USAARL-TECH-TR--2024-21

UNITED STATES ARMY AEROMEDICAL RESEARCH LABORATORY

The Effects of Simulated Hearing Loss on Aviator Performance and Cognitive Workload During Simulated Flight

Heath Jones, Jennifer Noetzel, Paula Henry, Kyle Hale,
Kevin Andres, Ryan Mackie, Xiaomin Yue, Riley McCormick,
Kichol Lee, & JR Stefanson

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) dated 8 November 2011.

IRB Determination and Number

This study, USAARL 2022-015, was approved by the Medical Research and Development Command Institutional Review Board on 19 January 2023.

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

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
12-04-2024	Technical Report	

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
The Effects of Simulated Hearing Loss on Aviator Performance and Cognitive Workload During Simulated Flight	
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
Jones, H.[1], Noetzel, J.[1], Henry, P.[1,2], Hale, K.[1,2], Andres, K.[1,2], Mackie, R.[1,3], Yue, X.[1,2], McCormick, R.[1], Lee, K.[1,2], & Stefanson, JR[1]	2022-015
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
U.S. Army Aeromedical Research Laboratory P.O. Box 620577 Fort Novosel, AL 36362	USAARL-TECH-TR--2024-21
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
U.S. Army Medical Research and Development Command Military Operational Medicine Research Program 504 Scott Street Fort Detrick, MD 21702-5012	USAMRDC MOMRP
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

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

13. SUPPLEMENTARY NOTES

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

14. ABSTRACT

Hearing loss can render an aviator more susceptible to the adverse effects of degraded communication signals and consequently lead to an increased allocation of mental resources to the task of processing radio communications (referred to as listening effort). The current study investigated the impact of simulated hearing loss on functional hearing assessments, flight performance, and cognitive workload of military-trained rotary-wing pilots. Subjects underwent current standard clinical audiometric testing and performed simulated rotary-wing flights. Two listening conditions, normal hearing and one of two simulated hearing loss, were tested. Clinical testing was conducted in a sound-treated audiometric booth using a tablet-based system and aviation communication earplugs. Simulated flight performance data were collected from pilots operating a full-motion UH-60 Black Hawk flight simulator at the U.S. Army Aeromedical Research Laboratory. Aviator performance was compared in high and low workloads across the different hearing conditions. Simulated hearing loss decreased all audiometric testing speech scores and decreased speech intelligibility in the flight simulator as well indicating that the larger the hearing deficit, the more missed or incorrect calls subjects had on average. Findings from this study will be leveraged in developing future protocols for aeromedical standards and provide data for the development of operator state monitoring capabilities.

15. SUBJECT TERMS

Hearing loss, workload, aviation

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
					19b. TELEPHONE NUMBER *(Include area code)*
UNCLAS	UNCLAS	UNCLAS	SAR	35	334-255-6906

This page is intentionally blank.

Summary

Hearing loss can render an aviator more susceptible to the adverse effects of degraded communication signals and consequently lead to an increased allocation of mental resources to the task of processing radio communications (referred to as listening effort). As such, Army aviation has implemented more stringent hearing standards for their aviators compared to other military occupations; however, these standards have traditionally been based primarily on pure tone and speech recognition test scores in quiet environments, which do not necessarily predict the functional impact of hearing loss. Recently, the Army adopted a new Military Operational Hearing Test (MOHT) to assess the functional impact of hearing loss. The current study aimed to evaluate the effects varying degrees of hearing loss have on flight performance and cognitive workload.

To investigate the impact of hearing loss on functional hearing assessments, flight performance, and cognitive workload, military-trained rotary-wing pilots in and around Fort Novosel, AL were recruited and self-screened for current active flight status. Subjects underwent current standard clinical audiometric testing and performed simulated rotary-wing flights. Two listening conditions, normal hearing and one of two simulated hearing loss according to Department of the Army Pamphlet (DA PAM) 40-502 and Aeromedical Policy Letters (APL) were tested. Clinical testing was conducted in a sound-treated audiometric booth using a tablet-based system and aviation communication earplugs. Simulated flight performance data were collected from pilots operating a full-motion UH-60 Black Hawk flight simulator at the U.S. Army Aeromedical Research Laboratory. Aviator performance was compared in high and low workloads across the different hearing conditions. Changes to workload were assessed by performance measurement on a secondary task, the National Aeronautics and Space Administration Task Load Index (NASA-TLX), administered following each route flown, and changes in pupil dilation as assessed using pupillometry.

Experimental results demonstrate that simulated hearing loss decreased all audiometric testing speech scores and increased the fail rate on the clinically adapted Modified Rhyme Test (MRT) portion of the MOHT. Degradation in speech intelligibility caused by the hearing loss simulator was seen in the flight simulator as well indicating that the larger the hearing deficit, the more missed or incorrect calls subjects had on average. Additionally, results from the NASA-TLX indicated the routes with a larger number of radio calls increased the subjects' workload, which exacerbated the degraded hearing loss condition as indicated by more missed radio calls. Findings from this study will be leveraged in developing future protocols for aeromedical standards, evaluating hearing loss mitigation strategies using various headset technologies and providing data for the development of operator state monitoring capabilities.

This page is intentionally blank.

Table of Contents

List of Figures

Table of Contents (continued)

List of Tables

Introduction

Army aviators require a level of hearing acuity to communicate in high operational tempos, which includes the use of multiple radios while performing flight operations. Military operations, including rotary-wing aircraft noise, present short-term risks to the communication abilities of Army aircrew and long-term risks to aviator hearing health in the form of hearing loss, which can be temporary or permanent. Hearing loss can render an aviator more susceptible to the adverse effects of degraded communication signal quality and consequently lead to an increased allocation of mental resources to the task of hearing, referred to as 'listening effort.'

Presently, Army aviation hearing standards, which are based on pure tone testing and speech recognition scores in quiet, do not predict the functional impact of hearing loss. The Army has recently adopted a new Military Operational Hearing Test (MOHT) to assess the functional impact of hearing loss to provide this data. Currently, no research exists that directly supports the predictive value of pure tone thresholds or word recognition in quiet on aviator-related performance. However, there is some evidence in the literature to suggest that there might be a synergistic relationship between the variables of hearing loss and aviator performance during portions of flight with high workload. The current study examined how various degrees of simulated hearing loss interfere with aviators understanding communication signals and its impact on flight performance.

Aviators Perform Hearing Critical Operations

Aviators use hearing and vision on a continuous basis performing their duties operating aircraft. Operating an aircraft requires that the operators use their vision to scan for issues inside and outside of the aircraft, monitor instruments, and look forward in the direction of travel. Operating an aircraft also requires the aviator to listen to radio communication, communication between personnel on the aircraft, communication with ground crews, and monitoring of warning signals. Aircraft are noisy environments; whether fixed-wing or rotary-wing, the noise generated by the engine and motors of aircraft exceeds safe levels during 100% of the time of operation. The noise generated within an aircraft makes hearing signals of interest, particularly speech communication and warning signals, challenging at times. Speech communication happens across headsets, which incorporate hearing protection but also occurs face-to-face.

Exposure to high levels of noise puts an individual at risk of experiencing hearing loss because of that noise exposure (Chen et al., 2020; Themann & Masterson, 2019). Hearing loss due to noise exposure typically begins in the high frequencies (between 2000 and 6000 hertz [Hz]) and then involves frequencies on either side of it as time goes on. The speech sounds that are of most importance to someone understanding speech are the consonants that occur in the high frequencies. Hearing loss makes speech communication difficult, and understanding speech in the presence of background noise is even more problematic.

Listening Effort Impacts Cognitive Workload

As a measure of mental effort, cognitive workload describes the level of mental resources required by an individual to complete one or more tasks. Putting this into the context of hearing-related tasks, it stands to reason that a hearing loss would impact the cognitive workload required to understand what is being heard. In fact, difficulties with hearing not only make understanding speech in noisy environments more challenging, but has also been shown to increase the amount of cognitive workload a listener needs to perform a listening task, which is referred to as 'listening effort' (Koelewijn et al., 2012; Kramer et al., 2013; Ohlenforst et al., 2017; Winn et al., 2018; Zekveld et al., 2010). With the impact degraded hearing conditions have on listening effort, it is important for Army aviation to understand any subsequent impacts degraded hearing has on the cognitive workload involved with other aviator-related tasks, such as flying the aircraft. Essentially, the greater the listening effort exerted by an aviator to understand what is being heard, the more mentally taxing all tasks being performed by the aviator becomes.

Growing support in the literature suggests listening effort be considered as an additional metric to complement speech intelligibility (i.e., the percentage of speech that a listener can understand) when quantifying functional impacts of various hearing conditions (Winn et al., 2018; Zekveld et al., 2014; Zekveld et al., 2010). For example, a listener might be able to obtain the same intelligibility score in two different listening situations but exert varying amounts of cognitive workload to do so. Indeed, studies have also shown that individuals with hearing impairment often employ cognitive strategies to compensate for the inability to hear (Peelle, 2018). With this being the case, even though an aviator with a hearing loss may be able to ultimately understand the same percentage of radio communications as expected with normal hearing, the amount of listening effort required to achieve that same level of performance is likely much greater.

In general, there are three main ways to measure cognitive workload. These include measurements of cognitive performance while completing multiple tasks (e.g., dual task paradigms), subjective ratings (i.e., direct scaling of mental effort collected from questionnaires), and measurement of physiological responses while performing tasks. While dual-task paradigms have been shown to be successful in measuring listening effort (Gagne et al., 2017), pupillometry has gained popularity as a tool for measuring cognitive workload (Winn et al., 2018; Zekveld et al., 2010). One reason for the increased popularity of pupillometry is that cognitive tasks across a wide range of domains can be differentiated into categories of more or less effortful based on changes in pupil diameter (Beatty, 1982). However, most all listening effort studies employing pupillometry have tightly controlled variables such as environmental luminance, stimulus delivery, reduced head movement and directed gaze. Implementing pupillometry as a tool to measure cognitive workload or listening effort in a dynamically changing environment, such as an aircraft cockpit, represents a major technical challenge. The current study represents a preliminary attempt to use pupillometry for assessing cognitive workload while flying a full motion, rotary-wing flight simulator.

Current Hearing Standards for Army Aviators

Hearing fitness-for-duty standards delineated in Department of the Army Pamphlet (DA-PAM) 40-502 (Department of the Army [DA], 2019a), Army Regulation (AR) 40-501 (DA, 2019b), and Aeromedical Policy Letter (APL) (United States Army, 2015) are based upon pure tone audiometric thresholds. The Army has recently updated its medical readiness standards in terms of hearing acuity and auditory fitness-for-duty (DA, 2019a; DA, 2019b). Hearing loss assessments and profiling conditions were changed dramatically under the revision. Consideration for performance on the new assessments, (i.e., MOHT) and profiling conditions should be investigated for updated APLs or inclusion criteria for any exception to policy or waiver. The APL hearing standard, by way of audiometric thresholds or waiver criteria, has not been updated since prior to 1984. Currently, for aviators who do not meet the pure tone standards, waivers are granted based on meeting the criterion value of 84% binaural word recognition in quiet. Currently, no research exists that directly supports the predictive value of pure tone thresholds or word recognition in quiet on aviator-related performance. A thorough explanation of current Army and APL hearing standards follows.

The initial entrance standards for the Army are governed by AR 40-501 *Standards of Medical Fitness*, while continued service and individual medical readiness are governed by AR 40-502 *Medical Readiness* and DA PAM 40-502 *Medical Readiness Procedures*. These regulations also outline various medical requirements for duties or jobs beyond initial entry, to include Special Forces, Survival, Evasion, Resistance, and Escape (SERE) training, divers, Ranger Regiments, and aviation. Additionally, Army aviation has its own APLs that guide medical entrance and retention standards for Army aviators, specifically. The APL criteria are stricter than regular Army medical standards outlined in any previous regulation. When an individual does not meet an established requirement, a waiver is required for them to enter or continue service.

As opposed to entrance into the Army, continued service is governed by DA PAM 40-502, *Medical Readiness Procedures* (DA, 2019a). All Army Service Members are enrolled into a hearing conservation program due to the innate noise exposure that accompanies military service. One requirement is that all Service Members must receive an annual hearing test to assess for hearing loss or any shifts in their hearing status. Annual hearing tests are automated and conducted in a group setting. All Service Members are assigned a hearing profile as part of their annual exam, which is an indication of their current hearing status. Hearing profiles are designated serially from H1 to H4. H1 suggests the best hearing, indicating a high level of medical fitness, or that the individual is fully qualified based on their hearing ability. There is an expectation that their hearing status will have no impact on their performance. As the number designator increases, so does the anticipated degree of impact on performance and limitations (i.e., H2, H3, H4). These higher number designators are assigned after a diagnostic audiological assessment.

