USAARL-TECH-TR--2025-40

UNITED STATES ARMY AEROMEDICAL RESEARCH LABORATORY

Pilot Performance in Multi-Talker Environments: Effects of 3D Audio and Active Noise Reduction

Jennifer Noetzel, Paula Henry, Ryan Mackie, JR Stefanson,
J. Kyle Hale, Kevin Andres, Danielle McDermott, Lauren Brunner,
& Heath Jones

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 Subject Use

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) and 32 CFR 219 (Protection of Human Subjects).

IRB Determination and Number

This study, USAARL 2022-016, was approved by the Medical Research and Development Command Institutional Review Board (IRB Protocol Number M-11063) on 30 March 2024.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)* 18-08-2025	**2. REPORT TYPE** Technical Report	**3. DATES COVERED** *(From - To)* 10/10/22-03/31/2025

4. TITLE AND SUBTITLE Pilot Performance in Multi-Talker Environments: Effects of 3D Audio and Active Noise Reduction	**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Noetzel, J.[1], Henry, P.[1,2], Mackie, R.[1,3], Stefanson, JR[1], Hale, J. K.[1,2], Andres, K.[1,2], McDermott, D.[1], Brunner, L.[1], & Jones, H.[1]	**5d. PROJECT NUMBER** M220077 **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 Rucker, AL 36362	**8. PERFORMING ORGANIZATION REPORT NUMBER** USAARL-TECH-TR--2025-40
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]Chickasaw, [3]Oak Ridge Institute for Science and Education

14. ABSTRACT

This study investigates the impact of three-dimensional (3D) spatial audio and active noise reduction (ANR) on aviators' cognitive workload and flight performance when monitoring multiple radio channels. Laboratory experiments demonstrated that 3D audio significantly enhanced speech recognition, particularly under high auditory workload, and reduced perceived workload, as confirmed by pupil dilation measurements. While increased auditory workload negatively impacted speech recognition, simulator-based experiments revealed no statistically significant differences in speech recognition across listening conditions. However, pilots showed a strong preference for 3D audio, indicating its potential for improving aviator performance and reducing cognitive burden. The study recommends prioritizing the integration of 3D audio into aircraft and suggests further research into optimizing ANR technologies for aviator headsets.

15. SUBJECT TERMS

spatial audio, 3-dimentional audio, active noise reduction (ANR), pupillometry, listening effort

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Loraine St. Onge, PhD
UNCLAS	UNCLAS	UNCLAS	SAR	55	**19b. TELEPHONE NUMBER** *(Include area code)* 334-255-6906

This page is intentionally blank.

Summary

Background

Aviators often operate in challenging listening conditions, regularly requiring a high level of sustained attention and cognitive effort. Implementing spatial audio, or three-dimensional (3D) audio, which distributes audio signals in virtual space around the listener, and utilizing active noise reduction (ANR), can potentially improve communication and reduce cognitive workload. These technologies aim to enhance situational awareness, optimize crew performance, and mitigate aviator cognitive workload in the context of Future Vertical Lift (FVL) aircraft operations, which demand advanced capabilities for high-altitude desert plateaus and urban canyons of megacities.

Purpose

The purpose of this study was to evaluate the use of 3D audio for radio communication and assess the possible synergistic effects with ANR on aviator performance in a UH-60 helicopter simulator. The experiment also evaluated objective and subjective measures of cognitive workload associated with the increased communication demands on the operator.

Methods

Both laboratory- and flight simulator-based experiments were conducted. Participants were asked to listen to multiple audio streams simulating different radio communication channels and to identify instructions specific to their callsign. For the laboratory-based experiments, the complexity of the audio streams increased as the experiment progressed by increasing the number of different voices in the stream to simulate a busy aviation radio environment. For the flight simulator-based experiments, flight missions were flown under heavy radio communications (i.e., radio communications on up to five channels were monitored). Speech recognition was scored for each participant for each listening condition.

Conclusions

Findings validated that increasing the number of auditory streams increased the perceived cognitive workload. Furthermore, the resultant increased perceived cognitive workload correlated with decreased speech recognition performance. However, pupillometry measures in the dynamic flight simulator environment did not produce anything consistent or significant due to the noisy nature of the data. Additional, more complex analyses will be required to investigate whether the presence of patterns indexing cognitive effort will emerge.

This page is intentionally blank.

Table of Contents

Table of Contents (continued)

List of Figures

Table of Contents (continued)

List of Figures (continued)

Page

List of Tables

This space is intentionally blank.

Introduction

As the Army advances its force-modernization priorities, the development of Future Vertical Lift (FVL) aircraft has become an area of concentrated effort. Tremendous focus placed on extending the reach of current rotary-wing aircraft and operating over new terrains will allow completion of novel mission sets. Given the ambitious flight profile of FVL aircraft operations (i.e., high-altitude desert plateaus and the urban canyons of megacities), it is imperative that the aviator is provided with state-of-the-art technologies and capabilities aimed at maintaining an operator’s situational awareness, enabling safe operations, optimizing performance, and potentially mitigating high cognitive workload. Two such technologies include three-dimensional (3D) spatial audio, or 3D audio, and active noise reduction (ANR). 3D audio uses signal processing to simulate over headphones how humans perceive sound in three-dimensional space, making it appear as if sounds are coming from different directions. ANR uses electronics in the headset to generate an anti-noise of equal amplitude and opposite phase and combines it with the primary noise, thus resulting in the cancellation of both noises. The present report documents an investigation into the impacts of these technologies on cognitive workload and flight performance among aviators monitoring multiple radio channels during simulated flight.

Rotary-wing aviators must monitor, and subsequently respond to, overlapping communication from up to four external radios as well as the internal communication system amidst hazardous ambient noise. In combat environments, as many as seven communication streams can be monitored simultaneously and continuously to perform coordinated activities such as search and rescue, casualty evacuation, or air traffic control. As such, aviators may have a difficult time understanding instructions given over the radio. They may also experience a mentally taxing endeavor, where they may be straining to discern incoming communication (i.e., exerting increased listening effort) while flying the aircraft. Given this, the goal of the present study was to assess the impact 3D spatial audio has on understanding several overlapping audio streams, which occurs when monitoring multiple radio communication channels. Performance was measured in both a laboratory environment and within a full-motion UH-60 Black Hawk helicopter simulator environment. Additionally, the flight simulator-based portion of the study also aimed to uncover any potential synergistic effects when 3D audio is combined with ANR technology.

Listening Effort and Cognitive Workload

Listening effort refers to the mental effort required to successfully perceive, cognitively process, and respond to auditory information (Zekveld et al., 2010). It is the deliberate allocation of mental resources to overcome obstacles in understanding speech or carrying out a listening task (Pichora-Fuller et al., 2016). Although speech recognition signifies simply understanding the context of speech, listening effort is about the mental resources expended to achieve that understanding. For example, when you are in a crowded room, you can likely understand what is being said, but filtering out the surrounding conversations and background noise requires increased listening effort. Factors like background noise (Picou et al., 2013), reverberation (Huang et al., 2022), and hearing loss (McCoy et al., 2005) all contribute to increased listening effort. Military aviators, often operating in noisy cockpits with communications from multiple sources, can experience increased listening effort, potentially impacting the cognitive resources they have available for critical flight tasks.

Researchers employ various methods to measure listening effort:

- Subjective measures: Directly asking individuals to rate their perceived listening effort (McGarrigle et al., 2014; Sheffield et al., 2017; Noetzel et al., 2025).
- Physiological indicators: Monitoring bodily changes like heart rate, pupil dynamics, electroencephalography, and skin conductance that correlate with effort (Bianchi et al., 2019; Zekveld et al., 2010; Winn et al., 2018)
- Behavioral paradigms: Assessing response time to secondary tasks and recall abilities, based on the principle that allocating cognitive resources to listening can impact performance on other tasks (Kahneman, 1973).

Researchers have long employed subjective measures and behavioral paradigms to investigate the concept of listening effort. Subjective measures, such as questionnaires or rating scales, directly ask individuals to report their perceived listening or workload effort. Behavioral paradigms, often involving dual-task scenarios, objectively measure how listening efforts impact performance on concurrent tasks. For example, slower reaction times on a secondary task while listening to speech in noise can indicate higher listening effort. Studies with military service members using these methods have shown that factors like hearing loss and high workload conditions can increase listening effort (Noetzel et al., 2025; Sheffield et al., 2017).

To provide a more objective assessment of listening effort, researchers have investigated physiological measures such as pupillometry. Pupillometry is the measurement of pupil size and reactivity and has been used in a variety of listening tasks to study its effect on speech recognition (Winn et al., 2018), listening effort (Zekveld et al., 2010), different acoustic masker types (Koelewijn et al., 2012; Zekveld et al., 2014), and cognitive function (Zekveld et al., 2011). Pupil dynamics (PD) refer to changes in pupil size that occur over time in response to changes in attended stimuli but are absent with changes in stimuli that are unattended. Such responses have also been associated with measures of memory load (Kahneman & Beatty, 1966), selective attention (Hillyard et al., 1973), motivation (Kahneman et al., 1968), and linguistic coherence of stimuli (Schluroff, 1983).

The pupillary response corresponds to an intuitive understanding of changing cognitive demands. For example, there is an increase in PD when tasked to memorize a long string of digits, compared to memorizing a shorter string. PD is also increased when participants exert greater effort to solve arithmetic problems (Kahneman et al., 1968). Conversely, decreased pupil size represents a reduction in effort. Bianchi et al. (2019) demonstrated that in individuals with hearing loss, using a new hearing aid technology results in a 36% decrease in pupil size as well as an increase of 21% in speech recognition scores. Pupillometry is a sensitive tool that can be used to measure differences across and within individuals; however, it is important to note that changes in signal quality for individuals could potentially result in differences in PD within the same experiment. Regardless, pupillometry has steadily become a well-studied measurement tool for indexing cognitive processing effort or cognitive workload.

More globally, cognitive workload refers to the amount of mental effort required to complete a single or multiple tasks. Given this and as mentioned above, an increased allocation of mental resources to the task of hearing is referred to as an increase in ‘listening effort.’ The present study aimed to capture increases in both cognitive workload and listening effort subjectively, by administering the National Aeronautics and Space Administration Task Load Index (NASA-TLX) and objectively, by measuring physiological indicators (i.e., PD) and a behavioral dual task paradigm. It was hypothesized that the signal processing techniques discussed, (i.e., 3D audio and ANR) increase speech recognition performance of overlapping radio communications and reduce both listening effort and cognitive workload in aviators compared to the standard split monaural configuration.

Spatial Audio (3D Audio)

Spatial audio, or 3D audio, uses signal processing to simulate over headphones how humans perceive sound in three-dimensional space, making it appear as if sound is coming from different directions. Presently, Army aviation personnel use a split monaural (diotic) presentation for radio communication, meaning that the same signal is routed to both ears (see Figure 1, left image). There are multiple problems with diotic presentation of speech signals. First, simultaneous presentation of two or more signals to the two ears masks each other, which makes each signal more difficult to hear. Second, presenting the same signal to both ears cause the perception that all of the signals are in one location, the center of the head, which adds to the difficulty of hearing one signal over another.

To combat sensory and mental overload during flight, aviation has long considered the implementation of 3D audio into cockpit design as this technology creates a natural, ecologically valid, egocentric representation of space where auditory signals behave realistically in terms of direction, distance, and motion. Potential aviation applications of 3D audio include threat location warning, aircraft location indication, collision avoidance, navigation guidance, and spatially separated multi-channel communications.

Previous research has shown that the use of 3D audio leads to an increased understanding of verbal messages over multiple communication channels (Brungart et al., 2002; Drullman & Bronkhorst, 2000; Ericson et al., 2004; McAnally & Martin, 2002) and quicker target acquisition (Bronkhorst et al., 1996; McKinley et al., 1994; McKinley et al., 1997; Simpson et al., 2005; Veltman et al., 2004). In addition, 3D audio displays aid in threat/collision avoidance (Begault & Pittman, 1996; Hartnett et al., 2020) and general navigation (Foyle et al., 1996; Milam et al., 2019). Despite the long-term interest in this technology, Army aircraft are not currently equipped with spatial audio displays due to the limitations in their current wiring configuration.

This space is intentionally blank.

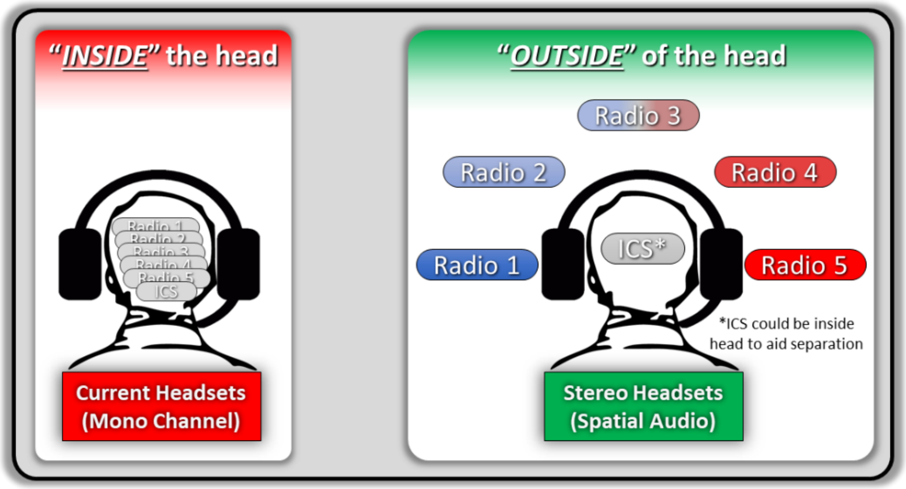

Figure 1. Graphic representation of current monaural radio headsets and potential stereo radio headset configurations.

Previous research demonstrates listening to more than one channel of speech presented diotically (i.e., inside the head) results in poor speech understanding (Abouchacra et al., 2001; Brungart at al., 2002; McKinley & Ericson, 1997; Kim et al., 2018). Brungart et al. (2002) demonstrated that speech recognition scores decreased when multiple signals were presented simultaneously compared to a single speech signal. When two speech signals were presented simultaneously, performance dropped from > 90% to 62%; when three signals were presented simultaneously, performance dropped to 38% and when four signals were presented simultaneously, performance fell below 25%. Given this significant drop in performance with the addition of speech signals, improvement would need to be accomplished by either reducing the number of speech signals presented or changing the delivery configuration of those signals.

If a listener needs to monitor two channels, the first improvement could be to direct one channel to one ear and the second channel to the other ear in a stereo configuration, also referred to as dichotic. Presentation through a dichotic configuration would improve the understanding of two speech signals by about 10-20% over diotic presentation (Ericson & McKinley, 1997); however, this improvement declines when needing to monitor three or more channels. Achieving further improvements to speech understanding for multiple channels is possible using spatial audio presentation.

Active Noise Reduction (ANR)

A second and complementary strategy to using 3D audio for improving speech communications in the cockpit is to integrate ANR into the headset. To mitigate noise hazards, Army aviators are required to use double hearing protection and often use the communications earplug (CEP), a device that couples passive hearing protection to pass-through communications. ANR offers a signal processing strategy in avionics headsets that reduces hazardous low-frequency noise; however, this technology is not yet in use in the current Army communication headsets. Mozo and Murphy (1997) demonstrated improved speech recognition with five commercially available ANR systems that were integrated into the Head Gear Unit-56/Personal

(HGU-56/P) aviator helmet over the standard HGU-56/P helmet configuration. Furthermore, Casto and Casali (2013) demonstrated the use of ANR through improved performance of flight-related tasks and decreased workload. Casto and Casali (2013) also showed that Army aviators with hearing loss performed better with ANR than with passive-earplugs paired with passive headsets.

The traditional passive earmuffs are valued for their high attenuation over a broad frequency range; however, they are less effective at low frequencies. ANR systems reduce the unwanted noise based on the principle of superposition (Hansen et al., 1997; Kuo & Morgan, 1996). Specifically, an anti-noise of equal amplitude and opposite phase is generated and combined with the primary noise, thus resulting in the cancellation of both noises. The ANR system efficiently attenuates low frequency noise, where passive methods may be less effective. A potentially better approach to using passive muffs or ANR alone is to use a combination of the two (i.e., providing ANR in addition to passive muffs and the CEP).

The combination of 3D audio and ANR in military aviation, particularly rotary-wing aircraft, is less studied. One study found ANR to be beneficial on speech recognition aboard fixed-wing aircraft (Bolia, 2003), and Ray et al. (2022) showed potential benefits of 3D audio on speech recognition. Ray et al. (2022) also found a significantly higher signal to noise ratio with a 3D audio and ANR audio system, suggesting improved speech intelligibility. A research gap exists concerning potential interaction effects of combining 3D audio and ANR in rotary-wing aircraft. Therefore, this study investigated the combined effects of 3D audio and ANR on speech understanding, performance of flight-related tasks, cognitive workload, and listening effort on aviators during simulated flight. The short-term outcome of this research is to describe the impact 3D audio and ANR technologies have on speech understanding and workload amidst multiple auditory streams (i.e., radio communication channels). The long-term outcome is providing a recommendation to guide the design of future Army aviation headsets. Below is a quick summary of the experiments conducted:

Laboratory-based Experiment I

This experiment aimed to quantify the potential benefits (i.e., improved performance and decreased workload) of 3D audio in managing multiple auditory inputs. It consisted of a speech recognition test where participants were presented with differing numbers of audio streams simultaneously (i.e., from 2 to 5 different streams) with and without 3D audio applied.

