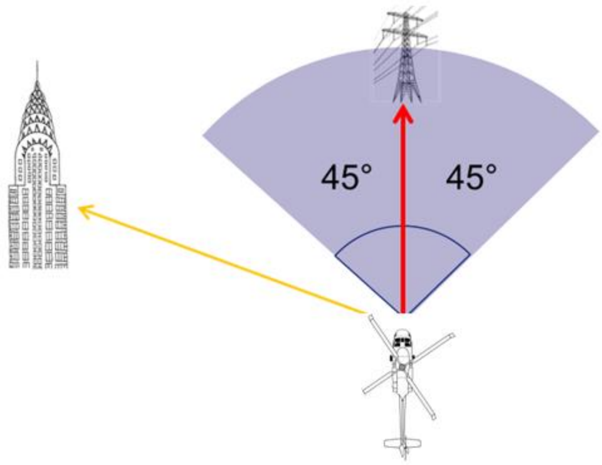


Figure B19. Cueing multiple obstacles.

Note. Angular rejection: ± 45º
An azimuth angular region about the most-urgent hit to omit from the search for the second-most urgent hit. Helps prevent multiple hits on the same obstacle and cueing the same region of space twice.

This space is intentionally blank.

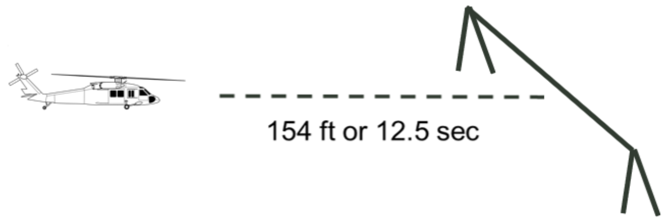


Figure B20. Power line database obstacle.

Note. Power line tower locations are stored in a database (they can also be detected by radar). The nearest power line segment is selected for cueing.
Below 7.3 knots, the cue occurs 154 ft from blades, above 7.3 knots, the cue is presented with a 12.5 second TTC.

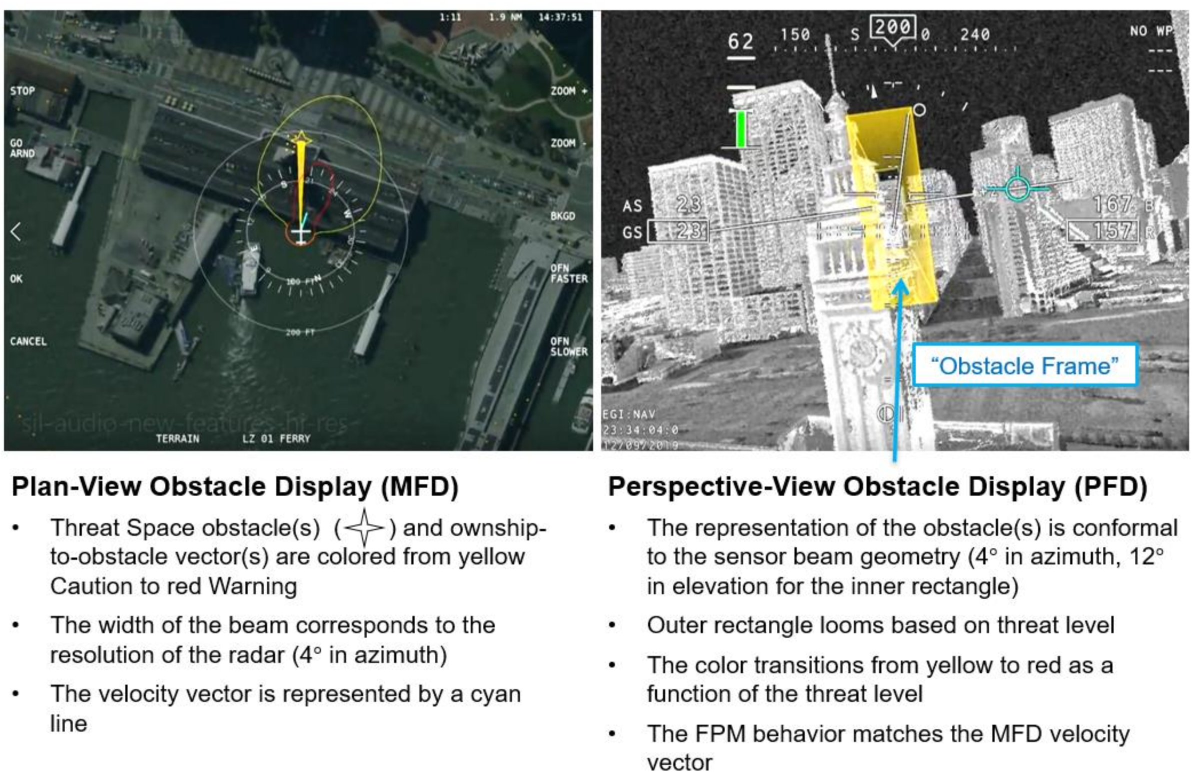


Figure B21. Bumper radar obstacle display: visual.