In 2019, AR 40-502 was updated, replacing AR 40-501 *Standards of Medical Fitness*, and clarifying the retention standard for hearing assessment with a diagnostic audiogram and for the first time, including functional performance on the MOHT in fitness-for-duty evaluations. Used in conjunction with AR 40-502, the updated DA PAM 40-502 outlines hearing profiles based on audiometric thresholds and MOHT performance.

This space is intentionally blank.

Table 1. DA PAM 40-502 Audiometric Thresholds (dB HL) for Hearing Profiles H1 and H2

Profile		**500 Hz**	**1000 Hz**	**2000 Hz**	**3000 Hz**	**4000 Hz**	**6000 Hz**
H1	**Better ear**	≤ 25	≤ 25	≤ 25	≤ 25	≤ 25	≤ 60
	Worse ear	≤ 30	≤ 30	≤ 30	≤ 35	≤ 45	-
H2	**Better ear**	≤ 25	≤ 30	≤ 25	≤ 40	≤ 60	≤ 70
	Worse ear	≤ 40	≤ 40	≤ 60	-	-	-

The update of hearing profiles for Army Service Members was implemented after years of research conducted at Walter Reed National Military Medical Center (Brungart et al., 2023). The previous profiling system was based primarily off of a Service Member's pure tone thresholds with minimal assessment into their functional performance, solely through binaural word recognition performance in quiet. The MOHT was designed to evaluate the functional auditory performance of individuals with elevated auditory threshold levels. Operational effectiveness is correlated with auditory thresholds; however, individual performance on mission critical auditory tasks can vary significantly. Therefore, if a Service Member's audiometric thresholds exceed the H2 profile pure tone threshold criteria, the MOHT is administered, and functional performance dictates profile designation. The MOHT is comprised of three components: 1) diagnostic audiological evaluation, including pure tone auditory thresholds and monaural word recognition in quiet scores for each ear, 2) an assessment of speech-in-noise performance by administering the clinical adapted 80-word Modified Rhyme Test (MRT_{80}), and lastly, 3) an evaluation of spatial awareness, which is obtained by administering the Spatial Digit Test (SDT) for any individual with significant hearing deficits in at least one ear. The next sections briefly describe the MRT_{80} and the SDT background and administration.

Aeromedical concerns for hearing loss include difficulty with in-flight communications, radio transmissions, and rapid and accurate assessment of warning tones in the cockpit. All these auditory tasks can impact flight safety and mission success. Inclusion into Army aviation requires even stricter adherence to auditory thresholds and referral criteria. Aviators are required to complete annual audiological evaluations adhering to the Department of Defense (DoD) hearing readiness referral criteria (i.e., significant threshold shifts averaging 10 decibels [dB], etc.). Additionally, aviators must adhere to the APL referral criteria, such as a 20 dB shift in either ear at a single frequency (1000 through 4000 Hz), which requires a full audiological evaluation. The APL outlines serial Class I categories 1 through 4 (C1-C4). All aviator applicants are C1, with exceptions being granted to pilots transferring from another Service or rated international pilots. C2 represents all rated aviators or front seaters. C3 and 4 are trained aviation personnel with a requirement for flight status and include flight surgeons, aircrew, air traffic controllers, and unmanned aerial system operators. Table 2 outlines the audiological thresholds for the C1 and C2/3/4 categories.

If the Class category is met with audiological thresholds alone, no further assessment is required. If an aviator does not meet the APL standard, an exception to policy (ETP) or waiver is considered. An ETP is for applicants wishing to enter aviation and a waiver is for individuals already in service and wishing to continue. For simplicity, the term waiver will be used in the remaining document. According to the APL, hearing loss designated with an H2 profile may or may not be a disqualifying condition. Most H2 profiles can apply for a waiver if their hearing status does not impact flight performance, which is determined by the flight surgeon. Waivers are not recommended for individuals with H3 hearing profiles. Waivers are considered on a case-by-case basis.

Table 2. APL Hearing Standards for Army Aviation and Air Traffic Control in dB HL (decibels Hearing Level)

	500 Hz	**1000 Hz**	**2000 Hz**	**3000 Hz**	**4000 Hz**	**6000 Hz**
Class 1	≤ 25	≤ 25	≤ 25	≤ 35	≤ 45	≤ 45
Class 2/3/4	≤ 25	≤ 25	≤ 25	≤ 35	≤ 55	≤ 65

The current audiological workup required for a waiver includes pure tone air and bone conduction testing, tympanometry, acoustic reflex testing, speech reception threshold (SRT) testing, and word recognition scores (WRS) in quiet in both monaural and binaural conditions. There is a requirement to score greater than or equal to 84% on WRS. If an aviator scores lower than 84%, the APL notes a requirement for in-cockpit/flight evaluation. This evaluation is determined by the aeromedical provider and is not standardized. The in-flight/cockpit evaluation is an attempt to ensure that hearing loss does not have a functional impact on their operational performance. However, in-flight/cockpit evaluations are fiscally and time intensive and rarely, if ever, performed.

MOHT with Clinically Adapted Modified Rhyme Test.

Clinically Adapted Modified Rhyme Test (MRT_{80}).

The MRT_{80} is a clinical adaptation of the MRT. The MRT is identified in MIL-STD1472H, DoD *Design Criteria Standard Human Engineering* (MIL-STD-1472H, 2020), as well as the Federal Aviation Administration (FAA) Human Factors Design Standard (HF-STD-001), as the method used to measure the communication performance of communication systems such as radios or headsets. The clinical adaptation of this speech-in-noise test measures a Service Member's ability to recognize a word from a list of six possible alternatives that differ only by either the initial or final consonant (Brungart et al., 2021). See Table 3 for word list examples. The MRT_{80} uses a mixture of speech stimuli that include high (+4 dB) and low (-4 dB) signal-to-noise ratios (SNR), meaning that at times, the speaker can be 4 dB higher in intensity relative to the background noise (+4 dB SNR) or 4 dB below the background noise (-4 dB SNR). The MRT_{80} is administered through a calibrated clinical audiometer set at 70 dB hearing loss (HL). However, to assess in various intensity levels, these SNRs are presented with loud (78 dB SPL) and quiet (70 dB SPL) speech stimuli through the audiometer. The primary function of the clinically adapted MRT_{80} is to assess an individual's speech-in-noise performance, providing a better understanding of their functioning in operational hearing tasks. Brungart et al. (2023) validated performance on the MRT_{80} using 288 Service Members with varying degrees of hearing loss.

This space is intentionally blank.

Table 3. MRT Word List Examples

Example list 1	DID DIG DILL DIM DIN DIP	Differs only by the final consonant
Example list 2	BARK DARK HARK LARK MARK PARK	Differs only by the initial consonant

Spatial Digit Test (SDT).

Military operations require a Service Member to be able to detect, localize, and react to potential threats in their environment. The SDT was developed as a clinical tool to assess a Service Member's ability to spatially localize/separate stimuli between the left and right ears. Administration of the SDT involves digit pairs presented simultaneously, but with an 800 µs interaural delay to the listener's left and right ears. This interaural delay creates the perception that the digits are heard separately in the left and right ears. Service Members are instructed to respond only with the digits that appear to originate on one side and ignore the digits that appear to originate on the other side (i.e., respond with the digit you hear in the right ear only). Normal hearing listeners have clear spatial separation between the digit pairs with the perception that one digit is heard in the left ear and the other digit is heard in the right ear.

Profile determinations and MOHT administration.

All Service Members are required to complete an annual hearing readiness audiogram conducted through the Defense Occupational and Environmental Health Readiness System (DOEHRS) system as part of their enrollment into the Army hearing program and ongoing monitoring. The annual hearing exam or audiogram is conducted through an automatic system in a group setting. If a Service Member's annual audiogram presents with thresholds that exceed the DA PAM 40-502 standards, then they are referred to an audiologist for a comprehensive audiological assessment. If the audiometric thresholds exceed H1 values, but are lower than H3 values, they are assigned a hearing profile as H2, no further assessment is required. If a Service Member's audiometric thresholds exceed H2 levels, they are not guaranteed an H3 profile because the MOHT is administered to make the profile determination.

Table 4 outlines the MOHT scoring criteria and profile designations. Monaural word recognition in quiet is the first criterion to consider with 78% correct designated as the cutoff. If the monaural word recognition score is equal to or better than 78% in both ears, the MRT_{80} is administered. A passing score is based on the Service Member's pure tone threshold in the better ear at 2000 Hz. Greater hearing loss requires the Service Member to perform better on the MRT_{80} to retain an H2 profile designation. If the Service Member meets the passing score after the first list, testing is complete; however, if the Service Member does not meet acceptable performance, a second list is administered, and a score is calculated out of a 160-word set (two administrations of the MRT_{80}) as opposed to a single 80-word set. The MOHT is standardized across all Army Military Treatment Facilities (MTF) via a tablet-based system with the MRT_{80} and SDT preloaded. The tablet is interfaced with the clinical audiometer for calibrated administration. The Service Member, under headphones, will complete the MRT_{80} and the SDT (if applicable).

The spatial awareness criteria are designed to ensure that Service Members who have audiometric thresholds that are at H3 levels are not reassigned to H2 unless they have some ability to identify and localize sounds. For individuals with good hearing in the low frequencies (i.e., thresholds for the worse ear lower than or equal to 40 dB hearing loss (HL) for 500 and 1000 Hz and lower than or equal to 60 dB HL for 2000 Hz), no further testing is needed. For individuals who score higher than or equal to 78% on monaural word recognition but present with significant low to mid frequency hearing loss in the worse ear (i.e., thresholds higher than 40dB HL at 5000 Hz, higher than 40 dB HL at 1000 Hz, or higher than 60 dB HL at 2000 Hz) further assessment is required. The Service Member must correctly identify the digit in the target ear in 8 out of 10 trials to obtain a passing score. If the Service Member's performance meets all the criteria under the following sections: 1) monaural word recognition score, 2) MRT_{80}, 3) SDT or low frequency thresholds, they can be reassigned from an H3 to an H2.

Table 4. MOHT Profile Determinations

	H2 Profile	**H3 Profile**	**H4 Profile**
Monaural Word Recognition Score (WRS)	≥ 78% in both ears	≥ 78% in better ear	< 78% in better ear
Modified Rhyme Test (MRT_{80})	Better ear ≤ 20 dB HL at 2000 Hz; MRT ≥ 55/80 or 104/160	≥ 80/160	< 80/160
	Better ear > 20 dB HL at 2000 Hz; MRT ≥ 59/80 or 112/160		
Spatial Digit Test (SDT) or Low Frequency Thresholds	SDT ≥ 8/10	N/A	N/A
	Worse ear ≤ 40 dB HL at 500 Hz, and ≤ 40 dB HL at 1000 Hz, and ≤ 60 dB HL at 2000 Hz		

This space is intentionally blank.

Methods

The current report consists of two studies. Both studies were conducted under the supervision of the U.S. Army Aeromedical Research Laboratory (USAARL) Regulatory Compliance Office and the U.S. Army Medical Research and Development Command Institutional Review Board (USAMRDC IRB). Military-trained aviators were recruited from in and around Fort Novosel, AL and were self-screened for current active flight status. Subjects underwent current standard clinical audiometric testing, including the MOHT. Two listening conditions, normal hearing and one of two simulated hearing loss conditions, were tested.

Clinical audiometric testing was conducted in a sound-treated audiometric booth using a tablet-based system and aviation communication earplugs (CEPs). Simulated flight performance data were collected from pilots operating a full-motion UH-60 Black Hawk flight simulator at USAARL. Aviator performance was compared in high and low workloads between the normal hearing and hearing loss conditions. In addition to audiometric performance data, performance measures such as flight task completion and subjective and physiological measurements of workload were collected.

Participants

For the experiments, 32 pilots (31 male, 1 female) were recruited from the Fort Novosel area. All pilots were between the ages of 25 and 61, native English speakers, and were fit to fly at the time of the study (i.e., had a current up-slip documented on a DD 2992). Ten subjects were dismissed for satisfying the exclusion criteria of hearing thresholds exceeding 25 dB HL at one or more test frequencies. A total of 22 subjects completed the study; however, 21 complete data sets (20 male, 1 female) were used for analysis because the experimental setup for the simulated flights (i.e., Phase II) changed slightly after collecting the data for the first subject.

Hearing Loss Simulator

The hearing loss simulator was comprised of two USB sound cards connected to a laptop. Stereo CEPs were connected as the output of one sound card, and the input to the opposing sound card was used to deliver the audio stimulus being tested. Using a software architecture based on the commercially available Hearing Loss and Prosthesis Simulator (HeLPS) headset, the system programming and real-time audio was managed. Depending on the audiogram profile being simulated, the algorithm passed in audio from the input of the sound card at select frequencies and raised the absolute detection thresholds (i.e., hearing levels) for other frequencies by attenuating the levels from the earphones and adding masking noise. The system is also intended to give the user a sense of loudness recruitment, which is the unusually quick increase in perceived loudness that is concurrent with a particular hearing loss. This rendered sounds that fall below the predetermined threshold at specific frequencies inaudible, while sounds that are presented well above the threshold are as loud as they would be to a listener whose hearing is not compromised.