- Hypothesis 1a: Speech recognition will be better in the 3D spatialized listening conditions compared to the non-spatialized conditions.
- Hypothesis 1b: Listening effort will be less in the 3D audio listening condition compared to the standard split monaural listening condition.

Simulator-based Experiment II

This experiment aimed to assess the impact of 3D audio and ANR on speech understanding, pilot performance, workload, and preferences, while determining the compatibility of these technologies. It consisted of pilots performing simulated flights with combinations of 3D audio and ANR technology configurations compared to flights without these technologies (i.e., current standard equipment).

- Hypothesis 2a: Using ANR and 3D audio will result in better flight-related performance measures than in the current CEP headset platform.
- Hypothesis 2b: Using ANR and 3D audio will result in better subjective measures of workload than in the current CEP headset platform.

The current study was reviewed and approved by the U.S. Army Medical Research and Development Command Institutional Review Board (MRDC-IRB Protocol M-11063) prior to execution. Participants were compensated for their participation (if participating in an "off-duty" status) and provided informed consent prior to study enrollment.

Study participants were current and qualified helicopter pilots with normal hearing, defined as thresholds within class 1 criteria of the Army Aeromedical Policy Letter (2021) (≤ 25 decibels of hearing level (dB HL) at 500, 1000, and 2000 hertz (Hz); ≤ 35 dB HL at 3000 Hz; and ≤ 45 dB HL at 4000 and 6000 Hz). Participants were recruited from the Fort Novosel, AL area. Exclusion criteria included hearing thresholds exceeding these limits, interaural asymmetry exceeding 20 dB at any frequency, ear abnormalities, alcohol consumption within 3 hours of testing, use of over-the-counter medications or anesthesia within 24 hours of testing, and non-native English speakers.

Methods - Laboratory-based Experiment I

Participants

A total of 21 pilots (all male) completed the first laboratory-based experiment. The average age was 39.1 years (*SD* = 7.3 years), and average flight time was 2389.3 hours (*SD* = 1766.7 hours). Twenty participants who completed the first experiment also completed the second simulator-based experiment. These participants reported their hearing as "good" or "excellent," while only one participant selected "a little trouble" when describing their hearing abilities.

Materials

Speech recognition test.

Participants performed a laboratory-developed word recognition task delivered over headphones in a sound booth. Multiple communication channels were simulated by systematically overlapping multiple audio streams (i.e., from two to five talkers). The *target* audio stream was layered to always be presented while all the *masking* audio streams were playing (i.e., if three streams were being presented, one being a *target* with two maskers, at least one *masker* would start before the *target* and end after the *target* ended). Two configurations (a monoaural [mono] signal and a 3D audio condition) were tested for each number of audio streams presented (2, 3, 4, and 5 streams). MATLAB was used to present the auditory stimuli to the listener over headphones. The stereo signals were passed through a set of generic head-related transfer function (HRTF) filters and presented to the participant over headphones. The HRTFs used for all participants were supplied by MATLAB's Audio Toolbox. Placement of audio streams in virtual space is detailed in Figure 2. Estimated presentation level under headphones was 87 dB sound pressure level (SPL).

The speech recognition task consisted of modified aviation-relevant phrases. An example of this modified aviation phrase is, “Grey Eagle 28, this is Sandman 76. Set heading to three two zero degrees.” This structure was used for all audio streams within each trial. Callsigns, numbers, instructions, and responses for calls serving as maskers were all randomized when presented within each trial. Target audio calls were the same for each participant.

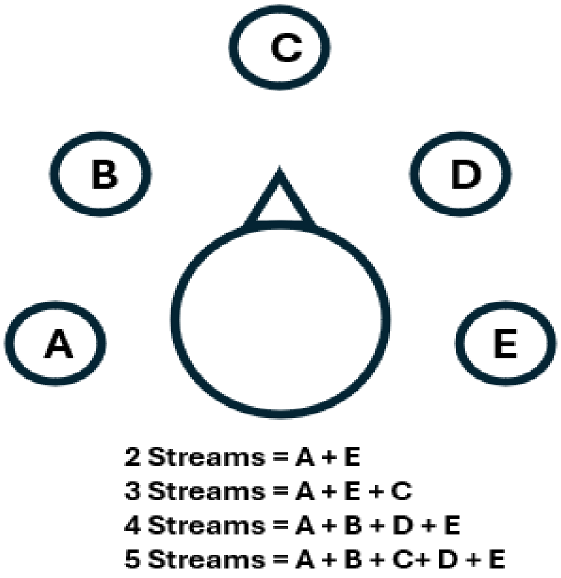

Figure 2. Virtual talker locations for 3D audio conditions.

Participants were required to repeat the target audio stream identified by their designated callsign. Speech recognition was scored by a researcher seated in the booth with the participant as verbal responses for five key components or numbers in each target phrase. These key components were: 1) the callsign of the sender (Sandman), 2) the first number associated with the callsign (7), 3) the second number associated with the callsign (6), 4) the requested action (set heading), and 5) the value requested (300 degrees). The speech recognition measure was calculated as the overall percent correct of key components out of all possible trials for a particular condition. See Table 1 for examples of scoring criteria.

This space is intentionally blank.

Table 1. Example of Scoring Breakdown for Laboratory Word Recognition Trials

Target Audio Call	**Scoring Criteria**				
	Callsign	**#**	**#**	**Instruction**	**Value**
Grey Eagle 28, this is **Titan 54.** Set **airspeed** to **100 knots**	Titan	5	4	airspeed	100 knots
Grey Eagle 28, this is **Shadow 69.** Set **altitude** to **16000**	Shadow	6	9	altitude	16000
Grey Eagle 28, this is **Big Brother 99.** Set **heading** to **0 degrees**	Big Brother	9	9	heading	0 degrees

Workload measurements.

Subjective workload.

The NASA-TLX is a subjective questionnaire that has the participant rate workload based on several aspects. It was used to assess participants' perceived workload based on six criteria, mental demand, physical demand, temporal demand, effort, frustration, and performance. For each of these categories, a 10-point scale was used with verbal anchors at the beginning and ending of the scale (e.g., low or poor at the beginning and high or good at the end of the scale). Participants were asked to rate their perception, using a slider scale on an iPad, for each of the categories at the completion of each condition. The NASA-TLX was tablet-based and was used during both the laboratory- and simulator-based experiments. The response scale ranges from 0 to 100 in increments of 5, essentially a Likert scale with 21 levels.

This space is intentionally blank.

Table 2. Descriptions of the NASA TLX Rating Scale Definitions (Hart, 1986)

Rating Scale Definitions		
Title	Endpoints	Descriptions
Mental Demand	*Low/High*	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	*Low/High*	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	*Low/High*	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow?
Performance	*Good/Poor*	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	*Low/High*	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration Level	*Low/High*	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

Objective workload.

Pupillometry, or the measure of pupil dilation, is a reliable index of cognitive load and has been used in a variety of listening tasks to study how listening effort is effected by speech recognition ability (Zekveld et al., 2010), varying acoustic maskers (Koelewijn et al., 2012; Zekveld & Kramer, 2014), and cognitive function (Zekveld et al., 2011). PD refers to changes in pupil size. Changes in PD occur in response to changes in attended stimuli, but the PD are not observed in changes to unattended stimuli.

PD were continuously recorded to measure cognitive workload using an Eyelink 2000 system (SR Research, LTD). This system employs an infrared emitter and camera to track the x-y positions and measure the pupil diameter of each eye. The camera was positioned approximately 50 centimeters (cm) in front from the participant and provided no interference with the tasks to be completed in this study. The camera systems required an initial calibration for each condition group. The calibration procedure is 1 minute in duration and requires participants to fixate on a series of dots or landmarks located in the primary field of view. The host computer handled the camera operation, data collection, and storage. Data acquisition was performed on a Dell Alienware desktop computer with MATLAB 2016b and PsychToolbox3 installed. Custom MATLAB code was created to generate the auditory and visual stimuli as well as record the participant's responses for scoring and future analysis.

Procedure

All participants provided written informed consent after which they were screened for inclusion and exclusion criteria to include hearing threshold levels. After eligibility was determined, participants completed the speech recognition test in a sound-isolation booth.

First, instructions were provided on the speech recognition test (see Table 1 for scoring examples). During the speech recognition test, pupillometry data was collected with the Eye Link system. Each participant sat in a height-adjustable chair, was fitted with open-back stereo headphones, and was asked to place their head in the head rest such that their forehead was against a brace and their chin was on a chin rest for the entirety of the experimental trials. Breaks were offered between trials. Eye tracking calibration occurred before every condition or when a participant's head was displaced from the head rest. During the speech recognition test, participants were instructed to look at a fixation point; a red cross was at the center of the screen, which aided in maintaining eye stability for more reliable pupillometry measurements. Once test trials started, participants were asked to maintain their gaze on the fixation cross while listening to the simultaneous audio streams and their designated callsign. Participants were instructed to wait until the fixation point changed from red to green (wait duration = 2 seconds) before responding to their callsign. Changes in pupil dilation were measured to monitor the listening effort.

Participants began with two simultaneous audio streams completing both mono and 3D audio conditions, and then progressed to three, four, and five simultaneous audio streams. Mono or 3D audio starting conditions were counterbalanced. Each condition presented 20 trials or sentences. In total, participants completed eight conditions (2, 3, 4, & 5 audio streams in both mono and 3D audio conditions). For every trial there was only one target audio while maskers were randomized. In the spatialized condition, the target call was randomized in terms of its 3D placement in virtual space (see Figure 2). At the completion of each condition, the NASA-TLX was administered via tablet.

Statistical analysis.

A 2 x 2 x 4 mixed factorial experimental design was used to assess speech recognition and workload with and without 3D audio. The within-subjects factors were listening condition (monaural vs. 3D audio) and the number of simultaneous audio streams (2, 3, 4, 5). The between-subjects factor was first listening condition (monaural or 3D audio). The dependent variables assessed were speech recognition scores and measures of workload as assessed using the NASA-TLX and PD. All statistical analyses were completed using R and R Studio with the following packages: tidyverse, rstatix, lmerTest, and emmeans. Statistical significance level was set at 0.05 for all statistical analyses. For pairwise comparisons, p-values were corrected for multiple comparisons using the Benjamini-Hochberg method to control the false discovery rate and balance controlling for Type-1 and Type-2 errors.

Speech recognition.

Percent correct was the dependent variable used for analysis. To simultaneously test listening condition and number of audio streams while also accounting for repeated measures, a mixed-effects linear regression model was used. The regression model contained fixed effects for listening condition (categorical: Mono, 3D), number of audio streams (categorical: 2, 3, 4, 5), and the interaction effect, and a random intercept for each participant. All data was visualized prior to analysis. The regression model was validated by checking normality and homoscedasticity of residuals. A Type-III analysis of variance (ANOVA) was used to test overall significance of the regression model fixed effects. Significant results for the interaction effect or number of audio streams were followed up with relevant pairwise comparisons.

Workload.

Subjective workload.

The workload difference between the two listening conditions (mono and 3D audio) for each participant and number of audio streams (2–5) was calculated and used in the analysis. The NASA-TLX subscale scores were evaluated by testing for a significant difference between the two listening conditions (i.e., mono vs. 3D audio). Difference scores were calculated by subtracting the 3D audio score from the monaural score for each of the number of streams. Wilcoxon signed-rank tests were used to determine if they were significantly different from zero for each level of audio stream. The non-parametric Wilcoxon test was chosen due to the nature of the scores being non-continuous. The six NASA-TLX sub-scales were analyzed as independent data. Four Wilcoxon tests (one for each of the audio streams) were analyzed for each NASA-TLX metric, and p-values were corrected for multiple comparisons using the Benjamini-Hochberg method.

Objective workload.

Pupil data were processed to remove blinks, other artifacts, and smooth out the data to obtain a rolling average pupil size. The audio event was divided into three phases for analysis: trial start, stimulus start, and response start (see Table 3 and Figure 3). PD during the stimulus start phase, when participants were actively engaged in listening (Bianchi et al., 2019) were modeled over time. These models were then compared between the 3D audio and mono conditions to evaluate differences in workload.

This space is intentionally blank.

Table 3. Laboratory Audio Stimulus Segmentation for Analysis

Audio Event Phase	**Event Description**
Trial Start	a. Computer displays a red cross b. 2 second duration
Stimulus Start	a. Simultaneous audio streams are played b. Audio streams vary in duration ranging from 11-15 seconds c. Include 2 second pause at the end of the audio where participants retain the information and wait to respond
Response Start	a. Computer screen displays green cross b. Participants give response to audio

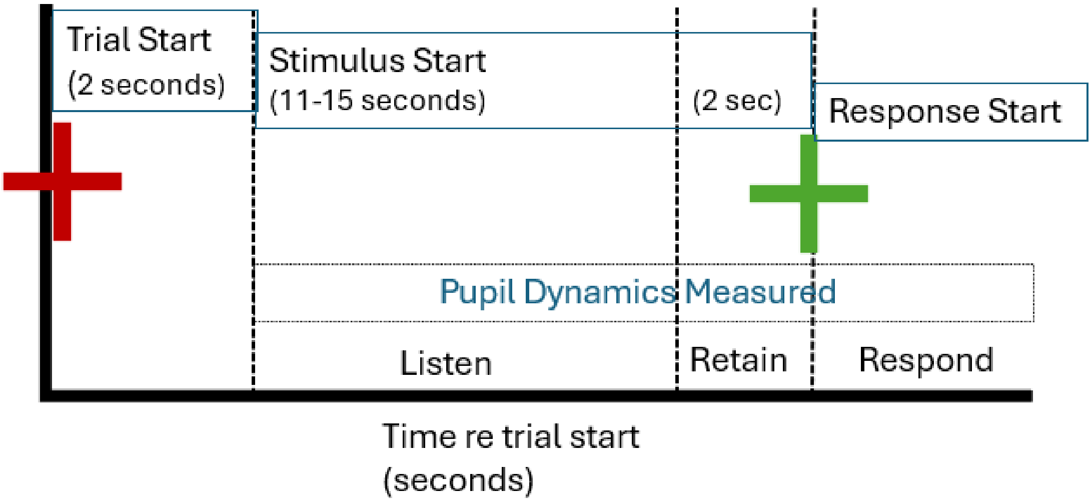

Figure 3. Visual representation of auditory stimulus segmentation.

Pupil data underwent quality control processing before analysis. Initially, data points with poor quality, identified through visual inspection, were removed from analysis. Visual inspection involved plotting and examining data for obvious issues, such as large spikes indicative of unremoved blinks missed in prior processing. In addition to visual inspection, summary statistics were calculated at each time point to flag potentially problematic data. If the data had any of the following issues, it was marked as poor quality: missing pupil size (blink), missing gaze position (looking away from screen), extreme mean absolute deviation (likely a blink), extreme pupil size (> 3 standard deviations from the mean), or extreme gaze position (> 3 standard deviations from the mean). If more than 25% of the data had poor quality, the entire event was excluded from analysis.

Following removal of poor-quality data, pupil sizes were normalized for each individual. Using only the good quality data during the stimulus start phase, the standard deviation of pupil size was calculated for each participant. Each participant's pupil size data was then divided by their respective standard deviation, effectively setting the starting pupil size to zero and transforming the units to "number of standard deviations."

The changes in pupil size over time were analyzed using a generalized additive model (GAM), treating the data as a time series. This approach allowed for the estimation of a line of best fit for the data, accounting for potential non-linear relationships between pupil size and time. GAMs were applied to all participant data for each listening condition to test the hypotheses. Specifically, the hypothesis that workload/listening effort would be less in the 3D audio condition compared to the mono (i.e., non-spatialized) condition was tested by examining whether increases in pupil diameter were smaller in the 3D audio condition compared to the mono condition.

Results - Experiment I

Speech Recognition Test

Participants performed significantly better in the 3D listening conditions than the monaural listening conditions regardless of the number of communication streams presented. Figure 4 shows the task performance as mean percent correct across each of the listening conditions. Performance decreased with the increase in number of audio streams for both listening conditions. The average drop in performance was greater for the mono listening conditions (47%) compared to the 3D audio condition (21%). Interesting to note is the 3D audio condition with five streams (70%) has about the same average speech recognition scores compared to the mono condition with two streams (77%).

This space is intentionally blank.

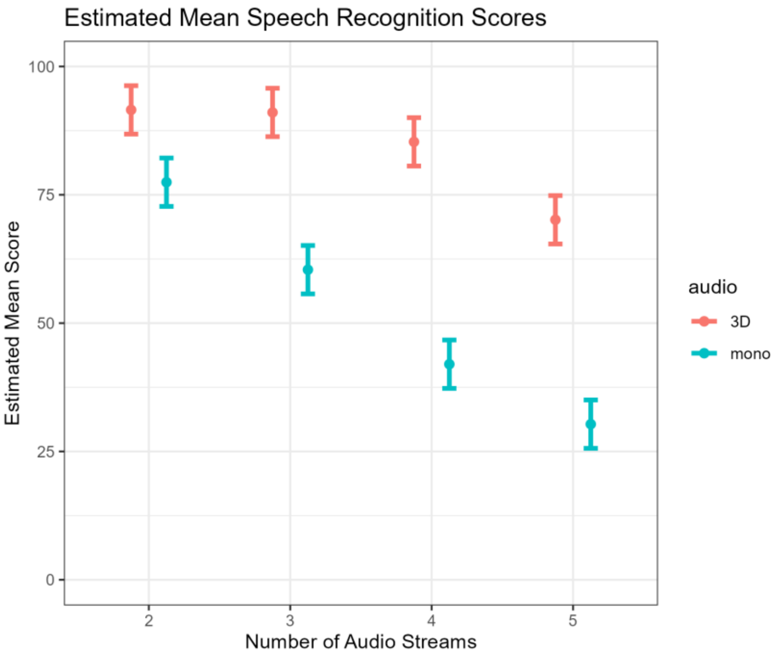

Figure 4. Comparison of mean speech recognition scores between monaural (mono) and 3D audio conditions with increasing number of simultaneous audio streams (2–5) estimated by the regression model. Error bars represent a 95% confidence interval (CI).

