AN EXAMINATION OF HEADSET, HEARING SENSITIVITY, FLIGHT WORKLOAD, AND COMMUNICATION SIGNAL QUALITY ON BLACK HAWK HELICOPTER SIMULATOR PILOT PERFORMANCE

by

Kristen L. Casto

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Industrial and Systems Engineering

John G. Casali, Ph.D., Chairman
William A. Ahroon, Ph.D., Member
Brian M. Kleiner, Ph.D., Member
Woodrow W. Winchester, III, Ph.D., Member

July 3, 2009
Blacksburg, Virginia

Keywords: aviation, hearing protective devices, communications earplug, active noise reduction

AN EXAMINATION OF HEADSET, HEARING SENSITIVITY, FLIGHT WORKLOAD, AND COMMUNICATION SIGNAL QUALITY ON BLACK HAWK HELICOPTER SIMULATOR PILOT PERFORMANCE

by

Kristen L. Casto

Chairman: John G. Casali, Ph.D.

Industrial and Systems Engineering

ABSTRACT

Among the many occupational hazards to which Army rotary-wing aviators are exposed is intense noise generated from the aircraft. The potential for permanent hearing loss and difficulty communicating in helicopter noise is well known; an appropriate way to evaluate a hearing-impaired pilot's safety risk due to hearing loss is not as well known. Previous research has studied communication ability in helicopter cockpit noise under different headsets, but there are not conclusive data on the combined effects of degraded speech intelligibility due to noise and flight workload under the headset technology currently available to Army helicopter pilots. In particular, there is a scarcity of information on pilots with hearing loss. Currently, Army Aeromedical standards stipulate audiometric threshold criteria for rated helicopter pilots to ensure their safe flying. If the standard is not met, a flight waiver for hearing is generally granted if the pilot demonstrates good (at least 84%) binaural word recognition ability in a quiet environment.

A research study was conducted to evaluate Army helicopter pilot performance with regard to flight workload, communication signal quality, headset configuration, and pilot hearing ability. Objectives of the study included the ability to refine current Army audiometric hearing waiver criteria, and to yield data on which to base flight and headset selection recommendations for pilots. In general, it was believed that flight performance and ratings of situation awareness (SA) would decrease as flight workload increased and communication signal quality decreased, and that assistive communication devices coupled with headsets would afford improved flight performance over their passive counterpart. It was also hypothesized that normal-hearing pilots would perform better than hearing-impaired pilots would.

Twenty Army helicopter pilots (one group of 10 pilots without a hearing waiver and one group of 10 pilots with a hearing waiver) participated in this study. The pilots flew three flights in a Black Hawk flight simulator, each with a different headset configuration and with varying flight workload levels and varying air traffic control (ATC) communication signal quality. Objective flight performance parameters of heading, altitude, and airspeed deviation and ATC command readbacks were measured. Additionally, measurements were taken on subjective measures of workload, SA, and headset comfort/speech intelligibility.

Experimental results partially supported the research hypotheses. Results indicated that flight performance and ratings of SA were negatively affected by increased flight workload and decreased communication signal quality for both groups of pilots. Results also showed that a passive headset/passive earplug combination use by the hearing-impaired group of pilots led to degradation of certain flight performance parameters and lower ratings of SA than the headsets equipped with assistive communication technology; however, the same headset effect was not seen with the group of normal-hearing pilots.

This study yielded results that support a conclusion that factors other than hearing thresholds and word recognition ability in a quiet environment should be considered when evaluating Army helicopter pilots flight safety with regard to hearing sensitivity. Rather, the synergistic effects of flight workload and communication signal quality with individual hearing levels should be considered when making continued flight recommendations and headset choice recommendations. Results also support a recommendation requiring hearing-impaired pilots to use assistive communication technology and not be permitted to fly with passive headset devices. Further research should include a functional hearing assessment in which pilot hearing requirements are determined and individual hearing abilities are compared to the requirements.

ACKNOWLEDGEMENTS

An enormous amount of thanks and appreciation go to Dr. John G. Casali for persevering with me throughout this endeavor. Even more than I am grateful to him for sharing his knowledge and skills, I appreciate his patience, optimism, and confidence – three qualities that I did not possess in abundance during the past few years. It is a pleasure to also acknowledge and thank my committee members Dr. William A. Ahroon, Dr. Jeff A. Lancaster, Dr. Brian M. Kleiner, and Dr. Woodrow W. Winchester. Their comments, criticism, and varied perspectives were genuinely helpful in shaping this project.

Simply put, I could not have completed this dissertation without support from the U.S. Army Aeromedical Research Laboratory (USAARL) at Fort Rucker, Alabama. Former USAARL commander Colonel James S. McGhee and current commander Colonel Joseph F. McKeon graciously supported my request to collect data at the USAARL. Among others, Mr. John Ramiccio, Mrs. Cathie Ramiccio Dr. Heber Jones, Mr. Paul Stricklin, Mr. Chuck Brown, and Dr. Efrem Reeves deserve special thanks (or medals) for assisting me with my data collection. I appreciate the assistance I received from Bose Corp. and from CEP, Inc., specifically Mr. Ben Mozo, who saved me from certain "experimental disaster". I am thankful for the pilots who participated in my study. They all took time out of their already over-scheduled lives to help me, and their abilities, time, and senses of humor are greatly appreciated.

I owe thanks to my family and friends for tirelessly listening to the ongoing saga of my dissertation. I wish my dad were here to see this come to fruition - I know he would have been proud. The largest debt of gratitude I owe is to my husband, Deron, and our children Leah and Ben. Deron has always been my biggest fan and my greatest support – and the past three years were no different. Leah and Ben have not had much choice in where we live or what schools they go to, but they rarely complain about what I put them through and have thrived anyway. Finally, I appreciate the loyalty of my little dog Maggie, who sat beside me and patiently waited for her walks throughout this entire process.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

United States Army rotary-wing aircraft pilots are required to perform many flight duties correctly in the midst of many challenges that may affect mission completion as well as aircraft and aircrew safety. Among the challenges that interfere with successful mission completion are noise levels, mental workload, communication demand, and individual hearing levels. When a pilot is overcome by these challenges while operating a helicopter, there is the potential for the obvious negative consequence of a flight accident and failure of the flight mission. There is also the potential negative consequence to the Army in the form of lost training costs and military disability compensation costs if the aviator is removed from flying duty due to inability to perform his or her job duties. U.S. Army medical personnel are charged with accurately assessing the health of pilots to ensure their safe flying abilities; however, the assessment for auditory performance is currently based on pure tone audiometric thresholds and binaural speech intelligibility in a quiet environment, which are arguably not the only factors to consider with respect to successful flight performance for a hearing-impaired aviator. Performance is also likely to be influenced by varying degrees of mental workload, increased communication demands, and hearing protective devices (HPDs) and communication system used in the cockpit.

As a result of the current hearing loss waiver criteria, as indicated by data from the U.S. Army Aeromedical Activity (USAAMA) and the U. S. Army Hearing Program at the Center for Health Promotion and Preventive Medicine (CHPPM) approximately 10% of Army aviators and aircrew have been granted continued flight status with a waiver for hearing loss for each of the past ten years, without any consistent evaluation of their flight performance as it relates to hearing sensitivity (M.D. Quattlebaum, AERO Program Manager and Chief, Epidemiology, U.S. Army Aeromedical Center, Ft. Rucker, Alabama, personal communication, 15 November 2007).

Army helicopter noise levels exceed the safe noise exposure limits set by the Department of the Army Pamphlet (DA PAM) 40-501, *Hearing Conservation*[1]. Without proper hearing conservation efforts, permanent noise-induced hearing loss (NIHL) will result from the high intensity, low frequency dominant noise to which aircraft pilots are exposed on a regular basis (Edwards, 1990). Additionally, Army helicopter noise levels are often at levels[2] that require double passive hearing protection in accordance with DA PAM 40-501 (i.e., wearing of earplugs underneath earmuff-style headsets and/or flight helmet systems). Double hearing protection may prevent hearing loss, however, it can create communication problems. For instance, this practice typically *decreases* pilot communication ability because the earplug worn under the aviator's helmet attenuates the speech signal from the earphones in the helmet as well as reducing the ambient noise (Mozo & Murphy, 1997b). The result is that the aviator must increase the volume of the earphone signal to overcome the insertion loss created by the earplug, typically resulting in distortion products in the earphone and a compromised speech output (Casali & Gerges, 2006). Trying to understand a distorted, nonlinearly-attenuated speech signal is a serious problem, especially for the hearing-impaired aviator and more so when other aspects of flight workload are demanding.

To overcome the communications difficulties associated with passive hearing protection in the high noise level helicopter environment, and communication difficulty related to double hearing protection, two different systems have been considered for Army aircrew, the Communications Earplug (CEP), and active noise reduction (ANR) headsets. The CEP was developed by the U.S. Army Aeromedical Research Laboratory (USAARL) at Ft. Rucker, Alabama and has been enthusiastically accepted by Army aviators (Ahroon, Gordon, Mozo, & Katz, 2000). The CEP consists of a small sound transducer coupled to an expandable, ported

---

[1] Personnel will be enrolled in a comprehensive HCP when they are exposed to steady-state noise with a time-weighted average (TWA) of 85 dBA or greater or impulse noise of 140 dBP or greater.

[2] Greater than 103 dBA TWA and up to and including 108 dBA TWA, personnel must wear double hearing protection.

foam earplug that is worn in combination with the standard flight helmet. The CEPs main advantage is that it prevents unwanted sound from entering the ear canal from the outside of the helmet, but permits communication signals to enter the ear without attenuation through the earplug (Ahroon et al., 2000). ANR systems are composed of electronic components that filter and reverse the phase of the measured sounds in the earcup, resulting in attenuated sound levels to the listener (Mason & Mozo, 1995), more so for low frequency signals than high frequency signals. This is an advantage of ANR when it is combined with the typical flight helmet, because it compensates for the lack of low frequency attenuation of the typical flight helmet.

The CEP has essentially become standard issue for aviators because this device has been shown to increase speech intelligibility, reduce noise hazards, be subjectively favored by pilots over ANR, meet crash worthiness evaluations, and it is less expensive than ANR systems (Casto, Ostler, & Ahroon, 2001; Mason & Mozo, 1995). In the past, ANR technology has not been practical for implementation in Army aviation, primarily due to weight of the systems, crash-worthiness issues, compatibility with flight helmets, and expense to the Army. However, recent technology developments have produced miniaturized ANR components, including implementation in earplugs, as well as crash-worthy earcup designs, suggesting that ANR may now be a viable option to CEPs for Army aviators. Further, an aviator with high-frequency noise-induced hearing loss may benefit from the low frequency attenuation characteristics of ANR that may reduce the amount of upward spread of masking on speech.

Despite the success with CEPs and other hearing conservation efforts, there are still many Army aviators with noise-induced hearing loss. To address the issue of aircrew safety in these cases, the Department of the Army Regulation (AR) 40-501, *Standards of Medical Fitness* specifies audiometric hearing thresholds, to be discussed later, for all active duty and civilian helicopter pilots employed by the U.S. Army ("Standards of Medical Fitness," 2007). If a pilot’s hearing levels exceed the audiometric standard, there is a requirement for further speech audiometry evaluation, intended to give the waiver-granting authority[3] an indication of the pilot’s functional hearing abilities necessary to fly a helicopter.

In addition to the hazardous noise levels, resultant noise-induced hearing loss, and artificial hearing loss caused by passive HPDs, mental workload is another factor that is influenced by communication device and can affect pilot performance. It is generally accepted that pilot mental overload can result if too many cognitive demands are presented (Svensson, Angelborg-Thanderz, Sjoberg, & Olsson, 1997) ; however, there exists no available research outlining pilot or crewmember mental workload, or resulting military flight performance, with CEP use in an operational environment. However, in analogous fashion there is evidence for the impact of communications earphones in at least two prior operational performance effectiveness experiments in other systems. Mission performance was shown to vary as a function of speech intelligibility in a simulated tank environment, with an increase in performance at levels of increased speech intelligibility (Peters & Garinther, 1990). Very recently, the increased speech intelligibility provided by certain ANR headsets over their passive counterparts was shown to decrease pilots’ rated mental workload, increase speech intelligibility, lower protected exposure levels (PELs) and result in improved simulated flight performance in civilian aircraft (Casali, Lancaster, Valimont, & Gauger, 2007).

This experiment examined robust metrics of flight performance, as well as other measures relating to isolated effects such as speech intelligibility, as a function of hearing loss, communication system type, pilot workload, and communication signal quality. The results are intended to provide more specific data on which to base future Army Aeromedical policy statements on hearing loss/flight waiver decisions and data on which to base flight and headset recommendations for hearing-impaired pilots. Ultimately, the results and subsequent recommendations may affect combat effectiveness realized through optimized performance, and reduction in aircraft critical incidents and accidents.

---

[3] The waiver-granting authority is the U.S. Army Human Resources Command, Chief, National Guard Bureau, or the local commanding officer

BACKGROUND

Aviation Noise Hazard

In the best of circumstances, operating a helicopter is a complex task that often places excessive sensory and perceptual auditory demands on the pilot. Helicopter noise is primarily low frequency, originating from engines, blowers, transmissions, vibration, and turbulence, with peak levels occurring near the blade passing frequency (James, 2005; Ribera, Mozo, & Murphy, 2004). The noise measured in a helicopter cockpit is mostly narrow band discrete tones with their associated harmonics, all superimposed on a broadband background noise. Helicopter noise is both aerodynamic and mechanical in nature. The main and tail rotors, as well as the interactions between the rotors and fuselage, generate aerodynamically-induced noise, and the mechanical noise originates from revolving systems that are connected to the rotors, such as gearboxes, transmission shafts, transfer gears, and drive shafts.

Because each model of helicopter is aerodynamically and mechanically different, there is an expected difference in unique acoustic signature across helicopters (James, 2005). For example, the Australian Army's S-70A-9 Black Hawk helicopter has at-ear noise levels in the cabin often exceeding 105 dBC, with the highest levels often at the rear of the aircraft. The spectral composition is highest in levels below 100 Hz, although higher frequency noise (above 980 Hz) is found at the front and middle of the helicopter. The rotor blade pass produces the low frequency peak at 17 Hz with harmonic peaks at 34, 51, 68, and 85 Hz (King, Saliba, & Brock, 1999). The gearbox main planetary gear is responsible for the mechanically-generated high frequency energy peak at 980 Hz as well as the first two harmonics peaks at 1960 and 2940 Hz. The main bevel gear generates a peak at 2012 Hz (King et al., 1999). It is expected that noise levels of the U.S. Army UH-60 Black Hawk helicopter are similar, since it is based on the same airframe. Other U.S. Army rotary wing aircraft vary in noise levels, presumably due to the individual aerodynamic and mechanical differences (see Table 1).

Table 1. Noise levels of sample U.S. Army rotary wing aircraft ("Noise Levels of Common Army Equipment" 2007).

Helicopter model / name	**Sound measurement location**	**Sound level (dBA)**
CH-47-D Chinook	Cockpit	102.5
UH-60 / Black Hawk	Pilot	106.0
	Co-pilot	106.0
AH-64 / Apache	Pilot	104.0
	Co-pilot	101.3
OH-58D / Kiowa	Right seat	101.6
	Left seat	100.3
UH-1H / Iroquois (Huey)	Pilot / co-pilot	101.9
	Max in rear	102.9

It is important to note that the sound levels a pilot experiences while flying a helicopter is a combination of the cockpit noise that bypasses or penetrates the hearing protector and the electrical communication signal that is directly delivered to the ear through the via the helmet earcups. Additionally, the cockpit intercommunications system (ICS) introduces acoustic distortion, which further confounds the auditory challenge (Ribera et al., 2004). For those with a noise-induced hearing loss, the hearing impairment adds yet another auditory demand on the piloting task. Because the noise levels inside the cockpit exceed Department of Defense (DoD) acceptable levels for single hearing protection offered by the flight helmet, many Army pilots are required to wear insert hearing protection (earplugs) under the flight helmet while flying. This practice introduces, through its attenuation of both external sounds and speech signals through the radio, a simulated hearing loss, especially if passive hearing protectors are worn.

When communicating in the aircraft, the aircrew's speech is converted into electrical signals by the microphone into which they are speaking. In addition to speech, the microphone is also picking up ambient cockpit noise, electrical noise, and radio interference that is transmitted to the listener, resulting in a combined signal of intended speech and contaminating noise. The interference of the intended signal often results in decreased speech intelligibility and increased noise dose to the listener, because the ICS volume is increased in an effort to improve speech intelligibility. For this reason, it is important to reduce the levels of contaminating noise entering the aviator's ear (James, 2005).

There are important implications related to the noise levels of rotary-wing aircraft. The high noise levels at the ear of the pilot resulting from aircraft noise can cause permanent hearing loss if the ear is not adequately protected. Additionally, high noise will interfere with speech communications and non-speech auditory warnings within the cockpit (James, 2005).

Human Hearing Anatomy and Physiology

The human ear consists of the outer, middle, and inner ears, as well as the central auditory pathways in the brain. The outer ear includes the auricle, or pinna and the external auditory canal, and its purpose is to funnel acoustic sound waves into the ear and to assist with sound localization. Due to the funneling effects of the pinna, and moreover, the resonant characteristics of the external auditory canal, the external ear effectively amplifies high-frequency (2000-5000 Hz) sound energy as much as 20 dB before it reaches the tympanic membrane (Shaw, 1974).

Additionally, the pinna and concha contribute to the amplification characteristics of the outer ear. The difference in sound levels measured at the concha and those just external to the external auditory canal, is known as the transfer function of the open ear (TFOE), and is due to factors such as distance between measurement microphones, auricle/concha resonance and diffractions effects (Casali, Mauney, & Burks, 1995). Casali, et al. (1995) measured TFOEs on subjects and found that the measurement varies between subjects, ears and frequencies, and is, in general, negligible at the low frequencies and peaks at about 10 dB at 4000 Hz.

The middle ear consists of the tympanic membrane, the ossicles (middle ear bones individually called the malleus, incus, and stapes), and the Eustachian tube. The purpose of the middle ear is to transform the acoustical energy that has been funneled into the ear canal into mechanical energy. Sound waves enter the outer ear through the pinna and cause the tympanic membrane to vibrate, which in turn transmits energy through the ossicles. The ossicles' lever action and impedance matching provide an additional signal amplification of up to 30 dB (Ward, Royster, & Royster, 2000). The function of the Eustachian tube is to equalize pressure in the middle ear space and is not part of the hearing process, per se, although proper function of the Eustachian tube is necessary for normal hearing sensitivity.

The footplate of the stapes fits into the oval window of the fluid-filled cochlea, the main hearing related component of the inner ear. Movement of the stapes against the cochlea causes pressure waves in the cochlear fluids and stimulates the basilar membrane and the attached organ of Corti, the sensory organ of hearing that runs the length of the spiraled cochlea. Within the organ of Corti are the inner hair cells, which are the primary sensory receptors for hearing, and the outer hair cells, which primarily facilitate the sensory response of the inner hair cells. Specific frequencies of sound vibrate specific locations along the length of the cochlea, resulting in maximum vibration at the base of the cochlea with high-frequency sounds, and maximum vibration at the apex with low-frequency sounds. The intensity of the sound also affects the basilar membrane vibration; as the intensity of the sound increases, the amplitude of the membrane vibration increases, although in a non-linear, compressive manner (Humes, Joellenbeck, & Durch, 2006).

Noise-Induced Hearing Loss

Noise, generally defined as any unwanted sound, is more specifically described as a waveform with a random change in instantaneous amplitude (Yost, 1985). The structures of the inner ear, specifically the hair cells, can be damaged by excessive levels of noise, and the effects of noise-induced hearing loss (NIHL) can be temporary or permanent. The magnitude of hearing loss resulting from noise depends on factors associated with the exposure; sound pressure level, duration, temporal pattern, type of noise, and spectral content of the noise (Humes et al., 2006).

Impulse and impact noise is described as high-level, short-duration noise. Impulse noise results from explosive devices, such as gunfire, and impact noise is generated by the forceful meeting of two hard surfaces, for example a hammer to a nail. Acoustic trauma resulting from impulse noise with peak levels greater than 140 dBP damages the cochlea by causing rapid mechanical failure that may lead to large cochlear lesions and significant hearing loss (Henderson & Hamernik, 1986). Acoustic trauma can occur from a single exposure to impulse or impact noise, or can result from extended periods of exposure over many weeks, months, or years (Humes et al., 2006). Some hearing recovery can take place following acoustic trauma, but too frequently, the individual is left with a permanent hearing loss.

Exposure to steady-state noise, either intermittent or continuous, can result in a temporary or permanent hearing loss. A noise-induced temporary threshold shift (TTS) is characterized by elevated auditory thresholds immediately following a high-intensity noise exposure that gradually improves within 24-48 hours after the exposure (Mills, Gengel, Watson, & Miller, 1970). For more intense noise exposures (greater than 90 dBA) or for longer duration exposures (greater than 24 hours), the amount of TTS is generally larger and elevated thresholds may continue for more than 30 days, and usually will not completely return to pre-exposure levels (Humes et al., 2006). In these cases, the individual is left with a residual permanent threshold shift (PTS).

There is increased risk of hearing loss when impulse and/or impact noise occur in combination with steady-state noise. Studies with laboratory animals have demonstrated that the combined effects of these simultaneous exposures are greater than the simple additive effects of the separate exposures to impulse and steady-state noise (Ahroon, Hamernik, & Davis, 1993; Henderson & Hamernik, 1982). Elevated audiometric hearing threshold levels (HTLs) that are a product of broadband noise and impulses that characterize most industrial and military settings generally show a ‘noise notch’ pattern, with HTLs reaching a maximum between 3000 and 6000 Hz, followed by a return towards normal hearing at higher frequencies (Humes et al., 2006).

Fitzpatrick (1988) found that 8.4% of a typical U.S. Army Aviation Brigade had a hearing loss by Army standards[4], and 29.7% of the same group had a significant threshold shift (STS)[5]. The author found no association between hearing loss and aircraft model; however, he

---

[4] Exceeding H-1 hearing criteria – an audiometric average of 500, 1k, and 2k Hz that cannot exceed 25 dB with no individual level greater than 30 dB. Threshold at 4k Hz cannot exceed 45 dB.

[5] A threshold shift of 25 dB calculated by subtracting the hearing threshold level at each frequency of a preflight reference audiogram from the level appearing on a current audiogram. The selection of STS was based on the American Academy of Otolaryngology-Head and Neck Surgery has used an average absolute hearing level of 25dB at 500, 1,000, and 2,000 Hz as the threshold of impairment.

did find relationships with other variables. Using the STS criteria, he found a progression of hearing loss with age, increased flight hours, and helmet-only hearing protection. The data were analyzed with an age-adjustment to determine if the hearing loss was a function of noise exposure or age, and the analysis revealed a significant association for total flight hours. Hearing loss by Army standards did occur predominantly in the slightly older, more experienced group, but hearing loss by STS criteria more often occurred in younger pilots with fewer flight hours, indicating that about 21% of those studied met acceptable hearing criteria, but had actually suffered an STS. (Fitzpatrick, 1988).

Recent (2007) audiometric data show that approximately 6% of Army pilots have hearing thresholds that exceed the Army's H1 hearing profile category, meaning that there is at least a mild hearing low frequency hearing loss in one ear or at least a moderate high frequency hearing loss in one ear (L.S. Domanico, USACHPPM, Army Hearing Program, personal communication, 20 September 2008). Similarly, a pilot with high frequency noise-induced hearing loss who likely exceeds the Class 2 flight standards at 3 and/or 4 kHz will also exceed the criteria for the Army's H1 hearing profile category. Based on this generalization, it can be inferred that at least 6% of active duty and Department of the Army Civilian (DAC) Army pilots do not meet Class 2 flight physical standards.

Other Auditory Effects of Noise Exposure

In addition to the auditory effect of hearing sensitivity loss, there are other functional auditory effects and possible non-auditory effects of noise exposure. An individual with high frequency noise-induced hearing loss (NIHL) will often experience associated tinnitus; a subjective perception of ringing, buzzing, or roaring sound in the ears. Most individuals with tinnitus have only minor associated problems; however, tinnitus can be distracting and functionally limiting to the person who experiences it (Humes et al., 2006). For those more severely affected, tinnitus has been reported to bring on fear, frustration, anger, irritability, stress, and anxiety, and the impact of tinnitus has been reported as more problematic than an associated hearing loss (Salmivalli, 1967; Axelsson and Barrenas, 1992; Mrena et al., 2002 as cited in Humes et al., 2006, pg 119). The Committee on Noise-Induced Hearing Loss and Tinnitus Associated with Military Service from World War II to the Present (Humes et al., 2006) reviewed at least five studies that examined the association between the prevalence of tinnitus and quantitatively measured hearing loss. Based on the review of the studies, and their pre-determined criteria for defining hearing loss, the committee concluded that there is sufficient evidence that hearing loss (i.e. hearing thresholds greater than 25 dBHL at one or more audiometric test frequencies between 250 and 8000 Hz) is associated with a higher prevalence of tinnitus compared to individuals with normal hearing sensitivity. Given the findings of Fitzpatrick (1988) and the data from USAAMA and CHPPM, it can be inferred that *at least* 10% of Army helicopter pilots are at high risk of suffering tinnitus.

Another possible auditory effect of noise exposure is hyperacusis, or markedly increased sensitivity to sound. Generally, this condition is associated with a traumatic impulse or impact noise exposure, but it can also arise from gradual, long-term steady noise exposure. Hyperacusis causes an individual to experience discomfort with sounds that most would find comfortable (Suter, 2002), but sufferers typically exhibit normal or near-normal audiograms. More commonly, individuals with high frequency NIHL experience recruitment, or a reduced dynamic range of comfortable listening levels, so that a loud sound may be more uncomfortable for the listener with NIHL than it is for a person without NIHL.

Arguably, the most limiting functional auditory challenge resulting from NIHL is speech misperception. Speech sounds, or phonemes, have differing frequency composition; the energy for vowels is primarily low frequency while consonants have predominantly higher frequency, but lower amplitude speech energy than do vowels (Ward, 2000). In the case of a listener with high frequency NIHL, the vowel sounds are heard correctly but, depending on the degree of hearing loss, the higher frequency consonants that give meaning to words are not audible. As a result, the individual must rely on speech reading or the context of the conversation to ‘fill in’ the missing consonants, and neither of these methods is as reliable as actually hearing the consonant.

As mentioned, the spectral energy of vowels is primarily found in the low and middle frequencies of the speech spectrum, while consonants are concentrated in the higher frequencies of the audible range. Further, vowel energy is considerably greater than the spectral energy of consonants. Pilots with high frequency NIHL essentially have an internal acoustic filter that inhibits high-frequency speech cues (i.e. consonants) from being audible. In an experiment in which the effects of high frequency filtering on speech discrimination was studied, Kryter

(1975) found that a "typical" high-frequency hearing loss causes an individual to effectively miss 13% of the total "discriminable units" that are available to normal-hearing individuals. This corresponds to 31% of the discriminable units available to normal-hearing listeners in everyday speech (as cited in Suter, 1985). Adding to the difficulty of hearing loss is the "upward spread of masking" that occurs in aircraft as the high-intensity, low frequency background noise efficiently masks the higher frequency consonants in speech. Martin and Pickett (1970) and Bess (1983) have speculated that hearing-impaired individuals experience more upward spread of masking than normal-hearing listeners do, but the relative effect appears to be similar, regardless of hearing sensitivity (as cited in Suter, 1985).

Helicopter pilots may be more at risk for speech misperception than fixed wing pilots are, due to the degree and configuration of helicopter pilots' hearing loss. Raynal et al. (2006) examined auditory thresholds in French military pilots of fighter aircraft, transport aircraft, and helicopters and found more high frequency hearing loss at 3000 Hz for helicopter pilots than for the fighter and transport pilots. Because low auditory thresholds at 3000 Hz are important for speech intelligibility, the authors point out that the helicopter pilots were at greater risk for compromised speech communication, particularly in the presence of background noise (Raynal, Kossowski, & Job, 2006).

Speech misperception is especially problematic for the pilot with NIHL because facial speech reading cues are not available when listening to radio transmissions and the context of the spoken message may not be known to the pilot, especially in the event of an emergency alerting message. In a study comparing performance of 20 hearing-impaired listeners with mild-to-moderate bilateral sensorineural hearing loss with 20 normal-hearing listeners, researchers found that the hearing-impaired group required a 5 dB greater signal-to-noise ratio to perform at the same level of the normal-hearing listeners. When provided visual cues (i.e. the listener could see the speaker's face), both groups of listeners displayed performance increases equivalent to a 5-6 dB increase in signal-to-noise ratio (Hawkins, Montgomery, Mueller, & Sedge, 1988). This study underscores the impact visual cues have on speech intelligibility; an advantage lost on military pilots communicating via radio or intercom transmission (Moore & von Gierke, 1991).

Susceptibility to NIHL

Susceptibility to NIHL apparently varies greatly between noise-exposed individuals, and therefore, is difficult to predict. Adding to the difficulty is the conflicting findings in the literature regarding NIHL susceptibility. Pyykko, Toppila, Zou, & Kentala (2007) offer possible explanations as to why the ears of some individuals are more easily damaged by noise and others are more resistant to the effects of noise. Among the individual noise susceptibility factors they found to be correlated with NIHL were environmental, or exogenous, variables such as smoking, analgesics, vibration and solvent exposure. Biological, or endogenous, variables studied included high cholesterol levels, pigmentation, age, and Reynaud's disease, which were also found to be risk factors of NIHL. Other endogenous, inherited variables that may influence susceptibility are detoxificating, potassium, or mitochondrial genes (Humes et al., 2006; Pyykko et al., 2007).