This method of simulating a hearing loss has been used previously (Dubno & Schaefer, 1992; Farrar et al., 1987; Humes et al., 1987; Sheffield et al., 2015; Sheffield, Brungart, et al., 2017; Sheffield, Ziriax, et al., 2017; Zurek & Delhorne, 1987). In the present study, the hearing loss simulator laptop received an audio output signal from a calibrated audiometer and routed to stereo CEPs with or without attenuation. For the MRT test via tablet, the hearing loss simulator received the audio output from the tablet and routed the signal to the CEPs. To ensure the hearing loss simulator was applying the appropriate levels of attenuation, sound pressure levels (SPLs) were measured on an acoustic test fixture (ATF). The measurements were conducted in three conditions: with the hearing loss simulator turned on but without applying a hearing loss profile, applying a mild hearing loss to simulate an H2 profile, and applying a more severe hearing loss to simulate an H3 profile. For the acoustic measurements, CEPs were donned on an ATF (GRAS 45CB) located in a sound-isolation booth and the audiometer was set to output 70 dB HL at 1000 Hz. Each condition was measured for approximately 10 seconds: no hearing loss, applying an H2 profile, and applying an H3 profile. Figure 1 shows the measured SPLs for each condition and demonstrates reduced levels for each hearing loss condition as expected.

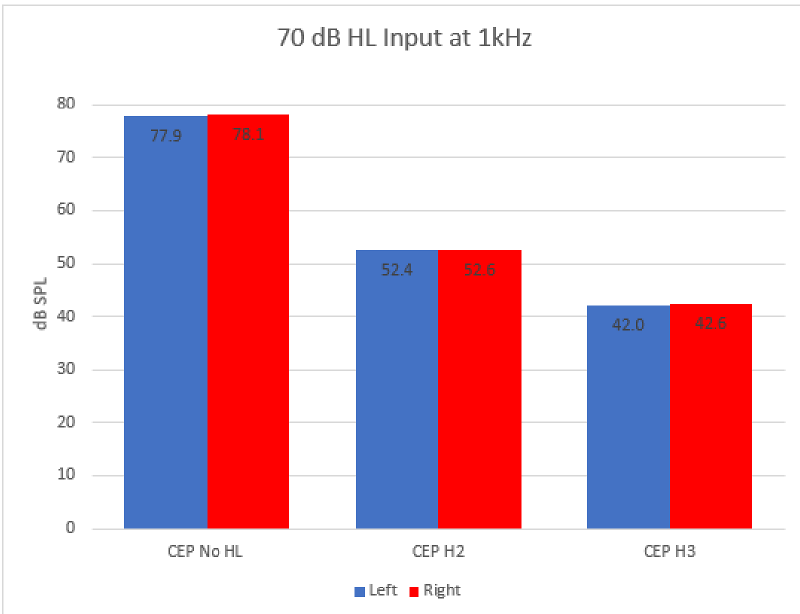

Figure 1. Reduction of sound pressure levels by the hearing loss simulator.

Phase I – Clinical Audiometric Testing

All subjects completed the study in approximately four hours during one visit to the Laboratory. Each volunteer was initially provided an informed consent document to read, ask questions, and sign if they agreed to participate in the study. Following the informed consent process, a brief questionnaire was administered to record demographic data to include age, sex, experience flying, and to ensure subjects were not under the influence of medication, alcohol, or recent anesthesia.

Next, an audiological evaluation consisting of otoscopy and audiometry was performed in a double-walled sound booth by a certified occupational hearing conservationist. The otoscopic exam ensured there were no ear canal abnormalities, excessive cerumen, irritation, or infection in the ear canal that may interfere with testing or proper fitting of the CEPs. A clinical audiologic exam was conducted after otoscopy to assess hearing sensitivity for two reasons. First, as screening to ensure subjects' hearing thresholds were normal—25 dB HL or less at each of the tested frequencies (0.5, 1, 2, 3, 4, and 6 kilohertz [kHz]). Second, qualified subjects' thresholds were used to quantify the amount of attenuation to apply during the simulated hearing loss conditions (i.e., to simulate an H2 or H3 hearing profile). All subjects were randomly assigned either a mild or severe hearing loss representing either an H2 or H3 profile, respectively.

Following the screening procedures, word recognition testing was conducted in quiet using the Northwestern University Auditory Test Number Six (NU-6) word lists with normal hearing and again with one of the two simulated hearing loss conditions. Both testing conditions (normal and simulated hearing loss) were conducted using CEPs, which were fit to each subject's ears with appropriately sized foam earplug tips. The CEPs were routed from the hearing loss simulator, which received audio signals from a calibrated audiometer. The hearing loss simulator either passed the audio signal received from the audiometer to the CEPs without attenuation or applied an appropriate attenuation per frequency to simulate a hearing loss based on the subject's thresholds and assigned hearing loss condition (i.e., H2 or H3). Word recognition tests were conducted monaurally at 70 dB HL and binaurally at each subject's preferred listening level for both hearing conditions (i.e., normal, and simulated hearing loss).

The MRT_{80} was administered next via tablet using the same CEPs as the previous test for each hearing condition (normal and simulated hearing loss). If subjects did not score high enough on the first MRT_{80} word list (see Table 4), they continued to complete the second word list for each listening condition. After the final test, subjects were escorted to USAARL's Flight System Branch, where they completed Phase II in the flight simulator.

Phase II – Flight Simulator Testing

Four routes were flown by each subject in the flight simulator. Flight simulator tasks consisted of maintaining heading, altitude, and air speed, listening for directions from the air traffic controller, and using a button on the hand controls to "clear" the master caution warning light whenever the pilot noticed that it appeared. Each route was a combination of workload level (low workload or high workload) and hearing level (no hearing loss or hearing loss).

- Route 1 – No hearing loss/low workload
- Route 2 – Hearing loss/low workload
- Route 3 – No hearing loss/high workload
- Route 4 – Hearing loss/ high workload

Low or high workload for the routes refers to the number of radio calls and master caution light instances that the subject had to respond to during each 10-minute flight. Workload was increased by delivering an increased number of radio calls, both target and maskers, increasing the number of master caution warning lights the pilot had to respond to, and increasing the level of turbulence such that the pilot had to continually adjust inputs to the controls to maintain heading and elevation.

Each subject was fit with eyeglass frames containing infrared cameras used for measuring pupillometry throughout the duration of the flight. The eyeglass frames do not impair the subject's visual field, but measure changes in pupil size over time.

The pre-run set up including the following procedures for each subject:

- Load the initial conditions for the route
- Calibrate pupillometry
- Start recording for both simulator and eye tracker
- Synchronization procedure (i.e., blink and pull trigger simultaneously three times)
- Set hearing loss simulator

Once the pre-run procedures were completed, the research pilot started the run. Each subject flew four 10-minute routes under instrument meteorological conditions (IMC) while simulating radar vectors from an air traffic controller (ATC) via pre-recorded voiceover injects. Instrument routes were flown in northern California (NORCAL) and pre-recorded ATC injects simulated a NORCAL approach controller. Subjects started each route at 4000 feet (ft) mean sea level (MSL), 110 knots indicated air speed (KIAS). Subjects were vectored to published instrument approach procedures at Metro Oakland International (KOAK) and San Francisco International (KSFO) under varied levels of workload and hearing loss conditions. Radar vectors were simulated to the following approaches:

- Route 1 – SFO ILS 28R – low workload
- Route 2 – OAK ILS 12 – low workload
- Route 3 – OAK ILS 30 – high workload
- Route 4 – OAK VOR 10R – high workload

All four routes were flown in a 2B60M full-motion Black Hawk helicopter simulator. All subjects flew route 1 first under normal hearing conditions as a baseline. Order of execution of subsequent routes were randomized (routes 2-4). Subjects were asked to maintain a selection of appropriate common standards (in-flight) from the H-60 Technical Manual (TM 1-1520-280-10) throughout the duration of the flight. Standards assessed are as follows:

- Maintain heading ± 10 degrees
- Maintain altitude ± 100 feet
- Maintain air speed ± 10 KIAS

Radio call procedure.

Subjects were given three attempts at each radio call within 15-20 seconds of the call occurring. If the subject correctly acknowledged the call, it was counted as correct, and no repeat was necessary. If the subject missed the call completely, it was counted as a miss, and the call was repeated. If the subject had an incorrect read back, it was counted as wrong, and the call was repeated. If the subject requested the tower to "say again," it was counted as repeat requested and the call was repeated. If the subject failed all three attempts, the Research Pilot would dial the correct heading/altitude into the uncoupled flight manager and notify the subject that that was the heading/altitude the subject needed to reach.

Flight performance assessments.

Flight performance in the full-motion simulator was quantified by calculating the root mean square deviation (RMSD) for three flight metrics: altitude, heading, and air speed. Research subjects were instructed to maintain a constant air speed of 110 knots for all four routes. Heading and altitude requirements were provided to research subjects via radio calls during the flights. Research subjects were instructed to follow a standard rate of climb (500 feet per minute [ft/min]) and a standard rate of turn (180 degrees per minute [degrees/min]) during all flights tested here. To account for individual differences in research subjects, a mixed-effects linear regression model (R function lmer from the lmerTest package) was used to analyze potential differences in RMSD for each flight metric (heading, altitude, and air speed). Each regression model contained a fixed factor for flight number (four categories: 1, 2, 3, 4), a fixed factor for hearing level (two categories: H2, H3), and a random intercept for each subject. Maximum likelihood was chosen as the estimation method so that the fixed and random effects could be estimated simultaneously.

Prior to the regression model, the data were visualized (with quantile-quantile plots) and tested (with a Shapiro-Wilk test of normality) to ensure that it approximately followed a normal distribution. An appropriate transformation was applied to any data that was not approximately normal. Outliers were removed from the data if research subjects did not adequately follow instructions, or if the assumptions of the linear model were violated (outliers defined as: Q3 + 3xIQR or below Q1 – 3xIQR, where Q1 and Q3 are the first and third quartiles, and IQR is the interquartile range). Each regression model was checked to ensure that the residual values were normally distributed and had a constant variance (homoscedasticity). When the regression model showed a significant interaction effect between route number and hearing level, the data was split by hearing level and new regression models were created as described above. When the data was split and re-tested, p-values were adjusted using the Bonferroni method.

Cognitive workload assessments.

As mentioned previously, there were three main types of measurements used to assess cognitive workload. The current study employed all three by measuring performance on a secondary task, measuring changes in pupil dilation, and scoring subjective ratings of perceived workload provided by the research subject following each route. Two levels of workload were used to determine if the increase in workload could be differentiated based on these cognitive workload assessments.

Secondary task – Warning light response.

Previous research has shown that reaction time on completing a secondary task is a suitable measurement of cognitive workload (Verwey & Veltman, 1996). Briefly, the amount of mental capacity demanded by a primary task will impact the reaction time for secondary task completion. Given this, subjects were asked to respond to a randomized illumination of the "Check EICAS" light on the panel (Figure 2). The Check EICAS light would illuminate for up to 5.5 seconds, and then extinguish regardless of subject response. Subjects were able to acknowledge this inject via pressing the VOX-CAUT switch on the cyclic or by pressing the Master Caution button, congruent with the H-60 Technical Manual (TM 1-1520-280-10), which states "The master caution can be reset from either pilot position by pressing the Master Caution button (Figure 2) or by pressing the VOX-CAUT button on either cyclic stick (Figure 3). If the subject responded to this stimulus, the Master Caution button light would extinguish.

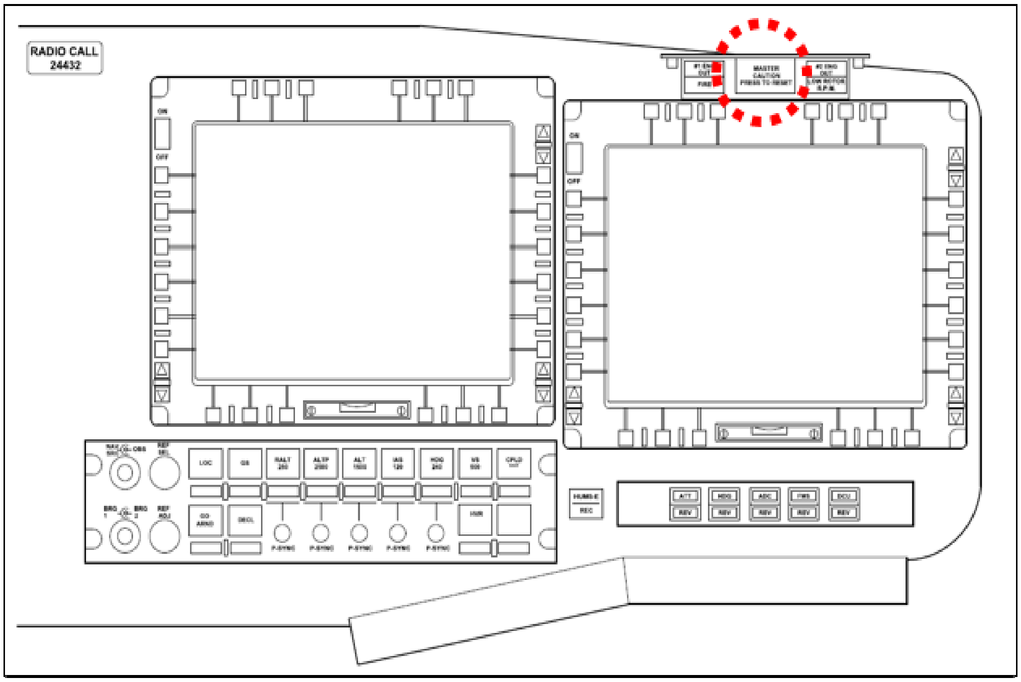

Figure 2. Master caution button.