A mixed-effects linear regression model was used to analyze the data. An ANOVA table was generated from this model to assess the main effects of number of audio streams and listening condition, as well as their interaction. Results of the ANOVA indicated significant main effects of number of audio streams (*F*(3, 147) = 145.73, *p* < .001) and listening condition (*F*(1, 147) = 671.19, *p* < .001) as well as a significant interaction (*F*(3, 147) = 28.02, *p* < .001). Pairwise comparisons were first made between the two listening conditions within each level of audio streams. Speech recognition performance in the 3D condition was significantly better than the mono condition for all levels of audio streams (2: t(147) = 5.71, *p* < .0001, 3: t(147) = 12.41, *p* < .0001, 4: t(147) = 17.55, *p* < .0001, and 5: t(147) = 16.14, *p* < .0001). Regardless of the number of audio streams, speech recognition performance was always significantly better in the 3D audio listening condition.

Pairwise comparisons were then made between the number of audio streams within each of the two listening conditions. Within the 3D audio conditions, the pairwise comparison between two audio streams vs. three was not significant. All other pairwise comparisons were significant [3 vs. 4, *t*(147) = 2.32, *p* < .05; 4 vs. 5, *t*(147) = 4.73, *p* < .0001]. Speech recognition performance for the 3D audio conditions did not change significantly between two and three audio streams but then decreased steadily with the addition of each additional stream of audio. Within the monaural listening conditions, there was no significant difference between two and three streams, but all other pairwise comparisons were significant [3 vs. 4, *t*(147) = 6.91, *p* < .0001; 4 vs. 5, *t*(147) = 4.73, *p* < .0001].

Subjective Workload

NASA-TLX results.

Figure 5 summarizes the NASA-TLX ratings assigned for each listening condition across the streams of audio from two to five. As shown in Figure 5, ratings of effort, frustration, mental demand, and temporal demand were all higher in the monaural listening condition and increased with increases in the number of audio streams presented. The performance scale provides low ratings when performance is good (perfect) and high ratings when performance is poor (failure). The ratings of performance were higher for the monaural listening conditions indicating that participants viewed their performance as worse in these conditions. Physical demand remained constant across all listening conditions and audio streams, which is logical given this was not a physically demanding task.

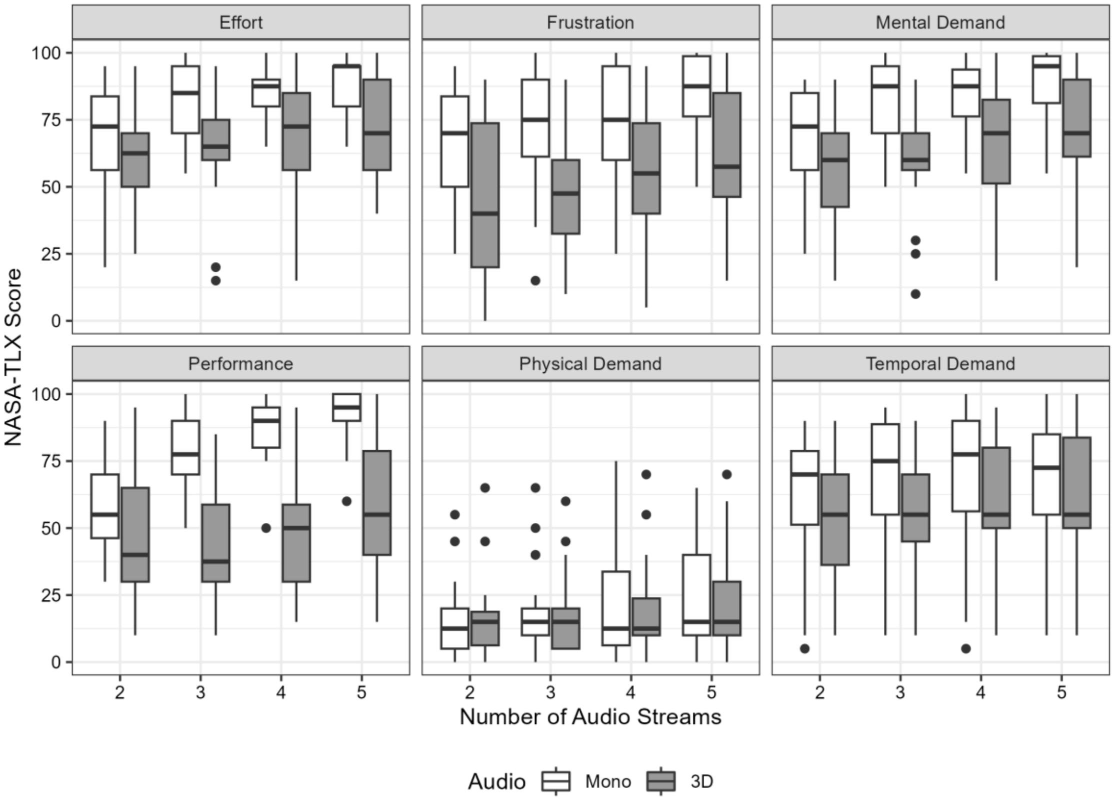

Figure 5. Boxplots summarizing scores from each subscale of the NASA-TLX in monaural (mono) and 3D audio conditions across increasing number of audio streams from two to five. Whiskers extend to the minimum and maximum values within 1.5 times the interquartile range.

Subjective workload was compared between the monaural and 3D audio listening conditions. Results are summarized in Table 4. The 3D audio condition was associated with significantly lower subjective workload compared to the monaural condition across several subscales. These included mental demand (for 2, 3, 4, and 5 audio streams), temporal demand

(3, 4, and 5 audio streams), performance (3, 4, and 5 audio streams), effort (3, 4, and 5 audio streams), and frustration (2, 3, 4, and 5 audio streams). These findings indicate a preference for the 3D audio condition. No significant differences were observed on physical demand (2, 3, 4, and 5 audio streams), temporal demand, performance, and effort subscales when only two audio streams were presented.

Table 4. Wilcoxon Signed-Rank Test Statistics for Each Subscale of the NASA-TLX Across Each Level of Audio Streams

NASA-TLX Subscale	**Audio Streams**			
	2	**3**	**4**	**5**
Mental Demand	171.5*	210.0****	166.0***	207.0***
Physical Demand	64.5	67.0	91.5	59.0
Temporal Demand	129.0	175.0**	174.0*	95.0*
Performance	167.5	253.0****	231.0****	245.5***
Effort	162.0	210.0****	206.0***	120.0***
Frustration	205.0**	205.5***	181.5***	171.0***

Note. Significant values: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$, $^{****}p < .0001$

In summary, the NASA-TLX scores showed significantly lower perceived workload ratings for the 3D audio condition compared to monaural listening when three, four, or five audio streams were presented. This difference was not observed when only two audio streams were presented. The lack of significant differences in the physical demand ratings across the conditions is likely due to the low physical demands of the task itself, because the participants were seated with their head supported by the chin rest. Furthermore, this consistently low physical demand score could add support to the idea that the participants were fully engaged and paying attention when completing the test.

Objective Workload

Pupillometry.

Across all participants, 11% of the events were marked as poor quality by either visual inspection and/or summary statistics. Pupil data was normalized for each participant and GAM regressions were applied to each listening condition for all participants. Example data from one subject is presented in Figure 6.

This space is intentionally blank.

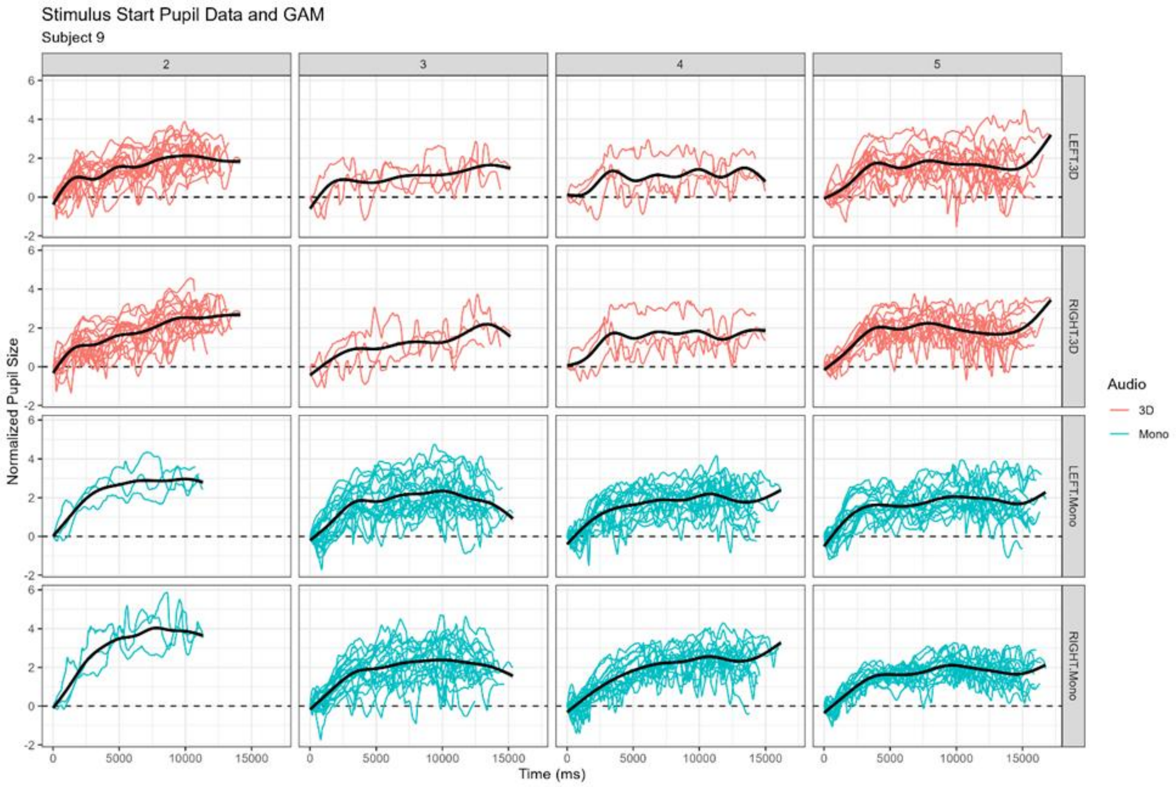

Figure 6. Study participant nine's raw pupil data from stimulus start time with the starting PD set to zero. Raw data is in red for 3D audio and blue for mono. Black lines are the GAM prediction using the raw data.

GAM models for all participants with 95% confidence intervals were constructed. See Figure 7 (top 2 rows). The estimated difference between the two listening condition GAM models (mono and 3D) were calculated.

$$\textit{Difference} = \textit{mono\_estimate} - \textit{3D\_estimate}$$

Figure 7 (bottom 2 rows) represents the calculated difference. The shaded region is the 95% confidence interval for the difference. If the confidence interval does not contain zero, then the two listening conditions (3D and mono) are statistically different at that point in time. Statistical significance was reached in every listening condition. The results support the perceived workload ratings in that increases in pupil diameter were less in the 3D audio listening condition when compared to the mono listening conditions. These results suggest that the 3D audio listening condition requires lower workload (listening effort) when compared to the mono condition. The conclusion also holds across all numbers of audio streams and both eyes.

An important consideration in the analysis is the variation in audio event durations, which ranged from 11 to 15 seconds. Consequently, data points beyond 11 seconds no longer represented all 20 events within a listening condition. Instead, these later time points reflect only the subset of events with durations exceeding that specific time, which may only be one or two events. Therefore, interpretation of the GAM models was limited to the first 10 seconds of data.

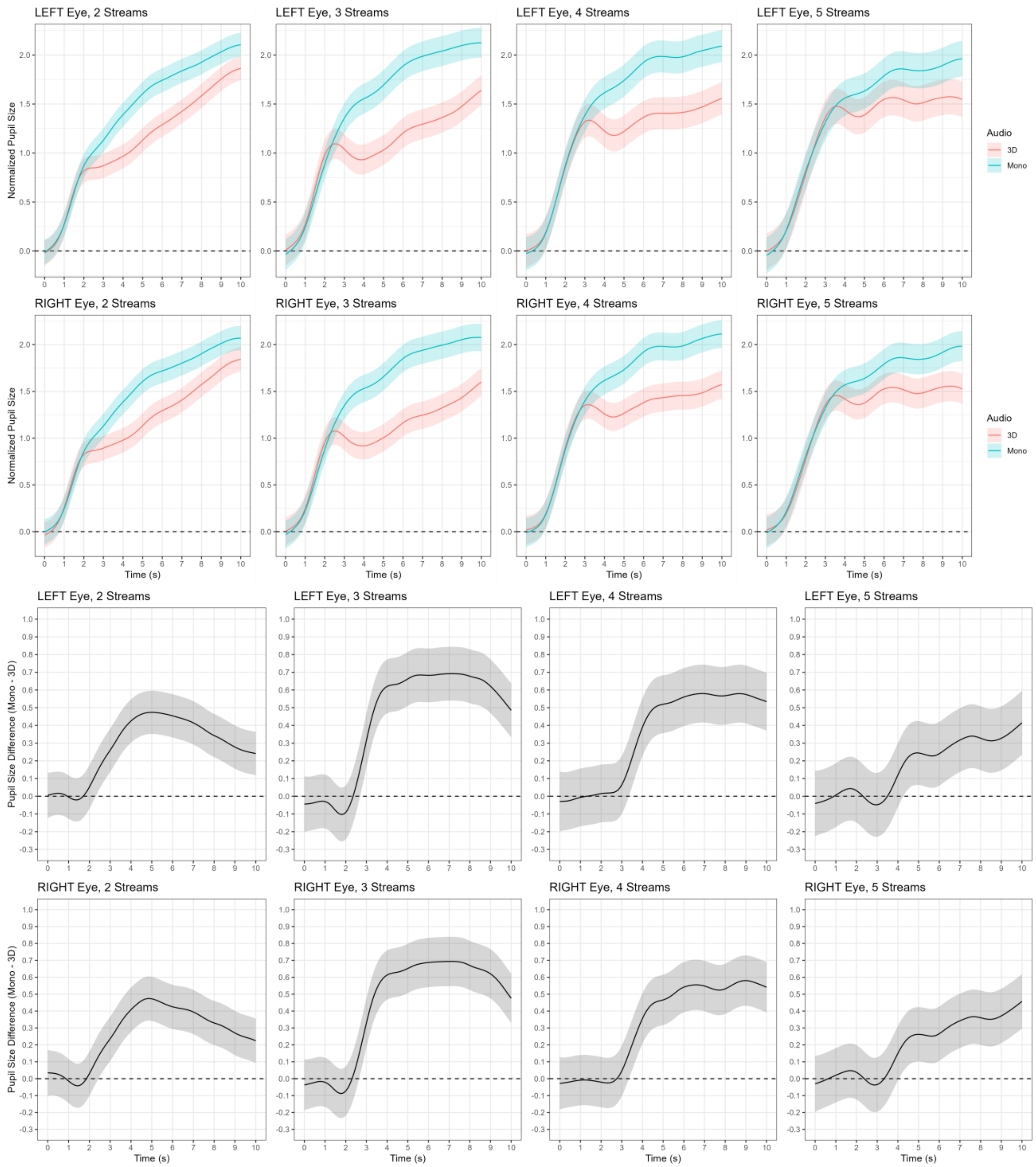

Figure 7. The GAM models for all participants (top). The calculated difference between the two listening condition GAM models (bottom). Solid lines show the GAM estimated value and shaded regions show 95% confidence interval.

Methods - Experiment II

Participants

A total of 20 pilots (all male) completed the second simulator-based experiment. All these pilots also completed the first experiment. The average age was 39.1 years (*SD* = 7.4 years), and average flight time was 2427.25 hours (*SD* = 1801.93 hours). These participants reported their hearing as "good" or "excellent," while only one participant selected "a little trouble" when describing their hearing abilities. In the simulator-based experiment, participants conducted simulated flights using the NUH-60 flight simulator.

Materials

NUH-60 flight simulator.

Flight performance was assessed within USAARL's NUH-60 research flight simulator. The NUH-60 consists of a simulator compartment containing a cockpit, instructor/operator station, observer station, and a six-degree-of-freedom motion system. It is equipped with a twelve-channel visual image generator system, ten-foot radius collimated optical display providing a 200 x 45 degree field of view and two chin displays. The collimated optical display system consists of seven projectors, each providing 2560 x 1600 pixels resolution for a combined resolution of 1.8 arcminutes/pixels. The visual system simulates the natural helicopter environmental surroundings for day, dusk, or night. The data collection system records aircraft/simulator state parameters at a 60 hertz (times per second) capture rate and over 200 variables.