Humes et al. (2006) report that non-acoustic factors such as aminoglycosides, cisplatin, and some solvents combine in an additive or synergistic manner with the effects of noise to increase the measured NIHL in laboratory animals. However, they explain that there is not enough evidence to conclude the same for humans, nor is there conclusive evidence for the effects of many other exogenous factors on hearing. Similarly, the authors indicate that there is insufficient evidence for humans to conclude that any endogenous factors, such as age, gender, race, pigmentation, or prior hearing loss predict susceptibility to NIHL.

Regardless of the conflicting information, there are differences in hearing loss among noise-exposed aviators that are evident on audiometric evaluations. The differences underscore the importance of considering possible susceptibility risk factors in addition to other factors when making decisions regarding hearing protective devices and communication systems for Army aviators.

Non-Auditory Effects of Noise Exposure

Smith (1991) defines non-auditory effects of noise as the following:

> All those effects on health and well-being which are caused by exposure to noise with the exclusion of effects on the hearing organ and effects which are due to the auditory

> information (i.e. communication problems). Such effects include performance effects, physiological responses and health outcomes, annoyance, and sleep disturbance. (p. 49)

In general, noise is a non-specific stressor that influences changes in the reticular activating system (RAS), which activates the higher cerebral centers, the sympathetic nervous system, the adrenal medulla and cortex, and the limbic system. Changes in these areas suggest that noise may affect cardiovascular function, increase catecholamine and cortisol, and lead to changes in emotion and mood. Additionally, health effects of noise can be indirectly related to noise exposure, such as health effects resulting from disrupted sleep due to noise (Smith, 1991).

Smith's (1991) review of research on non-auditory effects of noise concludes that there is evidence indicating that noise exposure leads to physiological responses that could adversely affect an individual's health, particularly if the noise exposure is prolonged. He cautions that much of the research has not controlled for factors other than noise that may have caused the changes in health. He also adds that some reactions to noise, like annoyance, are the result of a combined psychological reaction as well as objective exposure levels. Of the research reviewed, it appears that three non-auditory effects of noise are without serious research limitations; problems caused by communicating in noise, such as laryngitis, and vocal fold polyps and nodules, interference with sleep, and noise annoyance (Smith, 1991; Ward, 2000). Noise annoyance is often heightened by psychological factors such as the perception of noise as unnecessary or harmful to health (Smith, 1991), and can vary tremendously between individuals (Ward, 2000).

Noise can also affect operator task performance, mostly in negative way, but in some ways positively. Noise usually does not show effects of decreased performance for simple and repetitive tasks, in fact, noise may heighten operator arousal and actually improve simple task performance. For other tasks, noise can mask acoustic cues necessary to complete a task and contribute to task distraction (Poulton, 1978). Greater task degradation comes from unexpected or aperiodic noise than from predictable or periodic noise (Casali & Gerges, 2006).

Casali and Gerges (2006) point out several empirical studies that highlight the performance effects of noise. The first study is one in which Indian weavers were found to perform better with hearing protection than without hearing protection in high noise levels (Bhattacharya, Roy, Tripathi, & Chatterjee, 1985). Next, Weston and Adams (1935) found a

12% improvement in work efficiency when wearing hearing protection for weavers in England who alternated weekly between wearing and not wearing hearing protection. Another study by the same authors measured a 7.5% efficiency improvement when wearing hearing protection at work for a year compared to a similar group who did not wear hearing protection at work for the same year. The last study described reported a 20% increase in production in a plant after a bearing grinder was fitted with an enclosure that reduced the noise level by about 20 dB (Staples, 1981).

Presbycusis

In addition to NIHL, pilots also mature during their time in the military and may experience changes in their hearing due to aging known as presbycusis. Presbycusis is characterized by a loss of hearing threshold sensitivity, particularly at high frequencies, and related difficulty understanding speech in the presence of noisy environments (Kujawa & Liberman, 2006). Interestingly, Pauler, Schuknecht, & Thornton (1986) found that for some aging individuals, speech intelligibility ability decreases more than is expected based on threshold sensitivity declines (as cited in Kujawa & Liberman, 2006). As noted by Schuknecht (1993), this type of speech intelligibility deficit combined with primary neural degeneration in some aging ears (i.e. ‘neural presbycusis’) could have important consequences, even if auditory threshold changes are not large (as cited in Kujawa & Liberman, 2006; Schuknecht & Gacek, 1993). These age-related changes can also contribute to operational communication difficulty for an aviator, although not necessarily in an additive manner.

While the effects of presbycusis and NIHL are interrelated, their relationship is not purely additive, particularly when one of the causes of hearing loss or their sum produces a loss of 40 dB or greater (Ward et al., 2000). In a longitudinal hearing loss study, Gates, Schmid, Kujawa, nam, & D’Agostino (2000) found that individuals with previous damage from NIHL experienced exacerbated hearing loss at frequencies outside the original NIHL due to aging, suggesting that ears with noise damage experience age-related changes differently from those without noise damage.

Hearing Conservation

While hearing conservation efforts may lessen the combined effects of presbycusis and NIHL, the primary focus of hearing conservation programs is to prevent NIHL. To prevent the deleterious effects of noise, there are established guidelines and regulations that govern hearing conservation programs. According to the National Institutes of Health's Consensus Statement on noise and hearing loss, exposures of eight hours per day to sound levels above 85 decibels measured on the A weighted network of a sound level meter (dBA) are likely to result in permanent NIHL after many years of exposure (NIH, 1990). The Occupational Safety and Health Administration (OSHA) hearing conservation regulation (29 CFR 1910.95) for general industry specifies an action level (AL) of 85 dBA with a 5 dB exchange rate (the tradeoff between duration and intensity of a noise exposure) integrated over eight hours (time weighted average – TWA). When the AL criteria is met, one is required to be included in a hearing conservation program (HCP), which includes monitoring audiometry, health education training and availability of hearing protection, although use of hearing protection is optional. However, when OSHA's permissible exposure limit (PEL) of 90 dBA is met, the requirements include feasible engineering and/or administrative controls and mandatory hearing protection if levels are not reduced to below the criterion by those controls (Suter, 2000). The DoD requires the Army, Navy, and Air Force to establish HCPs that are at least as stringent as a PEL of 85 dBA with a 4 dB exchange rate. The Army requires enrollment in the HCP when personnel are exposed to steady-state noise with a TWA of 85 dBA or greater or impulse noise of 140 dBP (peak measurement) or greater and employs a 3 dB exchange rate ("Hearing Conservation Program," 1998).

Hearing Protective and Communications Devices in Army Aviation

The most desirable way to decrease an aviator's noise exposure is to reduce the noise levels of the aircraft, which is not always economical, or even possible. In this case, the best approach available is to reduce the level of the noise at the ear with hearing protective devices (HPDs). To prevent the effects of high intensity noise in aircraft cockpits, Army aviators have been required by Army regulation to wear hearing protective devices since 1978, when the first Department of Defense Instruction (DoDI) was published.

One approach to limit the noise levels that reach an aviator's ear is to modify the flight helmet to provide greater acoustic protection. Helmet modifications to change the attenuation characteristics can be made by using different materials for the earcup, using different sound-absorbing materials inside the earcup, or by increasing the earcup volume (James, 2005). Similar to aircraft modifications to decrease noise exposure, helmet modifications are not usually feasible for Army aviators, and double hearing protection is often required to meet the acceptable noise exposure requirements.

As mentioned previously, a helicopter's noise spectrum is dominated by low frequency sound. Most flight helmets are characteristically limited in low frequency attenuation due to their lightweight materials and leakage pathways around the ears, and therefore allow low frequency sound into the ear. Conversely, at the higher frequencies, where helicopters generate little noise and where the helmet attenuates a great deal, the noise level at the ear is lower. For this reason, double hearing protection is sometimes necessary to provide adequate hearing protection to the pilot. Although double hearing protection provides more attenuation than either device alone, the effect is not additive. The benefit of double hearing protection in a helicopter environment is that earplugs, particularly the hand-formed foam plugs, provide better attenuation than earmuffs below 500 Hz, thereby compensating for the lack of low frequency attenuation offered by the flight helmet, although their effectiveness in this regard is highly dependent on the quality of their fit in the ear canal. In this environment where the noise is dominated by energy below 500 Hz, "the attenuation of single HPDs will be the least and the potential gains from dual protection are the greatest" (Berger, 2000, pg. 424). It is important to note that when double hearing protection is worn, it is the earplug that is the most important part of the combination. When different earmuffs are used with the same earplug, there is very little change in attenuation; however, for a given earmuff, the attenuation with different earplugs is more variable (Berger, 2000).

Current Army regulation requires double hearing protection (earplugs worn underneath the flight helmet) for steady-state noise exposures greater than 103 dBA ("Hearing Conservation Program," 1998). While the UH-60 Black Hawk, CH-47D Chinook, and AH-1 Apache cockpit noise levels can exceed 103 dBA and require double hearing protection, Army audiologists typically recommend that all pilots use double hearing protection if possible, regardless of the specific aircraft flown. Prior to the practical implementation of alternative communication systems for rotary-wing aircrew, the only available options for double hearing protection were hand-formed, pre-formed, or custom-made passive hearing protection (without any electronic circuitry) to reduce the noise levels at the individual's ear. Such passive attenuation in earplugs is achieved primarily through use of high transmission loss materials and compliant materials that establish a hermetic seal of the device to the ear canal (Casali & Berger, 1996). If worn properly, passive hearing protection worn in combination with a flight helmet usually provide sufficient hearing protection in the cockpit. However, speech intelligibility can be negatively affected by this method of hearing protection.

Passive Hearing Protective Devices and Speech Intelligibility

Casali and Horylev's (1987) research on normal-hearing listeners suggests that passive HPDs have little or no negative effect on the wearers' ability to understand speech if the ambient noise levels outside the HPD are above about 80 dBA, in fact most studies report a slight improvement in speech intelligibility with certain HPDs (as cited in Casali & Berger, 1996; Howell & Martin, 1975). If the ambient noise is lower than about 80 dBA, HPDs do interfere with speech understanding compared to the unoccluded ear in the same environment (Casali & Berger, 1996). This condition may occur in Army aviation, when aircrew members wear HPDs as a precaution, for example, during preflight procedures when the noise is unpredictable. Suter (1989) reported that hearing-impaired HPD wearers, however, likely experience decreased speech communication ability with HPDs regardless of noise level (as cited in Casali & Berger, 1996).

Conventional passive hearing protective devices reduce speech and noise the same amount, and therefore do not change the speech-to-noise ratio (Casali & Berger, 1996). As reported in Suter (1985), this is especially problematic for those with NIHL because the ability of individuals with NIHL to discriminate speech is disproportionately influenced by a decrease in signal-to-noise ratio, compared to those with normal hearing (Suter, 1985). Similarly, others found that for normal-hearing individuals, HPDs have no detrimental effect on speech intelligibility at noise levels above 85 dBA (Abel, Alberti, Haythornthwaite, & Riko, 1982; Chung & Gannon, 1979; Kryter, 1946; Rink, 1979). The results for hearing-impaired individuals and for those with limited English proficiency show that HPDs interfere with speech intelligibility in both quiet and noise (Abel & Spencer, 1997).

In addition to the effects on speech intelligibility, passive hearing protective devices worn in a noise-hazardous workplace, such as the cockpit of a helicopter, have an effect on the wearer's overall functional hearing ability, which will be discussed in more detail later. Other important aspects of functional hearing include signal detection, signal identification/recognition, and signal localization. For safety reasons, it is critical that a pilot have the ability to maintain awareness through timely detection, identification/recognition, and localization of surrounding environmental events and alerting and warning signals (Soli, 2003). Clearly, signal detection ability is diminished by passive hearing protective devices. Nobel and Russell's (1972) studies have shown that localization ability is degraded when hearing protectors are worn, usually more so with circumaural protectors than with earplugs. The authors believe this is because pinna cues are lost with the earmuffs (as cited in Soli, 2003).

The use of double hearing protection may further decrease auditory localization ability. Results of an experiment (Simpson, Bolia, McKinley, & Brungart, 2002) show that double hearing protection may have a drastic negative effect on sound localization. This effect was demonstrated by measuring the response time to locate and indentify a visual target when an auditory cue at the target was provided. The response time was much quicker when participants wore single hearing protection than when they wore double hearing protection. Results of a 2003 experiment were consistent with previous experiments concerning effects of hearing protection on sound localization (Brungart, Kordik, Simpson, & McKinley, 2003). They found that when the stimulus was short, the use of either earplugs alone or earmuffs alone resulted in a slight degradation in left-right localization and a much larger degradation in front-back localization performance, and was worse with earmuffs than with earplugs. In addition, the researchers found that use of a combination of earplugs and earmuffs resulted in an even greater degradation in localization ability than with either earplug or earmuff use alone. They also found that even with a longer stimulus localization performance task, exploratory head motion cues provided a greater benefit with single hearing protection use than with double hearing protection use. In other words, and consistent with Simpson, et al. (2002), listeners required more time to localize continuous auditory stimuli with double hearing protection than they did with single hearing protection. The authors point out that decreased localization performance resulting from single hearing protection is different in character from the decreased localization performance resulting from double hearing protection. The single hearing protection conditions caused performance degradations mostly in the front-back localization tasks, and double hearing protection caused degradations in both left-right and front-back tasks, even with head movements aided in localization. They suggest that the explanation for this is not because of a larger disruption of high frequency pinna cues, but instead, that the acoustic cues that reached the cochlea through bone conduction contained interaural time and intensity differences that conflicted with the air conducted localization cues. This could be because in the double hearing protection condition, the bone conduction limit of sound perception was reached, but in the single hearing protection condition, it was not.

As discussed, decrements in speech intelligibility in noise are an important operational limitation of passive HPDs. An important factor affecting speech intelligibility is the signal-to-noise ratio, and is directly related to sound attenuation of a hearing protective device. A favorable signal-to-noise ratio for speech intelligibility is one in which the target signal is significantly elevated above the noise to allow sufficient auditory processing by the listener. In general, the greater the signal-to-noise ratio, the greater the intelligibility of the message communicated (Ribera et al., 2004). Two devices that are designed to improve the signal-to-noise ratio under the earcup of a flight helmet are the Communications Earplug (CEP) and active noise reduction (ANR) headsets.

Communications Earplug

The Communications Earplug (CEP) was developed by the U.S. Army Aeromedical Research Laboratory (USAARL) at Ft. Rucker, Alabama to overcome the communications difficulties associated with passive hearing protection in high noise helicopter environment, and has been enthusiastically accepted by Army aviators (Ahroon et al., 2000). The CEP consists of a miniature earphone transducer coupled to an expandable foam earplug using a threaded hollow tube that fits into a 2.5mm diameter hole that runs through the length of the foam plug (Figures 1, 2). This attachment method provides a pathway for the unimpeded, transducer-generated sound to enter the occluded ear canal. The opposite end of the CEP plugs into the helicopter's intercommunications system (through the helmet). In order to accommodate the CEP, a hole must be drilled into the earcup of the helmet allowing the insertion of the CEP cables (Mozo, 2004).

Figure 1. Communications earplug (CEP) assembly (left) and as worn in the ear (right).

Figure 2. CEP configured with flight helmet.

The main advantage of the CEP is that the device attenuates unwanted sound prior to it entering the ear canal from outside of the helmet, but permits communication signals to enter the ear without being attenuated (Ahroon et al., 2000). An added benefit of the CEP is that it is less susceptible to sound attenuation losses due to ancillary devices that are worn in combination with the device. Hearing protection afforded by the helmet alone (or an active circumaural device) is decreased when devices such as eyeglasses, protective masks, and cold weather hoods are worn at the same time. These ancillary devices often break the seal of the earcup and produce a sound path to inside the earcup (Mozo, Murphy, & Ribera, 1995; Robinette, Ahroon, & Ostler, 2003).

Evaluations of the CEP have demonstrated improved speech intelligibility compared to passive hearing protectors while affording adequate hearing protection in helicopter noise environments (see Table 2). Because of favorable objective and subjective findings, the CEP is standard issue for most rotary-wing Army pilots. Studies have used standardized speech intelligibility tests consisting of phonetically balanced words presented to the listener through CEPs in a simulated noise environment to evaluate individual speech intelligibility while wearing CEPs. Additionally, Ahroon et al. (2000) evaluated crew coordination using the Coordination Index Rating for Crew Linguistic Events (CIRCLE), a sequential analysis system that quantifies cockpit communications. The authors found that the CIRCLE sub-codes associated with pilots' request for communication to be repeated was reduced significantly with CEP use compared to foam earplug use under the flight helmet.

Table 2. Real-ear attenuation characteristics of the Communications Earplug per ANSI S3.19-1974 using standard and slim size eartips (fit as appropriate) (Mozo, 2004).

Test Frequency (Hz)	**Mean Attenuation (dB)**	**Standard Deviation (dB)**
125	29.7	5.3
250	34.2	5.9
500	39.7	4.7
1000	42.1	5.1
2000	38.0	3.2
3150	43.2	3.2
4000	43.5	3.5
6300	47.5	4.5
8000	46.1	4.4

Subjectively, the CEP was judged to have improved speech clarity (when listening to speech over the ICS) compared to the 'traditional' helmet/hearing protection combination by 80% of aviators and crewmembers assigned to the crash rescue unit (FLATIRON) at Ft. Rucker, Alabama. Respondents also rated the CEP as being generally acceptable and helpful to mission accomplishment (Mozo et al., 1995). Mozo et al (1995) also reported that 39 out of 40 responses indicated that the CEP reduced noise levels at the ear.

One disadvantage of the CEP discovered by Mozo et al. (1995) was difficulty donning the flight helmet with the additional wire required by the CEP. Ninety percent of the aviators surveyed reported that additional time and planning was required to put the helmet on due to the length of the wire with the device. Some aviators reported that the foam earplug came out of the ear while pulling the helmet down over the ear. The authors reported that this disadvantage was being corrected by routing the wire to the CEP from a point above the ear. The wire lengths were ultimately reduced to the current configuration of 12 inches on the right side and 17 inches on the left side, with the routing of the wire left to the user's discretion. This configuration has reduced the donning difficulty experienced by the early CEP prototype (Ben Mozo, personal communication, 5 March 2008).

The authors also reported that USAARL was developing a headband communication unit that combined the CEP and a noise-cancelling microphone that would remove this shortcoming. A subsequent experiment with a small dynamic microphone coupled to the CEP and into the communications system via a flexible headband found that the headband and the helmet were not compatible (Ribera, Mozo, & Murphy, 1996). As a result, the idea was not developed further (Ben Mozo, personal communication, 5 March 2008).

Another issue that was realized during the Mozo et al. (1995) evaluation was discomfort with a urethane foam earplug. In fact, fifty percent of the aviators who used the foam earplug with the CEP reported some degree of discomfort. Most of the respondents indicated that their discomfort with the foam tips began in the first hour, and it was subsequently determined that the source of the discomfort was the plastic core of the foam earplug, which can be attributed to improper insertion of the earpiece. Discomfort with the foam earplug continues to pose a problem with CEPs; however, newer foam earplugs used with the CEP have plastic cores that are softer and less rigid (W.A. Ahroon, personal communication, October 2, 2008). In addition, CEP, Inc. specifically addresses proper insertion techniques on its website.

Active Noise Reduction

Active noise reduction (ANR), also referred to as active noise cancellation, attenuation, or control refers to the attenuation of a primary sound by destructive interference with a controlled secondary sound that originates from a speaker within the ANR system (Harley, Shields, & Hendrix, 1994). The secondary sound, or 'anti-sound', is a sound wave that is of equal amplitude, but that is exactly out-of-phase at a given point in space from the primary signal that typically originates outside of the ANR system (Casali & Berger, 1996) (Figure 3).

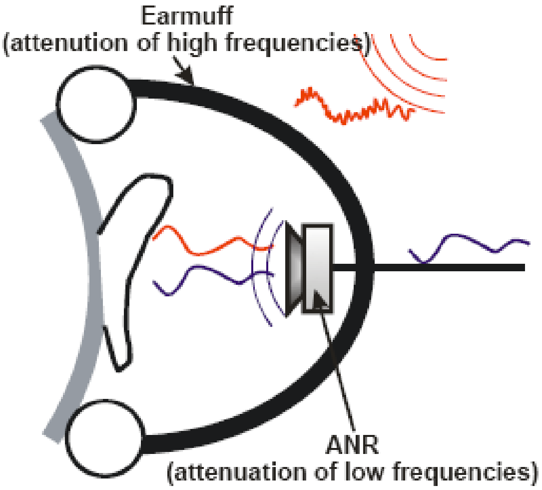


Figure 3. Principle of ANR under an earmuff (used with permission by Dr. Karl Buck, 2008).

Active Noise Reduction in Headsets

ANR has been used in systems designed solely for hearing protection as well as systems designed for one- or two-way communications. The latter systems are equipped with a microphone and earphone components and are referred to and ANR headsets. Both types of ANR systems have been implemented in both open-back and closed-back devices. Open-back systems contain an ANR microphone-earphone assembly that is surrounded by foam pads and rest on the outer ear. In a closed-back device, the system components are typically housed in a passive noise-attenuating earmuff. The main disadvantage of an open-back system is that there is no earcup to provide passive attenuation in the case of electronics failure (Casali & Robinson, 1994). The ANR headset chosen for the present study is a closed-back, circumaural headset in order to avoid the aforementioned shortcoming of the open-back variety.

In a typical ANR system used for hearing protection, a sensing microphone detects the signal that has passed through or around the external earcup of the HPD and feeds the signal through phase compensation filter that reverses the phase. The signal then passes through an amplifier before it is output as a signal inside the earcup that cancels the original signal (Casali & Berger, 1996). For ANR headphone to cancel a signal perfectly, the required opposite signal propagation would have to occur without any time delay between the different system transducers and the eardrum, which is not possible because the components and the eardrum are at different locations. Because of these phase shifts associated with the transducer location differences, as well as the delays in signal processing that occur throughout the system, establishing the correct relationship between the noise and the cancellation signal is more difficult as the bandwidth of the noise increases (Casali & Berger, 1996). For this reason, ANR is usually more effective for lower frequency noise, with maximal attenuation values of about 22 dB from 100-250 Hz and almost no attenuation above 1000 Hz (Nixon, McKinlely, & Steuver, 1992 as cited in Casali & Berger, 1996, pg. 177).

In past years, the Army has rejected ANR technology in flight helmets due to issues that include the weight of the electronics, lateral impact protection, and impulse noise attenuation. The weight of the flight helmet is an important factor for increased injury during a crash, and the added head-supported mass of systems that include night-vision goggles, head-up displays, and possibly ANR electronics is a significant contributor to the amount of injury in the case of an aircraft accident. Therefore, any weight reduction of the helmet is an advantage to the aviator (Mozo & Murphy, 1997a).

A more recent USAARL study has supported the use of ANR in Army Aviation due to its speech intelligibility advantage. Ribera et al., (2004) evaluated the speech intelligibility of aviators with normal hearing and aviators with hearing loss while wearing an SPH-4B flight helmet in its standard form, one equipped with CEPs, and one equipped with ANR. Due to the improved speech intelligibility with ANR and CEPs, the authors recommended retrofitting aviator helmets with the new technology (ANR or CEPs). This is likely due to the feasibility of ANR implementation in an insert earplug, rather than a circumaural headset.

As mentioned previously, ANR provides good low frequency attenuation, but it is not as protective for higher frequencies. ANR combined with passive devices does provide greater levels of hearing protection, especially since the low frequency attenuation of ANR compensates for the lack of low frequency attenuation of a typical flight helmet. A problem with ANR is that some of the passive attenuation, particularly in the 100-1000 Hz frequency range, provided by the volume of the helmet earcups, is lost by installation of ANR circuit boards and transducers that consume the volume of the earcup (Casali & Berger, 1996). However, as James (2005) points out, miniaturization of ANR electronics provides hope of offering the benefits of both passive and active technology more fully. Miniaturized ANR circuits would allow more of the passive attenuation characteristics of the circumaural earcup to be realized, thus providing more acoustic attenuation than can be currently achieved (James, 2005). Another potential limitation of ANR is that some analog ANR systems slightly amplify noise in the mid-frequency range. This can occur when the phase relationship is close to being in-phase and acoustic gain is unity (Casali & Robinson, 1994).

ANR Control Systems

ANR systems can be analog or digital, and one difference between the two is the ANR filter, the frequency dependent electronic amplification between the input to the ANR control system and the voltage output to the secondary sound source. For a digital system, the ANR filter consists of a digital filter in series with analog components. For an analog system, the filter is comprised of only analog components. The analog filter may have a variable, self-adapting, gain control, but the ratio of the gain at two different frequencies is fixed. In contrast, digital filters are able to adaptively adjust the frequency dependent gain to improve performance under varying conditions (Hansen, 2001; Harley et al., 1994).

There are two ways to implement ANR under a headset earcup. One is known as feed-forward, and the principle is based on the measurement of the signal outside the hearing protector to predict the signal inside the hearing protector (Figure 4). For this to be done, the measured signal is filtered and inverted before being reproduced with the speaker under the earcup. This process must be done with digital control schemes because the analog filters do not work well with the changing acoustical transfer function of the earcup due to factors such as fit of the headset, and location of the sound source with respect to the reference microphone.

Digital filters have adaptive algorithms that are able to obtain a minimal signal power at the place of the error microphone inside the earcup. The hearing protector will have its best performance if the external noise is stationary so that the error signal will converge to a minimum. When the external noise is not stationary, the digital filter will not be able to reach optimum effectiveness (Buck & Zimpfer-Jost, 2005). These feed-forward, adaptive digital ANR headphones and earplugs rely on an input microphone that samples the incoming signal and gives information on the frequency content of the signal to the electronic controller to produce the output signal. The controller effectiveness is measured by the error microphone, which gives a signal for the control algorithm to use in adjusting the controller output so that the error at the error microphone is minimized (Hansen, 2001; Harley et al., 1994)

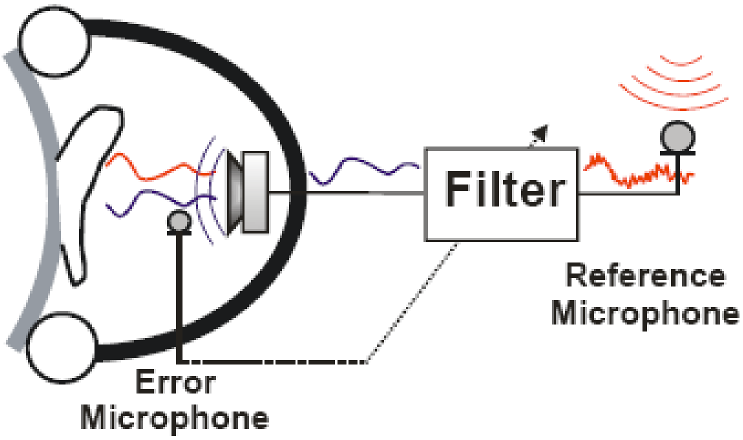


Figure 4. Principle of a feed-forward control system (used with permission by Dr. Karl Buck, 2008).

The second method for ANR implementation is through the feedback principle, which differs from the feed-forward principle primarily in that it works independently of the noise outside the hearing protector (Figure 5). Feedback-type ANR systems work by measuring the residual noise under the earcup, inverting its polarity, and then feeding this signal back into the earcup (Buck & Zimpfer-Jost, 2005).

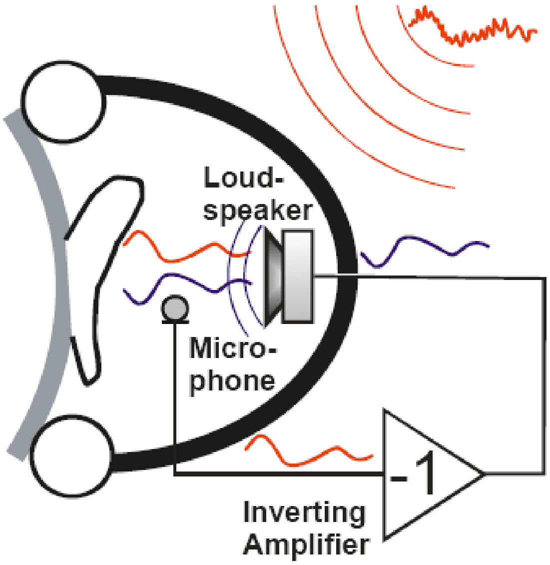


Figure 5. Principle of feedback control (used with permission by Dr. Karl Buck, 2008).

The difference between feed-forward systems and feedback systems is how the control signal is derived. In the feedback system, the signal has already passed the electronic controller by the time it is sampled by the error microphone; however, in a feed-forward system, the signal is sampled first. Because of the nature of this difference, a feedback system performs most effectively when the sound to be attenuated is predictable, or constant and periodic in nature. Feedback systems are well suited for situations where it is not possible to sample signals early enough to utilize a feed-forward system, but the performance of feedback technology is not as effective as feed-forward (Hansen, 2001).

Active Noise Reduction Performance

Although the principle of ANR is relatively straightforward, the practical application can be more difficult, particularly in the flight environment. ANR has not been accepted in Army aviation partly because the weight of the circumaural earmuff-type ANR systems pose a danger to the pilot in the case of a lateral impact to the helmet, and therefore has not withstood crashworthiness evaluations, and partly due to the high cost of the systems. However, there are current developments in ANR technology, including miniaturized components and earplug-style ANR devices, that could potentially eliminate the lateral impact risk.