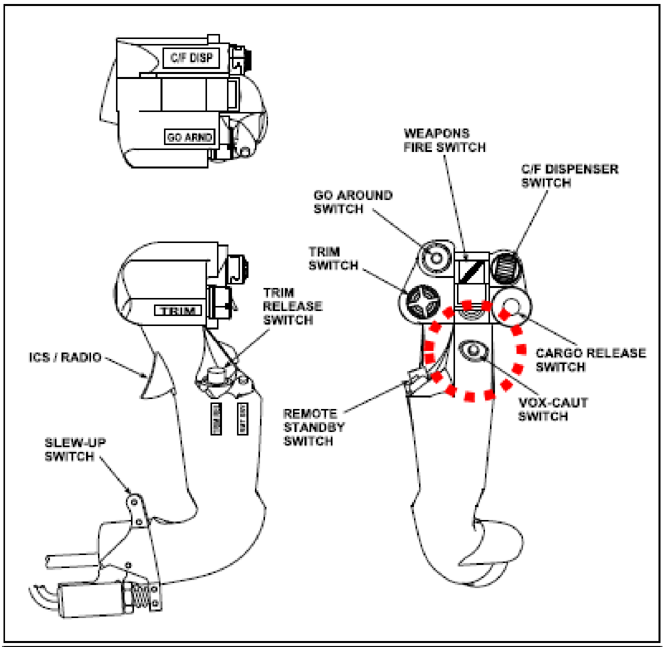

Figure 3. VOX-CAUT switch.

Research subjects were required to recognize and turn off the Master Caution warning light as quickly as possible during the simulated flights. The warning light came on 20 times during routes 1 and 2 (low workload) and came on 25 times during routes 3 and 4 (high workload). If the research subject did not turn off the warning light, it automatically turned off after 5.5 seconds. For each individual flight (of 20 or 25 warning lights), the warning light response times were averaged together. When research subjects did not respond to the warning light, a response time of 5.5 seconds was used. The average warning light response time was used for statistical analysis.

To account for individual differences in research subjects, a mixed-effects linear regression model was used to analyze potential differences in average warning light reaction times. The regression model contained a fixed-factor for route number (four categories: 1, 2, 3, 4), a fixed-factor for hearing level (two categories: H2, H3), and a random intercept for each subject. Maximum likelihood was chosen as the estimation method so that the fixed and random effects can be estimated simultaneously. The regression model assumptions and diagnostics were checked as described in the flight performance methods section.

NASA-TLX questionnaires.

Following each 10-minute route, subjects were asked to complete a subjective questionnaire, the NASA-TLX, which is a widely used, multidimensional assessment tool that asks individuals to rate their perceived workload (Hart, 1986). The workload assessment results in an overall workload score derived from subjective ratings according to six categories or subscales of 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). Subjects were asked to rate their perception, from 0 to 10, for each of the categories at the completion of each route while in the flight simulator. Within-subject differences were compared for each category of each route and averaged across subjects.

Pupillometry measurements.

Pupillometry measurements were conducted with Pupil Core eye tracking glasses. The initial set up was conducted in the control room with adjustments to the cameras to get a clear view of the pupils and a test calibration. The calibration procedure, as recommended by Pupil Core, was for the subject to hold the calibration card at arm’s length and move it in a circular motion while tracking the crosshairs in the center of the card with their eyes. After finishing the initial set up, the subject was briefed on the flight procedures by the research pilot and instructed on how to complete the NASA-TLX by a member of the research staff.

The calibration procedure before each route was the same as in the control room with subjects instructed to hold the calibration card at arm’s length and follow the crosshairs on the card with their eyes as they moved it in a circle. After calibration, the recording for both the simulator and eye tracker was initiated. The subject then performed a time synchronization procedure by blinking and pushing a button on the cyclic at the same time. Secondary synchronization points were acquired by selecting the frame number of the simulator and timestamp of the Master Caution light activating as seen in the Pupil Labs world camera.

Results

Phase I – Clinical Audiometric Testing

Pure tone thresholds and speech reception in quiet.

Monaural and binaural word recognition in quiet was collected with the participant's normal hearing and with simulated hearing loss. Word recognition was completed with the Northwestern University Auditory Test No. 6 (NU-6) word list consisting of 25 phonetically balanced words. Monaural presentation level was suprathreshold at 80 dB HL. Binaural presentation was at the participant's most comfortable listening level. The normal hearing condition maintained the highest average percent correct. As hearing loss increased, the percent correct decreased. These findings are expected; as we decrease audibility, speech recognition will also decrease. The current APL identifies greater than or equal to 84% binaural word recognition in quiet as the criteria for continued service in aviation. No participant in the normal hearing condition scored lower than 92%. In the H2/C2 simulated hearing loss condition, there were two participants who scored lower than 84%; both scored 80%. In the most severe hearing loss condition (H3), only two participants scored lower than 84%, two scored 72%. There were two participants who scored exactly 84% in the H3 condition.

Table 5. Word Recognition Scores in Percent Correct

Condition	Monaural Left	Monaural Right	Binaural
Normal Hearing (H1/C1)	89.5%	92.2%	97.3%
Simulated H2/C2	82.7%	81.0%	85%
Simulated H3	81.5%	79.6%	79.6%

MRT_{80} results.

The MRT_{80} was administered to each subject in their normal hearing condition and a simulated hearing loss condition. Twenty-one subjects completed the MRT_{80} in the normal hearing condition, 11 in the H2/C2 condition, and 10 in the H3 condition. As a reminder, there are different criteria for passing the MRT based upon the pure tone threshold at 2000 Hz; see Table 6. Two subjects in the H2/C2 condition did not complete the second MRT_{80} word list even though their scores were below the passing criteria, and they were excluded from the data below. It is unknown if administering the second word list would have resulted in a pass or fail of the MRT.

This space is intentionally blank.

Table 6. MRT Performance Criteria

Better ear 2000 Hz threshold	MRT_{80}		MRT_{160}	
≤ 20 dB	≥ 55/80	Pass	≥ 104/160	Pass
> 20 dB	≥ 59/80	Pass	≥ 112/160	Pass

Overall, the MRT pass rate was highest in the normal hearing condition (86%) and lowest in the severe, H3 simulated hearing loss condition, at 0% as shown in Figure 4. There were failures on the MRT in every hearing condition, including the individuals with normal hearing. In total, there were three MRT failures in the normal hearing condition, two failures in the mild or H2/C2 condition, and 10 failures in the severe or H3 condition.

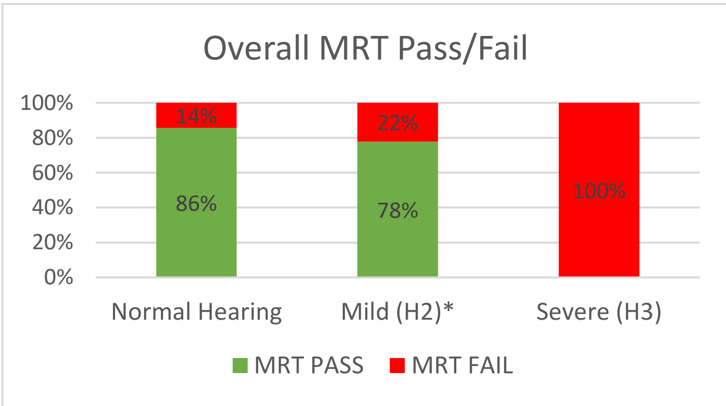

Figure 4. MRT pass/fail by hearing condition.

$MRT_{80/160}$ scores also varied as a function of simulated hearing loss. The MRT_{80} demonstrated the same pattern as word recognition in quiet with increasing hearing loss resulting in poorer performance. This pattern doesn't hold for the MRT_{160} at least in terms of percent score. An additional 80-word list in this instance is done to get a better statistical performance (Brungart et al., 2021). It is suspected that performance on the second MRT_{80} list, or the MRT_{160}, indicates the subject is moving closer to their true performance. For higher performers (normal hearing), scores went down slightly, lower performers (H3 or severe) performance improved, and there was no noticeable change for the H2/C2 (mild) group.

Table 7. Average MRT_{80} and MRT_{160} Overall Percent Correct

Condition	MRT_{80}	MRT_{160}
Normal hearing	69 ± 7.8%	65 ± 5.1%
Simulated mild (H2/C2)*	65.8 ± 6.4%	67 ± 5.4%
Simulated severe (H3)	53.3 ± 7.1%	63 ± 6.5%

*Two subjects were unable to complete the full MRT test as per DA PAM 40-502.

Although there were failures in every condition, operationally, only those individuals with elevated thresholds exceeding the H2 standard would be required to be evaluated with the MOHT. According to the current APL, only two participants did not meet the criteria for a waiver, having elevated thresholds greater than C2 and obtaining a binaural word recognition score of lower than 84%. Comparing this to the MRT_{80} results, there are vast differences. Every individual in the H3 condition failed the MRT_{80}, an operationally relevant test, suggesting that hearing loss may impact their ability to be operationally capable.

Phase II - Flight Simulator Testing

In-flight speech intelligibility.

Clinical speech testing showed simulated hearing loss decreased all speech scores and increased the fail rate on the MRT. Similar speech intelligibility scores were calculated based on the number of incorrect radio calls (i.e., either missed calls or calls for which participants asked for 'say again'). Figure 5 shows that the larger the hearing deficit, the more missed or incorrect calls participants had on average.

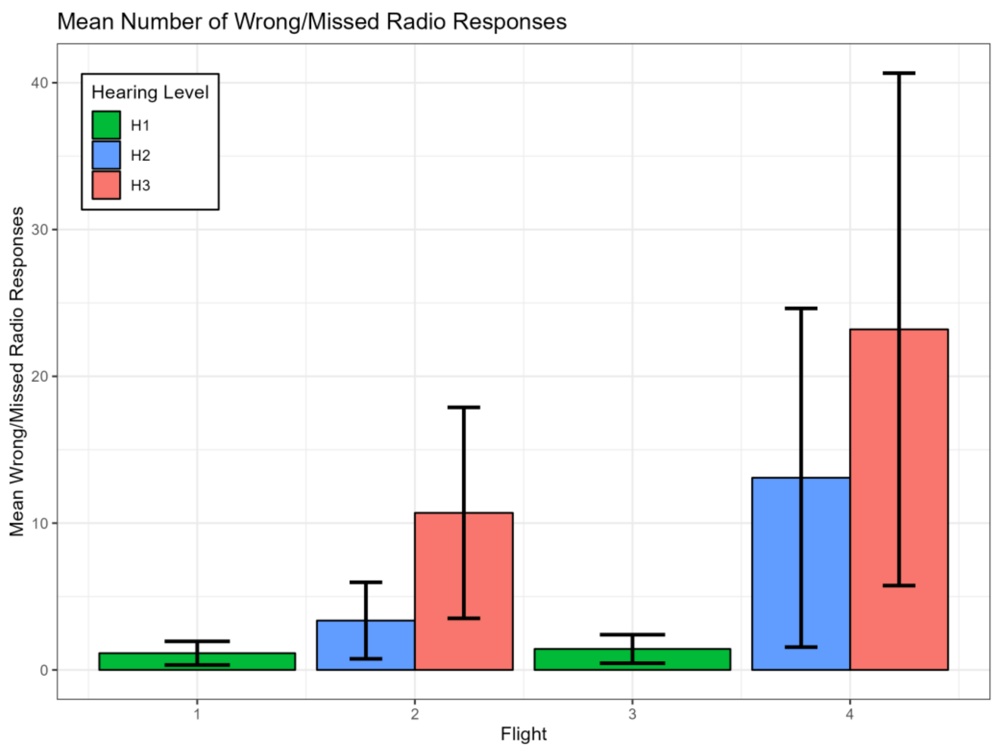

Figure 5. In-flight speech intelligibility. The total number of wrong radio responses was calculated as the sum of missed calls, wrong calls, and calls where the research pilot had to instruct the subject. The mean and 95% confidence interval was calculated for each route and hearing level. Routes 1 and 3 (i.e., flight number on the *y*-axis above) had no hearing loss applied (as indicated by the H1 green color) whereas routes 2 and 4 simulated either an H2 (blue) or H3 (red) hearing loss profile.