This study employed five distinct flight routes, each designated to simulate operational environments with varying levels of auditory workload throughout. Table 5 provides a brief description of each flight route. Each flight route consisted of a series of radio calls presented to the pilot (research participant), mimicking real world operations. These calls were designed to be operationally relevant and required the participant to perform specific tasks, such as aircraft maneuvers (adjusting heading, speed, or altitude), radio frequency changes, or an acknowledgement (verbally acknowledging instruction or information). Table 6 lists example radio calls used in this study.

This space is intentionally blank.

Table 5. Flight Routes and Descriptions

Flight Route	Description
Supply Movement	This route involves a single helicopter supply run, transporting supplies to a landing zone occupied by ground troops. The participants' callsign was FORGE25. A pre-programmed flight path guided the helicopter. Air traffic control (ATC) provided any necessary adjustments. In addition to routine ATC handling calls, this route required holding, rerouting, and entering into a restricted operating zone. After takeoff, the standard flight profile maintained an altitude of 800 feet and a speed of 100 knots.
Patient Transfer	This route involved a two-helicopter medical evacuation of one urgent patient from a base field to a Role 3 medical facility. The lead helicopter operated under the callsign FORGE25, with the accompanying flight medic using FORGE25D. The helicopters followed a pre-planned route, with the lead helicopter coordinating movements, and ATC providing any necessary adjustments. In addition to routine ATC handling calls, this route required patient-based speed and altitude changes, reroutes, and changes in landing direction. After takeoff, the standard flight profile maintained an altitude of 1000 feet and a speed of 100 knots.
Training Flight	This route involved a single helicopter training flight from a towered airport to an uncontrolled airfield, operating under the callsign FORGE25. The helicopter departed the airfield using the designated flight route. In addition to routine ATC handling calls, this route required communication with an operations center, deconfliction with other aircraft, and weather advisories. After takeoff, the standard flight profile maintained an altitude of 800 feet and a speed of 100 knots.
Passenger Movement	This route simulated a three helicopter passenger transport mission to an improvised landing zone. The lead helicopter, operating under callsign FORGE25 guided the formation. The flight generally adhered to a pre-programmed route with adjustments directed by ATC. Primary communications in this route were internal to the aircrew, the flight of three, and the ground commander. This route required multiple airspeed and altitude formations as well as landing zone coordination. Following takeoff, the standard flight profile maintained an altitude of 1500 feet and a speed of 100 knots.
Student Check Ride	This route simulated a single-helicopter student check ride in the Fort Rucker area using callsign FORGE25. The pilot was required to complete pre-taxi checks prior to takeoff. The flight followed a modified route with adjustments from ATC. In addition to routine ATC handling calls, this route required time-critical traffic calls to include ATC-directed aborted landings and deconfliction in congested airspace. After takeoff, the standard flight profile was 800 feet altitude and 100 knots airspeed.

To simulate realistic airspace, flight routes included two types of radio calls:

- Target Calls: These were directed at the participant/pilot and required a specific response, as outlined in Table 6. Each flight route included 20 target calls.

- Distractor Calls: These were representative of general air traffic chatter unrelated to the pilot's flight plan or mission. While not requiring a direct response, these calls required monitoring to ensure the pilot did not miss their callsign.

Table 6. Examples of Target Radio Calls During Simulated Flights With Correct Responses

Example Target Radio Call	Pilot Action or Response
Forge25, winds calm, clear for takeoff runway 36, straight out departure approved.	Depart the runway, maintain runway heading (360) during departure
Forge25, clear of my airspace to the north, contact departure on 131.5.	Frequency change/ask RP to change radio frequency
Forge25, Warhawk Xray requesting ETA to LZ.	Check ETA/ask RP to check time, verbal response of time
Forge25, Rhino Ops, what's your final destination?	Verbal response with location
Forge25, contact JTAC on 230.5, remain this frequency for FLT following.	Participant will change radio to 230.5
Can I get the numbers for the fuel check?	Verbal response with fuel numbers
Can we increase airspeed to 120?	Participant will increase airspeed
Can we descend to 200 feet above the ground?	Participant will descend 200 feet
Forge25, please circle to land to the east, avoid overflight of decontamination area, out.	Participant will perform flight maneuver

Note. ETA = estimated time of arrival; LZ = landing zone; JTAC = joint terminal attack controller; FTL = flight; RP = research pilot

Table 7. Example of Distractor Calls that Co-Occurred During the Simulated Flight Routes

Example Distractor Radio Calls
Raptor 21, good readback. Taxi via Alpha to runway 36, hold short and contact tower.
We've got traffic, 9 o'clock, level. Looks like it's headed toward us.
Copy all.
Anvil Xray, Dragoon 44 down on pad A3, please send out equipment for FARP Exxon.
Evac 01, Carter Radio copies, say type aircraft and callsign.
Rhino Ops, Evac 01 is Redcon 1.
Carter Radio, Evac 01 frequency change.

Note. FARP = forward arming and refueling point

Although all flight routes consisted of twenty target calls, they differed in the number and arrangement of distractor calls, creating varying levels of auditory workload. Percent correct was calculated from the twenty target calls as scored by the researcher in the simulator with the participant. The number of times the aircraft radio transmitted audio, referred to as "events," differed across routes. For instance, the "supply movement" route had twenty nine total events, with twenty five instances of overlapping radio calls. This overlapping involved two to five simultaneous audio streams, simulating challenging listening conditions and airspace chatter. Table 8 details the specific radio call types and event counts for each flight route.

The "supply movement" route featured three instances where the participant/pilot's callsign (target call) was presented concurrently with two distractor radio calls, resulting in three simultaneous talkers. This number of simultaneous streams for target call and flight route is outlined in Table 9. The table highlights the frequency with which target calls were presented in isolation versus being masked by varying numbers of distractor calls. Every route included at least one instance of a target call with four simultaneous distractor calls, representing a high auditory workload situation. The remaining target calls varied in their complexity depending on the specific operational environment being simulated. Furthermore, distractor calls were incorporated in instances where the radio transmitted audio, but no target call was present. These blank calls, sometimes presented in isolation and other times overlapping with other talkers, further contributed to the dynamic auditory environment.

All radio calls were recorded using voice actors, then filtered and spatially positioned using generic HRTFs within the simulated environment using SLAB 3D audio software, ensuring a realistic auditory experience for the pilots.

Table 8. Summary of Radio Call Types and Radio Events

Route	**Radio Calls**		**Radio Events**	
	Target	Distractor	Total	Simultaneous
Supply Movement	20	55	29	25
Patient Transfer	20	45	23	23
Training Flight	20	45	23	22
Passenger Move	20	42	28	20
Student Check Ride	20	48	31	21

This space is intentionally blank.

Table 9. Details of the Frequency and Distribution of Target and Distractor Radio Calls Across Flight Routes with Varying Numbers of Simultaneous Talkers

Route	**Target Calls**					**Blank Calls** (Distractor calls only; no target present)	
	Single	Simultaneous Talkers (Target call and distractor calls)				Single Talker	Simultaneous Talkers
		Two	Three	Four	Five		
Supply Movement	3	6	8	2	1	1	8
Patient Transfer	1	7	7	5	1	1	3
Training Flight	2	5	8	4	1	1	4
Passenger Move	3	7	7	2	1	5	3
Student Check Ride	5	7	4	3	1	5	5

Note. Target call is included in the number of simultaneous talkers (i.e., two represents the target call plus a distractor call).

Headset configuration for simulator.

Currently, Army UH-60s are configured with mono headsets in a standard HGU-56/P helmet without ANR. Stereo CEPs were used for all conditions that did not require ANR. To deliver ANR, a stereo headset configuration was assembled for use in the NUH-60 flight simulator to accommodate the spatial audio for this experiment. For this, a standard HGU-56/P helmet was modified and coupled with a Bose A20 Aviation Headset, which included ANR and stereo capabilities. The standard electronics (earcups and microphone) of the HGU-56/P helmet were removed. The soft inner lining and padding of the helmet was modified to accommodate the width of the headband of the Bose A20 headset. This allowed the headset to be recessed in the helmet for comfort.

The Bose A20 headset was connected to the full-motion simulator via connector type U-174, the same connection as the HGU-56/P. This supplied mono audio signals to the earcups, as well as power to the A20 microphone. The Bose A20 headset delivered stereo signal using a 3.5 millimeter (mm) tip ring sleeve, or 3-pole conductor, connection on the control module. The full-motion simulator's auxiliary connection supplied this module with the stereo output needed for spatial audio. The Bose A20 control module was powered on and set to the "Mix" position allowing the stereo signal to be passed into the earcups of the Bose A20 headset. This research helmet was used for any listening condition that required ANR.

Noise measurements.

The background noise of USAARL's NUH-60 Black Hawk flight simulator was recorded at the maximum volume setting using a GRAS 45CB acoustic test fixture (ATF) placed at the approximate location of a pilot's head in the right seat (participant/pilots' seat). Each of the three experimental listening conditions and an open ear condition (no hearing protection) was recorded for 30 seconds. During the noise recording, a Brüel & Kjær Type 2270 hand-held sound level meter was used to measure the noise inside the cockpit near the left ear position of the ATF.

One-third octave band and overall A-weighted sound levels were measured offline using BK Connect software.

Workload measurements.

Subjective workload.

The NASA-TLX was used and administered in the same manner as the laboratory-based experiment. That is, the NASA-TLX was completed after each flight and prior to starting the next flight and listening condition.

Objective workload.

To gain a better understanding of auditory workload throughout the flight, pupillometry data were collected to serve as an objective, continuous proxy measure of workload. This physiological measure offered a continuous and objective assessment of workload throughout the flight, complementing the subjective, end-of-flight NASA-TLX.

Although the Eyelink 2000 system was used to measure pupillometry in the laboratory portion of this project, Pupil-Labs pupillometry was selected for use in the simulator. Eyelink 2000 required immobilization of the head in order to calibrate and maintain focus on the eyes during the study. Immobilization would not work within the simulator as participants needed to have the ability to move their head to see around them. Pupil-Labs eye tracking system utilizes glasses worn by the participant and were therefore more suited to motion on the part of the participant.

The Pupil-Labs Core Binocular is a video-based infrared (860 nanometer) eye tracking system that features a pair of head-mounted eye cameras suspended just below the left and right eyes. To provide reference information, a plain video world camera is oriented in the gaze direction of the glasses frame form factor used. The pupil is identified and measured using a custom algorithm developed by Pupil Labs. The system and algorithms are detailed by Kassner, Patera, and Bulling (2014). The participant completed a five-point calibration process prior to each session to allow the Pupil Labs system to calibrate a 3D model of the eye relative to the participant's eye. Due to the eye camera being fitted to the participant's head (akin to wearing glasses), the distance between the participant's eye and the eye camera remained constant. This prevented measurement error of pupil size caused by movement closer or further from the camera.

Any change in this distance (such as the participant accidentally bumping the camera during task engagement) is adjusted for by the 3D model used in the algorithm and is flagged appropriately by the software. The eye cameras record the participant's pupil size using 400 x 400-pixel resolution eye images to derive a pupil diameter measure in pixels (and convert it into millimeters) at a sample rate of 120 hertz (Hz). The data recorded by the eye tracking system is pushed to Lab Streaming Layer (using an integrated Pupil Labs LSL driver) where it can be synchronized with task event markers in the flight simulator. The current study utilized the head-mounted eye camera system to measure pupil diameter (in mm for each eye).

Procedure - Experiment II

Twenty of twenty-one participants who completed the laboratory-based experiment returned to complete the flight simulator experiment. If more than two weeks had elapsed since their previous participation, participants underwent a new consenting and hearing screening. Participants received a briefing from the research pilot in the simulator control room. The briefing covered the five scripted flight routes: supply movement, patient transfer, passenger movement, training flight, and student check ride. These routes were explained again prior to execution in the simulator. Participants were instructed to prioritize adhering to flight instructions and responding to radio calls while completing a secondary task. They were informed that their decision making, emergency response, and aircraft management skills were not being evaluated. Participants monitored five radio channels simulating communications with ATC, internal aircraft systems, ground forces, other aircraft, operations cells, hospitals, and other route-relevant sources. The callsign "Forge25" could originate from any of these radio channels, requiring constant monitoring.

Once in the flight simulator, participants occupied the right seat of the NUH-60 flight simulator, while the research pilot occupied the left seat. A researcher was present to monitor pupillometry data and administer the NASA-TLX after each flight route. Participants were instructed to follow flight instructions, monitor and respond to relevant radio calls, adhere to appropriate flight traffic patterns, and perform the secondary task. They were informed that they could request support from the research pilot for unclear transmissions or assistance with flight management system operations. The participant was solely responsible for the secondary task and could not rely on the research pilot.

The secondary task in each flight route involved monitoring the navigation stack as part of their routine instrument scanning. Participants were instructed to immediately press the VOX-Caution reset button whenever the flight management system message (FMS MSG) light illuminated on the navigation stack. The light appeared randomly and extinguished when the button was pressed or after 5 seconds of illumination. Participants were reminded to integrate this monitoring task into their regular instrument scan. The light illuminated 25 times during each flight route.

Prior to each flight route, an audio test message was presented to ensure the participant’s preferred master volume was set and to demonstrate spatial locations if in a 3D audio listening condition. Once set, the master volume remained constant throughout the flight. Five listening conditions determined by headset configuration were counterbalanced across participants and flight routes. Each participant experienced all five listening conditions, with each condition tested on a different flight route. The 3D audio conditions utilized generic HRTFs to position virtual radios at -90 degrees (°), -45 °, 0 °, 45 °, and 90 ° in the horizontal plane. Flight route order was randomized for each participant using a random order generator. Eye tracking calibration occurred before each flight. During the flight routes, target call responses were scored by the research pilot within the simulator. Upon completion of each flight route, eye tracker recordings were stopped, and the NASA-TLX was administered. The next flight route and listening condition were completed until the participant had completed all five flight routes under each different listening condition as shown in Table 10.

Table 10. Flight Simulator Listening Conditions and Headset Configurations

Listening Condition	Headset Configuration	Helmet	Estimated Environmental Noise
Mono CEPs (traditional aviation set up)	HGU-56/P + CEPs	Participant's own helmet	49 dB A
Mono ANR	HGU-56/P + Bose A20 (no CEPs)	Research helmet	62 dB A
3D audio	HGU-56/P + stereo CEPs	Participant's own helmet	49 dB A
3D audio + ANR	HGU-56/P + Bose A20 (no CEPs)	Research helmet	62 dB A
3D audio + ANR + CEPs	HGU-56/P + Bose A20 + stereo CEPs	Research helmet	52 dB A

Statistical Analysis - Experiment II

Speech recognition.

Speech recognition testing was analyzed using a binary measure of accuracy (correct/incorrect). A mixed-effects logistic regression model, with a binary outcome of correct or incorrect, was conducted on the 20 target calls for each flight. The regression model included fixed effects for listening condition, flight route, run order, number of audio streams, target call order (Table 11), and a random intercept for each participant. This model accounted for the repeated measures design, as all subjects experienced each listening condition and flight route. A Type-II ANOVA was used to test overall significance of the regression model fixed effects. Significant results for any fixed effects were followed up with relevant pairwise comparisons. Pairwise comparison p-values were corrected for multiple comparisons as described in Experiment I methods. Predicted probability with 95% confidence intervals were reported. All data were visualized prior to statistical testing. The statistical significance level was set at 0.05 for all tests.

This space is intentionally blank.

Table 11. In-Flight Analysis of Speech Recognition Score Regression Factors, Levels, and Role in Analysis

Factor	Levels	Role in Analysis
Listening Condition	Mono CEP, Mono ANR, 3D audio, 3D audio + ANR, 3D audio + ANR + CEP	Independent variable
Number of Audio Streams	1, 2, 3, 4, 5 (categorical)	Independent variable
Flight Route	Supply movement, passenger movement, patient transfer, student check ride, training flight	Confounding variable
Run Order	1, 2, 3, 4, 5 (categorical)	Confounding variable, learning effects
Target Call Order	1 through 20 (numeric)	Confounding variable, learning effects
Speech Recognition Score	Binary outcome: correct or incorrect	Dependent variable

Workload.

Subjective Workload.

NASA-TLX.

Individual subject descriptive statistics for the NASA-TLX raw scores are presented visually to illustrate the subjective nature of the data and variability. A mixed-effects regression model was chosen to analyze this data to account for the repeated measures design and to simultaneously test several independent variables. Each NASA-TLX subscale was analyzed with a separate regression model containing fixed effects for listening condition (categorical: 5 levels), flight route (categorical: 5 levels), run order (categorical: 5 levels), the interaction between listening condition and flight route, and a random intercept for each participant. Backwards selection was used with each regression model to remove non-significant terms from the regression model before estimating means and confidence intervals. A Type-III ANOVA was used to test overall significance of the regression model fixed effects. Significant results for any fixed effects were followed up with relevant pairwise comparisons. Pairwise comparison p-values were corrected for multiple comparisons as described in Experiment I methods. The NASA-TLX data were assumed to be continuous for this analysis so that multiple independent variables could be analyzed simultaneously. Regression models were evaluated for normality and homoscedasticity of residuals.

Objective Workload.

Pupillometry.