From a performance perspective, ANR holds some promise in a helicopter environment. James (2005) found that within a flight helmet, the ANR working frequency range is between 50 Hz and 1000 Hz with peak levels of active attenuation between 20 to 23 dB. The author found that when the active attenuation is added arithmetically to the existing passive attenuation of the flight helmet, the operational flight trials showed reductions of about 6-10 dBA in rotary-wing aircraft noise. The author points out that a 10 dB reduction in noise realized by wear of an ANR flight helmet allows the pilot to fly 10 times as long with the same risk of hearing loss as those pilots using the standard helmet alone, and that their hearing loss risk is significantly reduced for the same number of flying hours. A performance limitation of some ANR devices is an inability to provide sufficient active noise reduction for sound levels at 120 dB or greater due to insufficient amplifier gain and/or device output (Casali & Berger, 1996).

An SPH-4B flight helmet equipped with ANR afforded improved speech intelligibility compared to the same helmet without ANR under laboratory conditions representing helicopter listening conditions (Chan & Simpson, 1990). Similarly, a study conducted by researchers at USAARL showed that ANR improved hearing protection and voice communications for Army aviators in both laboratory and field experiments; however, there were some unfavorable effects on speech intelligibility when ANR-equipped flight helmets were worn with spectacles and chemical/biological mask (Mozo & Murphy, 1997a; Staton, Mozo, & Murphy, 1997). In the laboratory experiment, aviators wearing the SPH-4B flight helmet demonstrated 40 percent speech intelligibility when the helmet was equipped with ANR compared to only 1% when the helmet was worn without any supplemental electronics. Additionally, maximum levels of speech intelligibility were reached at much lower intensity levels, thereby reducing the potentially hazardous effects of the speech signal (Staton et al., 1997). Overall, the operational test showed that ICS volume levels were reduced significantly form levels usually used for the standard helmet. The authors also reported that the effects on sound attenuation and speech intelligibility when wearing spectacles with ANR and the standard helmet were minimal and that subjects did not feel that any of the tested helmet systems (ANR, CEP, standard helmet) reduced their awareness of operations noises needed to ensure proper operation of the helicopter (Staton et al., 1997).

Gower and Casali (1994) investigated the attenuation and speech intelligibility characteristics of an ANR headset and a passive headset. They found that the ANR headset provided more attenuation than the passive headset did – up to 22 dB, depending on the frequency. Speech intelligibility results; however, showed that the passive headset outperformed the ANR headset. Over two experiments, the ANR headset required a significantly higher signal-to-noise ratio (12 dB on average) than the passive headset did to achieve the same level of speech intelligibility. The rationale is that the passive headset overcame its lower attenuation characteristics with a better frequency response in the critical bandwidth of speech (Gower & Casali, 1994).

Another study also found that ANR armor crew headsets provided more sound attenuation than their passive counterparts (Anderson & Garinther, 1997). Unlike Gower & Casali (1994) though, the speech intelligibility levels were better for the ANR headset than they were for the passive headset. More recently, the increased speech intelligibility provided by certain ANR headsets over their passive counterparts was shown to decrease pilots' rated mental workload, increase speech intelligibility, lower protected exposure levels (PELs) and result in improved simulated flight performance in civilian aircraft (Valimont, Casali, & Lancaster, 2006).

Aviator Hearing Loss Flight Waiver

Even with adequate hearing protection and communication devices, Army aviators continue to suffer hearing loss that can adversely affect flight safety. To assess such hearing loss and to address potential safety issues, the Army has a waiver process in place. The U.S. Army Aeromedical waiver process, developed to ensure the consistent and appropriate management of disqualified aviation personnel, is responsible for identifying those personnel with medical conditions that are not compatible with continued safe flying or their continued good health to maintain an effective fighting force. The ultimate goal is to keep aircrew flying safely and to ensure a "long and successful aviation career for each individual" (Woodson, 2006). However, there is limited information available to the waiver granting authorities on which to make this important decision regarding aviators with hearing loss.

Currently, aviators who meet the audiometric flight standards (as outlined in AR 40-501, *Standards of Medical Fitness* and the USAAMA policy letters), are considered to have hearing sensitivity that is compatible with safe flying (i.e., ‘normal-hearing’) and do not require further audiometric evaluation to maintain their current flight status. Those who do not meet the audiometric flight standards are considered to have hearing loss sufficient to require further testing and possible recommendation for a hearing waiver to maintain flight status. Audiometric flight standards for hearing thresholds for Class 1/1A and Class 2/2F/3/4 are displayed in Table 3. Flight Class codes are described below (Hightower, 1999):

1. Class 1 (warrant officer candidate)/Class 1A (commissioned officer or cadet) standards apply to applicants for aviator training.
2. Class 2 standards apply to student aviators after beginning training at aircraft controls, rated Army aviators, and Department of the Army civilian (DAC) pilots. Class 2F standards apply to flight surgeons and other medical personnel who have applied or are enrolled in the Army Flight Surgeon’s Primary Course or Army Aviation Medicine Orientation Course.
3. Class 3 standards apply to non-rated soldiers and civilians who are on flight status orders, but who do not operate flight controls.
4. Class 4 standards apply to Army and DAC air traffic controllers (ATCs).

Table 3. Acceptable audiometric hearing levels (dBHL) for Army aircrew members and ATC (Woodson, 2006).

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

** Isolated hearing loss at 6000 Hz will not require full audiology work-up unless recommended by the local flight surgeon or audiologist (i.e. new onset, etc.) and is not considered disqualifying; however, 6000 Hz hearing measurements will be reported for Aviation Epidemiology Data Register (AEDR) database and/or research and academic interest.*

If an aviator's pure tone audiometric evaluation does not meet the appropriate standard, a full audiometric evaluation is required. This audiometric evaluation consists of the following:

1. Pure-tone air conduction testing (and bone conduction if deemed necessary by audiologist or flight surgeon
2. Tympanometry
3. Acoustic reflex threshold testing
4. Speech reception threshold testing
5. Speech recognition (discrimination) testing in quiet under earphones, conducted monaurally and binaurally using the Northwestern University (NU6) word list materials. Monaural testing is conducted at 40 dB sensation level (SL) and binaural recognition is conducted at the aviator's most comfortable listening level (MCL).
6. An evaluation by an otolaryngologist and/or and in-flight evaluation may be required for significant hearing loss. The in-flight evaluation consists of speech audiometry using common aviation terms in the presence of noise conditions of the individual's primary aircraft.

Further, the Aeromedical Policy Letter (APL) on hearing loss states that:

> "Unrestricted waiver can be considered depending on amount of hearing loss and functional capability, provided a complete audiological evaluation indicates no underlying pathology, and binaural speech recognition score is 84% or higher. Aircrew members with a recognition score of less than 84% may receive a waiver, but are generally handled on

> a case-by-case basis. Patients who are H4[6] profile will inevitably be disqualified"

The local audiometric evaluation that either supports or fails to support a waiver recommendation is forwarded by the flight surgeon through USAAMA to the waiver granting authority, who makes the final determination based on the audiometric data and the audiologist and flight surgeon's recommendations (see Table 4). Depending on the status of the aircrew member, waivers are granted by U.S. Army Human Resources Command, Chief, National Guard Bureau; or by the local commanding officer. Most waiver requests are considered routine waivers, meaning that there is clear policy established concerning the medical condition, and require only review and endorsement before they are forwarded to the waiver authority with recommendations for follow up or restrictions. Some cases, such as those that are "unusual, potentially precedent setting, involve flight or other operations limitation, and all Class 1 exceptions to policy" are evaluated by the Aeromedical Consultants Advisory Panel (ACAP). ACAP's decision is either approved or disapproved by the Commander, U.S. Army Aeromedical Center (USAAMC) and forwarded to the appropriate waiver authority. The waiver authority issues either a waiver or a formal letter of termination from flight duties. Of important note in

---

[6] The "H" position, the 4th value in the profile series, is the designation for hearing and condition of the ears. The ability to hear, as well as the presence of any disorders of the ears, are noted based on the following categories:

> 1—The average hearing threshold for each ear at 500, 1000, and 2000 Hertz (Hz) is not more than 25 decibel (dB) with no individual value greater than 30 dB at these frequencies. The value should not be over 45 dB at 4000 Hz or over 35 dB at 3000 Hz.
>
> 2—The average hearing threshold for each ear at 500, 1000, 2000 Hz is not more than 30 dB, with no individual value greater than 35 dB at these frequencies. At 4000 Hz, the value should not be over 55 dB or audiometer level 30 dB at 500 Hz, 25 dB at 1000 and 2000 Hz, and 35 dB at 4000 Hz in better ear. If one ear has very poor hearing (can be deaf) the audiometer values for the better ear should be over 30 dB at 500 Hz, 25 dB at 1000 Hz and 2000 Hz, 45 dB at 3000 Hz, and 35 dB at 4000 Hz.
>
> 3—Speech reception threshold (SRT) in the better ear is not greater than 30 dB hearing level (HL), measured with or without a hearing aid, or acute or chronic ear disease.
>
> 4—Functional level below the standard of "3". ("Army Hearing Program," 2008)

the waiver process is that “the needs of the Army may occasionally supersede the medical recommendations” (Woodson, 2006).

Table 4. Categories of disposition for each waiver condition.

Qualified	**No waiver is required**
Qualified, information only	Condition will be tracked in database and reported on annual summary sheet, but is not disqualifying. No waiver is required.
Disqualified, waiver recommended	Waiver recommendation will be forwarded to Human Resources Command (HRC) for final approval.
Disqualified, waiver not recommended	Waiver/exception to policy is not recommended. If approved by HRC, will result in termination of aviation service.

Limitations of Hearing Waiver Process

The audiogram conducted as a part of the annual flight physical is intended to give the flight surgeon some assurance that the pilot is audiometrically fit for flying duty, implying that the pilot will have the functional hearing abilities that are required to safely fly his or her aircraft. Similarly, if a pilot meets the waiver criteria, and the waiver is granted, the same assurance is offered. In other words, meeting audiometric threshold requirements or hearing waiver criteria suggests that an aviator possesses functional hearing abilities adequate for signal detection, signal identification/recognition, signal localization, and speech perception – all in the aircraft environment.

Said another way, the hearing loss APL is implying that a pilot who meets the specified audiometric thresholds *or* demonstrates good binaural word discrimination at a comfortable listening level in quiet is able to perform his/her occupationally-required hearing critical tasks. However, the literature does not support either pure tone thresholds or binaural word discrimination in quiet for determining functional hearing ability, which likely directly affects pilot performance. A large body of research has consistently shown that audiometrically measured pure tone thresholds give an adequate picture of hearing impairments, but are a poor predictor of functional hearing ability (Soli, 2003). Laroche et al. (2003) attribute inappropriate reliance on the audiogram for job selection as a carryover from the use of audiograms in medical-legal cases to prove that noise-induced hearing losses is caused by occupational noise exposure (Laroche et al., 2003).

Another shortcoming of the current waiver-granting process is that communication handicap is poorly predicted by the audiogram. There have been many studies that investigate the effects of sensorineural hearing loss on speech communication functional hearing abilities. While the relationship between hearing impairment and reduced functional abilities is clear, the correlation between degree of audiometrically measured hearing loss and functional ability is not strong (Soli, 2003).

As described in the hearing loss APL, functional speech communication is inferred from binaural word discrimination ability at the listener’s most comfortable listening level and is only evaluated if the aviator fails to meet the specified audiometric criteria. A limitation of this practice is that individuals with normal audiograms but poor speech communication ability, particularly in noise, are not detected. This type of individual exists, and usually is diagnosed with functional hearing loss, discriminatory hearing loss, or obscure auditory dysfunction (Soli, 2003). Middelweerd et al. (1990) evaluated a group of individuals with normal pure-tone thresholds and speech reception thresholds (SRTs) in quiet who experienced subjective difficulty understanding speech in the presence of background noise. When these individuals were tested in noise, their SRT revealed significant differences from the normal-hearing control group. This effect was also evident when fluctuating masking noise was used during SRT evaluation instead of steady-state noise. The authors attribute this observation to an impaired temporal resolution in the experimental group (Middelweerd, Festen, & Plomp, 1990). Presumably, these individuals would experience difficulty flying safely and effectively, despite normal peripheral hearing sensitivity.

The last limitation of the current aviation waiver practice to be discussed is the assumption that normal speech communication ability is the only criterion for successful flight performance that should be measured. The current waiver process implies that hearing sensitivity and speech communication abilities are the most important factors to be considered to predict flight safety, while it may be possible that a pilot can fly effectively and safely with some degree of speech communication deficits. In fact, acceptable audiometric hearing thresholds (see

Table 3) are slightly higher (more allowable hearing loss) for a more experienced aviator (Class 2/3/4 compared to Class 1), suggesting that as hearing abilities decrease with age and effects of noise there are "compensatory skills, knowledge, and experience" (Soli, 2003, pg. 5) that aid the pilot. This assumption is probably correct; however, there is no required assessment of those skills, knowledge, or experiences in the flight waiver process.

Functional Hearing

It is important to clarify the differences between what is measured audiometrically in a clinical setting for an aviation hearing waiver, and what abilities are required of a pilot to function in the aviation environment – known as functional hearing abilities. The former focuses on the etiology, progression, and degree of peripheral hearing loss, measured monaurally in quiet, while the latter are situation-specific abilities that rely heavily on binaural hearing and central auditory processing. Further, functional hearing abilities in Army aviation are carried out in the presence of noise and are influenced by the type of communication system used. The literature indicates four fundamental functional hearing abilities required for successful operational performance: 1) signal detection, 2) signal identification/recognition, 3) signal localization, and 4) speech perception/intelligibility (Soli, 2003). These fundamental abilities are not specific to the aviation environment, but can be considered the foundation for many environments, including Army aviation.

Signal Detection

High levels of ambient noise in the aircraft, hearing protection, intercommunication system distortion, and individual auditory thresholds all interfere with signal detection, which is necessary before any other function related to that sound can be performed (Casali, 2006). A major problem encountered with ambient noise is masking, defined as an elevation in the signal auditory threshold caused by the presence of another sound. The lower the signal-to-noise ratio, or the greater the level of the background noise relative to the signal, the more challenging it is to detect the signal (Casali, 2006). Another aspect of masking that makes speech detection in noise difficult is that the masking effect of background noise tends to be greatest in the spectral vicinity of the masking tone, but also spreads upward and further masks signals of higher frequency as the intensity of the background noise increases (Casali, 2006). Because the majority of the energy in a speech signal is in the mid-frequency range, the upward spread of low frequency helicopter noise is especially problematic as it not only raises the signal-to-noise ratio, but also masks the speech signal. An additional problem with encountered with high intensity background noise is cochlear distortion, which occurs when the cochlea becomes overloaded and cannot properly differentiate between signal and noise (Casali, 2006).

As mentioned previously, the passive sound attenuation of a typical flight helmet is primarily high frequency attenuation, which can further mask a pilot's detection of speech signals originating outside of the helmet. Similarly, hearing losses resulting from noise exposure or presbycusis are characteristically high frequency in nature. High frequency hearing losses generally cause more difficulty with speech intelligibility rather than speech detection; however, it is conceivable that there is a synergistic detrimental effect between high-frequency hearing loss and the masking effects of a helicopter's noise on signal and speech detection in the operational environment.

Signal Identification/Recognition

Absolute signal identification and relative signal recognition are similar fundamental functional hearing abilities necessary in many operational environments. Signal identification refers to classifying a specific signal when there is no comparison signal available. A signal is recognized if it is differentiated between two or more signals presented in close auditory proximity (Sanders & McCormick, 1993). Both signal identification and signal recognition can only occur if the signal is detectable. Even if an auditory signal is detected, humans are limited in their ability to distinguish a signal from a group of similar items (Robinson & Casali, 2000). This could have implications for pilots who need to identify and recognize warning signals and speech in the midst of competing noise and verbal communications. As with signal detection, masking effects of background noise, hearing protection, and elevated auditory thresholds can interfere with a pilot's ability to properly identify and recognize auditory signals and speech.

Signal Localization

Signal localization is another of the fundamental auditory abilities necessary in a hearing-critical aviation environment and is dependent on a binaural auditory system (Soli, 2003). For an Army aviator, adequate signal localization may not be necessary in the cockpit, but is undoubtedly important during dismounted operations. Signal localization is described in terms of an individual's capacity to detect changes in the relative location of sound sources, or to determine the position of a single sound source (Soli, 2003). A normal-hearing listener with unoccluded ears is able to take advantage of time and intensity cues resulting from acoustical interaction with the external ear, head, and shoulders to localize sounds. When a sound generated in space creates a sound wave that is propagated and received by the ears of listener, low frequency wavelengths arrive at the ear closest to the sound source slightly before they arrive at the opposite ear and results in time cues as to the location of the source. Higher frequency wavelengths are localized by the head "shadowing" that occurs because of the slightly higher intensity of a sound on the side of the head from which the sound originated. In the middle range of about 1500-3000 Hz, neither phase nor intra-aural intensity differences exist, so localization is poor (Casali, 2006).

Localizing a sound source is not greatly affected by the presence of background noise, unless the noise and the signal source are co-located or the signal level is not adequately above the masked threshold (Robinson & Casali, 2000). However, localization is certainly affected when a circumaural headset or helmet effectively removes the pinna cues that provide important localization cues. Noble et al. (1994) studied hearing-impaired individuals' pure tone thresholds and their ability to localize sounds. They found that the decreased localization ability of hearing-impaired individuals was moderately correlated with the degree and configuration of their pure tone thresholds. The authors also found that individuals with conductive hearing loss contributed to the most localization ability deficits. They explain that in the case of a unilateral conductive hearing-impairment, the auditory pathway from the middle ear to the cochlea attenuates the signal ipsilaterally, but the attenuation is not comparably attenuated on the contralateral side, effectively reducing the "channel separation" that is required for binaural information to be received (Noble, Byrne, & Lepage, 1994). The authors indicate the need to measure aspects of hearing other than the pure tone audiogram to determine sound localization handicap.

Speech Perception/Intelligibility

Speech intelligibility in noise is paramount to successful military mission performance (Peters & Garinther, 1990). The last, but possibly most important functional ability for an Army aviator is that of speech perception/intelligibility – and is critical for the safety and the operational effectiveness of an aviator. It is vital for the pilot to comprehend verbal messages rapidly and thoroughly in order to safely maintain aircraft control. The inability to hear and understand auditory communication can compromise the mission and ultimately have unintended catastrophic results (Ribera et al., 2004). As with all hearing-critical tasks in the cockpit, speech perception is affected by hearing protective devices, communications systems, and hearing impairment. The effect of HPDs and communication systems on speech perception has been discussed – the following section describes the effects of hearing impairment on speech perception as it may apply to the aviation environment.

Plomp (1986) and colleagues have reported extensively on speech communication in noise. Of particular interest are their measurements of monaural and binaural SRTs in quiet and in noise in test conditions with and without spatial separation of the speech signal and interfering noise. The investigators used the elevation in the SRT above that obtained by normal hearing individuals tested in the same environment to determine the communication handicap caused by hearing impairment. The measured threshold elevation, in dB, is related to the speech intelligibility by using

> ...the slope of the empirical function relating sentence presentation level in dB to percent intelligibility. This function has a slope of 15-20% intelligibility per dB near threshold. Thus, a threshold elevation of 1 dB for a hearing-impaired individual would correspond to 15-20% poorer intelligibility than would be achieved by an unimpaired individual listening at the same presentation level (Soli, 2003, pg. 11).

Similarly, in a previous experiment, at an SRT level near 50%, a 1 dB increase in speech-to-noise ratio yielded a 20% higher intelligibility score for sentences (Plomp, 1986). Even a small difference in speech-to-noise ratio, whether it is from hearing loss or the masking effects of noise, results in a large difference in speech intelligibility.

It is important to note that these results indicate speech intelligibility measured at speech threshold levels, not at suprathreshold levels that one would expect in the cockpit of a helicopter. Nonetheless, it can be argued that the masking effects of the ambient noise in a helicopter effectively raise the pilot's threshold, thereby making speech heard through the ICS closer to threshold. Moreover, speech that is suprathreshold to a normal-hearing aviator may be at threshold for the hearing-impaired aviator. Another interesting issue this raises is that of the audiometric testing for hearing waivers. Audiologists evaluate and report a pilot's speech intelligibility at suprathreshold levels in quiet – therefore, this speech intelligibility disparity is not evident to the flight surgeon who recommends the flight waiver.

Functional Hearing and the Aviation Flight Waiver

Although the exact functional hearing abilities necessary for a pilot to operate an aircraft are complex and situation-dependent, the preceding discussion on functional hearing highlights the reality that many aspects of functional hearing are not evaluated diagnostically with a standard audiometric evaluation as outlined in the hearing loss APL. With the exception of the binaural speech discrimination requirement for waiver determination, the remainder of the audiometric evaluation is a monaural measure of peripheral auditory function, while a pilot's functional hearing is largely a binaural skill that is processed more centrally. Similarly, audiometric evaluations take place in a quiet environment, while a pilot's functional hearing occurs in the presence of high-level background noise. Lastly, an audiometric evaluation is conducted with the open ear, but the aviator's functional hearing ability is directly affected by hearing protective and communication devices (Soli, 2003). The difference between what is measured during a clinical audiometric evaluation and what functional hearing abilities are required creates a problem in defining the acceptable hearing standards for an Army pilot. It contributes to difficulty establishing an "objective relationship between hearing impairment as measured by the audiogram, and handicap caused by hearing impairment, as characterized by reduced functional hearing ability, and the conditions under which this handicap is disabling in the workplace" (Soli, 2003, pg. 4).

Mental Workload

In addition to the challenges imposed by noise, actual hearing loss, and simulated hearing loss due to hearing protection, performance is also challenged by the mental workload inherent in the flying task. There are several descriptions of mental workload found in the literature. Conceptually, mental workload is the interaction between the system and tasks, and the capabilities, motivation, and state of the operator (Kramer, 1991). The term workload also refers to the amount of mental effort needed to perform a task and is a function of supply and demand of attentional resources (Tsang & Vidulich, 2006). Others define workload as the difference between the human resources available and the task resources required (Sanders & McCormick, 1993). According to Tsang and Vidulich (2006) the two main factors that determine mental workload, and ultimately give insight to task performance are the exogenous task demands such as the priority of the task and the difficulty of the task, and endogenous factors that allow information processing such as perceiving, decision making, and response processing. The latter factor is further influenced by the individual differences in skill level or expertise (Tsang & Vidulich, 2006).

In the aviation environment, there is a distinct difference between a pilot’s experienced workload and the task demands imposed on a pilot while flying. The latter can be described as the demands placed on the pilot by the aircraft and the former can be described as the pilot’s response to those demands. Thus, the demands depend on the system (i.e., aircraft) requirements, and the workload is dependent on the pilot’s perception of the task demands. Overall system performance is the result of the combined task demands and workload (Corwin et al., 1989).

The concept of mental workload is based on resource models of divided attention, and suggests that a finite amount of resources is available for task completion. As such, an early theory of attention and mental workload was known as single resource theory (Kahneman, 1973). This theory indicated that attentional resources demanded by task processing were limited in supply and that the resources that were not being used by the task at hand (spare capacity) could be used for unexpected added demands at the will of the individual, until the supply of resources is insufficient to compensate for the demand, at which point performance declined. (Tsang & Vidulich, 2006).

Wickens (1980, 1987 as cited in Tsang & Vidulich, 2006) later proposed a multiple resource model suggesting that attentional resources are defined along three dichotomous dimensions; stages of processing (encoding and central processing or responding), processing codes (spatial or verbal), and input/output perceptual modalities (auditory or visual) (Tsang & Vidulich, 2006). Different tasks require varying amounts of different resources. For example, for a helicopter pilot to detect and interpret a flight control would require more encoding resources, while acknowledging a radio message from air traffic control would require more responding resources, and could be done because they draw on different resources (Pew, 1998). However, the pilot's performance would suffer if he or she were required to acknowledge the radio transmission while also communicating to his crew chief, particularly in the presence of high intensity background noise and elevated auditory thresholds.

Mental Workload Selection Criteria

Mental workload is multi-dimensional in nature; consequently, there is no single assessment technique that can measure all of the important aspects of mental workload. Selected measurement techniques may differ based on the given situation; however, there are important properties that must be considered regardless of technique selected. Among the most important characteristics of mental workload measurement techniques are sensitivity, diagnosticity, and intrusiveness (Eggemeier, Wilson, Kramer, & Damos, 1991).

Sensitivity refers to the degree to which a particular measurement tool can detect different levels of workload associated with the task performance. Employing a measurement technique with low levels of sensitivity will result in an inability to distinguish workload differences between tasks, which is a primary objective of workload measurement. Many primary task measures and subjective workload ratings are referred to as globally sensitive measures of operator workload because they are capable of discerning differences in workload across a broad range of operator information processing functions that influence workload (Eggemeier et al., 1991; Wierwille & Eggemeier, 1993).

Diagnosticity refers to the ability of a workload measurement technique to determine the type or source of the workload. When a workload measure is used for its diagnosticity, it is generally associated with multiple-resource models of information processing, and is said to be diagnostic if it gives insight to the different variety of resources employed by the operator. In general, secondary task workload measurement tools and certain physiological workload assessment techniques are said to be highly diagnostic and are most useful in discriminating the variations in load demands put upon specific types of operator information processing functions (Eggemeier et al., 1991; Wierwille & Eggemeier, 1993).

Another important property of workload assessment is intrusiveness, which refers to any disruption of the primary task performance caused by the application of a workload measurement technique. Secondary task measurement techniques, particularly in a laboratory environment, appear to be the most likely to be associated with intrusiveness (Eggemeier & Wilson, 1991; Kramer, 1991; Wilson & Eggemeier, 1991).

Implementation requirements and operator acceptance are also properties that should be considered when selecting a mental workload measurement technique. Implementation requirements consist of any equipment, instrumentation, or operator training that is required for proper application of the measurement technique. Eggemeier et al. (1991) caution that implementation requirements can limit the application of some measurement techniques, especially in a real-world environment such as a flight environment (Eggemeier et al., 1991; Wierwille & Eggemeier, 1993). Operator acceptance or operator perception of the usefulness of a workload assessment tool can affect the quality of operator responses, and whether they are even obtained.

Mental Workload Assessment Techniques

Workload assessment techniques are generally divided into three categories; performance-based techniques that evaluate the ability of the operator to perform specified tasks, subjective measures that rely on operator judgment of the workload experienced by a task, and physiological techniques that assess the biological response of the operator in response to the task (Wierwille & Eggemeier, 1993).

Performance-Based Techniques

Primary Task Measures

Primary task measures of mental workload are simply measures of operator task performance. This method involves monitoring an operator's performance and documenting what changes result from changes in task demands. The theory behind primary task measurement is that task performance is expected to decrease at some point as the task demands increase because humans have a finite capacity of resources to cope with the demands (Tsang & Vidulich, 2006). According to the single resource theory, mental resources that allow task completion are commodities of limited supply. As the task demands increase, the proportion of expended resources (or workload) associated with the task is assumed to increase. At some point, an upper threshold of mental resources is reached, and the operator can no longer perform the task at an optimal criterion level. Single resource theory holds that this lack of sufficient resources is the principle cause of operator performance decrements that occur in multi-task environments (Eggemeier & Wilson, 1991).

There are four important limitations of assessing mental workload with a primary task measures: 1) the primary task of interest may not require enough resources to overload the human system, resulting in perfect performance, 2) the measurement methods and meaning of measurements may differ between two primary tasks, 3) it may not be possible to obtain an adequate measure of primary task performance, and 4) two primary tasks may differ in their resulting performance due to differences in data limits rather than by mental workload (Wickens & Hollands, 2000).

Despite these limitations, primary task measures have been shown to be valuable tools that demonstrate high levels of sensitivity to variations in workload, given that they are specific to task demand (Wierwille & Eggemeier, 1993). As such they are always recommended as part a thorough workload evaluation (Tsang & Vidulich, 2006; Wierwille & Eggemeier, 1993).

Secondary Task Measures

While a system evaluator is likely to evaluate primary task performance for reasons other than mental workload inference, secondary task measurements are typically used only as an indicator of mental workload (Tsang & Vidulich, 2006). Secondary task measurement techniques assess workload through measurement of task performance when an additional task in performed at the same time as the primary task. (Wierwille & Eggemeier, 1993). Since the purpose of workload assessment is to determine if an operator is working within an acceptable information processing capacity while performing a task, it follows logically that the operator could perform another task if there is unused capacity (Tsang & Vidulich, 2006).

The single resource theory also provides the theoretical basis for secondary task measurement techniques (Eggemeier & Wilson, 1991). There are two general categories of the secondary task approach to mental workload measurement – subsidiary task paradigm and loading task paradigm (Knowles, 1963 as cited in Eggemeier & Wilson, 1991). The more commonly used secondary task approach is the subsidiary task paradigm, and requires the operator to perform a secondary task while maintaining maximum primary task performance, as if the primary task were the only task. If optimal primary task performance is maintained, but addition of a secondary task increases overall mental workload, differences in secondary task performance will reflect any discrepancies in the resource expenditures of the primary task. In this way, secondary task measurements can distinguish changes in mental workload that primary task measurements cannot (Eggemeier & Wilson, 1991). The loading task paradigm places emphasis on maintaining performance of the secondary task as if it were as important as the primary task. This emphasis is assumed to lead to degradations of the primary task, with more difficult primary task performance being degraded more than less difficult tasks (Damos, 1991).

Some of the most commonly used secondary tasks are rhythmic tapping, random number generation, probe reaction time, and time production or estimation (Wickens & Hollands, 2000). Because the secondary task performance is assumed to increase as the primary task resource requirement decreases, the investigator using this technique is interested in variations of secondary task decrement to infer differences in primary task demand (Wickens & Hollands, 2000).