Flight performance results.

Heading and altitude deviations from the ideal flight path were observed (see Figure 6). The ‘ideal flight path’ simply refers to the path the aircraft would be on if subjects accurately maintained all the instructed headings and altitudes throughout the flight route. Larger deviations occurred in the high workload routes compared to the low workload routes. Increased workload exacerbated the degraded hearing loss condition as indicated by more missed radio calls.

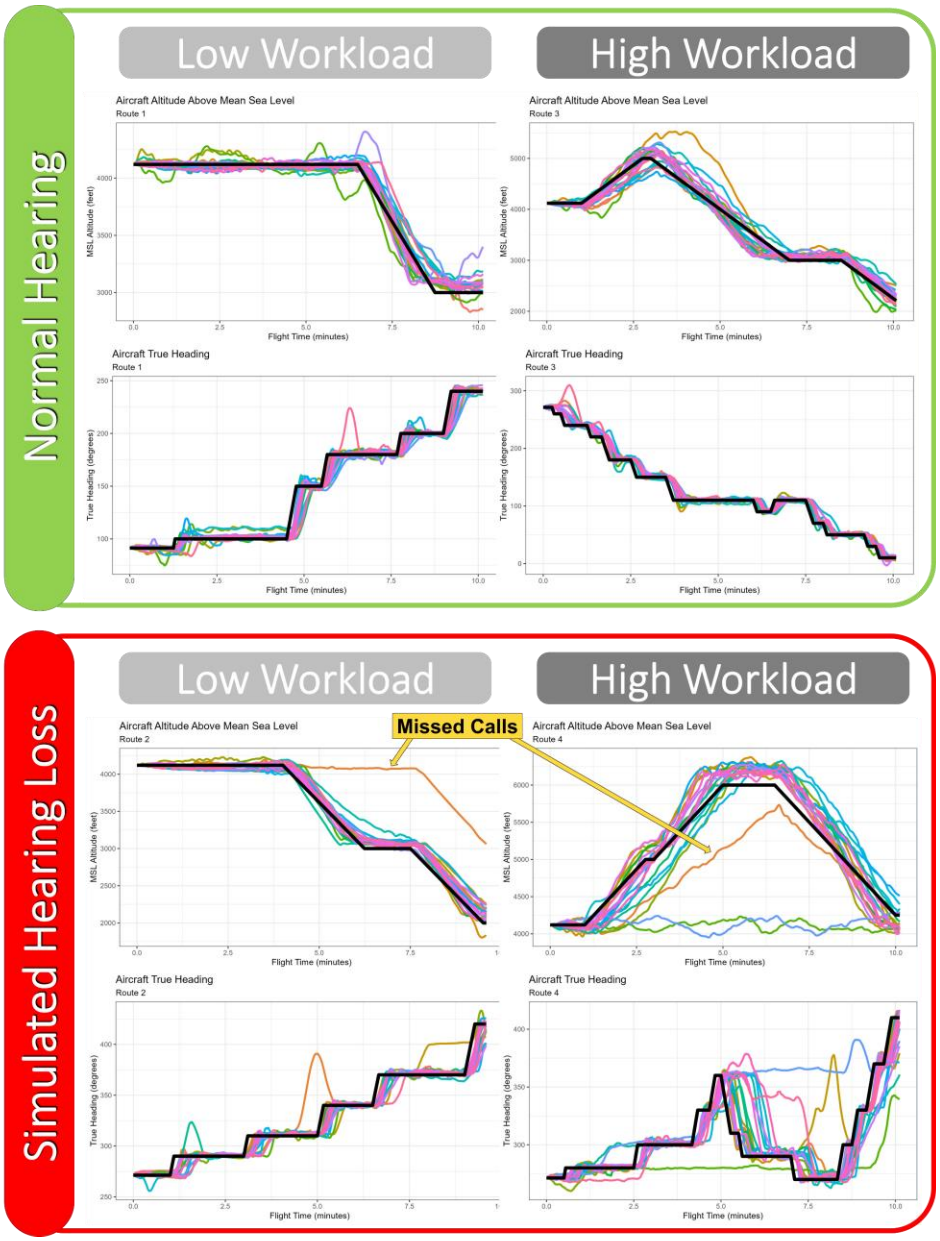

Figure 6. Raw flight path data. The thick black line represents the ideal flight path. Each colored line represents data from one subject. Flight performance was quantified by calculating the RMSD from the ideal flight path.

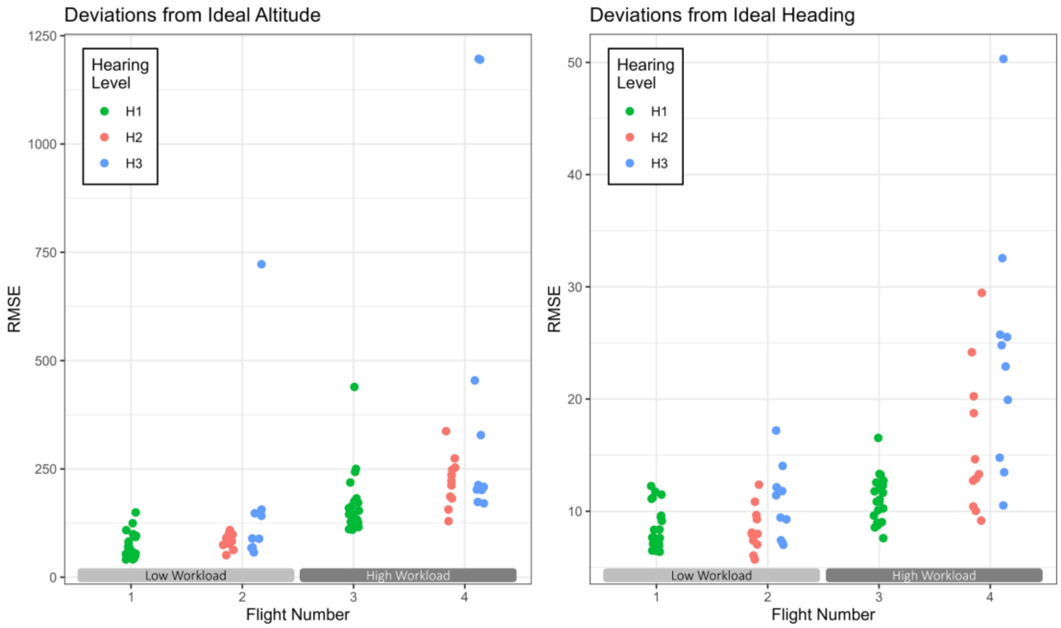

Figure 7. Summary of flight path deviations for each hearing and workload condition.

Altitude.

Prior to creating a regression model for altitude RMSD, three outliers were removed (due to research subjects not appropriately following radio call instructions to change altitude), and a natural log transformation was applied to the data. The mixed-effects linear regression model showed that route number was statistically significant ($F(3, 60.8) = 67.98$, $p < 0.001$). Hearing level ($F(1, 20.8) = 0.70$, $p = 0.41$) and the interaction effects ($F(3, 60.8) = 0.95$, $p = 0.42$) were not statistically significant. Pairwise comparisons of routes showed that altitude RMSD was statistically significantly different between routes 1 and 3 ($t(59.8) = 5.80$, $p < 0.001$), routes 2 and 4 ($t(59.8) = 7.77$, $p < 0.001$), and routes 3 and 4 ($t(59.8) = 2.82$, $p = 0.006$). Routes 1 and 2 did not show a statistically significant difference ($t(59.8) = 0.85$, $p = 0.40$). Table 8 provides summary statistics for each route.

This space is intentionally blank.

Table 8. Flight Performance RMSD Summary Statistics by Flight Number

	Flight	*N*	Median	Mean	SD	SE
Altitude RMSD	1	21	67.047	72.671	29.948	6.535
	2	20	84.626	89.043	29.839	6.672
	3	21	153.542	171.451	73.262	15.987
	4	19	211.941	230.950	75.806	17.391
Heading RMSD	1	21	7.657	8.590	1.935	0.422
	2	21	9.282	9.491	2.874	0.627
	3	21	10.873	10.972	2.114	0.461
	4	18	14.707	16.891	5.864	1.382
Air Speed RMSD	1	21	1.256	2.611	3.156	0.689
	2	21	0.867	1.798	2.342	0.511
	3	21	1.678	3.618	4.402	0.961
	4	21	1.701	3.178	3.422	0.747

Heading.

Prior to creating a regression model for heading RMSD, three outliers were removed (due to research subjects not appropriately following radio call instructions to change heading), and an inverse transformation was applied to the data. The mixed-effects linear regression model showed that route number ($F(3, 60.1) = 47.63$, $p < 0.001$) and the interaction between route number and hearing level ($F(3, 60.1) = 5.62$, $p = 0.002$) were statistically significant. Hearing level was not statistically significant ($F(1, 20.8) = 0.62$, $p = 0.44$). Due to the significant interaction effect, the data was split into two groups based on hearing level (H2 and H3) and the analysis was repeated to determine which routes showed significant differences in RMSD values.

Heading: Hearing level H2.

No additional outliers were identified, and an inverse transformation was applied to the data. The mixed-effects linear regression model showed that route number was statistically significant ($F(3, 31.9) = 30.80$, $p < 0.001$). Pairwise comparisons of routes showed that heading RMSD for the H2 hearing level group was significantly different between routes 1 and 3 ($t(31.8) = -4.39$, $p < 0.001$), routes 2 and 4 ($t(32.0) = -8.50$, $p < 0.001$), and routes 3 and 4 ($t(32.0) = -3.00$, $p = 0.010$). Routes 1 and 2 did not show a statistically significant difference ($t(31.8) = 1.28$, $p = 0.42$).

Heading: Hearing level H3.

No additional outliers were identified, and an inverse transformation was applied to the data. The mixed-effects linear regression model showed that route number was statistically significant ($F(3, 28.5) = 22.88$, $p < 0.001$). Pairwise comparisons of route showed that heading RMSD for the H3 hearing level group was significantly different between routes 1 and 2 ($t(28.1) = -3.07$, $p = 0.009$), routes 1 and 3 ($t(28.1) = -3.45$, $p = 0.004$), routes 2 and 4 ($t(28.9) = -5.37$, $p < 0.001$), and routes 3 and 4 ($t(28.9) = -5.02$, $p < 0.001$). Table 9 provides summary statistics for each hearing level and route.

Table 9. Flight Performance RMSD Summary Statistics by Hearing Level

	Hearing	Flight	*N*	Median	Mean	SD	SE
Altitude RMSD	H2	1	11	70.858	77.213	32.544	9.812
		2	11	83.053	81.498	17.886	5.393
		3	11	142.948	163.669	94.064	28.361
		4	11	222.868	221.523	58.048	17.502
	H3	1	10	54.029	67.673	27.627	8.736
		2	9	88.822	98.265	39.252	13.084
		3	10	169.674	180.010	44.053	13.931
		4	8	205.386	243.913	98.140	34.698
Heading RMSD	H2	1	11	8.343	9.008	2.231	0.673
		2	11	7.982	8.399	2.005	0.605
		3	11	10.873	11.401	2.268	0.684
		4	10	13.118	14.638	4.903	1.551
	H3	1	10	7.611	8.129	1.530	0.484
		2	10	10.426	10.692	3.291	1.041
		3	10	10.948	10.501	1.934	0.612
		4	8	21.414	19.706	6.023	2.129
Air speed RMSD	H2	1	11	1.788	3.211	3.702	1.116
		2	11	1.011	1.958	2.655	0.801
		3	11	1.664	2.700	3.765	1.135
		4	11	3.578	4.312	4.195	1.265
	H3	1	10	1.208	1.951	2.445	0.773
		2	10	0.560	1.622	2.072	0.655
		3	10	1.696	4.628	5.016	1.586
		4	10	0.975	1.931	1.780	0.563

Air speed.

Prior to creating a regression model for air speed RMSD, a natural log transformation was applied to the data. The mixed-effects linear regression model showed that route number was statistically significant ($F(3, 63) = 3.37$, $p = 0.024$). Hearing level ($F(1, 21) = 1.33$, $p = 0.26$) and the interaction effect ($F(3, 63) = 1.58$, $p = 0.20$) were not statistically significant. Pairwise comparisons of routes showed that airspeed RMSD was significantly different between routes 2 and 4 ($t(63) = 2.38$, $p = 0.020$). Routes 1 and 2 ($t(63) = -1.42$, $p = 0.16$), routes 1 and 3 ($t(63) = -0.18$, $p = 0.86$), and routes 3 and 4 ($t(63) = 1.14$, $p = 0.26$) did not show a significant difference.