Pupil data were recorded during simulated flight routes and synchronized with flight simulator events. The number of radio events ranged from 20 to 30 depending on the specific route, as detailed in Tables 8 and 9. For each radio event, segments of pupil data were extracted, similar to the methodology used in the laboratory experiment. See Table 12 for the flight simulator audio stimulus segmentation.

Table 12. Flight Simulator Audio Stimulus Segmentation for Analysis

Audio Event Phase		Event Description
Trial Start	a)	No radio/audio stream is audible
	b)	2 second duration prior to radio event(s)
Stimulus Starts	a)	Radio/audio streams are initiated. When multiple audio streams are presented, they did not start at the same time
	b)	Audio streams vary in duration pending the operational event
	c)	Included a 2 second pause at the end of the audio where participants retain the information
Response Start	a)	Participants respond according to radio call requests

The initial 2 seconds before the radio call serves as a baseline measurement. Stimulus start marks the onset of the radio event, with audio streams varying in number, duration, and start times, within 1-5 seconds of each other. In the dynamic flight environment, pilots were fully occupied with flying the aircraft, monitoring instruments, and responding verbally or behaviorally to radio calls and therefore eye movements were not constrained in this part of the experiment. Unlike in the controlled laboratory environment, pilots were not bound to wait for a specific "response start" and could promptly react once they grasped the task demands.

The raw pupil data underwent quality processing and normalization, following the procedures outlined in the laboratory experiment. Pupillometry testing employed a GAM, with details of this approach explained in the analysis for Experiment I. The normalized pupil data were plotted and used to estimate predicted pupil size using the GAM approach. The hypothesis tested was that workload/listening effort is lower in the 3D audio and ANR listening conditions compared to the standard mono listening condition. It was expected that ANR and 3D audio will result in smaller changes in pupil size compared to the current CEP headset.

Secondary task.

The secondary task measured accuracy in acknowledging the Flight Management System Message (FMS MSG) light. A mixed-effects linear regression analyzed the percentage of correct acknowledgments. Fixed effects included listening condition (categorical: 5 levels), flight route (categorical: 5 levels), and run order (categorical: 5 levels), and a random intercept for each participant. A Type-III ANOVA was used to test overall significance of the regression model fixed effects. Significant results for any fixed effects were followed up with relevant pairwise comparisons. Pairwise comparison p-values were corrected for multiple comparisons as described in Experiment I methods.

User preferences.

After the participant completed all flight routes and experienced all listening configurations, they were asked to provide feedback on which listening configurations they preferred most and least through a five-question survey. See Table 13. The survey was developed in-house and consisted of Likert scale questions and open responses. Summary statistics and open text responses were recorded.

Table 13. Five Question Survey Distributed to Participants

Survey Question	Answer Options
1a. Please select your MOST preferred headset condition out of the 5 conditions and state why. 1b. Please state WHY this is your MOST preferred headset condition.	**1/2a. Circle A, B, C, D, E** A. CEPs with HGU-56/P helmet (standard listening condition) B. CEPs with HGU-56/P helmet with spatialized audio (standard plus 3D audio) C. Active noise reduction (ANR) with HGU-56/P helmet (no CEPs) D. Active noise reduction (ANR) with HGU-56/P helmet and spatialized audio (no CEPs) E. All conditions together (CEPs with ANR and 3D audio) 1/2b. Open ended text response
2a. Please select your LEAST preferred headset condition out of the 5 conditions and state why. 2b. Please state WHY this is your LEAST preferred headset condition.	
3. On a scale of 1 (very ineffective) to 5 (very effective) how would you rate the overall effectiveness of the 3D audio technology?	1-Very ineffective 2-Ineffective 3-Neutral 4-Effective 5-Very effective
4. Please share any additional comments or feedback you have regarding your experience with 3D audio technology and/or ANR technology in military aircraft operations.	Open ended text response
5. Are there any specific improvements or enhancements you would like to see in future iterations of 3D audio technology and/or ANR technology?	Open ended text response

Results- Experiment II

Noise Attenuation for Listening Conditions

The noise levels for each listening condition and the open ear condition were recorded. The A-weighted sound levels for each listening condition are presented in Figure 8. At maximum volume, the simulator produced levels of about 88 dBA. Noise levels for the research helmet with ANR headset averaged 62 dBA, providing about 26 dB of attenuation. The listening conditions consisting of the CEPs (CEP and CEP with 3D audio) yielded the highest attenuation (39 dB) and quietest listening condition averaging 49 dBA. The listening condition with ANR and CEPs averaged 52 dBA sound levels (36 dB of attenuation). These are the estimated environmental noise the research participant was working in. See Table 10 for headset configurations.

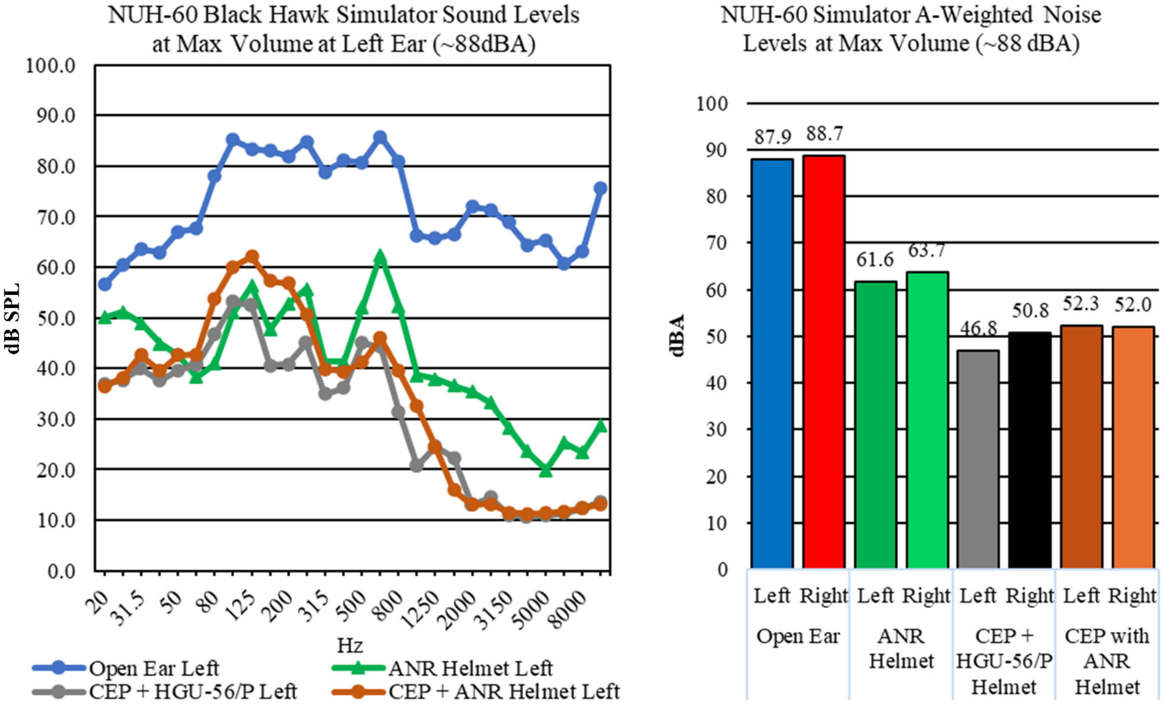

Figure 8. Noise measurements made in the NUH-60 Black Hawk simulator (Left) across frequency sound pressure level (dB SPL) of in-cabin simulated helicopter noise for each of the listening conditions; (Right) A-weighted sound levels at the ears for each listening condition.

A logistic regression model predicting binary response accuracy (correct/incorrect) was tested with five variables of interest. Those variables were listening condition, flight route, run order, number of audio streams, and target call order. Listening condition did not significantly predict response accuracy ($X^2(4) = 7.3$, $p = 0.11$). However, flight route ($X^2(4) = 46.2$, $p < .001$), run order ($X^2(4) = 16.4$, $p = .002$), number of audio streams ($X^2(4) = 51.5$, $p < .001$), and target call order ($X^2(1) = 16.4$, $p < .001$) were all significant predictors of correct responses.

Regression analysis of listening condition revealed no significant differences in predicted percent correct between the various 3D audio, ANR, and mono listening conditions. Although the 3D audio and ANR condition (0.724 [0.665,0.776]) and the 3D audio, ANR, and CEP condition (0.714 [0.655, 0.767]) yielded the highest predicted probabilities of correct responses, these differences were not statistically significant. See Figure 9. Therefore, the current data and experimental design do not provide sufficient evidence to conclude that any specific listening condition is superior in terms of enhancing accuracy. The study hypothesis that 3D audio and ANR will result in better speech recognition scores when compared to the standard aviation listening condition (mono CEP) did not hold true for this study.

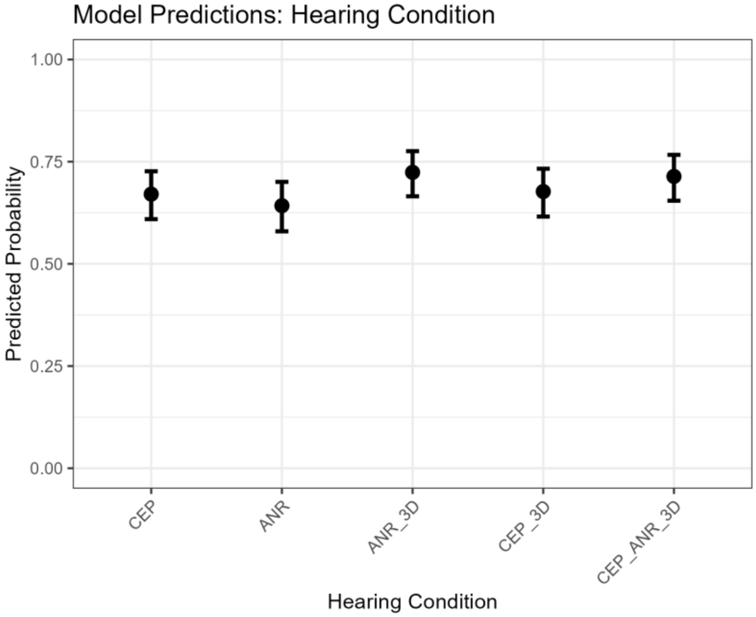

Figure 9. Predicted probability of correct radio responses by listening condition. Error bars indicate the 95% CI.

Regression analysis revealed a significant negative effect of the number of audio streams on the accuracy of radio responses, suggesting a decline in performance as audio stream complexity increases. This is the same pattern shown in Experiment I. As depicted in Figure 10, there is a clear downward trend in the probability of correct responses as the number of simultaneous audio streams increases. A single audio stream had the highest probability at 87% (0.870 [0.815,0.911]). Probability decreases with increasing audio streams: two streams at 70.7% (0.707 [0.662,0.750]), three streams at 64% (0.640 [0.592,0.686]), four streams at 61.1% (0.611 [0.544,0.675]), and five streams at 52.5% (0.525 [0.420,0.627]). These results are averages across all other variables in the regression model, including listening condition. Therefore, as the number of audio streams increases, speech recognition accuracy decreases.

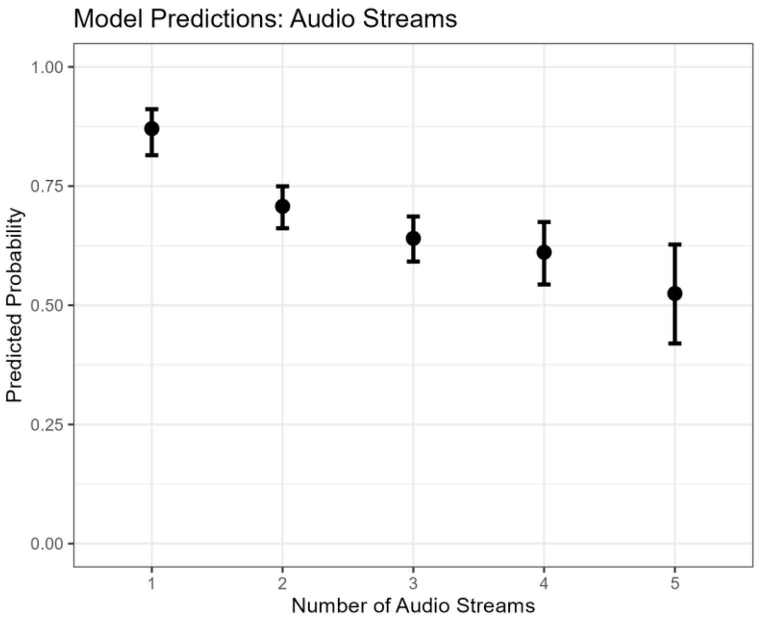

Figure 10. Effect of audio stream quantity on radio accuracy response. Error bars indicate the 95% CI.

Regression analysis for run order revealed a significant learning effect, with pilots demonstrating increased accuracy in radio responses over successive runs; see Figure 11. The predicted percent correct increases from 61% in run one to 72.7% by run three, after which performance stabilizes for runs four (72.6%) and five (70.3%). Statistically significant differences were observed only between the first run (run one) and the later runs (run three: *p* = .0053, run four: *p* = .0053, and run five: *p* = .032), suggesting that the most substantial learning gains occurred early in the task.

This space is intentionally blank.

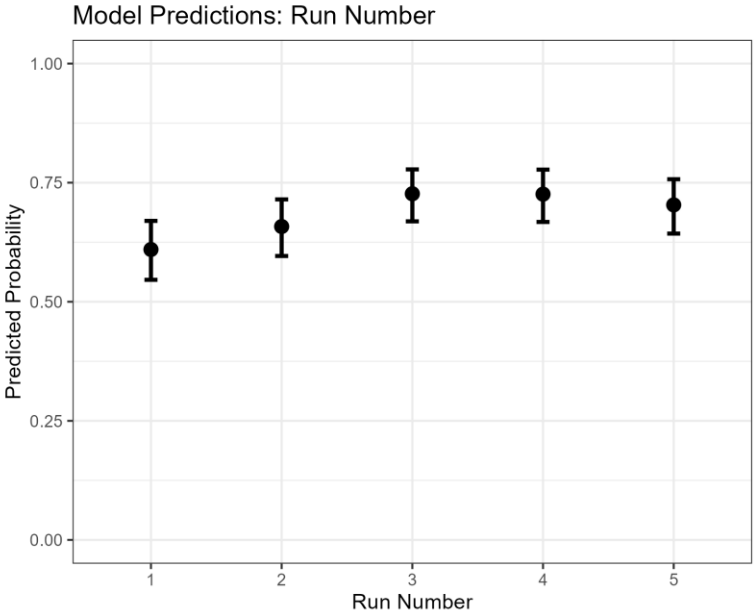

Figure 11. Learning effect on radio response accuracy across runs. Error bars indicate the 95% CI.

Target call order was a statistically significant predictor in the regression model, indicating that the probability of correctly answering a radio call increases over the course of a flight. The model predicts that pilots are more likely to respond accurately to calls later in the flight, with accuracy increasing by approximately 10% from the beginning to the end of each flight.

Regression analysis of flight route revealed that supply movement route (0.538 [0.475, 0.600]) was associated with significantly lower accuracy in responding to radio calls compared to other routes. See Figure 12. No significant differences in accuracy were found among the remaining four flight routes.

This space is intentionally blank.

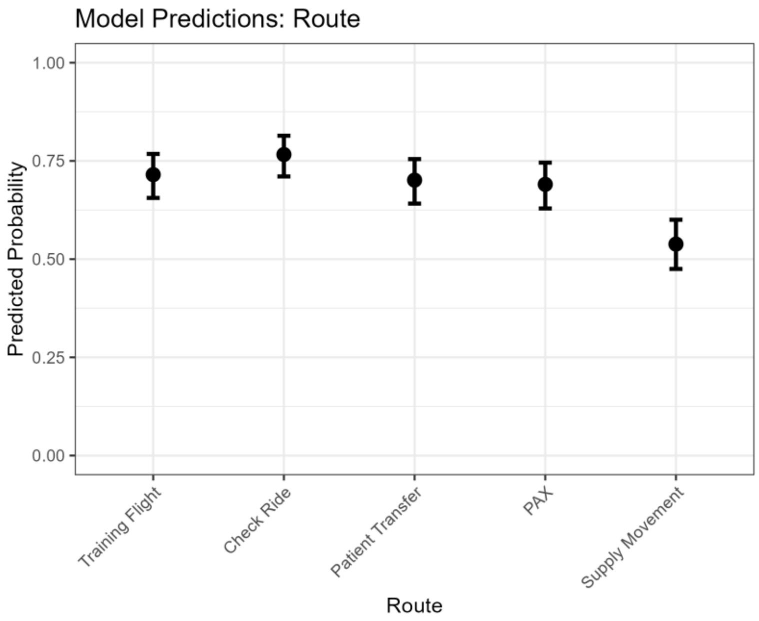

Figure 12. Effect of flight route on radio accuracy response. Error bars indicate the 95% CI.

Subjective Workload

Figure 13 presents the NASA-TLX scores for each participant, highlighting the subjective nature of experienced perceived cognitive workload. Considerable inter-individual variability is evident. For instance, participants 5, 10, 13, and 23 consistently reported low workload across all subscales, while participants 2, 4, 6, and 20 tended toward higher scores. Furthermore, the variance within each participant's responses differed. Participants 3, 7, and 15 exhibited a wider range of scores across subscales in contrast to participants 5, 6, and 18 who displayed more consistent ratings. This variability, while expected with a subjective measure, introduces complexity to the interpretation of the results. Figure 14 is the average NASA-TLX scores based on listening condition alone.