This space is intentionally blank.

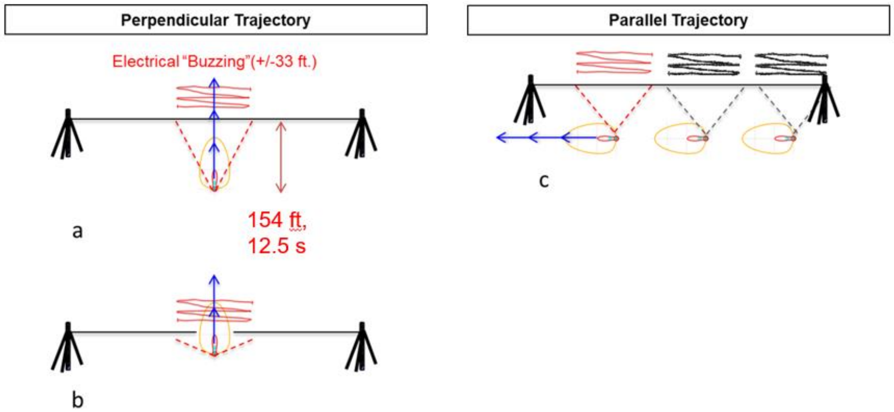


Figure B22. Power line obstacle display: auditory.

Note. The power line cue enables at 154 ft or 12.5 seconds (> 7.3 knots). A recording of a power line is used for the auditory icon. The icon's spatial location is swept up and down the power line.

Appendix C. Description of Training

Training on the ICE cueing system took place over the course of two days when possible; however, if a participant was unable to commit to two days of training, then procedures from days one and two were combined. Training activities are summarized in the table below. The initial training session consisted first of 90 minutes of slides with Flight Systems Branch research pilots covering all aspects of the ICE system (visual symbology, audio cueing and tactile cueing). After the slides were complete, a knowledge check was conducted. The participant then moved on to the initial hands-on portion of the training where they would fly through three training routes on a desktop trainer. The training routes were similar in duration and difficulty to the those completed on the testing day. A research pilot and/or a member of the research team was present to guide the participant through the route and answer any questions they had. Upon completion of the training routes on the initial visit, participants would then complete the Self-Assessment Training Survey to gauge their comfort and understanding of the various aspects of the cueing system. Following the survey, the initial training visit ended, and participants returned the next day to conclude the training portion of the study. On the second, training session, the participant was offered the opportunity to review the training slides before moving into the simulator to fly through the training routes an additional time with instruction from the research pilot. Additional iterations of the training routes were completed, as needed, or requested. After the participant completed sufficient training routes in the simulator, they completed a final novel training route. During this route, the research pilot filled out the Research Pilot Training Survey to document whether the participant was thoroughly trained to move on to the testing portion of the study. All participants were cleared after completing the final training route. Upon completion of the final training route, the participant flew an additional training flight to practice Crew Status Survey (CSS) workload ratings with the scale provided as a reference. The participant was instructed to update their workload rating throughout the duration of the route and their previous workload rating was displayed on the panel-mounted display as well as on the out-the-window view of the simulator. Participants completed the Self-Assessment Training Survey and were then cleared to move on to the final testing session.

This space is intentionally blank.

Table C1. Approximate Time Requirements for Activities

Activity	**Description**	**Approximate time**	**Team member**	**Session**
Classroom training/refresher for subsequent visits	Overview of slides	up to 90 minutes (30 minutes for refresher)	Research Pilot	Initial (refresher as needed)
Knowledge check/evaluation	Understanding of basic symbology	10 minutes	Research Pilot	All
Application of EEG	Application of EEG device to capture cortical activation patterns	45 minutes	Technician	Initial, final
Calibration of eye tracking system (glasses or desktop mounted system)	Calibrate camera(s) to eye movement	1-5 minutes	Technician	Initial, final
Hands-on training	Fly 3-4 training scenarios using advanced pilot cueing system	60-120 minutes	Research Pilot or Technician	All
Self-Assessment Training Survey	Participant assessment of comfort and knowledge with cues	2 minutes	Technician	All
Instantaneous Self-Assessment of Workload	Participant assessment of workload manipulations		Research Pilot or Technician	Final
Total time anticipated				Up to 5 hours

U.S. ARMY AEROMEDICAL RESEARCH LABORATORY

FORT NOVOSEL, ALABAMA

Optimizing

HUMAN PROTECTION AND PERFORMANCE

since 1962

All of USAARL's science and technical informational documents are available for download from the Defense Technical Information Center.

https://discover.dtic.mil/results/?q=USAARL

U.S. ARMY

FORGE THE FUTURE

FUTURES COMMAND

MRDC