Cognitive workload results.

Secondary task - Warning light response time results.

Prior to creating a regression model for average warning light reaction time, one outlier was removed, and an inverse transformation was applied to the data. The mixed-effects linear regression model showed that route number was statistically significant ($F(3, 62.1) = 7.63$, $p < 0.001$). Hearing level ($F(1, 21.1) = 1.27$, $p = 0.27$) and the interaction effect ($F(3, 62.1) = 0.48$, *p*

= 0.70) were not statistically significant. Pairwise comparisons of routes showed that average warning light reaction time was significantly different between routes 1 and 2 ($t(62.4) = 2.11$, $p = 0.039$) and routes 3 and 4 ($t(62.1) = -2.49$, $p = 0.016$); in both of these instances, adding hearing loss made a difference in response times. Additionally, when workload is increased and hearing loss is present there are differences as routes 2 and 4 ($t(62.1) = -4.17$, $p < 0.001$) were statistically different. There was no difference in reaction times in the normal hearing condition regardless of workload, as routes 1 and 3 did not show a significant difference ($t(62.4) = 1.70$, $p = 0.094$). Tables 10 and 11 provide summary statistics for average warning light reaction times. Although there were differences, the pattern indicated that hearing loss with low workload produced the fastest response time, however, hearing loss with high workload produced the slowest response times.

Table 10. Warning Light Reaction Time Summary Statistics by Route Number

	Route	*Hearing*	*Workload*	*N*	Median	Mean	SD	SE
Mean warning light response time	1	Normal	Low	20	1.705	1.820	0.476	0.106
	2	Hearing loss	Low	21	1.401	1.551	0.462	0.101
	3	Normal	High	21	1.513	1.687	0.559	0.122
	4	Hearing loss	High	21	1.698	1.842	0.438	0.096

Table 11. Warning Light Reaction Time Summary Statistics by Hearing Level

	Hearing	Route	*N*	Median	Mean	SD	SE
Mean warning light response time	H2	1	10	1.880	1.873	0.389	0.123
		2	11	1.471	1.637	0.408	0.123
		3	11	1.583	1.711	0.542	0.164
		4	11	1.698	1.917	0.445	0.134
	H3	1	10	1.582	1.766	0.566	0.179
		2	10	1.284	1.456	0.520	0.165
		3	10	1.489	1.660	0.605	0.191
		4	10	1.718	1.760	0.437	0.138

NASA-TLX questionnaires.

Participants completed the NASA-TLX following each route. Average subscale differences are plotted in Figure 8. As expected, the NASA-TLX scores confirmed an increase in perceived workload in both the hearing loss and workload conditions. There were larger perceived differences on the NASA-TLX within the simulated hearing loss condition when compared to the workload condition. NASA-TLX showed an increase in perceived workload on nearly every subscale and a decrease in perceived performance when hearing loss was introduced to the participant in both high and low workloads. The H3 participants typically reported worse scores than the H2. Changing workload alone produced deviations in perceived workload, with subjects, on average, experiencing higher amounts of frustration, mental effort, physical demand, temporal demand, and decreased perceived performance when flying under the high workload conditions in normal hearing. Increasing flight workload within the hearing loss conditions did not produce significant changes in the perceived workload.

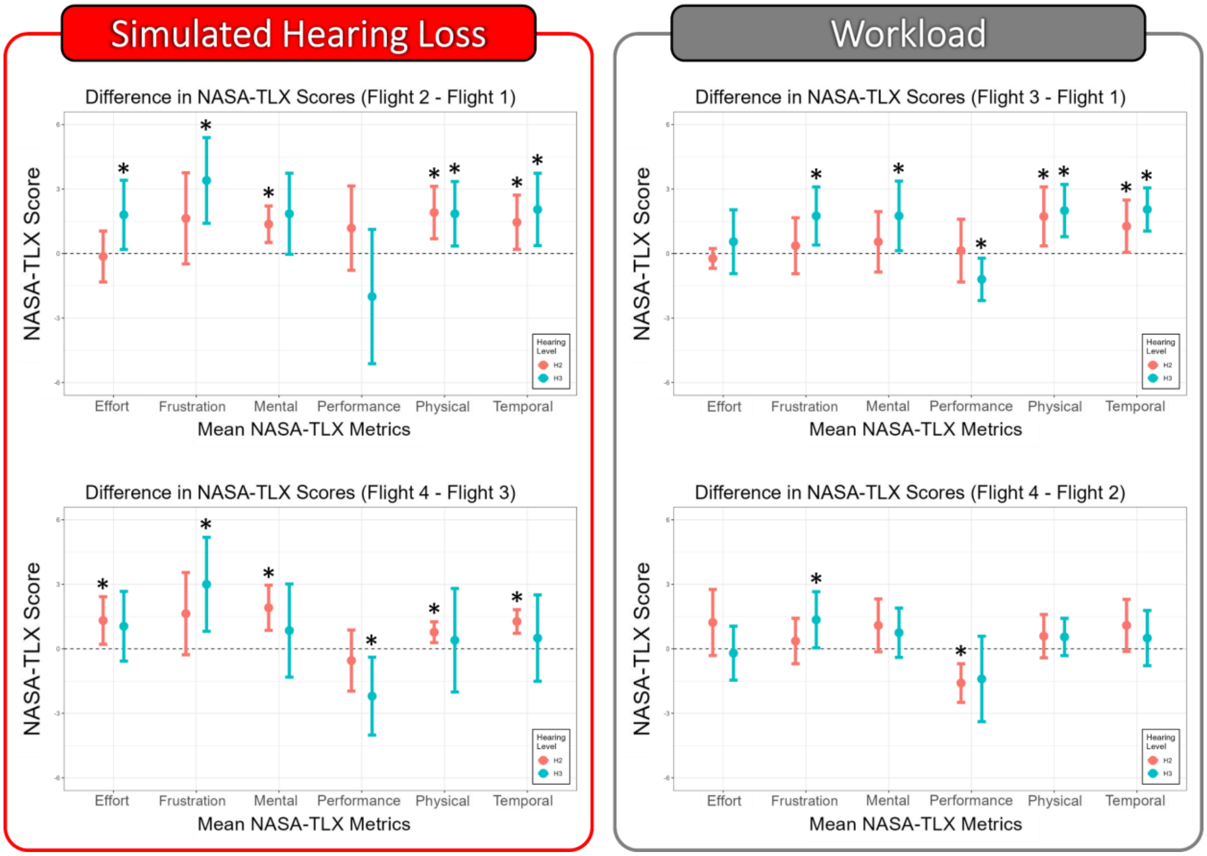

Figure 8. Summary of NASA-TLX scores as a function of hearing loss (left) and as a function of workload (right). Each point represents the across-subject average difference in score and standard deviation for each particular metric (i.e., effort, frustration, etc.) on the survey for each of the listening and workload conditions. Positive numbers above the dotted line indicate an increase in the metric listed on the *y*-axis and negative numbers respresent a decrease in score.

Pupillometry.

Data from the PupilCore eye tracking system underwent preprocessing using custom software. This processing involved removing aberrant data points through linear interpolation and de-spiking the data with an exponential moving average. The preprocessed data were then down sampled to 60 Hz from their original 200 Hz to align with the temporal resolution of the simulator data. Data from 11 subjects was excluded from subsequent analysis because the average confidence level calculated by the eye tracking system during the recording was below 0.6. In this scale, a value of one indicates that the system had no trouble detecting eyes, while a value of zero indicates the system couldn’t detect the eyes at all. Moreover, each data set was individually examined visually for usability and potential values. Only segments of the whole recording of the eye tracking data that synchronized with the simulator data were chosen for further analysis.

Pupillometry data can evolve in a non-linear fashion over time, making traditional linear statistical methods, like an analysis of variance (ANOVA), or generalized linear models, less suitable. To address this, we employed a generalized additive model (GAM). This advanced regression type performs well with time series data, accounting for natural variation over time and among individuals. This allows for a more precise assessment of experimental variable impacts. Figure 9 illustrates the time course for both the no hearing loss and hearing loss conditions (i.e., combined across H2 and H3 conditions) across all pilots, underling the data's non-linearity.

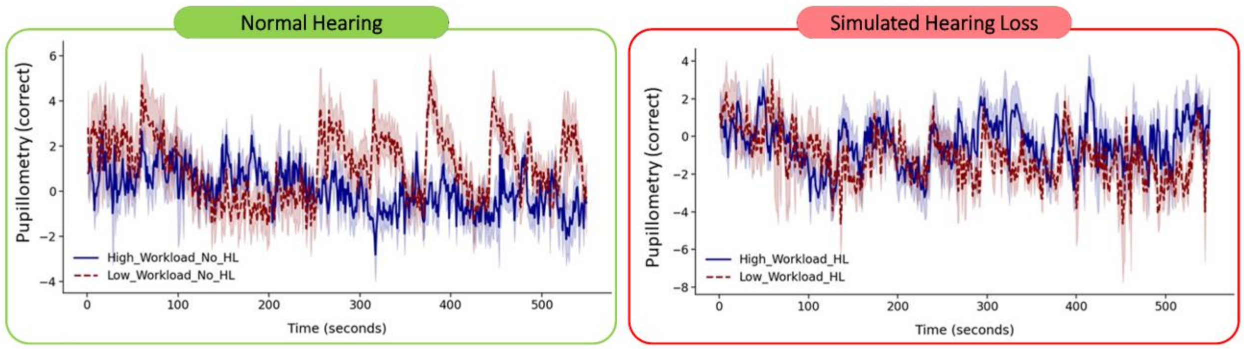

Figure 9. The time course of the pupillometry for the two different conditions.

In Figure 9, the no hearing loss condition is on the left and the hearing loss condition is shown on the right. In both plots, the x-axis denotes the time duration of each route, while the y-axis represents the pupillometry values with individual pilot means subtracted. The solid blue line indicates the high workload condition, the red dotted line represents the low workload condition, and the shaded regions signify the standard error of the mean.

The preprocessed data was input into a GAM to study the influence of the experimental variables on pupillometry. Initially, we developed a base model that considered time variation and the interaction between time variance and pilots. Given that pupillometry variation is largely influenced by mean illuminance and environment contrast, we expected this base model to account for significant variance, especially since our data was sourced from an uncontrolled and natural setting. Indeed, the base model explained 90.29% of the total variance, a statistically significant outcome ($\chi^2(96) = 2.08*106$, $p < 10\text{-}10$). The residual from this model fitting exhibited a normal distribution, indicating a good fit. Subsequently, we incorporated the experimental conditions and its temporal variation to create a full GAM, aiming to discern if these experimental variables could elucidate any additional variance.

In statistical modeling, particularly with complex models, overfitting is a common concern. Specifically, the base model already explains a large amount of variance (90.29% of the total variance). To determine if including the experimental condition might lead to overfitting, we relied on two widely used metrics, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). These measures assess the goodness-of-fit of models, with lower values indicating a better balance between model complexity and ability to fit the data accurately. Our full model achieved lower AIC and BIC values (AIC: 105600.43; BIC: 107104.80) compared to the base model (AIC: 108033.79; BIC: 108743.07). An ANOVA

(Table 12) further confirmed that the variance increase between the full and base models is statistically significant. This suggests that our full model, which incorporates the experimental conditions, is superior in fitting the data without overfitting.

Table 12. ANOVA Comparison Between the Full Model and the Base Model

Model	Residual degrees of freedom	Residual deviance	Degree of freedom	Deviance	Probability (>Chi)
Base model	20552.99	224265.55			
Full model	20522.56	204856.89	30.43	19408.66	0

Deviance, calculated as the difference in likelihood between models, increased by 19,408.66 with the addition of the experimental condition to the base model. This significant increase by the chi-square test highlights the superior fit of the full model over the base model to the data. To bolster confidence in the GAM as a fitting choice for this dataset, a visual inspection of the model's fit was conducted (Figure 10). Figure 10a reveals a high correlation of 0.95 between the fitted and actual values. Figure 10b demonstrates that the fitted values closely align with the average data across all routes and pilots. The residuals from the model fitting follows a normal distribution, as hypothesized, further affirming the model's suitability (Figure 10c). Additionally, the model's fitting for individual trials closely mirror the raw data (Figure 10d). This visual inspection aligns with the quantitative evaluation, underscoring the model's effectiveness.