Among the advantages of secondary task assessment techniques are high face validity and the ability to be applied to different primary tasks. A secondary task measure gives the appearance of relevance because it is designed to predict the amount of ‘leftover’ resources that will be available to an operator in the case of an unexpected event. Application of secondary tasks to very different primary tasks allows for comparison of workload measures in the same units, unlike measurement of primary tasks alone. There are other advantages to secondary task measurement of mental workload – many secondary tasks that vary in respect to resources demanded have been developed for use in different evaluations. This allows investigators to select secondary tasks that will compete for specific resources of the primary task (Gawron, 2000 as cited in Tsang & Vidulich, 2006 pg 250). A measure of secondary task performance is useful in situations where primary task performance is not possible, such as automobiles or aircraft that do not have built-in performance measuring capabilities.

An important limitation of employing secondary task techniques is that such techniques are not always sensitive, particularly if the resource demands of the secondary task do not match the resource demands of the primary task. According to the logic of multiple resource theory, performance of a secondary task will only be a sensitive workload measure of the primary task if the two tasks vie for the same processing resources. A low degree of interference between resources required for both tasks usually results in a higher degree of performance; however, for workload assessment, this is not desirable – the degree of interference is necessary to infer the level of mental workload (Tsang & Vidulich, 2006).

Another challenge associated with secondary task measures is that of intrusiveness; the degree to which the task interferes with the primary task (Wierwille, Rahimi, & Casali, 1985). Intrusiveness is evident when the addition of a secondary task causes an unintended change in the processing of the primary task as well as increasing the workload, and the resulting measurement is simply experimental artifact (Tsang & Vidulich, 2006). Embedded secondary tasks can be used to address the negative effects of intrusiveness. An embedded task is a normally occurring portion of the operator’s overall task, but it is of lesser priority than the primary task (Raby & Wickens, 1994). A naturally occurring embedded task in aviation is a lower priority task that, in a real-world flight scenario, can be “shed” if workload demands become excessive (Shingledecker, 1984; Vidulich and Bortolussi, 1988 as cited in Eggemeier & Wilson, 1991).

Subjective Measures

A number of reviews have documented subjective measurement sensitivity to a variety of task demand manipulations in multi-task environments such as an aircraft (Wierwille &

Eggemeier, 1993). Among the demonstrated sensitive techniques are the Modified Cooper-Harper (Wierwille & Casali, 1983), the Bedford Scale (Roscoe, 1987), the NASA Task Load Index (Hart & Staveland, 1988) and the Subjective Workload Assessment Technique (Reid & Nygren, 1988).

Subjective workload measurement techniques involve human operators to quantify their workload experience. Useful in categorizing subjective rating scales are three variables. The first variable is dimensionality and refers to the number of workload attributes on which the operator was required to rate his or her experience. The second variable, evaluation style, indicates whether the operator was required to provide an absolute judgment of workload or if they were required to compare the given experience to another and provide a relative rating. Finally, immediacy differentiates between subjective scales that are meant to be administered as soon as the experience has ended and those ratings that are conducted at the end of the entire session or experiment (Tsang & Vidulich, 1994; Vidulich & Tsang, 1987). Tsang & Vidulich (2006) report that the most commonly used combination of these variables are multi-dimensionality, absolute evaluation, and immediacy. The most common alternative combination is that of a unidimensional rating with a relative comparison that is collected retrospectively.

Tsang and Vidulich (1994) found that relative, immediacy-technique ratings were less sensitive to workload levels compared to relative delayed measures and relative redundant measures. They compared performance-based workload measures with several subjective measures and overall, their results were supportive of subjective workload ratings, particularly for highly complex or highly automated operational environments. The ‘bottom line’ of this experiment is that the authors recommend a retrospective subjective measure instead of an immediate measure for relative mental workload ratings. Further, the authors caution experimenters to not use the combination of immediate presentation with relative subjective ratings; instead, a retrospective approach is best suited for a relative workload judgment (Tsang & Vidulich, 1994).

Modified Cooper-Harper

One of the most popular and robust single dimension rating scales is the Modified Cooper-Harper (MCH) (Wierwille & Casali, 1983). The MCH is considered a unidimensional, immediate, and absolute scale. The MCH is an extensively modified version of the Cooper-

Harper scale (Cooper & Harper, 1969) aircraft handling rating scale intended for application outside of the aviation environment. The MCH preserved the flow diagram and 10-point decision tree format of the original scale, but the narrative descriptors were replaced to allow use of the scale for workload tasks other than aviation. For this scale, the range of ratings is from 1 to 10, and the scale provides an ordinal level of measurement. A rating of one represents easy task performance (operators' mental effort is minimal and desired performance is easily attainable) and a rating of 10 represents an impossible task (the instructed task cannot be accomplished reliably).

Several flight simulator experiments have demonstrated the sensitivity of the MCH scale. Casali & Wierwille (1983) applied the MCH as a measure of workload in an experiment that manipulated pilot communication load. The ratings also showed a monotonic relationship with communication demand levels. Additionally, the ratings differentiated the low level of demand from both the moderate and high levels of task demand (Casali & Wierwille, 1983). The same authors (1984) applied the MCH to pilot workload under different hazard detection conditions while in simulated flight. Again, the ratings exhibited a monotonic relationship with loading levels. The ratings also successfully discriminated between the low and high levels of hazard load as well as between the moderate and high levels of hazard load (Casali & Wierwille, 1984; Eggemeier & Wilson, 1991). Other flight simulation experiments have also supported the MCH for mental workload assessment in a cockpit environment (Itoh, Hayashi, Tsukui, & Saito, 1989; Skipper, Rieger, & Wierwille, 1986; Wierwille et al., 1985). These experiments all demonstrated that the MCH was sensitive to a range of task manipulations, and therefore supported the scales use in a multi-task cockpit environment (Eggemeier & Wilson, 1991).

The Bedford Scale

The Bedford Scale is another modification of the Cooper-Harper scale. Like the MCH, the Bedford scale retains the 10-point decision-tree format of the Cooper-Harper scale. This scale is based on operator judgments of workload and spare information processing capacity, although research on the ability of pilots to effectively judge spare capacity is lacking (Eggemeier & Wilson, 1991). Several simulator-based experiments have applied the Bedford scale and found it to be globally sensitive to various types of workload manipulations associated with the flight environment. Corwin et al. (1989) used the Bedford scale in a commercial flight simulator to assess pilot mental workload and obtained ratings, supported by videotapes that discriminated different levels of workload as well as differences between flight phases. Vidulich and Bortolussi (1988, as cited in Eggemeier & Wilson, 1991) found that the Bedford scale discriminated between workload associated with different flight phases in an experiment evaluating the effects of communications and phases of flight on pilot workload in a combat helicopter simulator.

NASA-Task Load Index (TLX)

NASA-TLX is a multi-dimensional, absolute, and immediate rating scale that is based on six scales (mental demand, physical demand, temporal demand, performance, effort, and frustration level) that are assumed to characterize important components of subjective workload. The operator is required to complete a series of 20-point ratings on each scale. Ratings on each scale are then weighted based on the operator's judgment of relative importance, and ultimately produces an overall workload rating between 0 and 100 for each experience rated (Eggemeier & Wilson, 1991; Tsang & Vidulich, 2006). Application of NASA-TLX has successfully demonstrated capability to differentiate task demand levels in flight simulator-based experiments. Battiste and Bortolussi (1988) found that NASA-TLX discriminated between both high and low workload flights, as well as a main effect for flight segment with lower ratings for the flight cruise and higher ratings for takeoff and landing. Similarly, Corwin et al., (1989) found that post-flight NASA-TLX ratings differentiated between low and high workload flight, as well as different flight phases in a commercial flight part-task simulator experiment. Data from experiments such as these support NASA-TLX as a globally sensitive rather than highly diagnostic workload assessment rating scale, due to its ability to identify task demand variations on different information processing resources (Eggemeier & Wilson, 1991).

Subjective Workload Assessment Technique (SWAT)

SWAT also is a multi-dimensional, absolute, and immediate rating scale that consists of three scales that are assumed important dimensions of mental workload (time load, mental effort load, and psychological stress load). The relative ratings of the scales are formulated by the operator's impression of the workload brought on by the combination of the workload variations in each of the three scales. Just as with NASA-TLX, the ratings are analyzed to ultimately produce an overall rating between 0 and 100 (Eggemeier & Wilson, 1991; Tsang & Vidulich, 2006). There are two phases involved in the application of SWAT; scale development and event scoring. The scale development phase involves combining the operator ratings from the three scales to generate an overall workload rating, which is unique due to its interval-scale properties. The event scoring phase consists of separate evaluation of operator workload ratings associated with the time, effort, and stress sub-scales (Reid & Nygren, 1988).

As with NASA-TLX, the global sensitivity of SWAT, rather than diagnosticity of mental workload has been demonstrated in a number of flight simulator experiments (Eggemeier & Wilson, 1991). Haworth et al. (1986, as cited in Eggemeier & Wilson, 1991) used SWAT in a nap-of-the-earth (NOE) simulator-based flight experiment and found that the rating scale discriminated between helicopter control configurations and single-pilot mission segments. Battiste and Bortolussi (1988) investigated SWAT's sensitivity with workload variations in a commercial aircraft flight simulator. Post-flight SWAT ratings successfully revealed significant differences between the high and low workload missions.

Comparisons of SWAT with NASA-TLX have indicated that data from SWAT would be more efficient in a prolonged activity because the SWAT has only three rating scales compared to NASA-TLX's six scales. In favor of NASA-TLX is the scale's paired comparison technique to determine rating scale weighting, which was found to be easier than the card-sorting technique that determines the weighting of SWAT's ratings (Tsang & Vidulich, 2006). Rubio et al. (2004 as cited in Tsang & Vidulich, 2006) reported that SWAT and NASA-TLX both provide diagnosticity and concurrent validity with performance. They also report that both have proved sensitive to difficulty manipulations; however, others have suggested that NASA-TLX tends to be more sensitive, particularly for low levels of workload (Battiste & Bortolussi, 1988; Hill et al., 1992).

Advantages of Subjective Workload Measures

The obvious advantage of subjective workload techniques is that they do not disrupt primary task performance. In addition, subjective workload measures have high face validity and are generally more direct than other measures. For example, physiological and performance measures require knowledge of the functional relationship between workload and those variables. With subjective measures, however, if an operator is asked how much workload he is experiencing, he can describe it in general, without requiring the examiner to interpret the information given. One must conclude an operator is experiencing high levels of workload if the operator says he is, despite what other measurements indicate. This logic leads to another advantage of subjective workload measure - usually the method used to validate objective measures of workload is to show that the objective measure correlates with the operators' subjective measure. Finally, subjective measurement techniques are easily administered and subjective data is easily obtained, making them adaptable to many real-world operational environments (Reid & Nygren, 1988). Caution should be taken when using subjective workload measure to ensure that an operator's judgment of workload is not influenced by other variables, such as dislike of task or unwillingness/inhibition to report any difficulty (Wickens & Hollands, 2000).

Physiological Measures

Physiological mental workload measurement techniques assess selected aspects of an operator's biological response to changes in task or system demands. Among the most frequently used physiological mental workload measurement techniques in a multi-task environment, such as the flight environment, are heart rate and heart rate variability, respiration, eye activity, brain activity, and hormone levels (Wilson & Eggemeier, 1991).

Heart Rate

Heart rate, or heart beat-to-beat variability, is another way to monitor cardiac changes during operator performance. In general, heart rate variability decreases as mental workload increases. Specifically, spectral analysis of inter-beat intervals has shown a decrease in the band of power centered around 0.10 Hz as mental workload increases (Wilson & Eggemeier, 1991).

There have been two general findings from studies that have used heart rate as an indicator of mental workload in both actual flight and in simulators. One is that heart rate provides a measure of specific flight-segment mental workload and the other is that the responsibility of controlling the aircraft produces higher heart rates than just flying the aircraft. However, some applications of heart rate measures in the flight environment have not shown reliable variations due to changes in mental workload. In three separate studies (Casali &

Wierwille, 1983, 1984; Wierwille et al., 1985) that looked at a variety of workload measures in a simulator, there were no significant differences found in heart rate or heart rate variability due to task difficulty manipulations incorporated into the simulated flight mission.

Respiration

Relatively few studies have examined the relationship between respiration and mental workload in a flight environment. In general, respiration rate tends to increase as mental workload increases. Among the difficulty in using respiration as an index of mental workload is that speech changes the pattern of breathing, so utilization of this metric is problematic in any situation where speech occurs (Wilson & Eggemeier, 1991).

Eye Activity

Endogenous eyeblinks, or eyeblinks that are not reflexive blinks in response to stimuli in the environment, have been found to change in activity as a function of the amount of visual attention required of a task. In terms of mental workload, blink rate has demonstrated a relationship to visual workload in a flight environment, and typically decreases as visual workload increases. Some studies within the context of the flight environment have shown blink duration to decrease as visual attention demands increase. However, other studies in a simulated flight environment have not shown reliable variations of eye function due to workload manipulations (Casali & Wierwille, 1983). Wierwille et al., (1985) examined blink rate as a function of arithmetic and geometric operations associated with navigational problems in a flight simulator and found that blink rate increased with increased mental workload, a finding inconsistent with trends found by other researchers previously.

It is important to note that the inconsistency in these findings could be due to differences in type of demand placed on the operator. In the earlier studies, the demand was visual, while the task demands by Wierwille et al. (1985) were cognitive in nature, suggesting that eyeblink measures would be more reliable indicators of mental workload with visual tasks (Wierwille et al., 1985; Wilson & Eggemeier, 1991). Lastly, eye point-of-regard (direction of gaze) measures have been used in simulated and actual flight environments to investigate how mental workload influences eye scanning. Wilson and Eggemeier (1991) caution that interpretation of eye point-of-regard is difficult because eye fixations are often a function of the available time regardless of the task, and when used with head up displays, it is sometimes impossible to know if the operator is looking at the display or out the window. Regardless of the potential difficulty, eye point-of-regard can be useful to determine visual workload or to make instrumentation comparisons (Wilson & Eggemeier, 1991).

Brain Activity

An electroencephalogram (EEG) is a recording of the brain's electrical activity that is recorded through the scalp with electrodes and is usually analyzed by decomposing the waveforms into its constituent frequency components (delta – up to 3 Hz, theta – 4-7 Hz, alpha – 8 – 13 Hz, and beta – 14-25 Hz). This measurement of brain activity has been applied to workload measurement in the flight environment, showing that EEG analysis, especially in the alpha and theta bands can demonstrate sensitivity to differences in mental workload in multi-task environments. Cortical evoked potentials (EP) have also been employed as a measure of mental workload in both flight simulations and real flight environments. Transient EPs consist of various positive and negative waveform peaks associated with stimulus presentation and occur in the first 750 milliseconds. The evoked response typically used for workload assessment is the positive peak that occurs 300 milliseconds following a stimulus presentation (Wilson & Eggemeier, 1991).

Hormone Levels

During periods of stress, including situations of high mental workload, the human body's adrenal glands release hormones, specifically catecholamines, into the blood system, and can be measured in the saliva, urine, or blood. The purpose of the hormones is to activate the body's resources to cope with the stressful situation. As such, these hormone levels can provide an indication of an operator's workload (Miller, 1968 as cited in Wilson & Eggemeier, 1991). Miller (1968) found that increased catecholamine levels were associated with the stresses of flight, long duration flights, level of experience, and degree of responsibility. Catecholamine levels are usually measured from urine samples following the task of interest; the levels indicate responses over a relatively long period and limit the ability to correlate the response to a specific event. Alternative hormone level measurements have been used, such as salivary cortisol levels following flight tasks (Kakimoto, et al. 1988 as cited in Wilson & Eggemeier, 1991). Despite the limitations of measuring hormone levels as an indicator of mental workload, this technique could be useful for complex task situations for which other measures are not possible.

Advantages and Disadvantages of Physiological Workload Measures

Physiological workload measurement techniques have distinct advantages over other tools. Among the advantages is the relative unobtrusiveness of physiological procedures. Most physiological measures require electrode or transducer placement on the operator's body, which does not introduce any additional task demands for the operator. The second advantage is that physiological measurements can be recorded in the absence of any specific behavior, thereby fulfilling the need to obtain workload metrics without the measurement of overt performance. Another advantage is related to the dimensions of mental workload. Mental workload is multi-dimensional by definition, and physiological measures are inherently multi-dimensional, thereby able to provide several perspectives of mental workload. A fourth advantage of physiological measures is the speed at which they can respond to periodic shifts in mental workload (Kramer, 1991).

A disadvantage of physiological assessment techniques is the relative difficulty in collecting, analyzing, and interpreting physiological data. Although the cost of this type of data collection has decreased over the years, there is still a need for specialized equipment that results in a more expensive method of workload measurement than other techniques. A second disadvantage is the amount of technical expertise required for the interpretation of physiological data. The third important disadvantage is related to artifact noise obtained through physiological measurement techniques and the difficulty discriminating between the desired signal and the artifact noise. There is also the potential problem of physiological measures being influenced by factors other than mental workload. For example, heart rate is affected by physical exertion as well as changes in mental workload. This problem makes direct interpretation of physiological measures sometimes difficult (Wilson & Eggemeier, 1991).

Situation Awareness

Just as the impact of task demands of flying a helicopter can be explained in terms of mental workload, the subsequent flight performance can be explained in terms of situation awareness. When viewed in terms of situation awareness (SA), the task demands are somewhat different, and encompass the perception and comprehension of an array of data that often changes very rapidly. The concepts of mental workload and SA are related, but also distinct from one another. Wickens (2001, p. 446) explains the difference in this way:

> Mental workload is fundamentally an *energetic* construct, in which the quantitative properties ("how much") are dominant over the qualitative properties ("what kind"), as the most important element. In contrast, situation awareness is fundamentally a *cognitive* concept, in which the critical issue is the operator's accuracy of ongoing understanding of the situation (i.e., a qualitative property).

In other words, one assesses the amount and type of workload, and the quality of the content of SA (Tsang & Vidulich, 2006). Mental workload and situation awareness are distinct concepts that appear to be interrelated, but it is not advisable to infer the status of one construct based on the measurement of the other. Rather, it is important to assess both mental workload and situation awareness when evaluation system effectiveness, since both concepts indicate valuable information about an individual's reaction to performing a complex task. In assessing both constructs, it is also important to focus on their interrelationship rather than their independence. Understanding the influence of system elements on both mental workload and situation awareness should aid in understanding how to help the pilot fly most effectively (Vidulich, 2003).

In the aviation environment, SA can be thought of as "an internalized mental model of the current state of the flight environment" (Endsley, 1999, pg. 257) that serves as the central knowledge base from which all decision making takes place (Endsley, 1999). This is especially important for the military pilot who must be aware of a multitude of factors concerning enemy and friendly aircraft as they relate to the assigned mission. A high level of SA is a critical aspect of successful aviation performance. In a review of military aviation accidents (Hartel, Smith, & Prince, 1991) and in a study of major air carrier accidents, 88% of the mishaps involving human error could be attributed to SA problems (Endsley, 1995).

Situation Awareness Defined

Situation awareness, in general terms, is “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future” (Tsang & Vidulich, 2006, p.247), and has been comprised in three levels.

Level 1 SA – Perception of the Elements in the Environment

The first level is the basic perception of cues, without which forming a correct picture of the situation decreases considerably. This level of SA is easily achieved by inexperienced pilots as long as there are not factors that distract from information acquisition, such as fatigue or distractions (Sarter & Woods, 1991).

Level 2 SA – Comprehension of the Current Situation

The second level (Level 2 SA) goes beyond basic perception and focuses on comprehension. In other words, how one combines, interprets, stores, and retains information. A person with Level 2 SA can use the Level 1 SA data perceived and obtain operationally applicable meaning from it. This ability requires expertise, defined as a “highly developed repertoire of pattern-oriented representations” (Chase & Simon, 1973, p. 51, as cited in Sarter & Woods, 1991). A lack of expertise can result in impaired ability to search for useful information.

Level 3 SA – Projection of Future Status

Finally, Level 3 SA is concerned with projection, or the ability to predict future situations based on current events. Attainment of this level represents a skilled expert in the domain of interest. From an aviation point of view, this level is the basis for “being ahead of the plane” and can be described as “the mental simulation of future system state and behavior to eliminate surprises” (Sarter & Woods, 1991, p. 51). Time is an important part of Level 2 SA and Level 3 SA in that a critical part of SA is understanding the amount of available time before an event occurs or a particular action must be taken (Endsley & Garland, 2000). In the dynamic aviation environment, minor departures from normal aircraft operation that are not critical at one time may progress or interact over time to become a major threat. Therefore, it is of utmost importance that pilots observe, integrate, and remember the minor deviations over time.

Isolated, recurrent situation assessments are insufficient for achieving SA (Sarter & Woods, 1991).

It is important to note the difference between the process and the product of SA. The former refers to the various perceptual and cognitive activities that lead to a state of awareness, and the latter refers to the state of awareness itself. The process of SA is supported by perception, working memory, and attention, while the product of SA is supported by knowledge of present events through the working memory, and by knowledge of the past and experience through the long-term memory (Tsang & Vidulich, 2006).

Situation Awareness Requirements

The elements a pilot (or an aircrew) needs to perceive, comprehend, and project are specific to the particular class of aircraft and the context of the mission. However, there are general classes of elements that are needed across many types of aircraft systems (Endsley, 1995).

Geographical SA

Elements important to geographical SA include the location of a pilot's own aircraft as well as other aircraft, terrain features and aircraft position in relation to designated features, airports, runways and taxiway assignments, and pathways to the desired location.

Spatial/Temporal SA

Elements such as aircraft readings of attitude, altitude, heading, velocity, and aircraft capabilities are essential to spatial/temporal SA.

System SA

For a pilot to have adequate system SA he or she must have command over the elements of system status, functioning, and settings; radio and ATC communications and impact of malfunctions or system degradations on system performance and flight safety.

Environmental SA

Of particular value for environmental SA is information about weather formations, instrument flight rules (IFR) versus visual flight rules (VFR) conditions, and projected weather conditions.

Tactical SA

Lastly, in order for a pilot to possess tactical SA, he or she must have knowledge the capabilities, location, and tactical status of other aircraft, as well as knowledge of his or her own aircraft capabilities in relation to other aircraft. The pilot must also be aware of aircraft detections, threat prioritization, and current and projected threat intentions.

Situation Awareness Measurement

Endsley (2000) explains that SA, as an intervening variable between stimulus and response, gives greater diagnosticity and sensitivity than typical performance measures. Endsley further explains that the relationship between SA and performance can be viewed as a probabilistic link. Good SA should increase the likelihood of good performance, but it does not guarantee it. On the other hand, poor SA increases the probability of poor performance, but does not always result in serious error. These issues with SA measurement suggest that behavior and performance only indirectly measure SA (Endsley, 2000).

Depending on the nature of the data collected, SA measurements can be divided into three general categories. As with mental workload assessments, SA assessments can be in the form of operator performance, subjective ratings, and psychophysiological measures. In addition to consideration of sensitivity, diagnosticity, intrusiveness, validity, reliability, ease of use, and operator acceptance properties that are common to workload measurements, one must also consider the process of building SA and the actual awareness that is the product of SA (Tenney, Adams, Pew, Huggins, & Rodgers, 1992). Different individuals may use different methods to obtain the state of knowledge, or may have a different state of knowledge based on the same circumstances because of their differences in comprehending and projecting acquired data. SA measures that are concerned about SA processes will provide information on how individuals acquire information needed for SA, but will not provide complete information regarding the level of SA obtained (Endsley, 2000).

Performance-Based Measurements

Performance-based measurement of SA is any assessment that infers SA from an individual's actions or system performance based on an individual's actions (Pritchett & Hansman, 2000). Real-time performance and testable responses have been used in SA measurement, but the recall-based memory probe technique is considered the prototypical measure of the theoretical concepts of SA (Tsang & Vidulich, 2006).

Memory Probe Measures of Situation Awareness

Some of the earliest SA research involved the use of memory probes to assess pilot SA, and this technique has met the threshold to be considered a feasible measure associated with a framework theory (reflecting the broad understanding of and consensus about a domain) of SA. The memory probe technique is considered the prototypical tool with which to measure a pilot's current understanding of a situation (Vidulich, 2003). The Situation Awareness Global Assessment Technique (SAGAT), as created by Endsley (1988) was the first popular and standardized procedure for measuring SA. SAGAT consists of a series of steps, the first of which is to identify the SA requirements for the given task. The second step is to collect the SA data during simulation of the task by stopping the simulation at random times and asking the individual to answer questions from memory about the status of information parameters related to the task. Not all of the SA requirements are sampled at each data collection points; rather, a randomly selected subset of requirements is assessed within and across participants until a sufficient data set is collected. The last step is to calculate the SAGAT score by comparing the responses to the actual state of the situation. The main limitation of this procedure is the intrusiveness of pausing the simulation to acquire the data; however, the advantages of the method are immediacy of measurement, globalness of information assessed, objective measurement, and face validity (Endsley, 1988). A well-designed memory probe evaluation can capture whether a pilot in a simulated flight has an accurate picture of the task situation with which they are confronted (Vidulich, 2003).

Situation Awareness Real Time Performance Assessment

Real-time performance assessment is based on the assumption that an operator will react appropriately to task demands in a timely manner if he or she is aware of the task demands. The Global Implicit Measure (GIM) (Vidulich & McMillan, 2000) is an example of such an assessment and is based on the assumption that a human-machine operator is attempting to accomplish goals at differing priority levels; therefore, it is possible to obtain performance-based SA measures by assessing the momentary progress toward accomplishing the goals (Tsang & Vidulich, 2006).

The GIM technique requires a detailed task analysis to link measureable behaviors to the accomplishment of mission goals, which vary depending on the mission phase. Measureable behaviors that affect goal accomplishment for each phase are then identified and scored. According to GIM algorithms, the proportion of mission-specific goals being accomplished indicates how well the pilot is meeting the goals of the mission. In addition, the behaviors that are not conducted during the mission provide and indication of the portions of the mission that the pilot was either unaware of or unable to perform. By evaluating the quality of task performance and a diagnosis of the problem if task performance differs from the ideal (as specified by the GIM task analysis and scoring algorithm), the GIM scores can potentially provide a real-time picture of a pilot's SA (Tsang & Vidulich, 2006).

Testable Responses

The uncertainty that arises when inferring SA from observable actions is an important concern when utilizing performance-based SA measurements. As mentioned earlier, good performance may result due to good luck or good procedural responses even in situations with poor SA. Likewise, poor performance may be the result of poor decision-making despite good SA (Pritchett & Hansman, 2000). Testable responses provide a way to reduce the ambiguity when determining one's SA based on their performance. Testable responses are isolated, experimentally controlled events that require a specific action that cannot be expected through any means other than good SA. Ideally, a testable response incorporates a proceduralized requirement for action, thereby minimizing the effect of decision-making, and making SA the largest factor in subject performance.

An example of a testable response was used in a flight simulator study in which the pilots were allowed to overhear radio communications indicating that another aircraft had not departed the runway on which the pilots were to land. In order to avoid a collision, the pilots had to take action. A strong reaction to the overheard information suggested good SA and the lack of a reaction indicated poor SA (Pritchett & Hansman, 2000). The use of testable responses may not be practical in situations that are complex and allow a large number of interpretations; however, for many types of environments, testable response events provide a non-intrusive, objective measure of SA (Pritchett & Hansman, 2000). The advantages of subjective assessments of SA are that they are inexpensive, easy to administer, nonintrusive, and can be collected in controlled real-world settings as well as in simulation studies (Endsley, 1996).

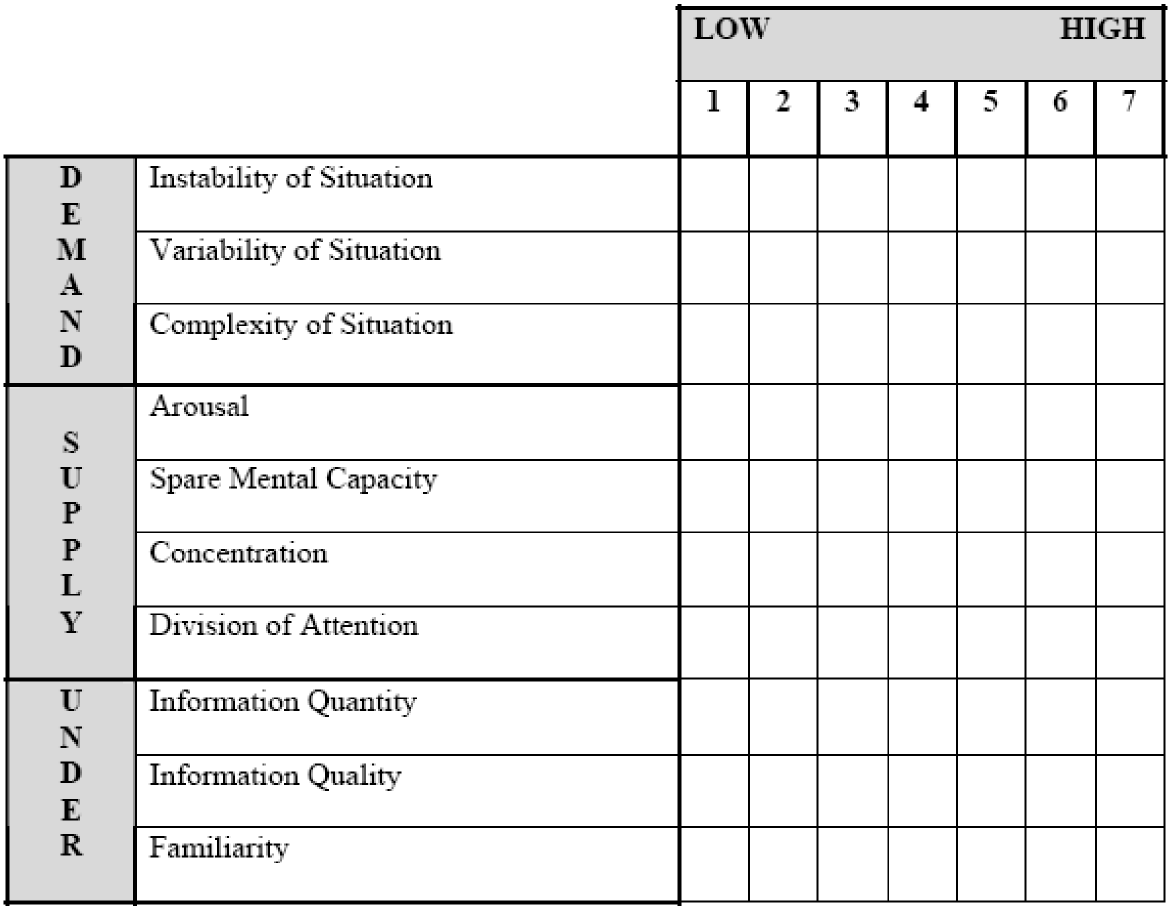
Subjective Ratings

Subjective assessment of SA usually involves assigning a numerical value to SA quality during a particular period during a task. Commonly, subjective rating techniques involve a linear scale with verbal descriptors at the endpoints. Many researchers have used multiple subscales within the rating due to the theory that SA is a multidimensional construct (Jones, 2000).