This space is intentionally blank.

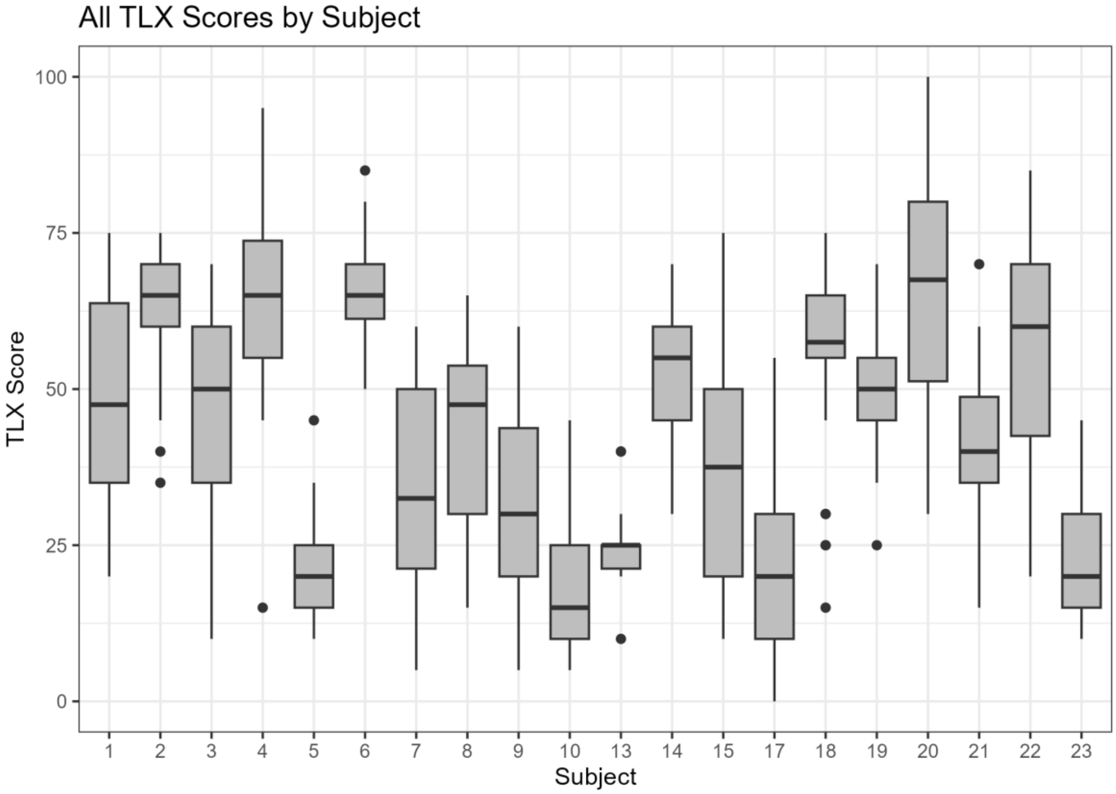

Figure 13. Boxplots showing NASA-TLX scores for each study participant.

This space is intentionally blank.

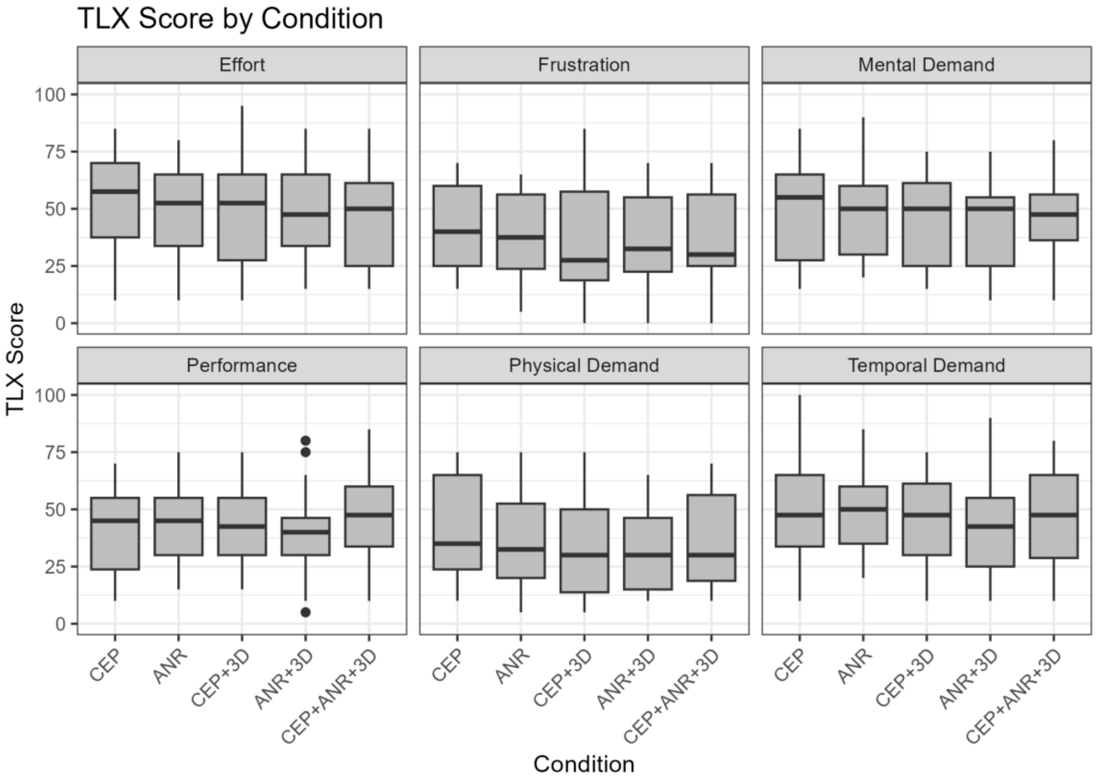

Figure 14. Boxplots showing NASA-TLX scores for each listening condition.

This space is intentionally blank.

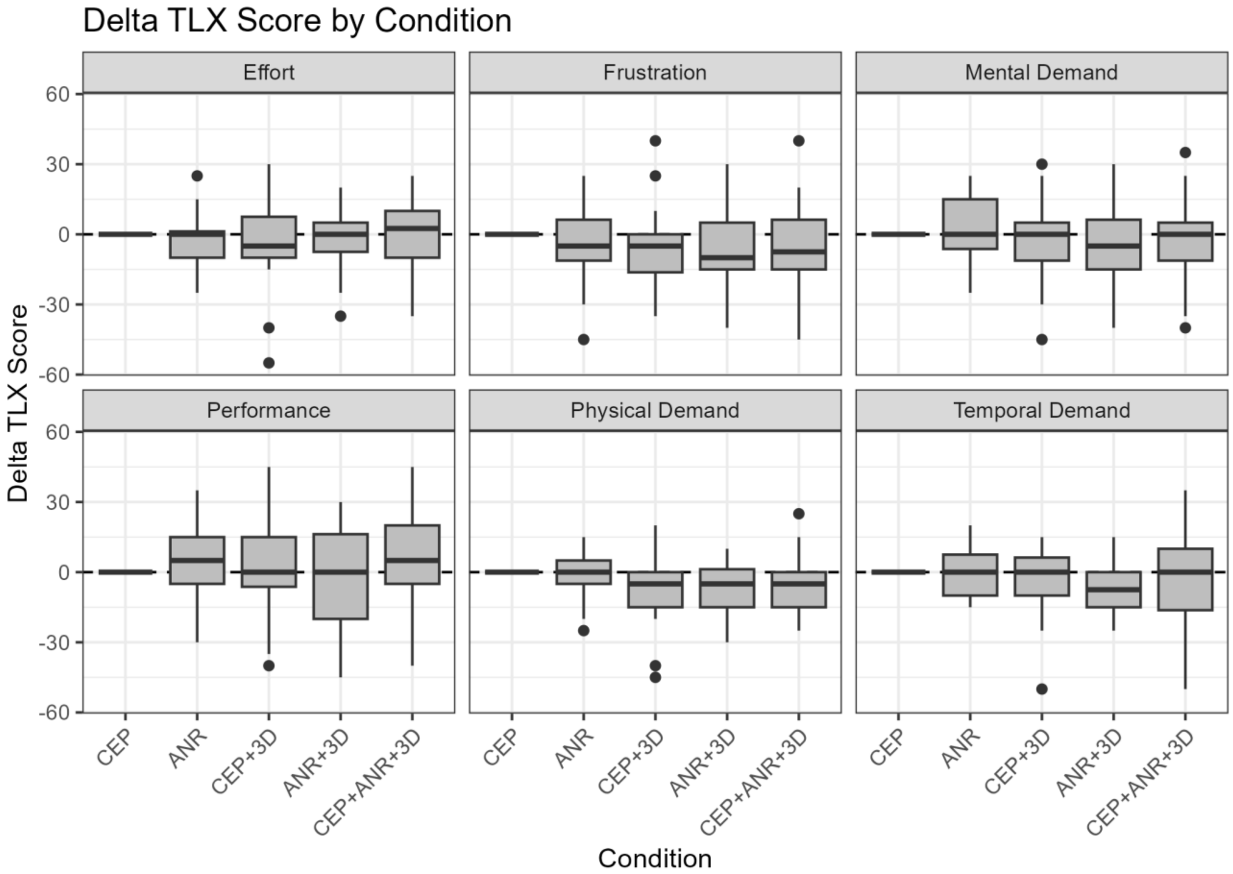

Figure 15. Boxplots showing calculated changes in perceived workload from the standard aviation listening condition (CEP). Positive values above the zero line indicate an increase in the NASA-TLX subscale and negative numbers represent a decrease.

Changes from baseline were calculated for listening condition, flight route, and run order. Positive values above the zero line represent an increase relative to the baseline condition, while negative values represent a decrease. The results suggest that listening condition (Figure 15) may influence workload, particularly frustration and physical demand. Flight route data (see Figure 16) suggest that the supply movement route involved a greater amount of workload. Run order appeared to have no influence on subjective workload.

This space is intentionally blank.

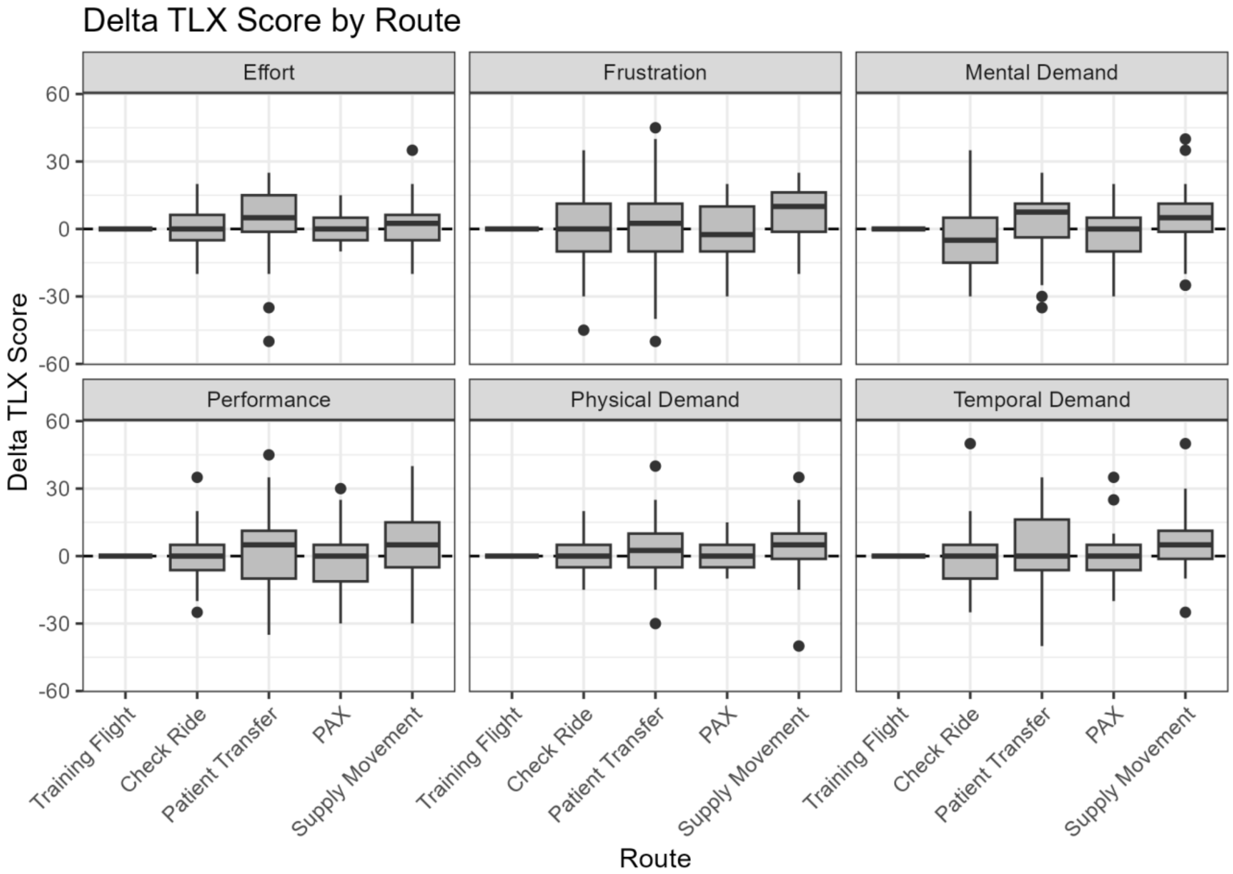

Figure 16. Calculated changes in perceived workload from training flight route. Positive values above the zero line indicate an increase in the NASA-TLX subscale and negative numbers represent a decrease. Error bars indicate 95% CI.

Regression analyses were conducted to examine the effects of listening condition, flight route, and run order on the NASA-TLX scores. A backward selection procedure was used to identify a statistically significant model. All non-significant predictors were removed to obtain a new model for regression. This standardized the approach to selecting which terms and interactions would be in the model. If there were any significant terms, relevant pairwise comparisons were made.

With regard to the mental demand subscale, the results indicated no statistically significant main effects for listening condition (*F*(4, 53.057) = 0.674, *p* = .613), or flight route (*F*(4, 53.105) = 1.8909, *p* = .126), or run order (*F*(4, 54.098) = 0.5008, *p* = .735), or the interaction between listening condition and flight route (*F*(16, 57.650) = 0.7326, *p* = .7497).

The results of the physical demand subscale revealed a statistically significant main effect for listening condition (*F*(4, 56.126) = 3.0825, *p* = .0230). Specifically, pairwise comparisons showed that participants rated the 3D audio and 3D audio with ANR conditions as significantly easier than the traditional CEP condition, with mean differences of 7.90 and 8.00 points, respectively (both *p* = .0.334). The 3D audio plus ANR plus CEP condition approached significance (*p* = .0791). Flight route was not statistically significant (*F*(4, 56.157) = 1.0635, *p* =

.3831). The interaction effect for listening condition and flight route was approaching significance (*F*(16, 85.490) = 1.6027, *p* = .097). These findings suggest that the inclusion of 3D audio may reduce perceived physical demand during flight tasks.

The results of the temporal demand subscale revealed a statistically significant interaction effect for listening condition and flight route (*F*(16, 59.764) = 1.8352, *p* = .04705). Pairwise comparisons only revealed one significant finding for the patient transfer flight route for 3D audio plus ANR plus CEP when compared to 3D audio alone (*p* = .0488). No other significant findings or trends were revealed. No main effects for listening condition (*F*(4, 56.435) = 1.4484, *p* = .2301) or flight route (*F*(4, 56.476) = 1.8053, *p* = .1405) were found.

The results of the performance subscale indicated no statistically significant main effects for listening condition (*F*(4, 53.569) = 0.7151, *p* = .585), flight route (*F*(4, 53.105) = 1.8909, *p* = .126), run order (*F*(4, 54.098) = 0.5008, *p* = .735), or the interaction between listening condition and flight route (*F*(16, 57.650) = 0.7326, *p* = .7497).

The results of the effort subscale indicated no statistically significant main effects for listening condition (*F*(4, 52.916) = 0.4336, *p* = .784), flight route (*F*(4, 52.859) = .2980, *p* = .878), run order (*F*(4, 53.693) = 0.6998, *p* = .596), or the interaction between listening condition and flight route (*F*(16, 56.605) = 0.4437, *p* = .9627).

The results of the frustration subscale indicated no statistically significant main effects for listening condition (*F*(4, 53.313) = 0.9057, *p* = .467), flight route (*F*(4, 53.361) = 1.7709, *p* = .148), run order (*F*(4, 54.098) = 0.3517, *p* = .8417), or the interaction between listening condition and flight route (*F*(16, 58.132) = 0.6556, *p* = .8239).

In summary, the NASA-TLX subscales of physical demand and temporal demand were the only ones that suggested any influence from listening condition or flight route. Specifically, physical demand ratings were significantly lower (indicating less perceived workload) in both the 3D audio and 3D audio plus ANR listening conditions compared to the standard CEP listening condition. Temporal demand ratings suggest an interaction between listening condition and flight route, specifically in the patient transfer flight. However, no clear patterns emerged. The remaining NASA-TLX subscales (mental demand, performance, effort, and frustration) were not significantly influenced by listening condition, flight route, or run order.