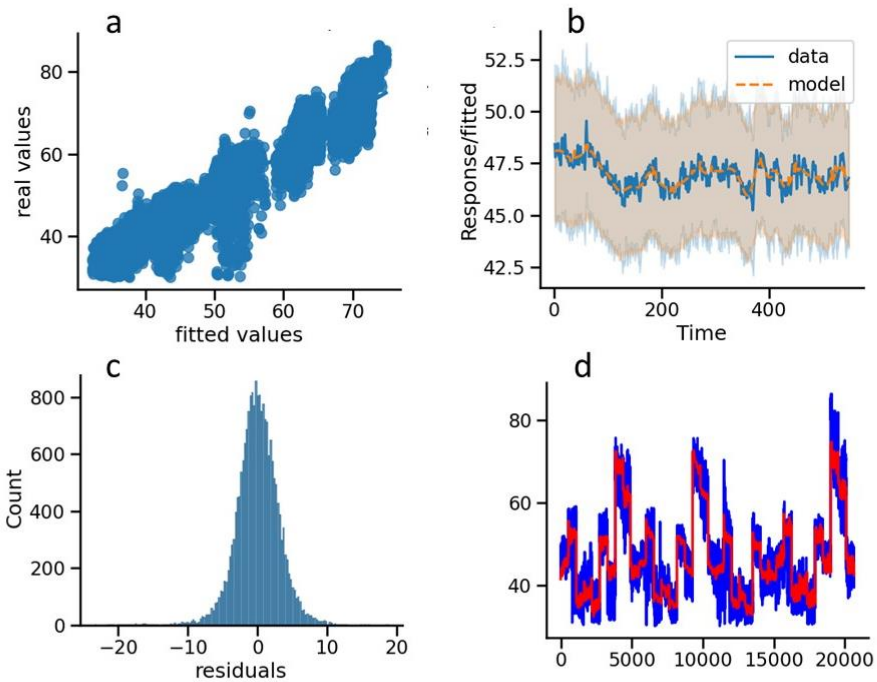

Figure 10. Evaluation of the GAM model fitting.

In Figure 10, the visual results presented here complement the quantitative assessment of the model, emphasizing the model's accuracy. Figure 10a is a scatter plot showcasing the correlation between fitted values (*x*-axis) and actual values (*y*-axis). The correlation is 0.95 and is highly significant. Figure 10b displays the time series of pupillometry data across all routes and all pilots. The orange depicts the model data while the light blue line represents the averaged data. The shaded regions indicate standard errors. Figure 10c shows the distribution of the residuals from the model fitting. Its normal distribution confirms the model's accuracy in fitting the data. Figure 10d displays the time series for individual routes, including all pilots. The blue line illustrates the actual data, and the red line shows the model's fitted values. The *x*-axis denotes the number of data points, while the *y*-axis represents the pupillometry value. This plot highlights the model's precision in tracking the actual data across all individual routes.

Such assessment underscores the success of utilizing advanced statistical approaches to analyze pupillometry data recorded in realistic settings that present challenges, like uncontrolled contrast and illuminance. This innovative method allows researchers to explore the relationship between workload and pupillometry in flight simulators, a task that was previously deemed almost impossible. The full model fitting was commendable, explaining 91.45% of the variance – an incremental 1.28% from the base model. While this variance increase appears modest, its significance is profound ($\chi^2(37) = 19838$, $p < 10^{-10}$) (see Table 12 for the ANOVA model comparison). This aligns with literature, indicating that cognitive processing influences 1 to 2% of pupillometry variance. As depicted in Figure 11, the coefficient for the high workload condition is significantly higher than that for the low workload condition when hearing loss is absent. This indicates that pupil diameter increased with heightened workload in the absence of hearing impairment (Beatty, 1982).

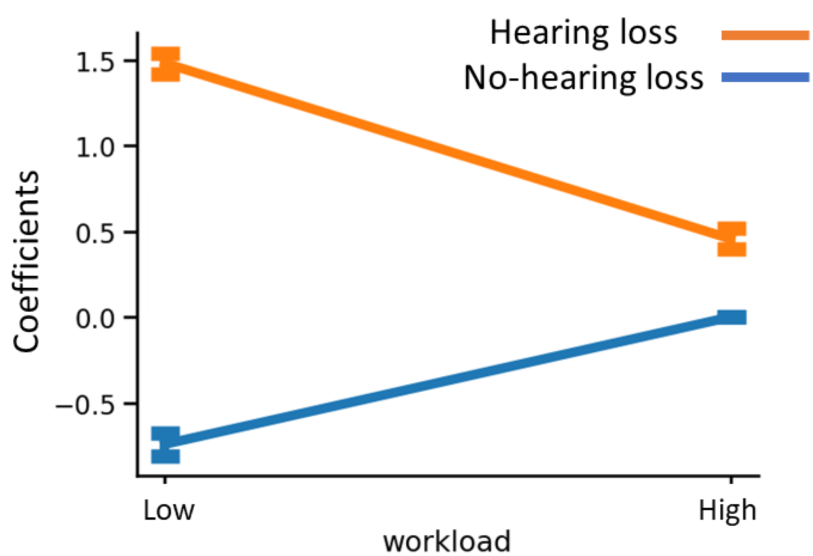

Figure 11. GAM coefficients for different experimental conditions.

Correlation analysis: MRT scores and flight performance.

Correlations between MRT percent scores and flight performance were evaluated using Pearson correlation coefficients. For routes 1 and 3, where subjects did not experience simulated hearing loss, the normal MRT percent score was used for analysis. For routes 2 and 4, the hearing loss MRT percent score was used. Altitude and heading flight performance (RMSE values) were analyzed separately. Four extreme RMSE values were removed before plotting and calculating correlation coefficients. These extreme values were at least twice as large as the next largest RMSE value. RMSE values were plotted against the subject's MRT percent score in Figures 12 and 13. Lines on the plot are simple linear regression lines. Pearson correlation coefficients and R^2 values are shown in Tables 13 and 14.

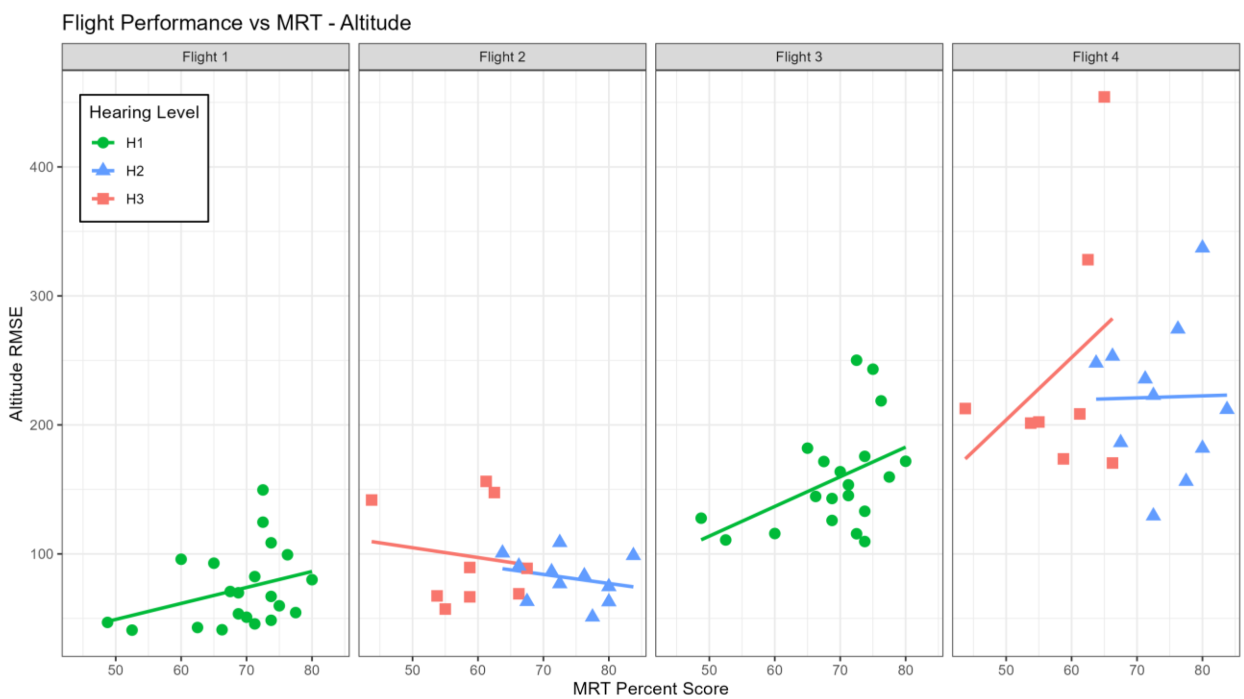

Figure 12. Altitude deviations from the ideal path as a function of MRT score. Solid lines are simple linear regression.

Table 13. Altitude Performance and MRT Correlations

Route	Hearing	*N*	Correlation	R^2	p
1	H1	21	0.320	0.102	0.154
2	H2	11	-0.250	0.062	0.460
	H3	9	-0.140	0.020	0.720
3	H1	20	0.440	0.194	0.052
4	H2	11	0.017	0.000	0.960
	H3	8	0.360	0.130	0.378

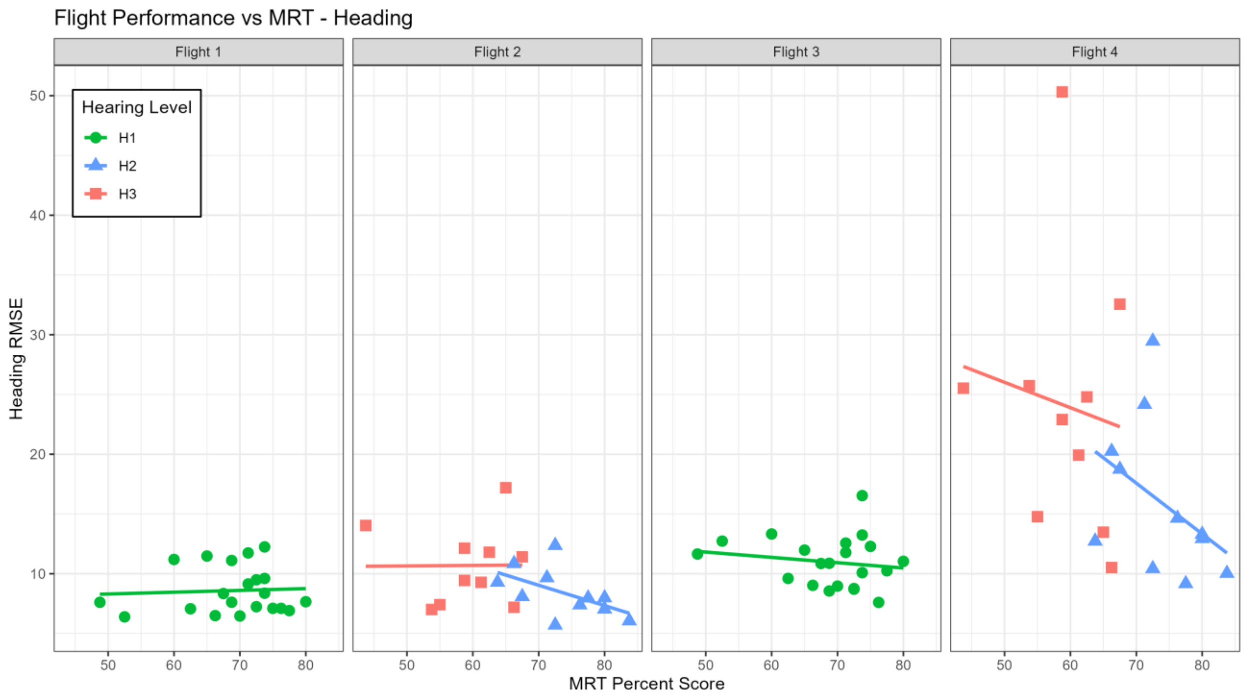

Figure 13. Heading deviations from ideal path as a function of MRT score. Solid lines are simple linear regression.

Table 14. Heading Performance and MRT Correlations

Route	Hearing	*N*	Correlation	R^2	p
1	H1	21	0.061	0.004	0.794
2	H2	11	-0.540	0.292	0.086
	H3	10	0.009	0.000	0.980
3	H1	21	-0.160	0.026	0.477
4	H2	11	-0.420	0.176	0.201
	H3	10	-0.130	0.017	0.715

Correlational analysis indicated that the variability in the outcome data could not be explained by the model and that there was no significant correlation between MRT score and flight performance. While the MRT does not appear to predict flight performance effects observed from increased workload or simulated hearing loss, it was the case that more radio calls were missed when subjects scored below MRT criteria. As shown above (see Figure 6), much of the difference observed in flight performance metrics were a result of not following the instructed flight paths due to missing radio calls.

This space is intentionally blank.

Discussion

The current study investigated the impact of current threshold requirements on in-flight performance data from pilots presented with simulated hearing loss. Study subjects were Army aviators with significant flight experience. Subjects completed aviation routes across two listening conditions (i.e., normal hearing and hearing loss) and in high and low workload conditions. Hearing loss was simulated through a computer program that decreased the audibility of signals within certain frequency ranges, provided masking noise, and simulated loudness recruitment that would be experienced by individuals with hearing loss. The simulation of the H2 and H3 hearing losses was successful, as seen in the changes in MRT_{80} scores across the hearing loss conditions and the increased number of missed calls in the flight routes completed. The increase in workload was generated by an increase in the number of radio communication signals throughout the flight as well as an increase in the frequency of a secondary task in which the subjects had to turn off a warning light.