Multidimensional Absolute Immediate Ratings

The most common multidimensional, absolute, immediate SA rating is the Situation Awareness Rating Technique (SART) (Taylor, 1990). During development of SART, 10 generic SA constructs were elicited from subject-matter expert interviews, which clustered into three broad dimensions: attentional demands (D), attentional supply (S), and understanding (U), as shown in Figure 6. These three dimensions make up an abbreviated 3-dimensional SART that can be used in cases where a shorter scale is advantageous. However, subjective estimates of SA can be collected using either the 3-dimensional or the full 10-dimensional scale, depending on the given application and the degree of intrusiveness permitted by the measured task (Taylor, 1990). The values for each dimension can be combined into a single SART value by the following formula: $SA = U - (D - S)$; however, there is an element of arbitrariness to this formula because it was derived from theoretical considerations of how the dimensions are related, rather than statistical evaluation. For that reason, one should not place undue emphasis on the single calculated value (Selcon & Taylor, 1990).

In addition to the advantages of general subjective SA measures, SART has high ecological validity because its dimensions were obtained from operational aircrew (Selcon & Taylor, 1990). SART also takes into account the workload constructs of supply and demand, and provides some measure of how changes in workload affect SA. This is an advantage to proponents who suggest that workload is a integral part of the multidimensional SA construct (Taylor, 1990); however, critics believe that including workload within the SA scale confound SA measurement by confusing it with workload (Jones, 2000).



		LOW						HIGH
		1	2	3	4	5	6	7
DEMAND	Instability of Situation							
	Variability of Situation							
	Complexity of Situation							
SUPPLY	Arousal							
	Spare Mental Capacity							
	Concentration							
	Division of Attention							
UNDER	Information Quantity							
	Information Quality							
	Familiarity							

Figure 6. SART rating scale

The Situational Awareness Rating Scale (SARS) measurement technique is another subjective measure of SA (Bell & Waag, 1995). The SARS scale was developed by interviewing experienced F-15 pilots to reach 31 behavior elements of SA that are important to mission success, which represent eight categories of mission performance. To complete the SARS scale, each participant rates each of the 31 elements on a six-point scale with the anchors of ‘acceptable’ and ‘outstanding’ (Jones, 2000; Waag & Houck, 1994). Based on validity and sensitivity evaluations of SARS, the scale appears to assess ‘perceiving what is important’ and then ‘using that perception to guide the selection and performance of appropriate behaviors’, as was intended. However, SARS does not provide the same type of subjective measure of SA as other subjective scales; rather, SARS considers SA as more of an innate ability than a changeable state of knowledge (Jones, 2000). Another disadvantage of SARS is the scale considers dimensions other than SA in its measurement, such as decision-making abilities, flight skills, and performance (Endsley, 1996).

Unidimensional Relative Retrospective Judgments

The Subjective Workload Dominance (SWORD) technique was adapted to measure SA and is known as SA-SWORD (Vidulich & Hughes, 1991). The adaptation required no changes in the data collection process or in the analysis; it only required a change in the instructions to the operators. As with SWORD, SA-SWORD utilizes pairwise comparisons for data collection and then implements a geometric means rating calculation algorithm. The SA-SWORD technique has demonstrated sensitivity and inter-rater reliability and has shown promise that the technique is a useful tool for assessing subjective SA. The main disadvantage of the method is that it is a single metric of SA and as such, is not likely to show more than part of the picture regarding SA in complex tasks (Jones, 2000).

Psychophysiological Measures

The concept of physiological measures of SA is relatively new, and its theoretical and methodological progress is not at the advanced stage as physiological measures of mental workload (Tsang & Vidulich, 2006). According to Tsang and Vidulich (2006), SA is supported by many of the same processes as mental workload, but there are not hypothesized cortical regions or other physiological responses associated with SA as there are with mental workload.

RESEARCH OBJECTIVES AND HYPOTHESES

Army pilots are faced with several challenges that can interfere with mission success. The literature supports the notion that in order to fly safely, pilots need to possess *some* degree of functional hearing, probably supplemented by adequate communication devices, and have manageable flight workload levels. However, there are serious limitations in the literature in generalizing results to Army aviators. There is a strong need to provide an empirically based justification for quantitative hearing criteria for Army aviators, as well as evidence for communication device selection to support their mission performance.

There are no performance data that are applicable to hearing-impaired pilots, and as such, there is no commentary on how the choice of headset may interact with an Army aviator's hearing impairment that does not meet Class II audiometric flight standards, which may result in a flight waiver. Further, neither pilot workload nor flight/mission performance has been evaluated with CEPs or with passive headset/earplug combination use. Lastly, there has been no performance comparison on operational metrics (i.e., flight control, speech intelligibility, pilot workload) among CEPs, ANR headsets, and headset/earplug combination use.

With these voids in consideration, the objectives of this project were as follows:

1. To refine audiometric waiver criteria, in part based on flight-related performance data to ensure the maintenance of a readily mobile, effective fighting force of Army aviators.

The current waiver criteria for those who do not meet the audiometric standards are based largely on binaural speech intelligibility in quiet conditions, which is arguably not the only factor to consider with respect to successful flight performance for a hearing-impaired aviator. Because the degree and pattern of hearing loss of a hearing-impaired aviator can vary greatly, and since performance is likely influenced by varying degrees of mental workload and increased communication loads, it was expected that performance among hearing-impaired aviators would vary as well, and possibly exhibit interaction effects with workload. This project examined flight performance as a function of hearing loss and under varying flight workload levels, with the intent of providing more specific data on which to base future aeromedical policy statements on hearing loss and subsequent flight waiver decisions.

2. To yield data on which to base flight recommendations and headsets for hearing-impaired pilots.

The experiment looked at differences in pilot performance with different communication headsets/earphones, different levels of pilot workload, and different degrees of communication challenges. The results were expected to provide audiologists and flight surgeons with more information about anticipated individual flight performance as well as recommendations for the most appropriate communications system based on individual hearing sensitivity and the pilot's mission, communication, and workload requirements.

3. Ultimately, to provide selection guidance for communication headsets, considering aviators' hearing abilities, that will improve combat effectiveness as realized through optimized performance.

Favorable aircrew performance is a force multiplier. An aviator who has the communication system that affords optimized performance in light of his or her hearing abilities and limitations, as well as workload and communication requirements, will be operating at a higher level of operational effectiveness and safety—all of which affect overall mission success.

The experimental goals of this project were:

1. To determine the level of *flight mission-induced workload* at which performance decreased with a CEP-equipped headset, ANR-equipped headset, and a passive headset/ passive earplug combination between aviators with normal hearing and aviators with varying degrees of hearing loss.
2. To determine the level of *communications signal quality* at which performance decreased with a CEP-equipped headset, ANR-equipped headset, and a passive headset/passive earplug combination between aviators with normal hearing and aviators with varying degrees of hearing loss.
3. To determine the level of *hearing loss* at which performance decreased with a CEP-equipped headset, ANR-equipped headset, and passive headset/passive earplug combination beyond acceptable levels as a function of workload and communication signal quality.

As such, formal hypotheses related to the research were as follows:

H1: As flight workload increased and communication signal quality decreased, normal-hearing aviators would perform better than hearing-impaired aviators, regardless of the communications system used.

H2: For hearing-impaired aviators, as flight workload increased and communication signal quality decreased, performance would decrease first with passive headset/passive earplug combination, then the ANR-equipped headset, followed by the CEP-equipped headset.

H3: For normal-hearing aviators, as flight workload increased and communication signal quality decreased, performance would decrease first with the passive headset/passive earplug combination, then the ANR-equipped headset, followed by the CEP-equipped headset.

H4: As flight workload increases and communication signal quality decreased, ratings of situation awareness (SA) would decrease first with the passive headset/passive earplug combination, then the ANR-equipped headset, followed by the CEP-equipped headset, regardless of hearing loss category.

METHODOLOGY

Experimental Design

The experimental design was a 3×3×3 within-subjects, randomized complete block design, displayed in Figure 7. There were three independent variables for this experiment and five groups of dependent measures. Participants' hearing status was a blocking variable in which participants were sorted by hearing loss category into two homogenous groups (blocks). The treatment effects were randomly assigned to subjects within the blocks and were considered fixed effects. The block effects (pilot participants within the groups) were considered random since inferences made from the experimental results were made to a larger set of pilots than just those who participated in the study.

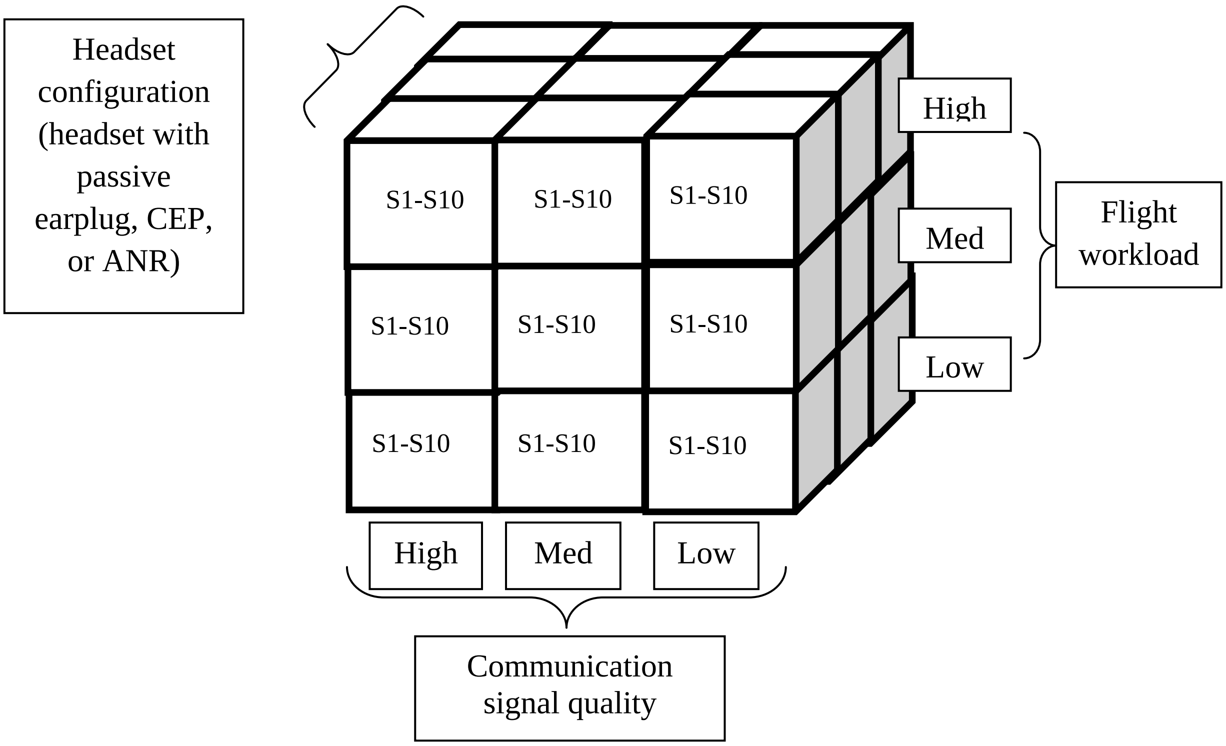


Figure 7. Experimental design for one representative homogeneous group of pilots (2 groups of 10 pilots participated in this study, with the groups differing by hearing ability as described in text).

Independent Variables

Headset Configuration

The first independent variable was the type of headset configuration - a within-subjects variable consisting of three levels. The three headset configurations used for this study were a Bose Aviation X ANR headset (B-ANR), a passive David Clark model H10-66 headset equipped with CEPs (DC-CEP), and a passive David Clark model H10-76 headset with no modifications, but with a foam earplug worn under the headset (DC-FOAM) (see Figure 8).

Headsets were chosen instead of the typical flight helmets because of the prohibitive time and expense of modifying several different sizes of flight helmets to fit all participants. Thus, a headset was a better choice toward fitting a variety of participant head sizes without extensive fit modifications. The Bose Aviation X ANR technology was chosen because it is commonly used in civilian aviation, and it has been successfully used in a similar experiment (Valimont et al., 2006). CEP devices were chosen because of the popularity of their use among Army helicopter pilots, and the passive configuration was used because a passive earplug under the issued flight helmet is the current “default” recommendation in the absence of assistive communication devices for the flight helmet.

Each headset was worn, one at a time, for a 45-minute (approximate) flight. Therefore, three flights were required for this experiment, one per headset configuration. Pilots were permitted to adjust the volume controls on the helmets and the ICS to their most comfortable listening level, as they would in actual flight.

Bose Aviation X headset (B-ANR)

David Clark model H10-66 headset with CEP (DC-CEP)

Passive David Clark headset model H10-76 (DC-FOAM)

Figure 8. Headset configurations.

Workload

The second independent variable was flight workload – a within-subjects variable that consisted of three levels (high, medium, and low), encompassing perceptual, psychomotor, and communication aspects of workload.

Perceptual workload refers to the detection, perception and interpretation of visual cues within the cockpit that provide information on actual and/or potential flight situations. In this experiment, perceptual workload was manipulated by varying the simulated visibility conditions, wherein low perceptual workload was 6 statute miles (SM) visibility, medium was 1.75 SM visibility, and high was 0 SM visibility (see Figure 9).

Low workload – 6 statute mile (SM) visibility

Medium workload – 1.75 statute mile (SM) visibility

High workload – 0 statute mile (SM) visibility

Figure 9. Visibility conditions for low, medium, and high workload flight.

Psychomotor workload for this study refers to the motor skills necessary for flight. Psychomotor workload was manipulated by increasing the number of maneuvers required in each flight segment, wherein low psychomotor workload was straight-and level flight with heading changes, medium was straight-and-level flight with heading and altitude changes, and high was straight-and-level flight with heading, altitude, and airspeed changes.

Communication workload was varied with respect to the amount of information contained in recorded radio messages presented at each workload level and corresponding to the psychomotor workload level. In a low workload flight segment, the radio message consisted of a directive comprised of one task (e.g. “turn right or left to a specified heading”). In a medium workload flight segment, the radio message consisted of a directive comprising two tasks (e.g. “turn right or left *and* climb or descend to a specified altitude”). For high workload flight segments, the radio message consisted of a directive with three tasks (e.g. “turn right or left *and* climb or descend, *and* accelerate or decelerate to a specified airspeed”).

Additionally increasing the communication workload was a radio transmission that required the pilot to read back specific flight parameters during the execution of the flight maneuver, for example, altimeter and wind speed. This readback-only task did not require that the pilot manipulate any controls; rather, simply read back the transmission. The readback-only task consisted of the same amount of information in each workload level. The workload parameters and levels described here were chosen after discussion with research pilots at the USAARL who indicated that these levels were representative of what would be experienced in actual flight. The workload/communication signal scenarios are listed in Table 5.

Table 5. Perceptual, psychomotor, and communication flight workload representation.

Workload level	Perceptual aspect / flight conditions	Psychomotor aspect / flight task	Communication aspect / sample commands
Low workload	Day Visual Meteorological Conditions (VMC) No ceiling, unlimited visibility	Straight and level flight followed by turns to various headings	One task radio command Ex. "Army 748 turn right heading 270°." Read back task during maneuver Ex. "Army 748, altimeter 29.90, cleared to land runway 06."
Medium workload	Dusk Low visibility, fog No ceiling, 1 ¾ mile visibility	Straight and level flight followed by turns to various headings combined with altitude changes	Two-task radio command Ex. "Army 748. turn right heading 290°, climb and maintain 3000'." Read-back task during maneuver Ex. "Army 748, winds 050@10, information delta current."
High workload	Night Instrument Meteorological Conditions (IMC) No ceiling, zero visibility	Straight and level flight followed by turns to various headings combined with altitude and airspeed changes	Three-task radio command Ex. "Army 748, turn right heading 270°, climb and maintain 3500' while decelerating to 100 knots." Read-back task during maneuver Ex. "Army 748, runway 36 in use, new altimeter 29.95."

Communication Signal Quality

The third independent variable was communication signal quality – a within-subjects variable consisting of three levels: good, average, and poor. The earphone-output signal quality of each headset was manipulated and quantified with the Speech Intelligibility Index (SII) (ANSI, 1997) method of predicting speech intelligibility for consistency across headsets and participants. The recorded radio messages were combined with pink noise to achieve a signal-to-noise ratio that yielded an SII level of 0.8 (good), 0.6 (average), and 0.4 (poor). The SII method was a convenient means with which to quantify the communication signal quality levels in such a way that they were consistent in presentation across all headsets and all participants.

As mentioned previously, pilots were permitted to the volume controls on the headsets and the ICS volume to their most comfortable listening level. The expected earphone-output signal-to-noise ratio based on SII value was the result of the noise mixed with the ATC commands prior to presentation to the pilot through the headset. It is important to note that even with an individual participant's volume control manipulations, the signal-to-noise ratio as stated above was the same because the pink noise was mixed with the ATC command and could not be adjusted independently. There was additional aircraft noise in the simulator, but it was not considered as part of the SII calculation.

Hearing Loss Categories

Stratification of pilot participants was by degree and configuration of hearing loss. It was originally expected that the following categories would be represented by participants in this study:

1. Audiometric pure-tone air-conduction thresholds at 1000, 2000, 3000, and 4000 Hz not exceeding 25 dBHL in either ear (normal hearing sensitivity in both ears).
2. Audiometric pure-tone air-conduction thresholds not exceeding 25 dBHL at 500, 1000, and 2000 Hz, not exceeding 35 dBHL at 3000 Hz and not exceeding 45 dBHL at 4000 (slight to mild high frequency hearing loss).
3. Audiometric pure-tone air-conduction thresholds exceeding 25 dBHL at 500, 1000, and 2000 Hz, exceeding 35 dBHL at 3000 Hz, and exceeding 45 dBHL at 4000 Hz in either ear (mild to moderate or greater high frequency hearing loss).

The first hearing sensitivity group listed above represents normal hearing sensitivity, in accordance with the American Academy of Otolaryngology-Head and Neck Surgery. The second group is the current audiometric standard for continued flight status without a waiver for a Class 1 flight physical and generally corresponds to the U.S. Army's H-1 hearing profile category, in accordance with Army Regulation 40-501 *Standards of Medical Fitness*. The third group generally corresponds to the Army's H-2/H-3 hearing profile categories.

Due to difficulty in recruiting pilots that had slight hearing loss that did not require a hearing waiver (the second category in the original stratification), the hearing loss categories were changed to generally reflect the hearing waiver criteria for a Class 2 flight physical.

1. Audiometric pure-tone air-conduction thresholds not exceeding 25 dBHL at 500, 1000, or 2000 Hz, not exceeding 35 dBHL at 3000 Hz, and not exceeding 55 dBHL at 4000 Hz in either ear
2. Audiometric pure-tone air-conduction thresholds exceeding 25 dBHL at 500, 1000, or 2000 Hz, exceeding 35 dBHL at 3000 Hz, and exceeding 55 dBHL at 4000 Hz.

For the final grouping of pilots, Group 1 had hearing thresholds that were better (lower hearing threshold levels) than the Class 2 flight standards and Group 2 met or exceeded the Class 2 flight standards as outlined in the APL.

Dependent Measures

The five groups of dependent measures that were used in this experiment are described below.

Primary Task Performance

The UH-60 Data Analysis System (HAWK) sampled and collected data on 26 flight parameters; however, the only measures used in this analysis were heading, altitude, and airspeed. These flight performance parameters were used because they had predetermined assigned values that the pilot was expected to achieve (given in ATC commands) to compare with actual values controlled by the pilot. Calculations were made in Microsoft Excel to obtain the deviation of actual heading, altitude, and airspeed relative to the assigned heading, altitude, and airspeed. Two research pilots at the USAARL reviewed and approved of these flight performance parameters.

Subjective Workload

Workload was measured subjectively via the MCH workload rating scale (shown in Appendix D) at the completion of each flight, for a total of nine MCH administrations (nine workload-signal quality combinations) throughout the course of each flight and 27 total MCH administrations across the course of the experiment. The simulator was stopped after each flight segment for form completion.

Situation Awareness

Situation awareness was measured subjectively via the SART (shown in Appendix E) with participant scores obtained at the completion of each flight segment for a total of nine SART administrations (nine workload-signal quality combinations) throughout the course of each flight, and 27 total SART administrations across the course of the experiment. The simulator was stopped after each flight segment for form completion.

Speech Intelligibility

Verbal ATC commands were realistic commands that a pilot would normally experience during flight. Because correct command execution was possible without perfect speech intelligibility due to contextual cues and pilot experience, inferring speech intelligibility by flight performance was not the preferred option for this experiment. Rather, participants were required to repeat radio commands prior to command execution as the measure of speech intelligibility. If a radio command was not heard or was not understood fully, the pilot was instructed to request a readback, and the recorded command was presented again. Pilots were not required to repeat ATC commands verbatim; rather, they were required to repeat the critical information (direction, heading, altitude, and/or airspeed) given in the command in accordance with accepted flight language, which varied between pilots (e.g. ATC command “Army 748, turn right heading 240”, pilot responded, “Army 748, right turn 240” or “Roger, 748 turning right to 240”.) If the critical information in the ATC command was repeated incorrectly, the command was repeated, and one readback was recorded.

Communications intelligibility was measured by recording the number of pilot readbacks during each flight segment. Readbacks were recorded by marking a timestamp via the data collection computer keyboard when a pilot requested a repetition of the ATC command, or if a incorrect readback initiated an ATC command repetition, or if a pilot requested a repetition of a readback-only message. A different character timestamp was used if a pilot requested an ATC command to be repeated following the readback-only message.

Headset Comfort and Speech Intelligibility

Headset comfort and speech intelligibility data were collected using a validated bipolar rating scale based on a method that has been shown to be sensitive to comfort parameters in studies incorporating variables of considerably different contexts (Casali & Grenell, 1990; Casali, Lam, & Epps, 1987). The use of the bipolar rating scale also had practical advantage to this study in that it was easy to use and required no training with only a minimal amount of instruction.

Participants

Twenty active duty, instrument-rated Army, Department of the Army Civilians (DAC) and contract helicopter pilots assigned to Fort Rucker, Alabama participated in the experiment. Pilots were recruited through permanent units located at Fort Rucker, various Army training courses, and through contacts at Lyster Army Health Clinic, also located at Fort Rucker. All pilots recruited were deemed healthy via documentation of a current “full flying duties” (FFD) flight physical, and had flown a military helicopter or military flight simulator within the past year. One volunteer participant was excluded due to active middle ear pathology and was referred for medical management. Age and experience demographics are displayed in Table 6. Mean hearing threshold levels are displayed in Table 7 and Table 8.

Table 6. Participant demographics.

	Group 1	Group 2	Groups Combined
Age range	20-51	33-66	20-66
Age median	31	52.5	38
Age average	33.4	50.2	41.8
Flight hours range	75-12,000	1100-11,000	75-11,000
Flight hours median	200	4350	1900
Flight hours average	1678.50	4770	2678.75

Table 7. Mean right ear hearing level thresholds (dBHL)

	250Hz	500Hz	1000Hz	1500Hz	2000Hz	3000Hz	4000Hz	6000Hz	8000Hz
Group 1	9.0	7.0	6.0	5.5	4.0	9.0	10.5	14.0	11.0
Group 2	13.0	14.5	19.0	27.0	29.5	41.0	51.5	51.5	54.5

Table 8. Mean left ear hearing level thresholds (dBHL)

	250Hz	500Hz	1000Hz	1500Hz	2000Hz	3000Hz	4000Hz	6000Hz	8000Hz
Group 1	8.5	7.5	7.5	6.0	6.0	12.0	17.0	15.5	15.0
Group 2	18.5	19.5	22.5	25.5	30.5	54.5	67.5	69.5	69.0

Apparatus

This experiment took place at the USAARL at Fort Rucker, Alabama. The USAARL's NUH-60FS Black Hawk simulator is routinely used to objectively measure aviator performance under controlled environmental conditions. The simulator is a six-degree-of-freedom motion-based system that includes an operational crew station, computer-generated visual display, environmental conditioning system, and a multichannel data acquisition system (Figure 10).

Figure 10. USAARL NUH-60FS Black Hawk flight simulator.

Combinations of pitch, roll, yaw, lateral, longitudinal, and vertical movement were provided by the simulator's motion system that consists of a moving platform assembly driven and supported by six hydraulic actuators, which operate synergystically. Motion was controlled to simulate the effects of pilot input as well as the effects of rotor operations, turbulence, and changes in aircraft center of gravity (McGrath, Estrada, Braithwaite, Raj, & Rupert, 2004).

The simulator compartment contains a cockpit with a pilot and copilot station (Figure 11), an instructor operator (IO) station (Figure 12), and an observer station. In the cockpit are all the controls, panels, and indicators that are found in the actual aircraft, although some controls are not functional in the simulator. The cockpit station contains forward, left, and right side window displays on which the digital image generator system provides a full color visual display that depicts day, night, and dusk scenes, as well the effects of the aircraft searchlight/landing light (McGrath et al., 2004).

Figure 11. USAARL NUH-60FS Black Hawk flight simulator cockpit.

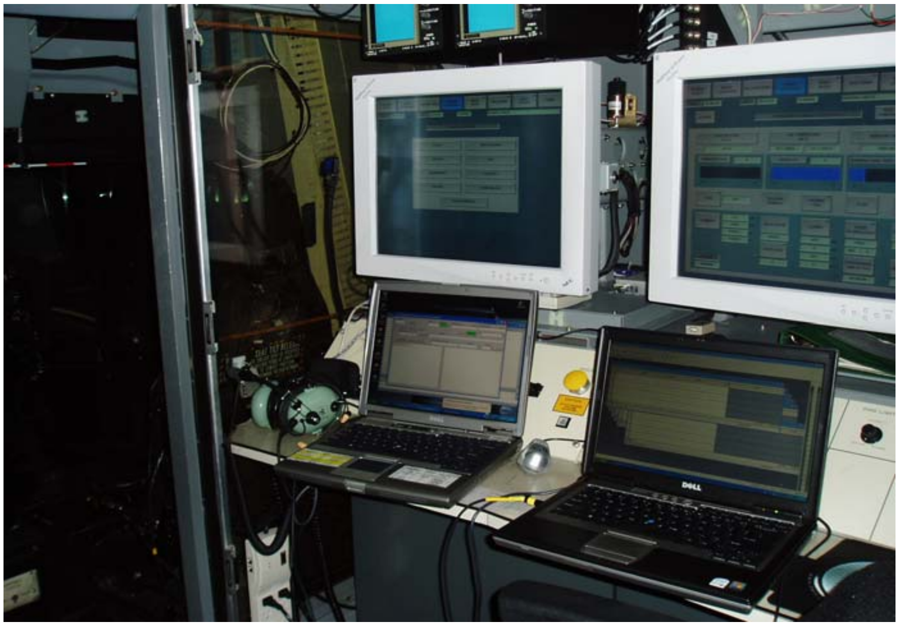


Figure 12. USAARL NUH-60FS Black Hawk flight simulator instructor operator station.

Loudspeakers in the simulator compartment provided auditory cues similar to an actual aircraft; however, the ambient noise level inside the NUH-60FS flight simulator at the USAARL as it is normally configured is lower than the ambient noise level inside an actual UH-60 helicopter. Because communication, speech intelligibility, and hearing levels were an important aspect of this study, the ambient noise level in the flight simulator was increased to more closely replicate that of the actual helicopter, as described later. This created a noise level that was potentially hazardous to participants' hearing; however, all headset configurations used in the experiment provided an adequate amount of hearing protection in accordance with OSHA and Army hearing conservation standards. Furthermore, in no case was the noise level greater than noise level found in UH-60 helicopters that the pilots normally fly.

Procedure

Simulator Noise Levels

The noise level of the USAARL's JUH-60 helicopter was measured on the A-scale, slow response of a Type I sound level meter (Brüel and Kjær Investigator 2260 Type I sound level meter running Enhanced Sound Analysis software #BZ7206) prior to the experiment. At the pilot seat location (pilot door open and co-pilot door closed) with the helicopter on the ground with the rotors turning, the noise level was 103 dBA. Correspondingly, the noise level in the NUH-60FS simulator was supplemented to 103 dBA to match the real-world engine noise. To obtain a noise level spectrum inside the simulator as close to the actual helicopter as possible, an equalizer and two amplifiers were added to the simulator computer control room and routed to larger speakers and a subwoofer inside the simulator (see Figure 13 for a schematic). Figure 14 shows the comparison of the helicopter's noise spectrum and the simulator noise spectrum after equalizer adjustments (linear measurements were converted to dBA).