It is important to note a limitation of the study design, in which both the order of the listening conditions and flight routes were randomized. Therefore, any observed changes in the NASA-TLX scores could be attributed to either factor or an interaction between them. This complicates the interpretation of the results. Furthermore, the NASA-TLX provided workload ratings at the conclusion of each flight; it did not capture potential fluctuations during the flight itself, particularly as the number of audio streams varied, auditory workload varied.

Objective Workload

Pupillometry.

Pupillometry data collected in a dynamic flight simulator environment, characterized by movement, variable lightening, and vibration, presents significant challenges and results in noisy data to interpret. The graphs below (Figure 17) show the pupil size data for all audio events; raw data is represented with the colored lines. Each row shows a different listening condition, and each column shows the number of simultaneous audio streams. The audio event, or stimulus start, begins at the vertical dashed line. All GAMs predicted effectively flat pupil dynamics (solid black lines) across all listening conditions and audio stream counts, and no discernable patterns in response to radio calls during simulated flights. While the observed data did not reveal significant changes in pupil dynamics, this may be attributed to inherent noise within the data and/or the already high workload experienced by participants throughout the simulated flight, potentially masking the impact of the listening task. Future analysis should incorporate more advanced noise reduction and analysis techniques to try and mitigate these limitations.

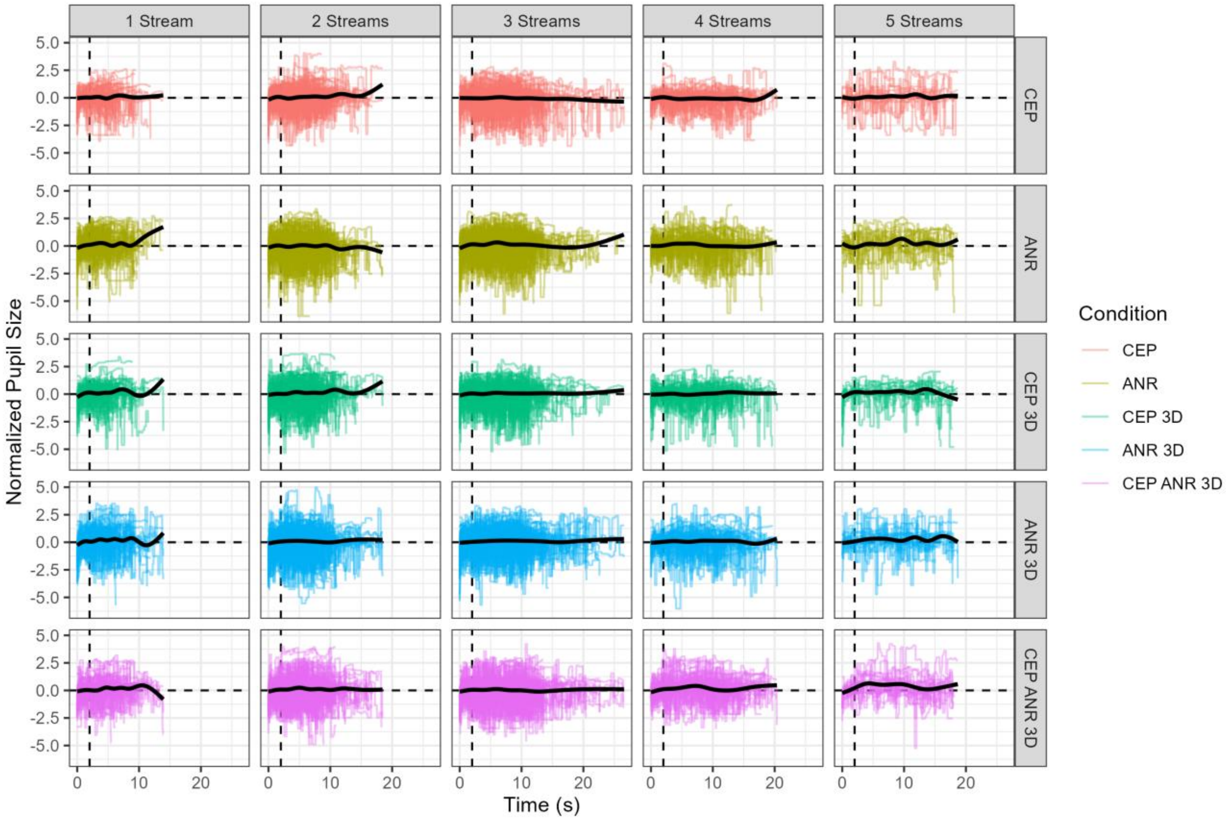

Figure 17. Pupil response to audio events during simulated flight. Graphs show pupil size changes across listening conditions and audio stream counts. Vertical dashed line indicates stimulus onset. Solid black lines represent GAM predictions.

Secondary Task

Results revealed run order (*F*(4, 50.525) = 10.31, *p* < .001) significantly affected response accuracy. Flight route also showed a significant effect (*F*(4, 50.362) = 4.17, *p* = .005). However, the listening condition did not significantly impact the outcome (*F*(4, 50.362) = 4.17, *p* = .005).

Similar to the speech recognition results, regression analysis revealed a significant learning effect for acknowledging the FMS messages. Pilot accuracy improved by about 30% from the first run to the fourth run; see Figure 18. The predicted percent correct increased from 15.9% in run one to 45.5% by run four. Statistically significant differences (*p* < .05) were observed between run one and all subsequent runs: run two: (33.1%, *p* = .0113), run three: (37.8%, *p* = .0004), run four: (45.5%, *p* = < .0001), and run five: (43.0%, *p* = .0001). Participants frequently neglected to perform the secondary task, particularly during the initial flights of the experiment. This likely contributed to the observed learning effect, as compliance with the secondary task increased over time, resulting in more consistent and reliable data collection in later trials.

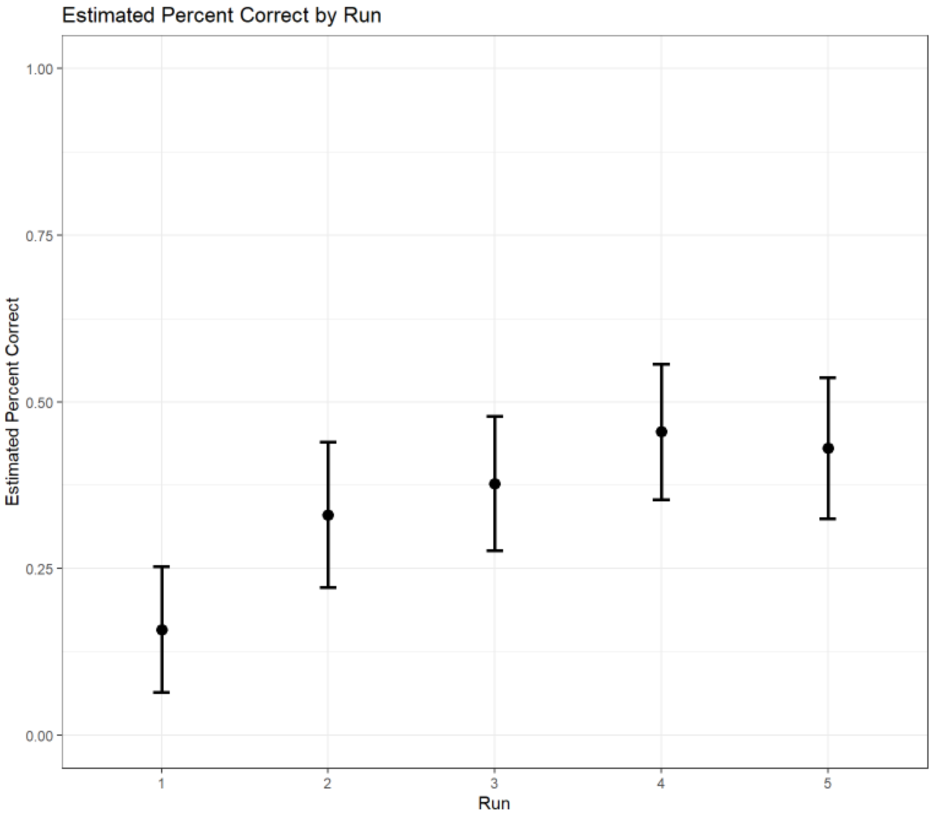

Figure 18. Learning effects on FMG MSG light acknowledgement accuracy across runs. Error bars indicate the 95% CI.

Passenger flight route had the highest FMS acknowledgement accuracy at 45.9%. Table 14 summarizes the average accuracy for each flight route. Pairwise comparisons noted differences between passenger movement and both student check ride (*t*(49.06) = -3.813, *p* = 0.0038) and supply movement (*t*(49.38) = 2.738, *p* = 0.0428). No other flight comparisons reached significance.

Table 14. Mean Accuracy (%) of FMS Message Acknowledgement for Each Flight

Flight Route	FMS MSG Accuracy
Passenger Movement	45.9%
Patient Transfer	38.3%
Training Flight	34.4%
Supply Movement	31.6%
Student Check Ride	25.0%

User Preferences

After completing all five flight routes, participants provided feedback on their experiences through a five-question survey. Although there wasn’t a single preferred listening condition or headset setup, a clear trend emerged. The top three most preferred combinations all incorporated 3D audio: 1) 3D audio plus ANR plus CEP, 2) 3D audio plus CEP and 3) 3D audio plus ANR. See Figure 19. This suggests that participants favored the experience provided by 3D audio. Participants cited these configurations as the clearest, most natural listening experience and it helped “reduce the overlapping voices and made it easier to prioritize the most relevant radio calls." Appendix B outlines all the responses from subjects on the survey.

Interestingly, the standard listening condition (CEP mono) was overwhelmingly considered the least preferred, with no participants ranking it as their favorite. ANR without 3D audio was also among the least preferred options. Pilots described significant difficulty in understanding communications with these configurations, using phrases like "constantly getting stepped on," "radios too meshed together," and "hardest to hear.”

Participants were asked to respond using a scale of 1 (very ineffective) to 5 (very effective) how they would rate the overall effectiveness of the 3D audio technology. The average response was 4.4 with all participants selecting either 4 or 5. Open-ended comments, and feedback were positive for 3D audio and ANR technology. Pilots responded enthusiastically to the potential of 3D audio and ANR technology in military aircraft, recognizing its capacity to significantly enhance communication and reduce overload. Pilots specifically praised the "natural" interaction provided by 3D audio, highlighting its ability to mimic real-world conversations and improve clarity during multi-person communications. Suggestions for future iterations include adjustable volume controls for individual radios, integration with head-tracking for more immersive spatial audio, and Bluetooth connectivity. Concerns regarding the potential for ANR to mask critical aircraft sounds were also raised, emphasizing the need for careful consideration in its implementation.

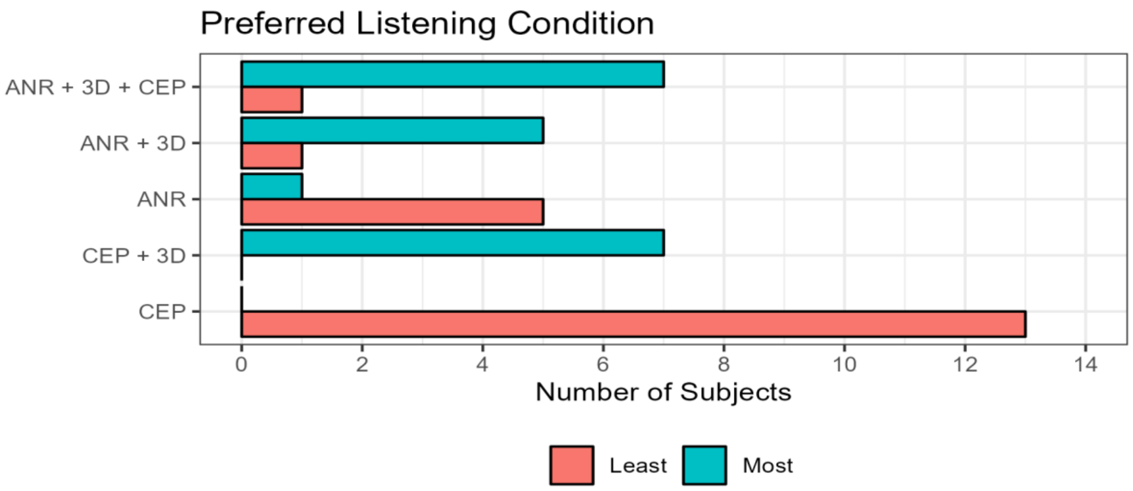

Figure 19. Pilot-reported most and least preferred headset configurations. The length of each bar represents the total number of participants who selected a given configuration, with the blue bars indicating "most preferred" selections and the orange bars indicating "least preferred" selections.

Discussion

This study investigated the impact of 3D audio and ANR technologies on speech recognition and cognitive workload in aviators, utilizing both a controlled laboratory experiment and a realistic flight simulator environment. The hypotheses were that speech recognition performance would be significantly better with 3D audio in both environments, that perceived workload would be lower with 3D and ANR and that flight performance in the simulator would be better with 3D audio. While the laboratory experiment provided compelling evidence for the benefits of 3D audio, the flight simulator results presented a more complex picture, highlighting the challenges of translating laboratory findings to operational settings.

Speech Recognition Performance

The results of the laboratory experiment demonstrated the superiority of 3D audio for both improved speech recognition and lower perceived workload. This advantage of spatial separation enabled listeners to selectively attend to relevant speech streams and filter out distracting ones, aligning with previous research (Brungart et al., 2002; Drullman & Bronkhorst, 2000; Ericson et al., 2004). Spatial audio speech recognition performance in high workload (five streams) was similar to performance of monaural audio in low workload (two streams); this highlights its benefit in complex scenarios.

However, the flight simulator failed to replicate these findings. No specific listening condition performed superior on speech recognition in the flight simulator environment. Although the 3D audio plus ANR and 3D audio plus ANR plus CEP listening conditions yielded numerically higher predicted probabilities of correct responses, regression analyses did not reveal statistically significant differences. The absence of statistically significant differences in speech recognition performance across listening conditions in the simulator contrasts with that found by Mozo and Murphy (1997) who demonstrated improved speech recognition with five commercially available ANR systems that were integrated into the HGU-56/P aviator helmet over the standard HGU-56/P helmet configuration. Casto and Casali (2013) also demonstrated the use of ANR improved performance of flight-related tasks and decreased workload.

One plausible explanation for this is the presence of learning effects compounded by a lower token count. In an effort to enhance “operational” realism of the simulator experiment, the number of communication events available for analysis was significantly reduced compared to the laboratory setting. The controlled laboratory setting provided 100 data points per condition. The simulator experiment data points were reduced and more variable. Each flight included a limited number of high workload communication events (one instance of five simultaneous radio communications, and two to four instances of four simultaneous communications), interspersed with a greater proportion of lower workload scenarios. This, coupled with the varying flight routes and tasks intended to reduce procedural learning, may have inadvertently masked the potential benefits of 3D audio

Workload: Subjective and Objective Measures

In the laboratory, 3D audio reduced perceived workload, particularly when managing multiple communication streams. Participants reported significantly lower perceived effort and workload ratings compared to monaural listening when managing three or more simultaneous communication streams. In contrast, the impact of 3D audio and ANR on perceived pilot workload in the flight simulator environment was less clear. While there was a reduction in the physical domain, the effects on other workload dimensions were inconsistent. Specifically, 3D audio, with or without ANR, was associated with lower physical demand ratings compared to standard CEPs. However, no statistically significant effects were observed for mental demand, performance, effort, or frustration across the different flight routes.

Pupillometry, intended as an objective measure of workload, supported the subjective ratings in the laboratory, with reduced pupil dilation observed under 3D audio conditions. However, the pupillometry data from the simulator were inconclusive. Two different devices were used in the two experiments. The decision to use a different device for the simulator resulted from the need to have an eye tracking system that would move with the participant rather than one that is stationary. The lack of observed differences in the simulator is likely due to two reasons; first, the noise levels in the pupillometry data were very high from the dynamic flight simulator environment which could have made the data uninterpretable; second, there may not have been observed changes in the pupil dynamics of the pilot due to the already high load associated with the task of flying.

User Preference

Despite the lack of statistically significant performance differences in the flight simulator, participant preferences revealed a strong preference for 3D audio integration. Although individual preferences vary, the top three listening configurations all incorporated 3D audio technology. The standard listening condition CEP (without 3D audio) was overwhelmingly ranked as the least preferred option. This highlights the importance of considering user acceptance and perceived value when evaluating new technologies.

Limitations and Future Directions

This study has several limitations that should be considered when interpreting the results. The relatively small sample size may have limited the statistical power to detect subtle differences between listening conditions. The randomization of both listening conditions and flight routes introduces potential confounding factors, making it difficult to definitively attribute observed changes in NASA-TLX scores solely to the listening conditions. The complexity of the flight simulator environment introduced variability that may have obscured the effects of the audio technologies. Furthermore, the use of different pupillometry devices in the laboratory and simulator experiments may have introduced confounding factors.

Future research should address these limitations by employing larger sample sizes, refining the experimental design, and utilizing consistent pupillometry equipment across both experimental settings. Advanced noise reduction techniques should also be explored to improve the quality of pupillometry data collected in dynamic environments.

Recommendations

- **Implementation of 3D Audio:** This study reinforces the documented benefits of 3D audio, including enhanced ability to attend to multiple audio streams, a reduction in perceived and objective workload, and that it is preferred by aviators. Therefore, the U.S. Army should prioritize the integration of 3D audio technology into the FVL program and, when able, legacy aircraft, to improve aviator performance and reduce cognitive burden.