It was hypothesized that increases in workload and the introduction of a hearing loss would influence response times to the secondary task. Increases in workload produced changes in response times, but in different directions depending on the hearing loss and workload condition. The subjects in the severe hearing loss condition (H3) with a low workload had quicker reaction times than the subjects in the mild hearing loss conditions (H2). Hearing loss may be causing response times to be shorter simply because the subjects are unable to hear the radio communications, and thus able to focus more on visual information being presented to them. It may be the case that the subjects are too well trained to respond to the warning caution light and that it would take a much larger increase in workload than was used in this study to start to see response times increase. As expected, the high workload conditions with hearing loss showed the slowest response times. However, it appears the secondary task was unable to distinguish either the hearing loss or workload conditions based on the flight profiles used in the current study.

Both subjective (NASA-TLX) and objective data were analyzed to determine the degree to which the increases in workload were manifested and the degree to which hearing loss contributed to perceived workload. The NASA-TLX is a measure of the subject’s perception of cognitive workload and the scores reported here indicated that the simulated hearing loss increased the aviator’s effort, frustration, and temporal demands over those perceived in the condition with normal hearing. Results of the analysis of the pupillometry data supported both an increase in workload through the hearing loss simulation as well as with the addition of the warning light task. Across the two levels of workload, the perceived changes in cognitive workload were minimal. This may have been due to the nature of the tasks not resulting in a large enough difference between the two workload levels. Across the hearing levels, no hearing loss versus hearing loss, the perceived levels of workload increased, as shown in the data, implying that hearing loss itself increases perceived workload.

The analysis of the pupillometry data further supports increases in cognitive workload for both conditions involving the additional warning light and conditions of hearing loss. In his seminal paper titled *Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources*, Beatty (1982) demonstrated that pupillary dilation occurs during various cognitive activities, such as short-term memory tasks, long-term memory retrieval, and problem solving. The more cognitively challenging a task, the greater the pupil dilation. In Figure 11, the orange line corresponds to the hearing loss condition, and the light blue line to the no hearing loss condition. The error bars denote the standard errors of the model-estimated coefficients. The significantly higher coefficients in the hearing loss condition, irrespective of workload adjustments, underscores the profound influence of hearing on pilots' workload.

Similarly, in this study, pupil diameters were considerably larger under hearing loss conditions than under no hearing loss conditions, irrespective of workload variations. This suggests that the hearing task in the experiment is likely more cognitively demanding for those with hearing loss than for those without. Combined with advanced data analysis techniques, our findings reveal that pupillometry is sensitive to the interplay between workload and hearing loss. This underscores the value of eye tracking as a pivotal tool for studying hearing loss in realistic settings, especially when pilot movements and responses can't be strictly controlled.

Interestingly, when hearing loss is present, the pupil diameter is notably smaller under high workload conditions than under low workload conditions. The exact reason for this observation remains uncertain. However, one possibility is that the hearing task, when introduced in the high workload condition with hearing loss, became exceedingly difficult, leading pilots to disengage. There's a complex interplay between pupil dilation and task difficulty. While it's established that task difficulty typically results in increased pupil dilation, this relationship isn't linear. Tasks perceived as insurmountably challenging can lead to stabilized or even reduced pupil dilation. For instance, pupil response to task difficulty during a digit reversal task has been examined and it was found that pupil dilation peaked at intermediate difficulty levels, then declined as tasks became increasingly challenging (Ahern & Beatty, 1979). Similar trends with a visual tracking task have been observed in that pupil dilation increased with task demands up to a point, after which added difficulty didn't cause further dilation, potentially indicating a cognitive limit (Granholm et al., 1996).

Previously, Casto & Casali (2013) examined the effects of hearing loss and flight workload and communication signal quality on aviator performance. The authors concluded that factors other than hearing thresholds and speech recognition in quiet should be included in evaluating an aviator's hearing fitness-for-duty (Casto & Casali, 2013). One of the analyses in the current study looked at whether the MRT_{80} tests could be used to predict aviator performance. Correlational analysis between MRT_{80} performance and flight performance measures indicated that the variability in the outcome data could not be explained by the model and that there is no significant correlation between MRT_{80} score and flight performance. With this being the case, the current study findings suggest that based on the design of the experiment, the operationally relevant performance tests results exhibit no predictive value on assessing flight performance. However, there are some noteworthy trends going on with the data even though it did not reach statistical significance. Similar findings have been reported previously.

The correlations for altitude do not appear to be very predictive, but the correlations for heading are interesting. It was observed in routes 1, 2, and 3 that performance was about the same across the different MRT_{80} scores. For route 2 there is a noticeable trend, but only for the H2 group. Of note, route 4, hearing loss with high workload, shows a negative correlation. For heading deviations, for route 4 only, the data does show that MRT_{80} may be predictive of performance, although it was not found to be statistically significant. It should be noted that this could potentially be a sample size issue. It also makes sense that better predictive value from the heading data were observed compared to altitude, because there were more changes in heading compared to changes in altitude during these flight plans. Consequently, there are more opportunities for pilots to miss heading calls compared to altitude calls. Further studies are needed to investigate whether the value of the MRT_{80} for predicting functional performance in aviators during flight is only evident at higher workload, or in more complex flights.

The hearing loss simulator used in the current study is a viable method for simulating hearing loss across H2 and H3 profiles. Increasing workload proved to be challenging in the current study and needs to be investigated further to provide conditions to aviators that test changes in hearing acuity in difficult flying conditions. Although hearing loss contributed to increases in cognitive workload, the attempts to increase workload without hearing loss were unsuccessful. Results of both the NASA-TLX and pupillometry support the notion that hearing loss contributes to perceived cognitive workload. This increase in cognitive workload can have a detrimental effect on aviators and should be investigated further.

Conclusion

The overall objective of this study was to measure the impacts of hearing loss in Army aviators on flight performance and cognitive workload. The experiment in this study aimed to identify the impact of fitness-for-duty standards more accurately for aviators. Understanding the speech audibility, and thus intelligibility requirements needed in the aircraft can also guide aircraft communication system designs to maximize performance.

The Army updated their *Medical Readiness Procedures* in 2019 in an effort the fill the operational gap of ensuring auditory fitness-for-duty and that any Service Member with hearing loss was able to perform auditory tasks at an acceptable level. This regulation update will have some impact on the aviation community but fitness-for-duty standards for Army aircrew and air traffic control are delineated in the APL and are based upon pure tone thresholds. The current Army aviation hearing standards do not necessarily predict the functional impact of hearing loss, but evidence suggests a synergistic relationship between the variables of hearing loss, aviator performance, and increased workload. The MOHT is now a standardized auditory assessment available at every MTF. Operationally relevant clinical assessments, such as the MOHT, should be considered in making the determination for acceptable hearing performance in aviators. The MRT_{80} ensures adequate auditory performance in the communication space while the SDT ensures appropriate localization abilities and therefore situational awareness; both are relevant to the aviators' auditory tasks. It should be noted the SDT was not tested in the current study due to technical compatibility issues with the hearing loss simulator. Here, we directly tested in-flight speech intelligibility of aviators with various hearing loss profiles and determined the impact of such hearing loss on flight performance and a pilot's listening effort.

This space is intentionally blank.

Major findings from this study were as follows:

- Hearing loss resulted in binaural speech recognition in quiet scores that were lower than the no hearing loss condition but would still be considered acceptable according to the APL.
- An increased number of failures on the MRT_{80} for the simulated hearing loss condition indicated the test did a better job of detecting functional deficits induced by the hearing loss simulator than word recognition in quiet testing.
- For flight performance, the larger the hearing deficit, the more missed or incorrect calls subjects had on average.
- Increased workload during flight exacerbated the degraded hearing loss condition as indicated by more missed radio calls and larger deviations in flight performance metrics.
- The longest response times were recorded for the secondary task, but only in the high workload condition with hearing loss. Low workload conditions resulted in some of the fastest response times.
- Assessment of pupillometry underscores the success of utilizing advanced statistical approaches to analyze pupillometry data recorded in realistic settings.
- The significantly higher coefficients in the pupillometry data for the hearing loss conditions, irrespective of workload adjustments, underscore the profound influence of hearing on pilots' workload.
- The hearing task in the experiment is likely more cognitively demanding for those with hearing loss than for those without.
- Combined with advanced data analysis techniques, our findings reveal that pupillometry is sensitive to the interplay between workload and hearing loss.

Findings from this study will be leveraged into future research for evaluating hearing loss mitigation strategies using various headset technologies.

References

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science, 205*(4412), 1289–1292.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*(2), 276.

Brungart, D. S., Makashay, M. J., & Sheffield, B. M. (2021). Development of an 80-word clinical version of the modified rhyme test (MRT80). *The Journal of the Acoustical Society of America, 149*(5), 3311–3327.

Brungart, D. S., Sheffield, B. M., Galloza, H., Schurman, J. R., Russell, S., Barrett, M. E., Witherell, K., Makashay, M. J., & Heil, T. (2023). Developing an evidence-based military auditory fitness-for-duty standard based on the 80-word modified rhyme test. *Ear and Hearing, 44*(1), 209–222.

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.

Chen, K.-H., Su, S.-B., & Chen, K.-T. (2020). An overview of occupational noise-induced hearing loss among workers: Epidemiology, pathogenesis, and preventive measures. *Environmental Health and Preventive Medicine, 25*, 1-10.

Department of the Army. (2019a). DA-PAM 40-502: *Medical readiness.*

Department of the Army. (2019b). AR 40-501: *Standards of medical fitness.*

Department of Defense. (2020). MIL-STD-1472H: *DoD design criteria standard human engineering*.

Dubno, J. R., & Schaefer, A. B. (1992). Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners. *The Journal of the Acoustical Society of America, 91*(4), 2110–2121.

Farrar, C. L., Reed, C. M., Ito, Y., Durlach, N. I., Delhorne, L. A., Zurek, P. M., & Braida, L. D. (1987). Spectral-shape discrimination. I. Results from normal-hearing listeners for stationary broadband noises. *The Journal of the Acoustical Society of America, 81*(4), 1085–1092.

Gagne, J.P., Besser, J., & Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: A review. *Trends in Hearing, 21*.

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology, 33*(4), 457–461.

Hart, S. G. (1986). NASA Task load Index (TLX). *Paper and Pencil Package*, *Volume 1.0*. NASA Ames Research Center.

Humes, L. E., Dirks, D. D., Bell, T. S., & Kincaid, G. E. (1987). Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers. *The Journal of the Acoustical Society of America*, *81*(3), 765–773.

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.

Kramer, S. E., Lorens, A., Coninx, F., Zekveld, A. A., Piotrowska, A., & Skarzynski, H. (2013). Processing load during listening: The influence of task characteristics on the pupil response. *Language and Cognitive Processes*, *28*(4), 426–442.

Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., Lunner, T., & Kramer, S. E. (2017). Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review. *Ear and Hearing*, *38*(3), 267.

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*(2), 204.

Sheffield, B., Brungart, D., Tufts, J., & Ness, C. J. (2015). The relationship between hearing acuity and operational performance in dismounted combat. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.*

Sheffield, B., Brungart, D., Tufts, J., & Ness, J. (2017). The effects of elevated hearing thresholds on performance in a paintball simulation of individual dismounted combat. *International Journal of Audiology*, *56*(sup1), 34–40.

Sheffield, B., Ziriax, J., Keller, M. D., Barns, W., & Brungart, D. (2017). The impact of reduced speech intelligibility on reaction time in a naval combat environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.*

TM 1-1520-280-10. (2016). *Technical manual: Operator's manual for helicopters, tactical utility transport with change.* 8.http://www.dtic.mil/whs/directives/corres/pdf/500002_dodi_2015.pdf.

Themann, C. L., & Masterson, E. A. (2019). Occupational noise exposure: A review of its effects, epidemiology, and impact with recommendations for reducing its burden. *The Journal of the Acoustical Society of America*, *146*(5), 3879–3905.

United States Army (2015). Army aeromedical policy letter and aeromedical technical bulletins. http://www.cs.amedd.army.mil/usasam/

Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied*, *2*(3), 270.

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.

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

Zurek, P. M., & Delhorne, L. (1987). Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment. *The Journal of the Acoustical Society of America, 82*(5), 1548–1559.

U.S. ARMY AEROMEDICAL RESEARCH LABORATORY

FORT NOVOSEL, ALABAMA

Optimizing

HUMAN PROTECTION AND PERFORMANCE

since 1962

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

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

U.S. ARMY

FUTURES COMMAND

MRDC