The helicopter sound effects in the simulator originated from a digital recording that was linked to the cockpit controls so that appropriate sound effects (e.g. noise level and spectrum change as pilots actuated the collective control) occurred with the corresponding simulator action. The sound effects were produced by computers in the simulator computer room, located outside of the simulator area. The left (JBL Model 4312 Control Monitor SN 80120) and right (Nady Audio PM-100 Stage Monitor SN31027110320) simulator speakers were driven by two QSC Audio (PLX3402 Pro 3400 Watt SN120151715 and SN120151726) amplifiers, and an EV (EQ-231 Stereo Graphic SN013390061) equalizer. The subwoofer (JBL Control 10 SNJ804-A019897) and the front simulator speaker were driven by a Crown International D-75 Dual Channel Power Amplifier (SN035875) amplifier. Calibration of these sound levels was accomplished using a Brüel and Kjær Investigator 2260 Type I sound level meter running Enhanced Sound Analysis software #BZ7206 to approximate the 1/3 octave band levels recorded from USAARL's JUH-60 helicopter.

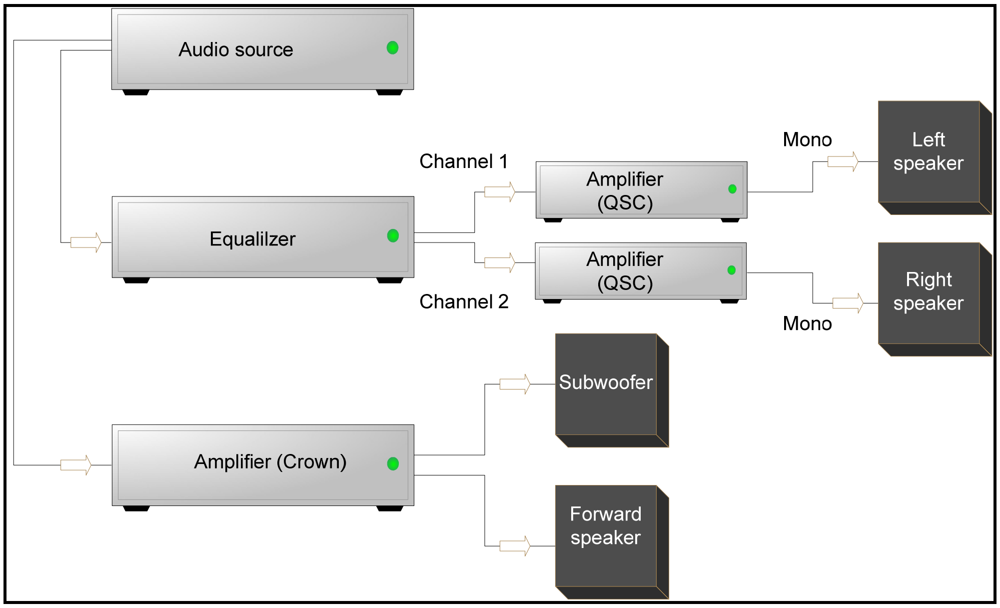


Figure 13. Audio equipment configuration.

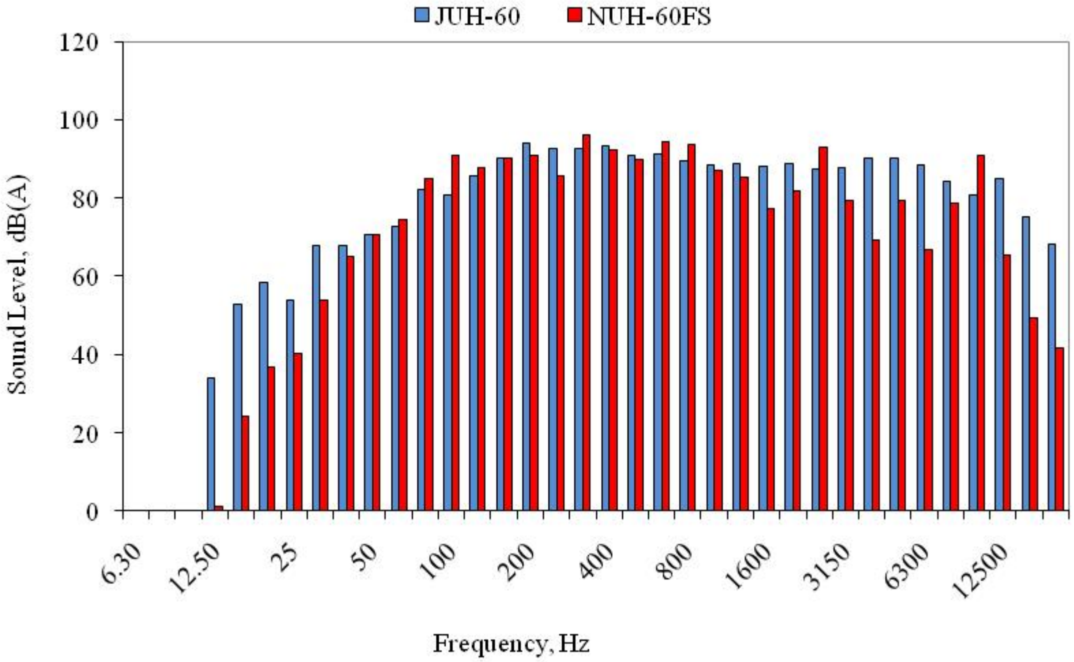


Figure 14. Noise spectrum levels of JUH-60 helicopter and NUH-60FS simulator.

ATC Command Recordings

A female speaker who is a former air traffic controller recorded the ATC commands. Recordings were made in an anechoic chamber using Sound Forge 6.0 software with a Sony electret condenser microphone placed approximately six inches from the speaker's mouth. In order to calculate the amount of pink noise to mix with the ATC commands to yield the predetermined SII levels, it was first necessary to convert the recorded ATC .wav files to 1/3 octave band data using MATLAB. The MATLAB code produced 26 1/3 octave bands, however, calculations to determine SII only required 18 bands, beginning at 160Hz, so the first 8 bands (25, 31.5, 40, 50, 63, 80, 100, and 125Hz) from the 1/3 octave band calculations were not used for SII computation.

Trial and error was used to calculate the noise levels necessary to mix with the 1/3 octave band levels to achieve the predetermined SII levels of 0.8, 0.6, and 0.4 using a program downloaded from http://www.sii.to ("SII: Speech Intelligibility Index,"). After determining the levels of pink noise necessary to achieve the desired SII levels (0.8 for good signal quality, 0.6 for average signal quality, and 0.4 for poor signal quality), the pink noise intensity level was manipulated in Sound Forge 6.0 and verified with the 1/3 octave band code in MATLAB. After verification, the appropriate pink noise level was mixed in Sound Forge 6.0 with each ATC command.

Recorded ATC commands were presented to pilots through the three headsets using a Dell Latitude D620 with Intel Centrino Duo running Windows XP (SN HF974A02) computer connected to the ICS jack at the observer's station in the simulator. The SII levels used in the experiment were chosen after input from an experienced research pilot at the USAARL judged the levels consistent with good, average, and poor signal quality in terms of speech intelligibility.

Introductory Session

Participants were expected to participate in four sessions: an introductory session and three flight sessions. Most participants completed all four sessions in one day; however, four of them completed the sessions in two different days.

Prospective participants reported to the front entrance of USAARL and were met by the experimenter who escorted them to the Acoustics Branch wherein the purpose of the experiment was explained. Before conducting any portion of the experiment, including the audiometric evaluation, all participants were required to read and sign an informed consent, which was approved by the U.S. Army Medical Research and Materiel Command's (USAMRMC) Human Subjects Review Board. Participants were given all the time that they needed to review the informed consent, and any questions were answered before the forms were signed. Pilots were queried to ascertain their interest in participating. If they decided not to participate, they would have been released without penalty; however, none chose not to participate.

Pilots were shown and were allowed to handle the headsets to be used in the experiment. Their record of flight hours and most recent flight physical were reviewed to ensure that participants were instrument-rated and had flown a military helicopter in the past year. The most current flight physical was reviewed only to determine if the examining flight surgeon recommended full flying duties (FFD). All pilots were instrument-rated and had flown a military helicopter or military helicopter simulator in the past year and had a current FFD flight physical. No pilot was excluded from participation for those reasons.

Pilots who met the flight requirements were then subjected to an audiometric qualification procedure, which consisted of 1) a brief audiometric case history (to gather information about noise exposure, ear disease, hearing aid use, and known hearing loss), 2) an examination of the outer ear canal and eardrum using a lighted otoscope (to rule out excessive ear wax [cerumen: more than 50% blockage] and/or a perforated eardrum, or any other visible outer ear abnormality), 3) a pure-tone audiogram to ascertain hearing threshold levels in each ear at the audiometric test frequencies of 250, 500, 1000, 1500, 2000, 3000, 4000, and 6000 Hz, and 4) a tympanogram to rule out middle ear abnormalities (such as fluid, retracted ear drum, or abnormalities of the bones in the middle ear). If any active ear disease was present or suspected, the pilot was referred to a specialist and was excluded from the experiment. One pilot was excluded from participation for this reason. If any pilot's ear canal had been more than 50% blocked with cerumen, the pilot would have been referred to a specialist to have the cerumen removed prior to continuing with the experiment. No pilots were referred for cerumen removal.

Headset Familiarization and Experiment Instruction

Each pilot was familiarized individually by the experimenter on the proper fitting and use of each headset configuration (ANR-equipped headset, CEP-equipped headset, and passive headset/passive earplug combination). Fitting instructions were presented verbally and if requested by the pilot, a demonstration of proper fitting was performed by the experimenter. Emphasis of familiarization was placed on design aspects of the headset configuration, how to properly fit the headset to obtain a noise-blocking seal, and how to obtain a comfortable fit. All pilots except one had experience wearing CEPs, which, due to its design, was the headset most difficult to fit properly.

Once the fitting and use was completed satisfactorily, pilots were briefed on the following flight rules for the experiment:

1. Pilot immediately reads back all ATC commands to verify that the directive is understood.
2. Pilot immediately complies with ATC commands by beginning maneuver upon receiving and understanding the message.
3. Pilot maintains last assigned airspeed, heading, or altitude unless directed to change by ATC.
4. Pilot will request ATC to “say-again” if he/she does not fully understand each ATC command.
5. When clearly understood, the pilot acknowledges ATC information in accepted ATC phraseology (e.g. ATC command: “Army 748, winds 360@5, altimeter 29.92.” Pilot response “Roger, winds 360@5, altimeter 29.92”).
6. The pilot does not actually physically change the altimeter setting, but instead only reads back the ATC command.
7. The pilot will make all turns at half-standard rate (black over white on the turn needle).
8. All climbs and descents will be at 500 feet per minute.
9. Pilot will attempt to fly exactly the correct heading, airspeed, and altitude; common standards do not apply.

Pilots were given a safety briefing. Because the simulator was on full motion-base operation, they were reminded to keep their seat belt harness fastened throughout each flight and to inform the investigator of any signs of simulator-induced sickness (no pilots reported any symptoms). Prior to entering the simulator, pilots were briefed on the location of the emergency exit ladder that would be used in the event that the exit ramp could not be lowered. Inside the simulator, they were familiarized with the simulator's emergency stop and freeze functions that were to be used in the event that the investigator was unable to stop or freeze the simulator.

Simulator Familiarization

This experiment was open to all instrument-rated military aviators, rather than only UH-60 rated aviators. The rationale for this decision was that recent experiments at the USAARL had heavily recruited UH-60 pilots, thereby reducing the available pool of participants for this experiment. There is not an Army or USAARL requirement to be a UH-60 pilot to fly the NUH-60FS simulator, and because this experiment was open to instrument-rated pilots that were not all UH-60 rated aviators, a familiarization session was available to those pilots who felt that they needed it. Research pilots at USAARL indicated that the major cockpit controls necessary for flying do not differ significantly across military helicopters, but there are some minor cockpit control differences that were addressed during the familiarization period, so this was not considered a confounding factor in this experiment.

All pilots were offered time to familiarize themselves with the cockpit controls in the simulator. The simulator was put on motion and a daylight flight condition was set on the simulator. Pilots were permitted to fly the simulator as long as they needed to – the experiment did not begin until the pilot verbalized to the experimenter that he was ready to begin. Approximately one-quarter of the pilots used the familiarization time, and all of those who did said that they were comfortable with the controls after 15-20 minutes.

Simulator Flight Missions

Each pilot flew three flights (one with each headset configuration). Each flight consisted of the identical representation of each workload level and communication signal quality levels (nine segments representing the nine workload/signal quality combinations, as shown in Table 9); however the segments were counterbalanced across participants via a balanced Latin square design as described in USAARL Report 99-14 to control for order effects (Wildzunas, 1999).

As mentioned, there were three workload levels representing three particular types of workload - perceptual, psychomotor, and communication demand. Perceptual workload was manipulated by changing the visibility conditions in the simulated flight, psychomotor workload was manipulated through the number of maneuvers required for the flight segment, and communication workload was addressed by changing the amount of information given via radio communication.

Also as mentioned, there were three levels of communication signal quality (good, average, and poor) throughout the mission that were manipulated by mixing the recorded ATC commands with pink noise in accordance with SII guidelines to degrade the expected speech intelligibility. Each level of communication signal quality was presented with each workload level, and affected all radio communications within that particular flight segment. There were three maneuvers (and three readback messages) in each flight segment – meaning that there were 27 maneuvers in each flight – or 81 total maneuvers.

Table 9. Flight segments (counterbalanced for experiment).

Flight 1: ANR headset condition	Flight 2: CEP headset condition	Flight 3: Foam plug headset condition
1) Low workload (WL) – good signal quality (SQ)	1) Low WL – good SQ	1) Low WL – good SQ
2) Low WL – average SQ	2) Low WL – average SQ	2) Low WL – average SQ
3) Low WL – poor SQ	3) Low WL – poor SQ	3) Low WL – poor SQ
4) Medium WL – good SQ	4) Medium WL – good SQ	4) Medium WL – good SQ
5) Medium WL – average SQ	5) Medium WL – average SQ	5) Medium WL – average SQ
6) Medium WL – poor SQ	6) Medium WL – poor SQ	6) Medium WL – poor SQ
7) High WL – good SQ	7) High WL – good SQ	7) High WL – good SQ
8) High WL – average SQ	8) High WL – average SQ	8) High WL – average SQ
9) High WL – poor SQ	9) High WL – poor SQ	9) High WL – poor SQ

DATA ANALYSIS OVERVIEW

Analysis Approach by Dependent Measure Category

Flight control performance mean data (heading deviation, altitude deviation, and airspeed deviation) were analyzed individually with a repeated measures mixed effects analysis of variance (ANOVA) model. All data were tested for significance at an alpha level of 0.05, chosen to determine whether any observed differences in the means were due to chance, or were due to systematic manipulations of the experimental variables. Post-hoc comparisons were conducted on any significant results using Tukey's multiple comparison test. *Workload data* collected with the MCH scale were analyzed with a nonparametric Friedman's *F*-test since the data represented non-parametric indexes. Applicable post-hoc testing was conducted using the associated Rank Sum multiple comparison test described by Conover (1999). *Situation awareness data* were analyzed with a three-way mixed effects ANOVA model. Post-hoc comparisons of the SA data were conducted using Tukey's multiple comparison procedure. *Speech intelligibility data*, as inferred by the number of command readbacks, were analyzed with a repeated measures mixed effect ANOVA model. Post-hoc comparisons were conducted on significant ANOVA results using Tukey's multiple comparisons procedure. *Subjective headset comfort/speech intelligibility rating scale data* were converted to a numerical score ranging from one to six and analyzed using a repeated measures ANOVA procedure for the three headset levels. Post-hoc comparisons on ANOVA main effect variables were conducted with Tukey's multiple comparisons procedure.

The experimental data have been grouped for presentation herein by objective and subjective measures. The results of the flight performance parameters (heading, altitude, and airspeed deviation) and the communication intelligibility (ATC readbacks) will be presented first. The subjective measures (MCH, SART, and headset comfort/speech intelligibility rating scale) analysis and results will follow.

Fixed- vs. Random-Effects Variables and Analysis Approach

Recall that this experimental design was a randomized complete block design (see Figure 7) in which hearing waiver category was the blocking variable, and all treatments were randomly assigned to subjects within the blocks. In this design, the treatment effects were considered fixed because the particular independent variables used in this experiment were the only ones to which inference was made. The block effects (pilot participants within the groups) were considered random because inference was made to a larger set of pilots than those that participated in the study. In order to estimate and compare treatment means with precision and levels of statistical significance that were valid to a larger set of pilots, the statistical methods needed to accommodate the random effects as well as the fixed effects of the treatments. Therefore, a mixed effects model using Statistical Analysis Software's (SAS) MIXED procedure was used for the data analysis.

SAS's more commonly used generalized linear model (GLM) procedure assumes that all model effects are fixed and that the only source of variation is error and therefore all standard errors are some function of only error variance. In making this assumption, the GLM analysis can compute the wrong standard error. The MIXED procedure; however, also takes into consideration the variance of the blocking factor and uses a restricted maximum likelihood (REML) method to show estimates of the variance component parameters. In the case of a balanced data set, the REML estimates are the same as estimates shown in the more traditional GLM ANOVA output (Littell, Milliken, Stroup, & Wolfinger, 1996).

Another difference in these procedures is that the MIXED procedure calculates the *F*-statistics with a general Wald-type quadratic form (see equation below), while GLM uses the ratios of mean squares (Wolfinger & Chang, 1998).

$$F = \frac{\hat{\beta}' L[L'(X'\hat{V}^{-1}X)^{-}L]^{-1}L'\hat{\beta}}{rank(L)}$$

These values are the same in the case of a balanced design, as this one was. The advantage of the MIXED method is that different types of covariance structures can be used; however, a disadvantage is that the denominator degrees of freedom are assigned using other methods. In this analysis, a "containment method" was used, which obtains degrees of freedom by partitioning the residual degrees of freedom into between and within-subjects parts and assigning them to the corresponding between and within-subject effects (Wolfinger & Chang, 1998).

Assumptions of ANOVA include residuals that are normally distributed, independent, and with constant variance across groups. The second two assumptions are often not met, particularly when data come from subjects that are somehow correlated to each other within blocks, such as in this study. The advantage of using the MIXED procedure was that it could model data with non-constant variances across groups, it automatically used the correct error terms' values, and it allowed more flexibility to model the variance/covariance matrix (the within-subject correlations) in a manner that minimized concern about the lack of sphericity condition (High, 2001). The latter advantage obviates the necessity of manually adjusting the degrees of freedom to protect against heterogeneity of covariance for repeated-measures variables, such as with a Greenhouse-Geisser or Huyhn-Feldt correction (Milliken & Johnson, 1984).

For these reasons, the ANOVA summary tables herein appear different from what is typically displayed with the GLM or other more traditional procedures. The sums of squares and mean squares values are not part of the standard SAS output for the MIXED procedure; however, a statement was added to the model to output the mean squares. The degrees of freedom also appear different from what one would expect with a traditional generalized linear model, and there is not a corresponding error degree of freedom per the explanation above. On the following ANOVA summary tables, the mean square value for each variable and the mean square value for the error were listed for the reader's reference, and these can be used to calculate *F*-ratios that correspond to those of the MIXED SAS model output because this model is a balanced design.

RESULTS AND DISCUSSION

Flight Control Performance

Heading Deviation

Heading deviation data points were calculated by subtracting the absolute value of the actual heading flown from the ATC-assigned heading. Visual inspection of the univariate normality plots of the residuals revealed a non-normal distribution. A log transform was performed which resulted in a normal probability plot of the residuals that closely conformed to the normality values calculated by SAS. The ANOVA summary table for mean heading deviation is shown in Table 10.

Table 10. ANOVA summary table for mean heading deviation.

Source	df	MS	*F*	*p*	
Between					
Group (G)	1	0.512	0.07	0.7896	
Error		6.974			
Within					
Headset (H)	2	1.486	0.73	0.4829	
Error		2.038			
Workload (W)	2	11.914	5.85	0.0031	*
Error		2.038			
Signal Quality (SQ)	2	2.902	1.42	0.2417	
Error		2.038			
H × W	4	3.522	1.73	0.1424	
Error		2.038			
W × SQ	4	2.023	0.99	0.4111	
Error		2.038			
H × SQ	4	4.219	2.07	0.0836	
Error		2.038			
G × H	2	0.617	0.30	0.7391	
Error		2.038			
G × W	2	0.498	0.24	0.7833	
Error		2.038			
G × SQ	2	1.407	0.69	0.5019	
Error		2.038			
G × H × W	4	0.757	0.37	0.8288	
Error		2.038			
G × W × SQ	4	4.312	2.12	0.0777	
Error		2.038			
G × H × SQ	4	0.815	0.40	0.8088	
Error		2.038			

*indicates significant result ($p < 0.05$)

Main Effects

The mixed effects ANOVA for mean heading deviation showed a significant main effect of workload ($F = 5.85$, $p = 0.003$). Mean absolute heading deviations were 2.1, 3.2, and 2.4 degrees for low, medium and high workload conditions, respectively, and post-hoc Tukey's multiple comparison test showed that heading deviation was significantly lower in the low workload flight conditions compared to the medium and high workload conditions, which did not differ with respect to this metric. Pilots experienced greater difficulty in maintaining their heading as the workload condition increased (see Figure 15).

Interaction Effects

Statistical analysis showed no significant interaction effects of variables on pilots' mean absolute heading deviation.

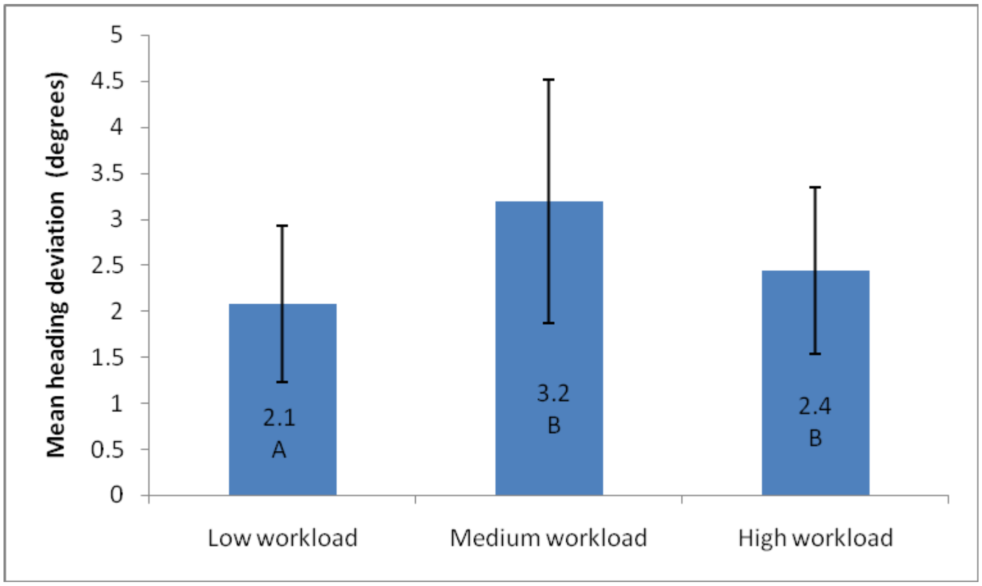


Figure 15. The effect of workload level on pilots' mean heading deviation. Different letters represent significant differences due to workload at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

Altitude Deviation

Mean sea level (MSL) altitude deviation data were calculated by subtracting the absolute value of the actual altitude flown from the ATC-assigned altitude. Visual inspection of the univariate normality plots of the residuals revealed a non-normal distribution. A log transform was performed which resulted in a normal probability plot of the residuals that closely conformed to the normality values calculated by SAS. The ANOVA summary table for mean altitude deviation is shown in Table 11.

Table 11. ANOVA summary table for mean altitude deviation

Source	df	MS	*F*	*p*	
Between					
Group (G)	1	0.330	0.05	0.8179	
Error		6.050			
Within					
Headset (H)	2	0.749	0.40	0.6680	
Error		1.854			
Workload (W)	2	19.777	10.67	< 0.0001	*
Error		1.854			
Signal Quality (SQ)	2	10.336	5.57	0.0040	*
Error		1.854			
H × W	4	3.584	1.93	0.1038	
Error		1.854			
W × SQ	4	5.189	2.80	0.0256	*
Error		1.854			
H × SQ	4	1.415	0.76	0.5495	
Error		1.854			
G × H	2	0.032	0.02	0.9829	
Error		1.854			
G × W	2	1.976	1.07	0.3452	
Error		1.854			
G × SQ	2	1.571	0.85	0.4291	
Error		1.854			
G × H × W	4	0.207	0.11	0.9785	
Error		1.854			
G × W × SQ	4	1.021	0.55	0.6987	
Error		1.854			
G × H × SQ	4	3.217	1.74	0.1410	
Error		1.854			

*indicates significant result ($p < 0.05$)

Main Effects

The ANOVA revealed significant main effects of workload ($F = 10.67$, $p < 0.0001$) on mean absolute altitude deviation. The mean altitude deviation for low, medium, and high workload conditions was 23.1, 50.3, and 63.6 feet, respectively.

Post-hoc analysis of the mean altitude deviation for each of the three workload conditions using Tukey's multiple comparison test indicated a significant difference in altitude performance between the medium and high workload conditions compared to the low workload condition. Pilots experienced greater difficulty in maintaining their altitude in the higher workload conditions. The data suggest that workload can degrade altitude maintenance by as much as 27 feet, which for a helicopter conducting terrain flight, is significant from a mission performance standpoint (Figure 16).

ANOVA also revealed a significant main effect of signal quality ($F = 5.57$, $p = 0.0040$) on the mean altitude deviation during each flight across all headset configurations. The mean altitude deviations in good, average, and poor signal quality conditions were 35.9, 29.7, and 71.5 feet, respectively. Post-hoc analysis using Tukey's multiple comparison test indicated a significant difference in altitude performance between the good and average signal quality conditions compared to the poor signal quality condition. Pilots experienced difficulty maintaining their altitude when communication signal quality was poor. The data suggest that signal quality can degrade altitude maintenance by as much as 40 feet, with the same ramifications as mentioned earlier (Figure 17).

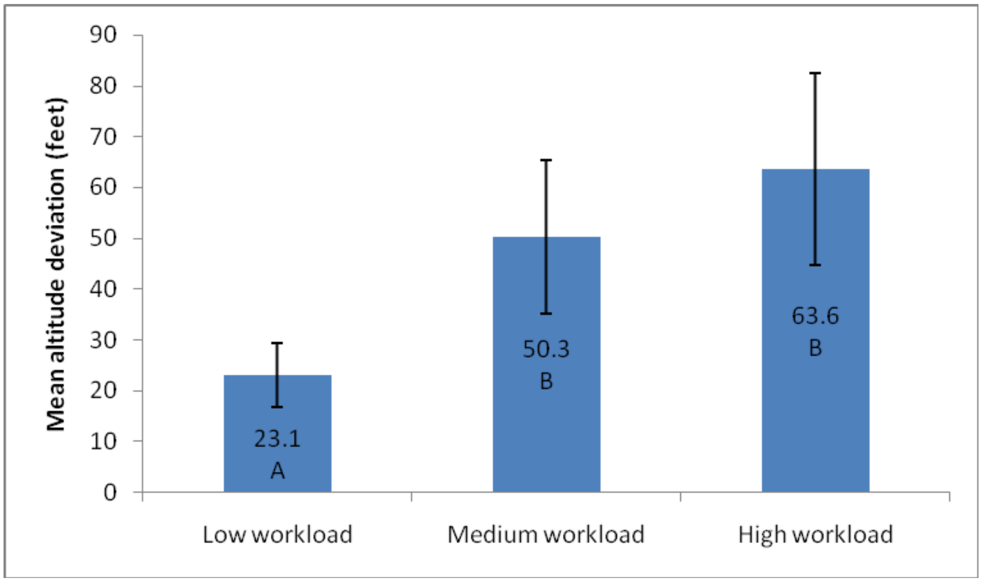


Figure 16. The effect of workload level on pilots' mean altitude deviation. Different letters represent significant differences due to workload at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

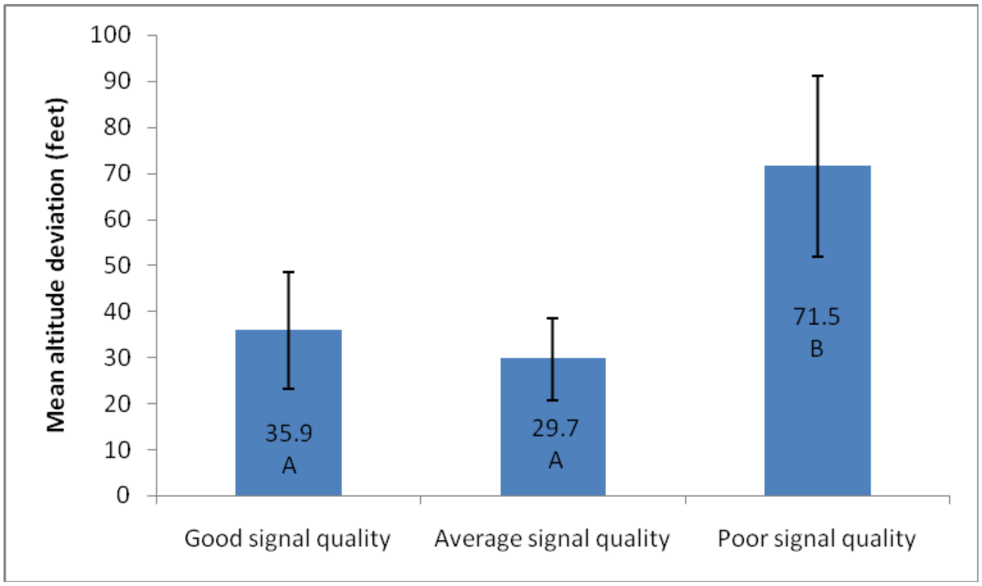


Figure 17. The effect of signal quality on pilots' means altitude deviation. Different letters represent significant differences due to signal quality at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

Interaction Effects

ANOVA indicated a significant interaction between workload and signal quality (F = 2.80, p = 0.0256) on mean altitude deviation, graphically demonstrated in Figure 18 and Figure 19. Post-hoc comparisons were first conducted by isolating the signal quality variable and performing Tukey's pairwise comparison test for workload on the particular signal quality level (Figure 18). This comparison showed that there was a significant difference in the mean altitude deviation in the poor signal quality condition when the workload was high compared to the low and medium workload conditions, with significant differences indicated between low and medium workload levels in the good and average signal quality conditions. Both groups had difficulty maintaining altitude as workload increased in all signal quality conditions.