- **Further Research into ANR Technology for Aviator Headsets:** The present study did not demonstrate a significant noise reduction benefit from ANR technology compared to the current standard aviator listening condition. Acoustic measurements and participant preferences did not support the addition of ANR. Further research is needed to explore and optimize ANR technologies specifically for aviator helmets, potentially focusing on alternative ANR algorithms or headset designs. The possibility of the current aviator listening condition providing sufficient noise attenuation without the added complexity of CEPs should also be investigated.

Conclusions

- **3D audio significantly enhanced speech recognition, particularly under high auditory workload in laboratory conditions:** Accuracy was consistently higher in 3D versus monaural listening, even with increasing communication streams. Notably, speech recognition scores for 3D audio with five streams was similar to monaural audio with only two, highlighting its benefit in complex scenarios.

- **3D audio significantly reduced perceived workload, particularly in high auditory workload laboratory conditions:** Subjects reported significantly lower effort and workload ratings in 3D audio compared to monaural listening when managing three or more simultaneous communication streams. This advantage of 3D audio was not observed when only two streams were presented.

- **Pupil data objectively confirmed lower workload during 3D audio listening in laboratory settings:** Analysis of pupil dilation patterns consistently indicated reduced workload in 3D audio compared to monaural listening conditions in the laboratory experiment, supporting the subjective ratings. This finding held true across all levels of communication workload (number of audio streams).

- **Increased auditory workload, characterized by a higher number of simultaneous audio streams, negatively impacts speech recognition accuracy across both laboratory and simulated operational settings:** Regression analysis confirmed that an increasing number of simultaneous audio streams negatively impacts speech recognition accuracy, corroborating previous laboratory findings and extending this effect to a simulated operational environment.

- **No specific listening condition performed superior on speech recognition in the flight simulator environment for this experimental design**: While the 3D audio plus ANR and 3D audio plus ANR plus CEP listening conditions yielded numerically higher predicted probabilities of correct responses, regression analyses did not reveal statistically significant differences.

- **Limited impact of 3D audio and ANR on perceived pilot workload, with a reduction in physical demand but no consistent effects on other workload dimensions in the flight simulator environment:** Only the NASA-TLX subscales of physical and temporal demand showed potential influence from listening conditions. Specifically, 3D audio, with or without ANR, was associated with lower physical demand ratings compared to standard CEPs. However, no statistically significant effects were observed for mental demand, performance, effort, or frustration across the different flight scenarios.

- **Pupil dynamics didn’t change in response to different listening conditions or the number of audio streams in the simulator environment**: High noise levels in the pupillometry data from the dynamic flight simulator environment likely masked potential pupil dynamics in response to audio events, necessitating the consideration of advanced noise reduction and analysis techniques in future studies.

- **Pilots show a strong preference for 3D audio integration in headset configurations:** While individual preferences vary, the top three listening configurations all incorporated 3D audio technology. The standard listening condition CEP (without 3D audio) was overwhelmingly ranked as the least preferred option.

References

Abouchacra, K. S., Breitenbach, J., Mermagen, T., & Letowski, T. (2001). Binaural helmet: Improving speech recognition in noise with spatialized sound. *Human Factors, 43*(4), 584–594. https://doi.org/10.1518/001872001775870368

Begault, D. R., & Pittman, M. T. (1996). Three-dimensional audio versus head-down traffic alert and collision avoidance system displays. *The International Journal of Aviation Psychology, 6*(1), 79–93.

Bianchi, F., Wendt, D., Wassard, C., Maas, P., Lunner, T., Rosenbom, T., & Holmberg, M. (2019). Benefit of higher maximum force output on listening effort in bone-anchored hearing system users: A pupillometry study. *Ear and Hearing, 40*(5), 1220–1232. https://doi.org/10.1097/AUD.0000000000000699

Bolia, R. S. (2003). Effects of spatial intercoms and active noise reduction headsets on speech intelligibility in an AWACS environment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 47, No. 1, pp. 100-103). SAGE Publications.

Bronkhorst, A. W., Veltman, J., & Van Breda, L. (1996). Application of a three-dimensional auditory display in a flight task. *Human Factors, 38*(1), 23–33.

Brungart, D. S., Ericson, M. A., & Simpson, B. D. (2002). Design considerations for improving the effectiveness of multitalker speech displays. In *Proceedings of the 2002 International Conference on Auditory Display, 1,* 424-430.

Casto, K. L., & Casali, J. G. (2013). Effects of headset, flight workload, hearing ability, and communication message quality on pilot performance. *Human Factors, 55*(3), 486–498.

Drullman, R., & Bronkhorst, A. W. (2000). Multichannel speech intelligibility and talking recognition using monaural, binaural, and three-dimensional auditory presentation. *Journal of the Acoustical Society of America, 107(*4), 2224–2235. https://doi.org/10.1121/1.428503

Ericson, M. A., Brungart, D. S., & Simpson, B. D. (2004). Factors that influence intelligibility in multitalker speech displays. *The International Journal of Aviation Psychology, 14*(3), 313–334.

Foyle, D. C., Andre, A. D., McCann, R. S., Wenzel, E. M., Begault, D. R., & Battiste, V. (1996). Taxiway navigation and situation awareness (T-NASA) system: Problem, design philosophy, and description of an integrated display suite for low-visibility airport surface operations. *SAE transactions*, 1411-1418.

Hansen, C. H., Snyder, S. D., Qiu, X., Brooks, L. A., & Moreau, D. J. (1997). *Active control of noise and vibration* (p. 1267). E & F.N. Spon.

Hart, S. G. (1986). *NASA task load index (TLX): Paper and pencil package-volume 1.0.*

Hartnett, G., Hicks, J., Durbin, D., Godfroy-Cooper, M., Miller, J., Feltman, K. A., St. Onge, P., Aura, C., & Stewart, J. (2020). *Pilot cueing for 360 obstacle awareness during DVE missions* (USAARL-TECH-FR--2020-026). U.S. Army Aeromedical Research Laboratory.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science, 182*(4118), 177–180.

Huang, H., Ricketts, T. A., Hornsby, B. W. Y., & Picou, E. M. (2022). Effects of critical distance and reverberation on listening effort in adults. *Journal of Speech, Language, and Hearing Research*, *65*(12), 4837–4851. https://doi.org/10.1044/2022_JSLHR-22-00109

Kahneman, D. (1973). *Attention and effort* (Vol. 1063, pp. 218-226). Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154(*3756), 1583–1585.

Kahneman, D., Onuska, L., & Wolman, R. E. (1968). Effects of grouping on the pupillary response in a short-term memory task. *The Quarterly Journal of Experimental Psychology*, *20*(3), 309–311.

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

Kim, S., Miller, M. E., Rusnock, C. F., & Elshaw, J. J. (2018). Spatialized audio improves call sign recognition during multi-aircraft control. *Applied Ergonomics*, *70,* 51–58. https://doi.org/10.1016/j.apergo.2018.02.007

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*, *33*(2), 291–300.

Kuo, S. M., & Morgan, D. R. (1996). *Active noise control systems* (Vol. 4). Wiley.

McAnally, K. I., & Martin, R. L. (2002). Variability in the headphone-to-ear-canal transfer function. *Journal of the Audio Engineering Society*, *50*(4), 263–266.

McCoy, S., Tun, P., Cox, L., Colangelo, M., Stewart, R., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *The Quarterly Journal of Experimental Psychology Section A*, *58(*1), 22–33.

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, E., Moore, D. R., Barry, J. G., & Akeroyd, A. (2014). Listening effort and fatigue: What exactly are we measuring?. *International Journal of Audiology*, *53*(S1), 433–440.

McKinley, R. L., D'Angelo, C. W. R., & Ericson, M. A. (1997). Flight demonstration of an integrated 3-D auditory display for communications, threat warning, and targeting. *Audio Effectiveness in Aviation*, (596).

McKinley, R. L., & Ericson, M. A. (1997). Flight demonstration of a 3-D auditory display.

McKinley, R. L., Erickson, M. A., & D'Angelo, W. R. (1994). 3-dimensional auditory displays: Development, applications, and performance. *Aviation, Space, and Environmental Medicine, 65*(5 Suppl), A31–A38.

Milam, L., Akins, E., Simpson, B., Williams, H., & Jones, H. (2019). *Techniques to explore spatial audio cues for aiding helicopter navigation in degraded visual environments* (Report No. 2019-13). U.S. Army Aeromedical Research Laboratory.

Mozo, B. T., & Murphy, B. (1997). *Evaluation of the communications earplug in the H-53 and CH-46 helicopter environments* (Report No. 97-36). US Army Aeromedical Research Laboratory.

Noetzel, J., Henry, P., Mackie, R., Cave, K., Stefanson, J. R., Hale, J. K., ... & Jones, H. (2025). Simulated Hearing Loss on Speech Recognition, Flight Performance, and Workload in Aviators. *Aerospace Medicine and Human Performance, 96*(4), 269-278.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing, 37(*Suppl 1), 5S–27S. https://doi.org/10.1097/AUD.0000000000000312

Picou, E. M., Ricketts, T. A., & Hornsby, B. W. Y. (2013). How hearing aids, background noise, and visual cues influence objective listening effort. *Ear and Hearing, 34(*5), e52–e64. https://doi.org/10.1097/AUD.0b013e31827f0431

Ray, J., Maw, E., & Muqolli, E. (2022, September 22). *F-16 3D audio localization final report*. Georgia Tech Research Institute, Georgia Institute of Technology.

Schluroff, M. (1983). In the eye of the beholder: Cognitive effort during sentence processing. *Meaning, Use, and Interpretation of Language*, 302–323.

Sheffield, B., Ziriax, J., Keller, M. D., Barns, W., & Brungart, D. (2017, September). The impact of reduced speech intelligibility on reaction time in a naval combat environment. In Proceedings of the *Human Factors and Ergonomics Society Annual Meeting* (Vol. 61, No. 1, pp. 1570-1574). SAGE Publications.

Simpson, B. D., Brungart, D. S., Dallman, R. C., Joffrion, J., Presnar, M. D., & Gilkey, R. H. (2005, September). Spatial audio as a navigation aid and attitude indicator. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, No. 17, pp. 1602-1606). SAGE Publications.

U.S. Army Aeromedical Activity. (2021). *Aeromedical policy letters and aeromedical technical bulletins.* U.S. Army Aeromedical Activity.

Veltman, J., Oving, A., & Bronkhorst, A. W. (2004). 3-D audio in the fighter cockpit improves task performance. *The International Journal of Aviation Psychology, 14*(3), 239–256.

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in Hearing, 22*, 1–32. https://doi.org/10.1177/2331216518800869

Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., & Kramer, S. E. (2014). The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. *Neuroimage*, *101,* 76–86.

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology, 51*(3), 277-284.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing*, *31*(4), 480–490.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing, 32*(4), 498–510.

Appendix A. Acronyms and Abbreviations

3D	3-Dimensional
ANR	Active Noise Reduction
ANOVA	Analysis of Variance
ATC	Air Traffic Control
CEP	Communication Earplug
CI	Confidence Interval
dB	Decibel
ETA	Estimated Time of Arrival
FVL	Future Vertical Lift
GAM	Generalized Additive Model
HL	Hearing Level
HRTF	Head Related Transfer Function
Hz	Hertz
NASA-TLX	National Aeronautics and Space Administration Task Load Index
PD	Pupil Dynamics
RP	Research Pilot
SPL	Sound Pressure Level
USAARL	U.S. Army Aeromedical Research Laboratory

Appendix B. Survey Open Responses

1. Please select your MOST preferred headset condition out of the 5 conditions and state why.

ANR (No CEP)

- The listening was simplified but easy to understand, allowing for pilots attention to be focused elsewhere. Also comfort over a prolonged use seems to be worth noting.

3D audio + ANR (No CEP)

- Most clear and easiest to understand.
- The most comfortable configuration. I was also able to hear radio calls in this configuration better than the others.
- Quieter, easy to identify/isolate the radio calls that were pertinent to me.
- Calls came in clearer and clean with less fluctuation in noise.
- Seemed like the best clarity. CEPs were a little intrusive.

3D audio + CEP

- Provided radio differential without muffling, able to distinguish.
- Seemed to be able to determine who was talking to me more clearly and didn’t get drowned out by other calls.
- The 3D audio helps separate the voices allowing for better listening.
- Felt the cleanest with most understanding.
- Familiar with fit and feel.
- Clarity of transmissions. Most normal sound that I'm used to but spatialized.

3D audio + ANR + CEP

- It gave me a better noise attenuation and gave me a different quadrant. Overall, any noise reduction will be helpful.
- Comfort, noise reduction, and ease of differentiating calls.
- Active noise reduction reduces ambient noise, engine noise, and background distraction. 3D spatial audio helps to distinguish between radios and helps to separate them, so they don’t get "stepped on."
- Reduced the overlapping voices and made it easier to prioritize the most relevant radio calls.
- Seemed more 'natural' to flying.
- I felt it had the best clarity and noise protection. Everything sounded clearer to me.
- Easiest to hear radio calls and identify what was being said. CEP, HGU seemed to provide most reliable and "closer" to ears for better hearing.

2. Please select your LEAST preferred headset condition out of the 5 conditions and state why.

CEP

- Took a little more time to decipher and I think I had to say "say again" more times, call overlap = frustration or inability to hear.
- Constantly getting stepped on and unable to clearly hear/understand communications.
- All the radios "step on" each other, no spatial audio, and far more ambient noise.
- It was the hardest configuration to hear which radio was transmitting.
- I couldn't isolate the radio call in my mind and the background noise was louder.
- Had the most confusion and understood the least.
- It made it harder to distinguish which radio and who was talking. Radios were too meshed together.
- Seemed difficult to sort through the noise.
- Ambient noise would override transmissions/receptions.
- After having the noise reduction and spatial audio I do not prefer CEP and helmet. It was obvious that it wasn’t as clear and crisp.
- CEPs were less comfortable. The audio was fair during the trial run.
- Harder to tell the voices apart. Less comfortable with CEPs.

ANR (no CEP)

- Everything seemed muddled together with radio calls.
- The ANR was uncomfortable with the helmet.
- Most background noise. Hard to hear transmissions when not spatialized.
- Difficult to identify who/channel was talking and could not hear entire communication.
- Harder to get a seal on ears to hear.

3D audio + ANR

- I felt the spatialized listening required the most focus, having to listen and focus on multiple quadrants can be task saturating. The quality and clarity were overall enhanced but may take some time getting used to.

3D audio + ANR + CEP

- ANR modes made distinguishing between radios difficult; radios were blending in ANR.

3. [Open ended question] Please share any additional comments of feedback you have regarding your experience with 3D audio technology and/or ANR technology in military aircraft operations.

- The inability to change volumes in the radios makes it more difficult to distinguish. Not changing volume made all radios blend together despite spatial differences.
- I think it will be super beneficial.
- active + passive + CEP + 3D audio would be great.
- As a pilot I am usually deciphering voices and who is talking from what seat. 3D audio allows me to interact like a normal conversation.
- Was great. Could understand much more clearly.
- Never experienced outside this study. However, I believe it will be beneficial once crewmembers get used to the zones of listening and prioritizing radios and their volumes.
- I felt I was able to focus easier on incoming calls with the 3D audio. Although when there was 3 or more people talking it was still hard to hear.
- Would really like to see the 3D audio implemented soon.
- Spatial audio greatly helps isolate/concentrate on the pertinent radio calls.
- ANR combined with 3D audio gave me an experience that I've never had in a UH-60. It significantly increased my ability to multitask effectively.
- It was interesting and I would like to see it evolve.
- The ability to isolate certain radios to have 3D audio would be beneficial. I like the ANR as a constant. It is possible for 3D audio to know the location of the incoming transmission.
- Will the headset be durable for maintenance? How available will the parts be to the units?
- I truly believe this would be beneficial.
- First time using it. Seemed effective.
- This is very useful technology and should be implemented ASAP.
- This would be VERY beneficial to apply additional information during a mission and reduce overload.

4. [Open ended question] Are there any specific improvements or enhancements you would like to see in future iterations of 3D audio technology and/or ANR technology?

- Differing volumes or artificial pitch change could assist in this, like auto tune or a pitch layer.
- I honestly believe any noise reduction would be amazing. I also believe having a molded CEP.
- Use ANR with passive noise reduction.
- ICS 3D spatial with headtracking like Apple air pods.
- Overall great experience, I would like to see its use in an actual aircraft in an active airspace.
- Spatial audio and noise reduction would be a great improvement to the headset.
- Would recommend the sim be the subjects' primary aircraft.
- Better helmet audio technology improvements such as the ANR + 3D audio.
- Include Bluetooth capability to cell.
- I would like ANR technology in the UH-60. I would like the 3D audio to be able to come from the front if possible.
- Would definitely utilize the 3D audio; however, ANR may make it difficult to hear channels in engines/rotor and other components within flying that could detract from awareness of issues that could arise.
- Implement it into aircraft sooner rather than later.

U.S. ARMY AEROMEDICAL RESEARCH LABORATORY

FORT RUCKER, ALABAMA

Optimizing

HUMAN PROTECTION
AND PERFORMANCE

since 1962

U.S. ARMY

FORGE THE FUTURE

FUTURES COMMAND

MRDC