The signal quality variable was then analyzed as a function of workload in a similar fashion – the workload level was isolated and the effect of signal quality was evaluated with Tukey's pairwise comparison test (Figure 19). This analysis showed that there was a significant difference in the mean altitude deviation in the high workload condition when the signal quality was poor compared to the good and average signal quality conditions; however, there were no significant differences in the mean altitude deviation between signal quality levels when the workload was low or medium. Both groups had difficulty maintaining altitude as signal quality decreased in high workload conditions. Mean altitude deviation values for all workload and signal quality level combinations and 95% confidence interval values are shown in Table 12.

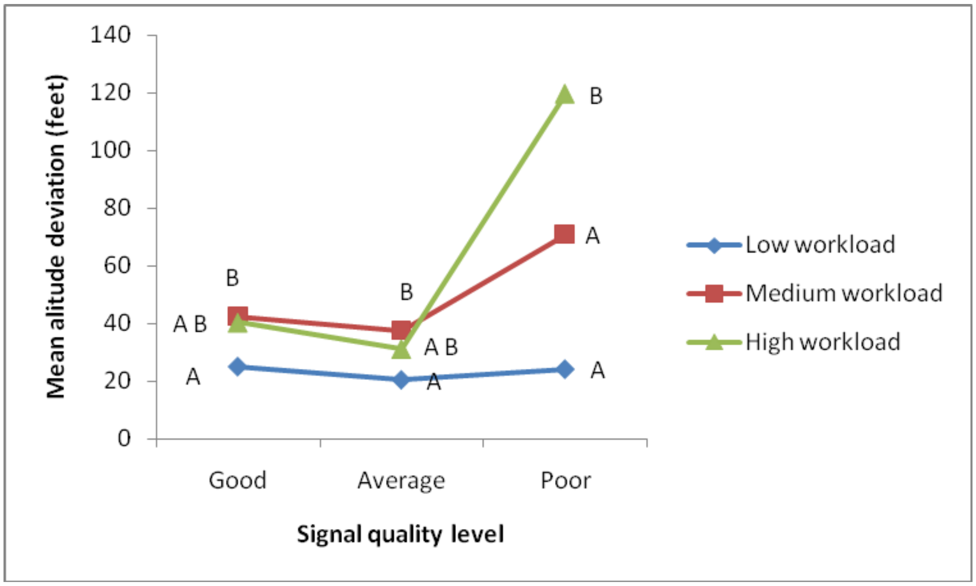


Figure 18. The effect of workload on pilots' mean altitude deviation for all signal quality conditions. For each signal quality level, different letters represent significant differences due to workload at $p < 0.05$ using Tukey's pairwise comparison test.

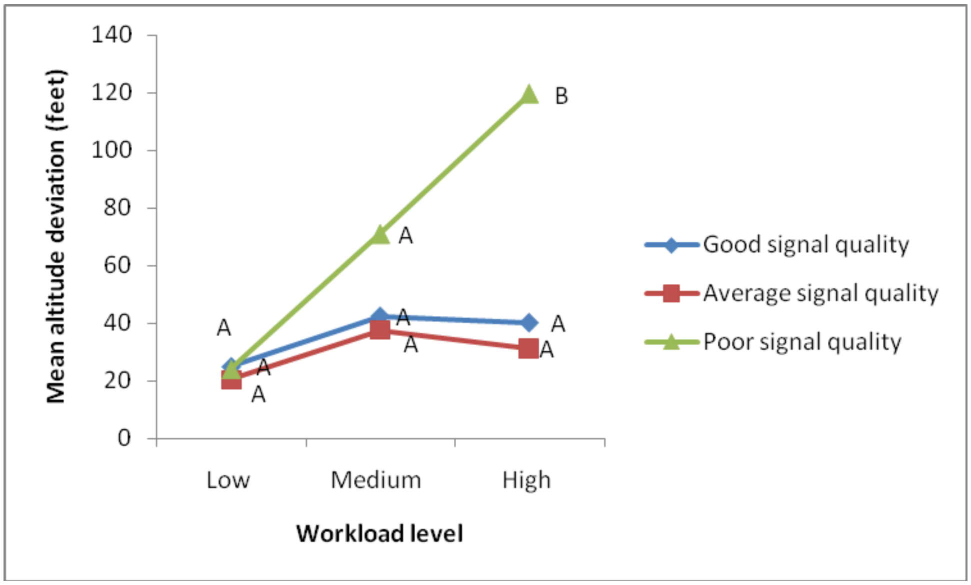


Figure 19. The effect of signal quality on pilots' mean altitude deviation for all workload conditions. For each workload level, different letters represent significant differences due to signal quality at $p < 0.05$ using Tukey's pairwise comparison test.

Table 12. Mean altitude deviation and 95% confidence interval values for workload and signal quality levels.

Workload level	Signal quality level	Mean altitude deviation (feet)	Lower 95% confidence interval	Upper 95% confidence interval
Low	Good	24.9	8.7	41.1
Low	Average	20.3	14.1	26.6
Low	Poor	24.0	16.0	32.1
Medium	Good	42.4	15.7	69.2
Medium	Average	37.6	15.3	59.8
Medium	Poor	70.9	41.4	100.4
High	Good	40.2	17.8	62.7
High	Average	31.1	17.1	45.0
High	Poor	119.5	71.2	167.8

Airspeed Deviation

Airspeed deviation data were calculated by subtracting the absolute value of the actual airspeed flown from the ATC-assigned airspeed. Visual inspection of the univariate normality plots of the residuals revealed a non-normal distribution. A log transform was performed which resulted in a normal probability plot of the residuals that closely conformed to the normality values calculated by SAS. The ANOVA summary table for airspeed deviation is shown in Table 13.

Table 13. ANOVA summary table for mean airspeed deviation

Source	df	MS	*F*	*p*	
Between					
Group (G)	1	3.649	0.57	0.4615	
Error		6.403			
Within					
Headset (H)	2	0.485	0.34	0.7099	
Error		1.412			
Workload (W)	2	44.449	31.47	<.0001	*
Error		1.412			
Signal Quality (SQ)	2	1.426	1.02	0.3610	
Error		1.412			
H × W	4	1.120	0.79	0.5338	
Error		1.412			
W × SQ	4	5.509	3.91	0.0039	*
Error		1.412			
H × SQ	4	0.392	0.28	0.8918	
Error		1.412			
G × H	2	4.591	3.27	0.0390	*
Error		1.412			
G × W	2	1.616	1.14	0.3214	
Error		1.412			
G × SQ	2	1.329	0.95	0.3875	
Error		1.412			
G × H × W	4	1.098	0.78	0.5357	
Error		1.412			
G × W × SQ	4	1.120	0.79	0.5305	
Error		1.412			
G × H × SQ	4	1.572	1.12	0.3465	

*indicates significant result ($p < 0.05$)

Main Effects

The ANOVA revealed a significant main effect of workload ($F = 31.47$, $p < 0.0001$) on mean absolute airspeed deviation. The mean airspeed deviation for low, medium, and high workload conditions was 1.2, 2.1, and 2.5 knots, respectively. Post-hoc analysis of the mean airspeed deviation for each of the three workload conditions using Tukey's pairwise comparison test showed a significant difference between all three workload conditions (low vs. medium $p = 0.0135$, low vs. high $p < 0.0001$, medium vs. high $p < 0.0001$), shown in Figure 20. Both groups had increasingly more difficulty maintaining airspeed as the workload level increased. Again, this could have implications for low level and contour terrain flight, in which constant airspeed is important, and nap-of-the-earth (NOE) terrain flight, in which terrain features dictate variable airspeed. For all terrain flight, attention to airspeed is critical to mission success.

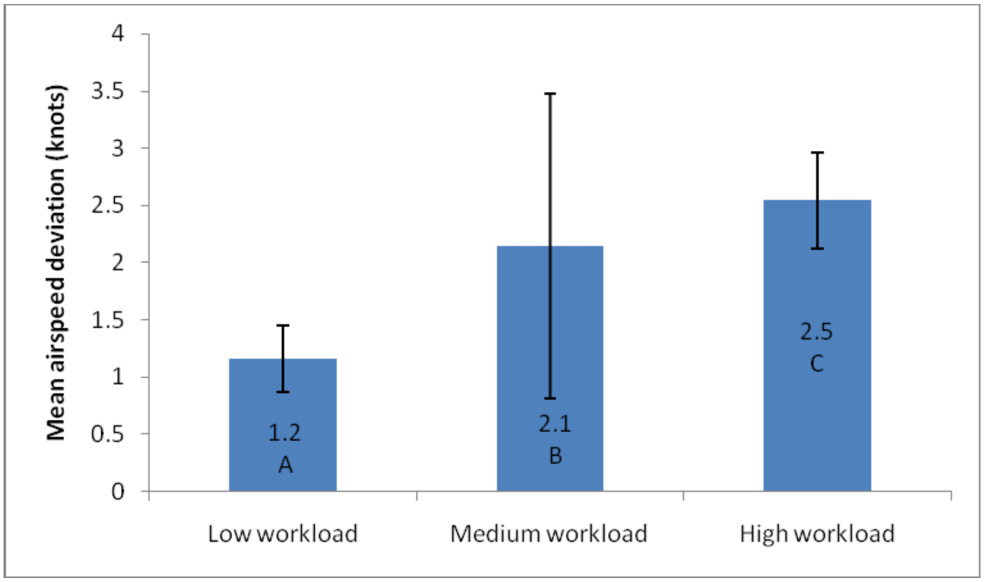


Figure 20. The effect of workload level on pilots' mean airspeed deviation. Different letters represent significant differences due to workload at $p < 0.05$. Vertical bars indicate 95% confidence intervals about the means.

Interaction Effects

ANOVA indicated a significant interaction effect between workload and signal quality ($F = 3.91$, $p = 0.0039$) on mean airspeed deviation, graphically demonstrated in Figure 21 and Figure 22. Post-hoc comparisons were conducted by isolating the signal quality variable and performing Tukey's pairwise comparison test for workload on the particular signal quality level (Figure 21). This comparison showed that there were significant differences in workload level on the mean airspeed deviation in all signal quality conditions. In the good signal quality condition, low workload resulted in significantly lower mean airspeed deviation than in the medium and high workload scenarios. In the average signal quality condition, all three workload levels contributed to significantly different mean airspeed deviations. In the poor signal quality condition, low and medium workload levels resulted in significantly different mean airspeed deviations than the high workload level did. Regardless of the signal quality level, pilots demonstrated differences in airspeed maintenance as the workload increased.

The signal quality variable was then analyzed as a function of workload in a similar fashion – the workload level was isolated and the effect of signal quality was evaluated with

Tukey's multiple comparison test (Figure 22). This analysis showed that there were significant differences in the mean airspeed deviation for the low and medium workload levels only. In the low workload condition, poor signal quality influenced significantly higher airspeed deviations than the good and average signal quality conditions did. Similarly, in the medium workload conditions, poor signal quality resulted in significantly higher airspeed deviations than did the good and average signal quality levels. There were no significant differences in mean airspeed deviation for different signal quality levels in the high workload condition. Mean airspeed deviation values for all workload and signal quality level combinations and 95% confidence interval values are listed in Table 14.

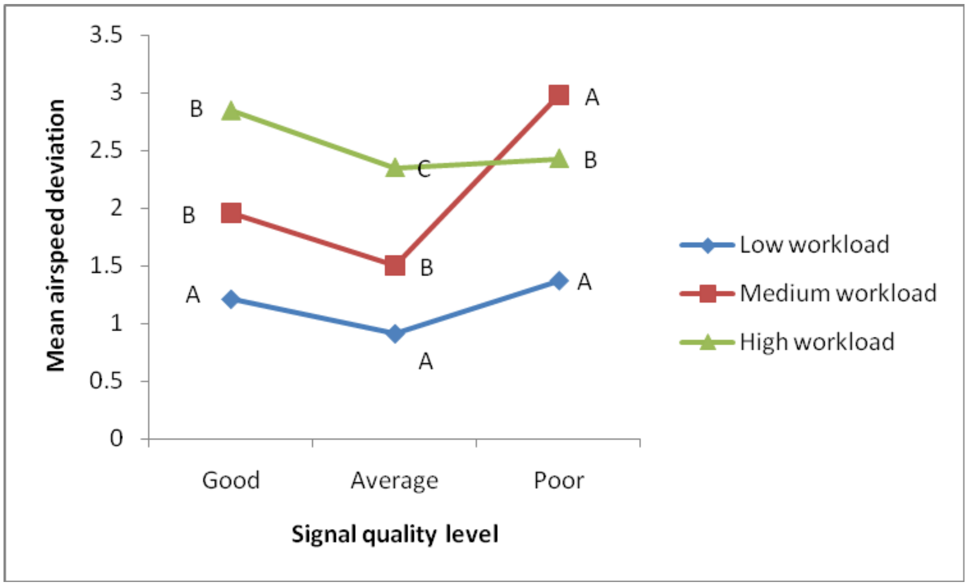


Figure 21. The effect of workload on pilots' mean airspeed deviation for all signal quality conditions. For each signal quality level, different letters represent significant differences due to workload at $p < 0.05$ using Tukey's multiple comparison test.

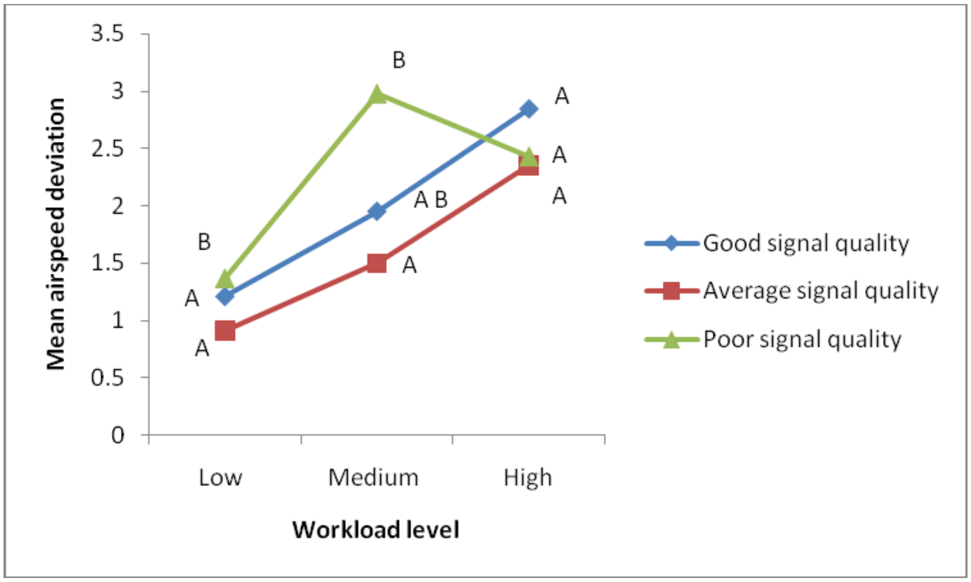


Figure 22. The effect of signal quality on pilots' mean airspeed deviation for all signal quality conditions. For each workload level, different letters represent significant differences due to signal quality at $p < 0.05$ using Tukey's multiple comparison test.

Table 14. Mean airspeed deviation and 95% confidence interval values for workload and signal quality levels**.**

Workload level	Signal quality level	Mean airspeed deviation (feet)	Lower 95% confidence interval	Upper 95% confidence interval
Low	Good	1.2	0.6	1.9
Low	Average	0.9	0.6	1.2
Low	Poor	1.3	0.9	1.9
Medium	Good	1.9	1.3	2.6
Medium	Average	1.5	1.0	2.0
Medium	Poor	2.9	-1.0	7.0
High	Good	2.8	2.0	3.7
High	Average	2.4	1.7	3.0
High	Poor	2.4	1.7	3.2

Lastly, for airspeed, there was a significant interaction found between headset configuration and group, depicted in Figure 23 and Figure 24. However, when the headset variable was isolated, there were no main effects for group discovered. There was also not a main effect seen when the group was isolated. Although the point of significance was not revealed in the breakdown of the interaction at 0.05, further investigation of the data show that the effect of the DC-FOAM headset was seen at $p = 0.1593$, which suggests that Group 2 pilots had difficulty maintaining airspeed when wearing the DC-FOAM headset configuration. Mean airspeed deviation values for all headset configurations and group combinations and 95% confidence interval values are listed in Table 15.

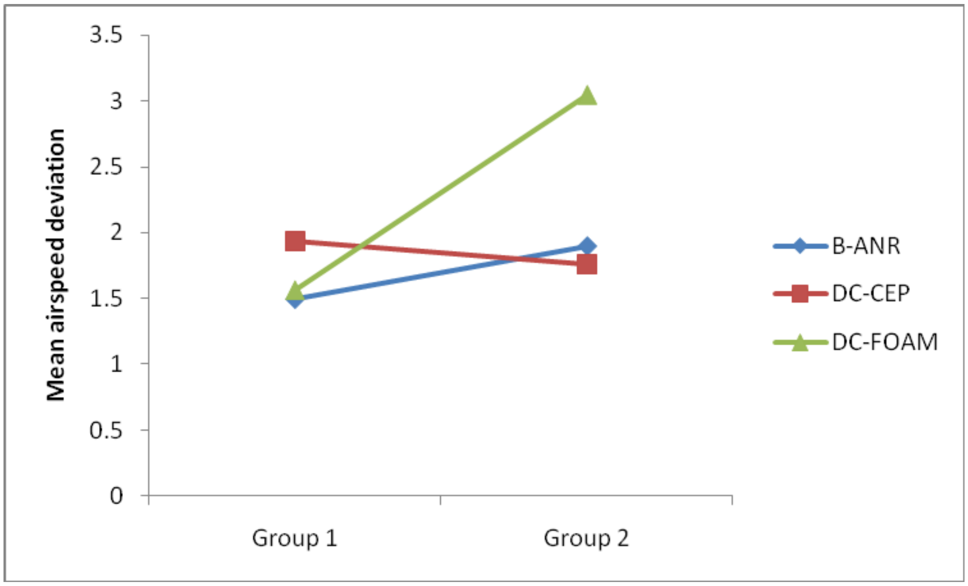


Figure 23. The effect of headset on pilots' mean airspeed deviation for both pilot groups. For each group, the absence of letters represents no significant differences due to the headset at $p < 0.05$ using Tukey's multiple comparison test.

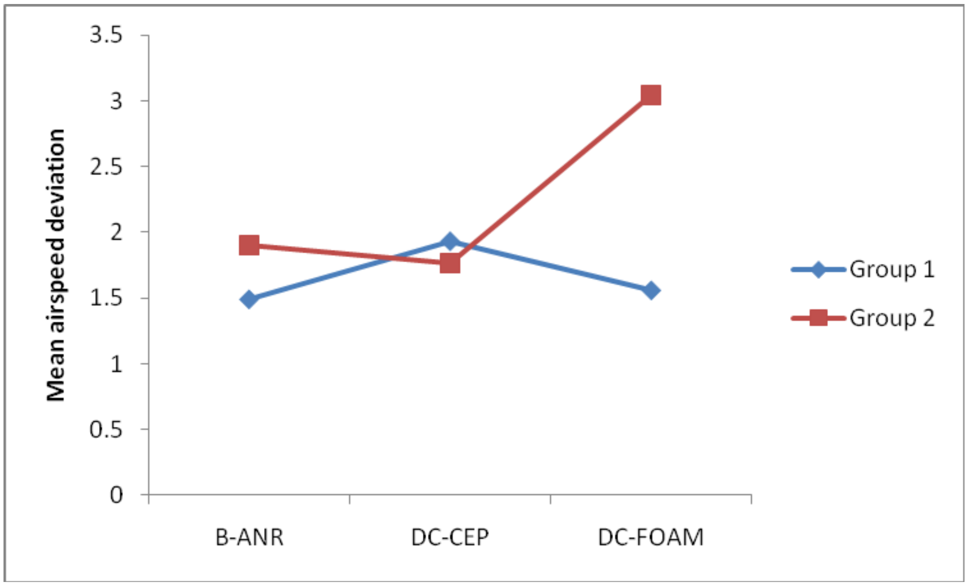


Figure 24. The effect of group on pilots' mean airspeed deviation for all headset levels. For each headset, the absence of letters represents no significant differences due to the group at $p < 0.05$ using Tukey's multiple comparison test.

Table 15. Mean airspeed deviation and 95% confidence interval values for group and headset.

Headset	Group	Mean airspeed deviation (feet)	Lower 95% confidence interval	Upper 95% confidence interval
B-ANR	1	1.5	1.1	1.9
DC-CEP	1	1.9	1.3	2.6
DC-FOAM	1	1.6	1.1	2.0
B-ANR	2	1.9	1.4	2.4
DC-CEP	2	1.8	1.2	2.3
DC-FOAM	2	3.0	0.4	5.7

Flight Performance Discussion

For both groups of pilots, flight performance decreased as workload increased and signal quality decreased. For flight performance measures of absolute heading and altitude deviations, the point at which performance decreased was between low and medium workload. Recall that the parameters that differentiated low and medium workload in this study were visibility conditions (6.0 SM vs. 1.75 SM), complexity of flight maneuvers (heading changes only vs. flight heading and altitude changes), and number of tasks in ATC commands (one task vs. two tasks). These results show that regardless of signal quality, low visibility combined with complex flight maneuvers and complex ATC commands resulted in degraded flight index maintenance.

For airspeed deviation, all three workload groups were significantly different from each other, but in low and medium workload conditions, the most airspeed deviation occurred with poor signal quality. This suggests that the workload levels were manageable in the low and medium incident workload until the signal quality degraded to at least below 0.6 SII. In the high workload condition, increased workload negatively affected performance, regardless of signal quality level.

Similarly, analysis of altitude deviation suggests that the point at which flight performance decreased was between poor and average signal quality (between 0.4 and 0.6 SII). This supports previous studies that showed performance degradations once speech intelligibility dropped below 50% (Whitaker, Peters, & Garinther, 1989).

One hypothesis for this study was that for normal-hearing aviators, as flight workload increased and communication signal quality decreased, performance would decrease first with the passive headset/passive earplug combination, then the ANR-equipped headset, followed by the CEP-equipped headset. This was partially supported by the finding that hearing-impaired aviators' performance, as measured by airspeed deviation, was negatively affected by the DC-FOAM headset before decrements were seen with the B-ANR headset or the DC-CEP headset.

Communications Intelligibility

Readbacks

The communications intelligibility metric was represented by the number of times a pilot required recorded ATC commands to be replayed in order to understand the command and to repeat it back to the experimenter correctly.

Visual inspection of the univariate normality plot of the residuals was conducted and revealed a slope that followed the normality slope output by SAS. A scatterplot of the residuals versus the predicted values was visually analyzed and revealed a pattern that supported the homogeneity of variance assumption. The ANOVA summary table for the mean number of readbacks is shown in Table 16.

Table 16. ANOVA summary table for mean number of ATC command readbacks

Source	df	MS	*F*	*p*	
<u>Between</u>					
Group (G)	1	2.017	0.29	0.5968	
Error		6.951			
<u>Within</u>					
Headset (H)	2	0.801	1.08	0.3407	
Error		0.743			
Workload (W)	2	10.985	14.79	< 0.0001	*
Error		0.743			
Signal Quality (SQ)	2	15.202	20.46	< 0.0001	*
Error		0.743			
H × W	4	0.216	0.29	0.8842	
Error		0.743			
W × SQ	4	3.999	5.38	0.0003	*
Error		0.743			
H × SQ	4	0.657	0.88	0.4727	
Error		0.743			
G × H	2	5.838	7.86	0.0004	*
Error		0.743			
G × W	2	0.600	0.81	0.4465	
Error		0.743			
G × SQ	2	0.050	0.07	0.9349	
Error		0.743			
G × H × W	4	0.447	0.60	0.6614	
Error		0.743			
G × W × SQ	4	0.192	0.26	0.9048	
Error		0.743			
G × H × SQ	4	1.206	1.62	0.1673	
Error		0.743			

*indicates significant result ($p < 0.05$)

Main Effects

An ANOVA revealed a significant main effect of workload ($F = 14.79$, $p < 0.0001$) on communications intelligibility. The mean number of readbacks for low, medium, and high workload conditions was 0.5, 0.5, and 0.9, respectively. Post-hoc analysis of the number of readbacks for each of the three workload conditions using Tukey's pairwise comparison test showed a significant difference between the low and medium workload conditions compared to the high workload condition (Figure 25). Pilots in both groups required more readbacks to correctly repeat ATC commands in the high workload condition.

ANOVA also revealed a significant main effect of signal quality ($F = 20.46$, $p < 0.0001$) on communications intelligibility. The mean number of readbacks in good, average, and poor signal quality conditions was 0.5, 0.4, and 0.9, respectively. Similarly, post-hoc analysis of the number of readbacks for each of the three signal quality conditions (Figure 26) showed a significant difference between the good and average signal quality conditions compared to the poor signal quality condition. Pilots required significantly more ATC command readbacks in poor signal quality conditions.

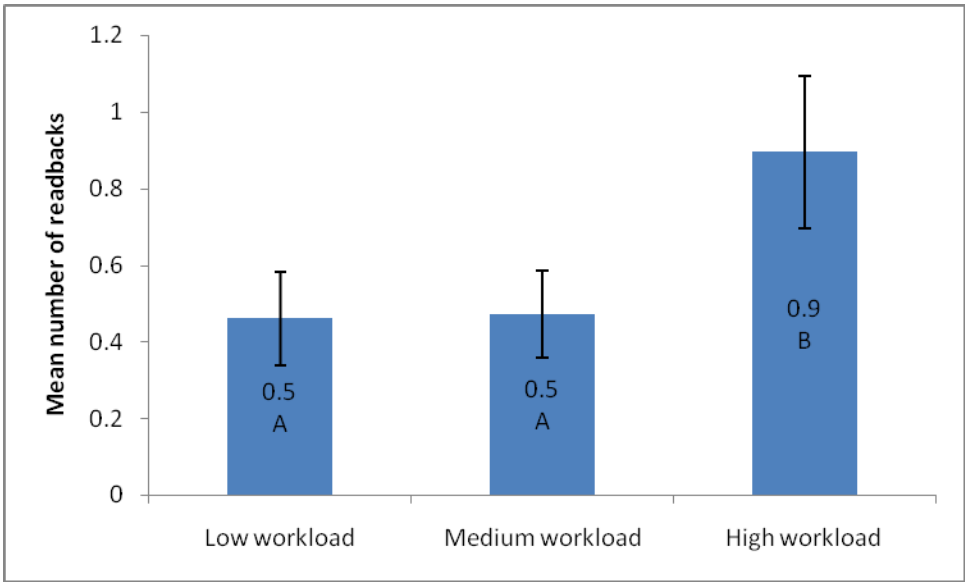


Figure 25. The effect of workload level on the mean number of ATC command readbacks required for a pilot to understand and correctly repeat the command. Different letters represent significant differences due to workload at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

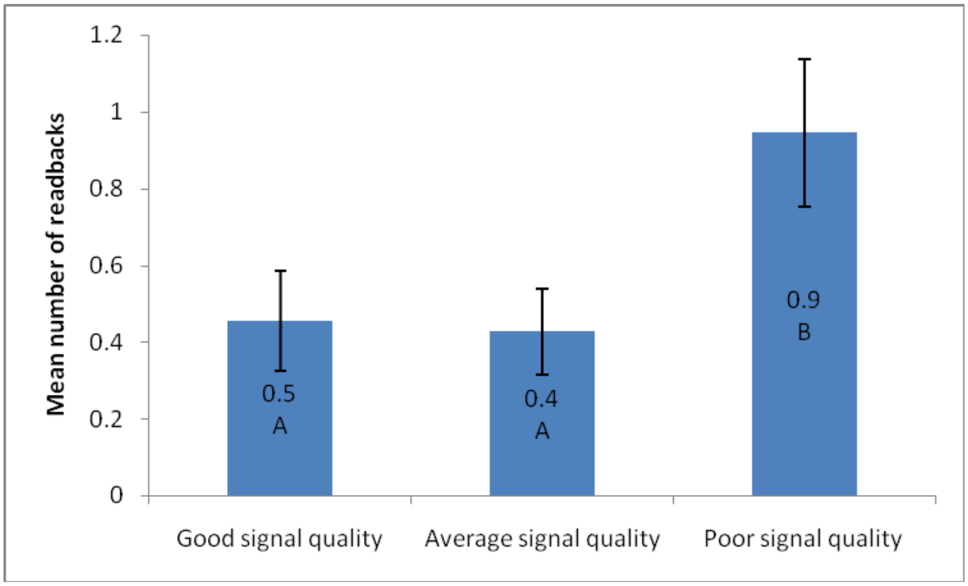


Figure 26. The effect of signal quality on the mean number of ATC command readbacks required for a pilot to understand and correctly repeat the command. Different letters represent significant differences due to signal quality at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

Interaction Effects

An ANOVA indicated a significant interaction between workload and signal quality ($F = 5.38$, $p = 0.0003$) on communications intelligibility, graphically demonstrated in Figure 27 and Figure 28. Post-hoc comparisons were conducted by isolating the workload level and conducting Tukey's multiple comparison test for signal quality on each workload level (Figure 27). This analysis indicated that there was a significant difference in the mean number of readbacks required in the low and high workload conditions when the signal quality was poor.

The workload variable was then analyzed as a function of signal quality. The signal quality level was isolated and the effect of workload was evaluated with Tukey's multiple comparison test (Figure 28). This analysis indicated that there was a significant difference in the mean number of readbacks required in the good and poor signal quality conditions when the workload was high. Both analyses show that pilots in both groups required the most ATC command readbacks when workload was high and signal quality was poor. The mean number of readback values for each workload/signal quality combination and 95% confidence interval values are shown in Table 17.

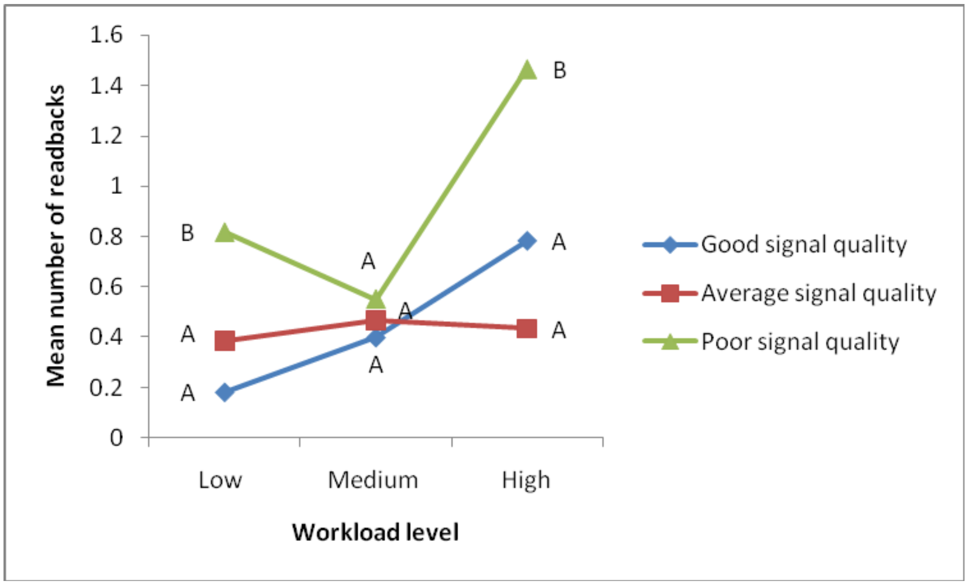


Figure 27. The effect of signal quality on pilots' mean number of ATC command readbacks for all workload levels. For each workload, different letters represent significant differences due to signal quality at $p < 0.05$ using Tukey's multiple comparison test.

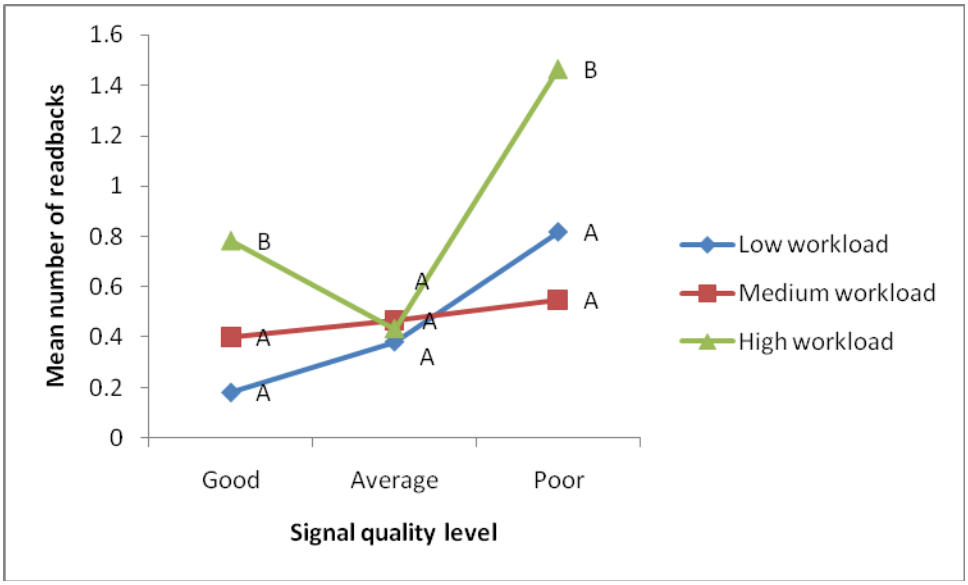


Figure 28. The effect of signal quality on pilots' mean number of ATC command readbacks for all workload levels. For each signal quality, different letters represent significant differences due to workload at $p < 0.05$ using Tukey's multiple comparison test.

Table 17. Mean number of ATC command readbacks and 95% confidence intervals for workload and signal quality level combinations.

Workload level	Signal Quality	Mean number of readbacks	Lower 95% confidence interval	Upper 95% confidence interval
Low	Good	0.2	0.1	0.3
Low	Average	0.4	0.2	0.6
Low	Poor	0.8	0.5	1.1
Medium	Good	0.4	0.2	0.6
Medium	Average	0.5	0.3	0.7
Medium	Poor	0.6	0.3	0.8
High	Good	0.8	0.5	1.1
High	Average	0.4	0.2	0.7
High	Poor	1.5	1.0	1.9

ANOVA also showed a significant interaction of pilot group and headset (F = 7.86, p = 0.0004) on communications intelligibility, shown in Figure 29 and Figure 30. Post-hoc comparisons were conducted by isolating the group and conducting Tukey's pairwise comparison procedure for headset on each group (Figure 29). This analysis showed that pilot Group 2 required significantly more readbacks when wearing the DC-FOAM headset compared to DC-CEP and B-ANR. There was not a significant difference of headset on communications intelligibility for Group 1.

The group variable was then analyzed as a function of headset (Figure 30). The headset type was isolated and the effect of group was evaluated with Tukey's multiple comparison test. This analysis showed that Group 2 required significantly fewer ATC command readbacks than Group 1 when using the DC-CEP headset. There were no significant differences between groups with B-ANR or DC-FOAM. The mean number of readback values and 95% confidence intervals for workload and group combinations are listed in Table 18.

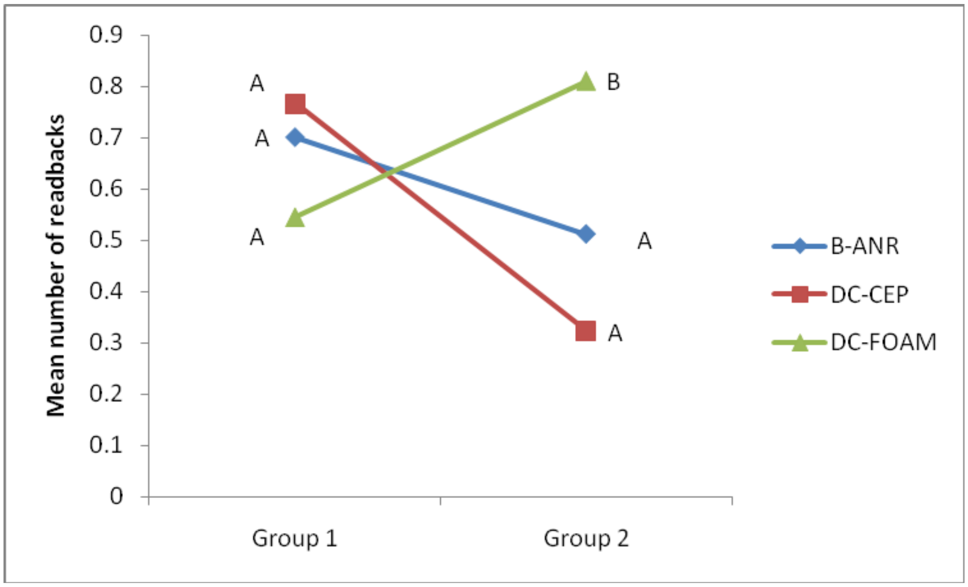


Figure 29. The effect of headset on the mean number of ATC command readbacks for both groups. For each group, different letters represent significant differences due to headset at $p < 0.05$ using Tukey's multiple comparison test.

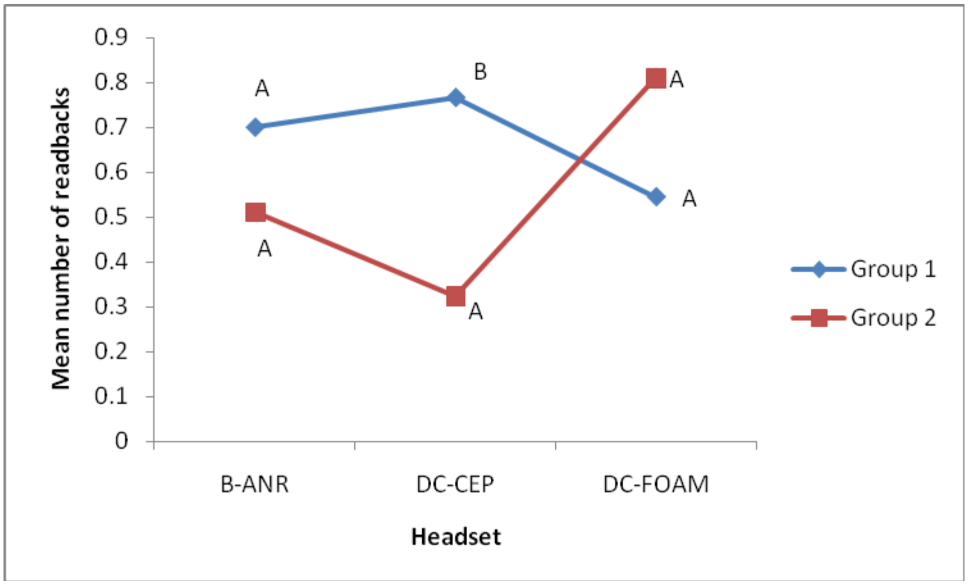


Figure 30. The effect of group on the mean number of ATC command readbacks for all headsets. For each headset level, different letters represent significant differences due to group at $p < 0.05$ using Tukey's multiple comparison test.

Table 18. Mean number of ATC command readbacks and 95% confidence intervals for workload level and group combinations.

Workload level	Group	Mean number of readbacks	Lower 95% confidence interval	Upper 95% confidence interval
Low	1	0.7	0.5	1.0
Medium	1	0.8	0.6	1.0
High	1	0.5	0.4	0.7
Low	2	0.5	0.3	0.7
Medium	2	0.3	0.2	0.5
High	2	0.8	0.5	1.1

Discussion

The point at which communications intelligibility suffers in regards to workload is between medium and high workload. Just as with the flight performance measures, the point at which signal quality had a significant negative effect on the communications intelligibility is between average and poor signal quality (0.6 and 0.4), coinciding with the Whitaker, et al. (1989) finding that performance begins to suffer when speech intelligibility drops below 50%. Also consistent with the previous flight performance measurements, the combination of high workload and poor signal quality resulted in difficulty understanding and repeating ATC commands, regardless of hearing ability. As for the group/headset interaction, pilots in Group 2 required more readbacks when wearing the DC-FOAM headset, similar to the finding with airspeed deviation.

Workload Measures

Modified Cooper-Harper

Workload ratings collected from pilots after each flight segment were classified into 10 categories, which were further grouped into the following four groups, consistent with the Modified Cooper-Harper (MCH) Scale:

- 1 -3 = an acceptable level of workload

- 4-6 = high workload, workload should be reduced
- 7-9 = major design deficiencies, system redesign is strongly recommended
- 10 – major design deficiencies, system redesign is mandatory

Workload ratings were analyzed with a Friedman’s *F*-test, chosen as most appropriate because it was designed to compare more than two categorical variables with ordinal type data (Conover, 1999). Results of the Friedman’s *F*-test analysis are shown in Table 19.

Table 19. Friedman's *F*-test summary table for mean MCH score

Source	df	MS	*F*	*p*	
Between					
Group (G)	1	0.000	0.00	1.0000	
Error		69071			
Within					
Headset (H)	2	771.838	0.36	0.6983	
Error		2147.502			
Workload (W)	2	326126.051	151.86	< 0.0001	*
Error		2147.502			
Signal Quality (SQ)	2	28219.172	13.14	< 0.0001	*
Error		2147.502			
H × W	4	1145.516	0.53	0.7112	
Error		2147.502			
W × SQ	4	6388.494	2.97	0.0191	*
Error		2147.502			
H × SQ	4	3322.745	1.55	0.1873	
Error		2147.502			
G × H	2	6310.388	2.94	0.0539	
Error		2147.502			
G × W	2	11353.251	5.29	0.0054	*
Error		2147.502			
G × SQ	2	521.372	0.24	0.7845	
Error		2147.502			
G × H × W	4	762.795	0.36	0.8404	
Error		2147.502			
G × W × SQ	4	1196.515	0.56	0.6939	
Error		2147.502			
G × H × SQ	4	1186.949	0.55	0.6971	
Error		2147.502			

*indicates significant result ($p < 0.05$)

Main Effects

Friedman's *F*-test revealed a significant main effect of workload ($F = 151.86$, $p < 0.0001$) on the mean MCH score with a post-hoc comparison revealing a significant difference between all three workload levels, shown in Figure 31. This result shows that ratings of workload increase as imposed experimental workload conditions increase. This finding is indicative of effective experimental workload manipulation that is near linear in effect.

A main effect of signal quality ($F = 13.14$, $p < 0.0001$) on the mean MCH score was also revealed (Figure 32). A post-hoc comparison indicated a significant difference between poor signal quality compared to good and average signal quality. This is consistent with previous results in that the point at which differences appear is between poor and average signal quality.

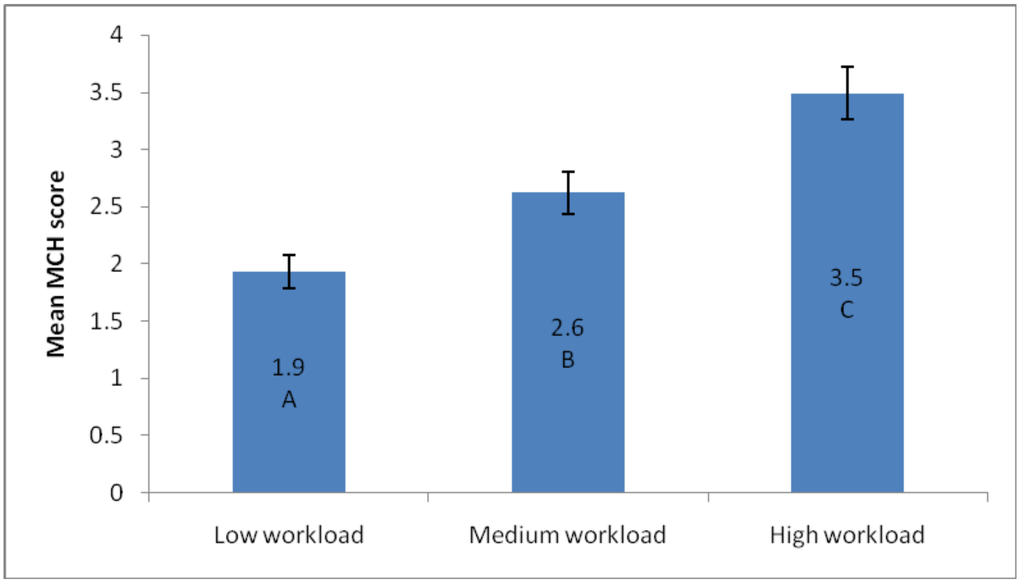


Figure 31. The effect of workload level on pilots' mean MCH score. Different letters represent significant differences due to workload at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

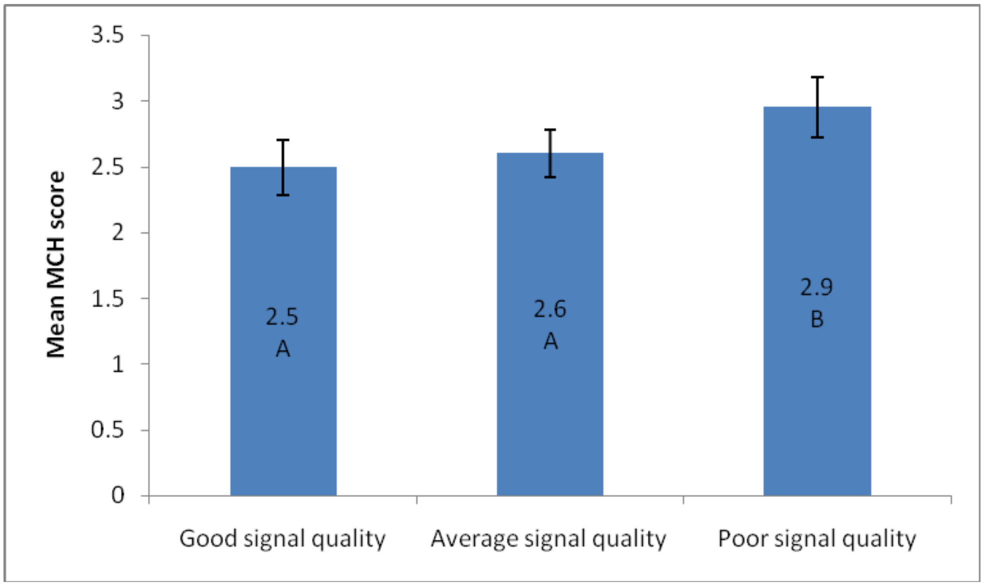


Figure 32. The effect of signal quality on pilots' mean MCH score. Different letters represent significant differences due to signal quality at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

Interaction Effects

ANOVA analysis indicated a significant interaction between workload and signal quality ($F = 2.97$, $p = 0.0191$) on the mean MCH score, graphically demonstrated in Figure 33 and Figure 34. A post-hoc comparison was conducted by isolating the signal quality variable and conducting Tukey's multiple comparison test for workload on the particular signal quality level (Figure 33). This comparison showed that there were significant differences in workload level on the mean MCH score in all signal quality conditions. In the good and average signal quality condition, all three workload levels were rated significantly different. In the poor signal quality condition, the high workload level resulted in a significantly higher mean MCH score than the low and medium workload levels did. Regardless of the signal quality, pilots in both groups rated the high workload condition as significantly higher than the other workload conditions.

The signal quality variable was then analyzed as a function of workload (Figure 34). The workload level was isolated and the effect of signal quality was evaluated with Tukey's multiple comparison test. This analysis showed that there were significant differences in the mean MCH score at the low and high workload levels. In the low workload condition, all three signal quality levels were rated significantly different. In the medium workload condition; however, no signal quality level was rated significantly different from the others. For the high workload condition, the high signal quality level was rated significantly higher on the MCH scale than the average and poor signal quality levels were. Mean MCH score values for all workload and signal quality level combinations and 95% confidence interval values are listed in Table 20.

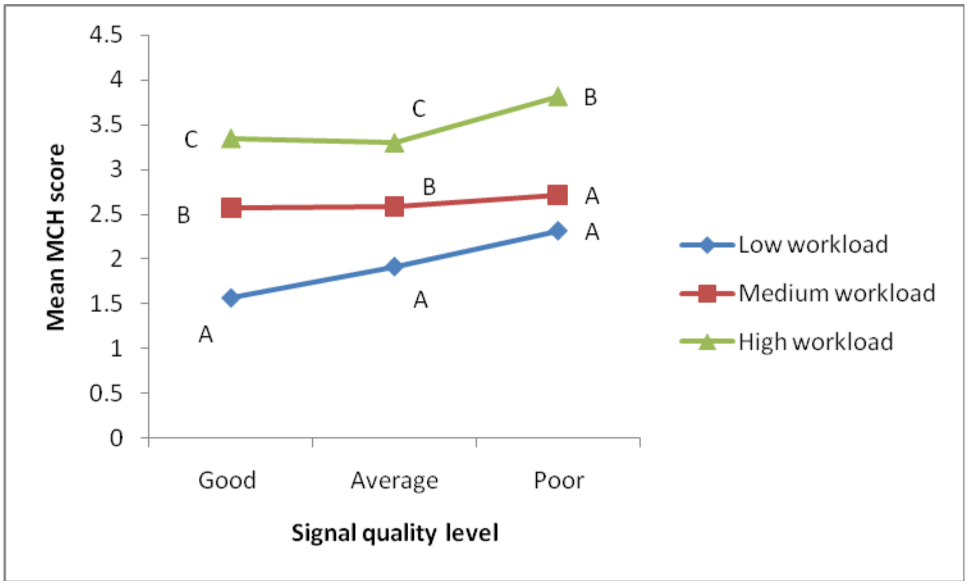


Figure 33. The effect of workload on pilots' mean MCH score for all signal quality conditions. For each signal quality level, different letters represent significant differences due to workload at $p < 0.05$ using Tukey's multiple comparison test.

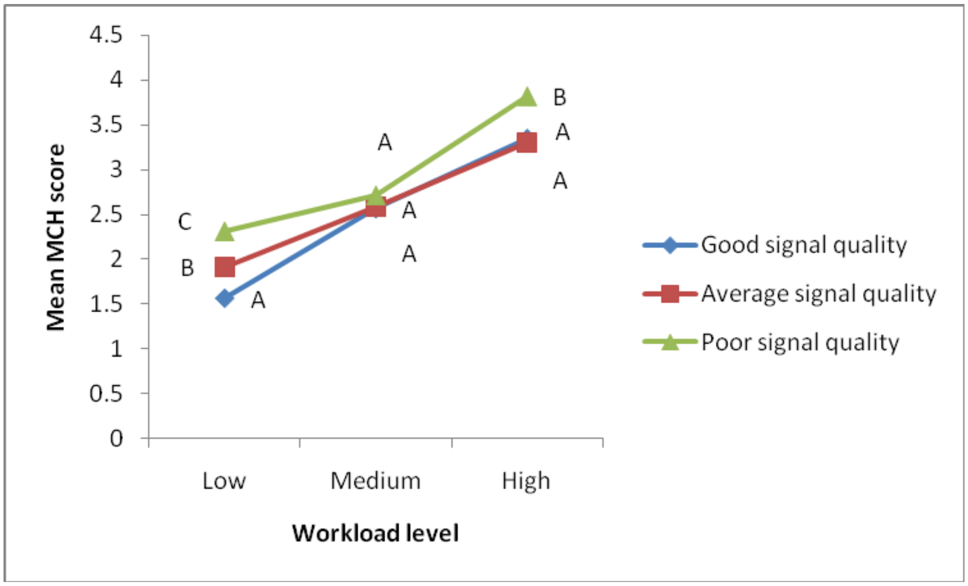


Figure 34. The effect of signal quality on pilots' mean MCH score for all workload conditions. For each workload level, different letters represent significant differences due to signal quality at $p < 0.05$ using Tukey's multiple comparison test.

Table 20. Mean MCH score and 95% confidence interval values for workload and signal quality level.

Workload level	Signal quality level	Mean MCH score	Lower 95% confidence interval	Upper 95% confidence interval
Low	Good	1.6	1.4	1.8
Low	Average	1.9	1.7	2.1
Low	Poor	2.3	2.0	2.6
Medium	Good	2.6	2.2	2.9
Medium	Average	2.6	2.3	2.9
Medium	Poor	2.7	2.4	3.0
High	Good	3.4	3.0	3.7
High	Average	3.3	2.9	3.7
High	Poor	3.8	3.4	4.3

ANOVA also indicated an interaction that approached significance between group and headset ($F = 2.94$, $p = 0.0539$) on the mean MCH score, graphically demonstrated in Figure 35 and Figure 36. Post-hoc comparisons were conducted by isolating the group variable and conducting Tukey's multiple comparison test for headset on the particular group level (Figure 35). This comparison showed that for Group 1, the MCH rating was not significantly different with any of the three headset configurations. For Group 2; however, pilots rated workload significantly higher when they were wearing the DC-FOAM headset configuration compared to when they were wearing the DC-CEP headset.

The group variable was then analyzed as a function of headset (Figure 36). The headset type was isolated and the effect of group was evaluated with Tukey's multiple comparison test. There were no main effects of group in this analysis, indicating that the interaction was within Group 2 instead of within the headset variable. Mean MCH score values for all headset and group combinations and 95% confidence interval values are listed in Table 21.

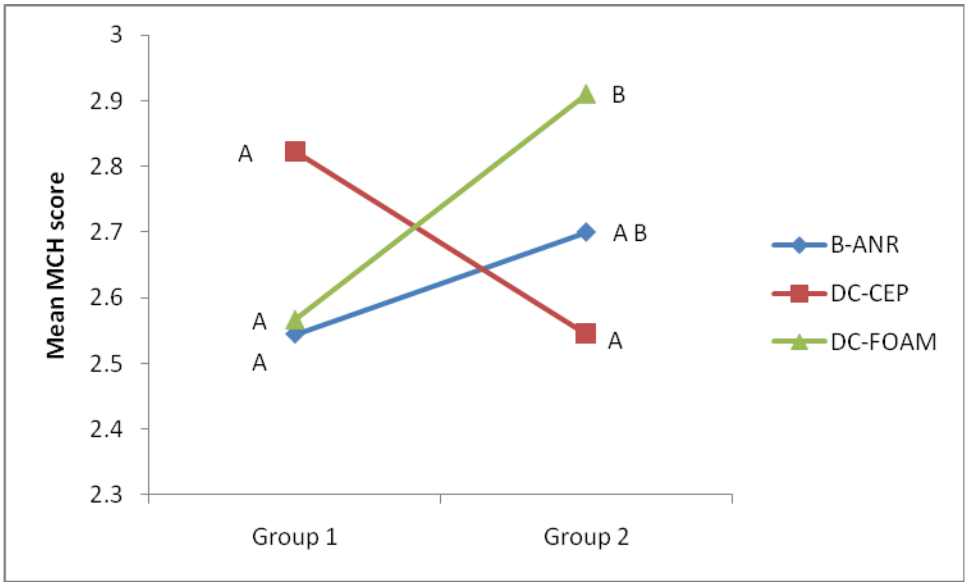


Figure 35. The effect of headset on pilots' mean MCH score for both groups. For each group, different letters represent significant differences due to headset at $p < 0.05$ using Tukey's multiple comparison test.

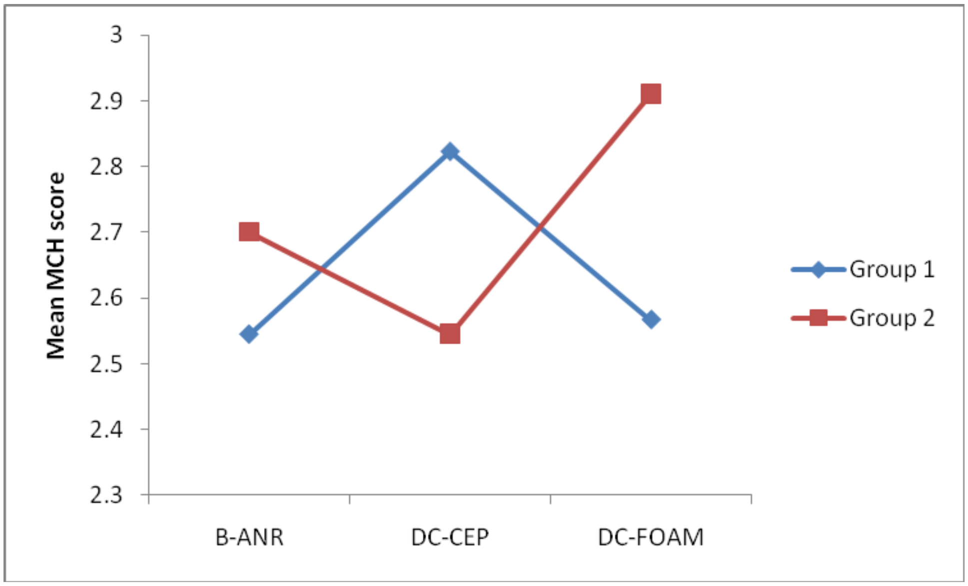


Figure 36. The effect of group on pilots' mean MCH score for all headset levels. For each headset, the absence of letters represents no significant differences due to group at $p < 0.05$ using Tukey's multiple comparison test.

Table 21. Mean MCH score and 95% confidence interval values for group and headset.

Headset	Group	Mean MCH score	Lower 95% confidence interval	Upper 95% confidence interval
B-ANR	1	2.5	2.3	2.8
DC-CEP	1	2.7	2.4	3.0
DC-FOAM	1	2.8	2.5	3.2
B-ANR	2	2.5	2.3	2.8
DC-CEP	2	2.6	2.3	2.8
DC-FOAM	2	2.9	2.6	3.3

ANOVA analysis indicated a significant interaction effect between group and workload ($F = 5.29$, $p = 0.0054$) on the mean MCH score, graphically demonstrated in Figure 37 and Figure 38. Post-hoc comparisons were conducted by isolating the group variable and conducting Tukey's multiple comparison test for workload on the particular group level (Figure 37). This comparison showed that for both groups, the MCH rating was significantly different for each workload level.

The workload variable was then analyzed as a function of workload (Figure 38). The workload level was isolated and the effect of group was evaluated with Tukey's multiple comparison test. There were no main effects of group in this analysis, indicating that the interaction was due to the differences in workload instead of the differences between groups. Mean MCH score values for all group and workload combinations and 95% confidence interval values are listed in Table 22.

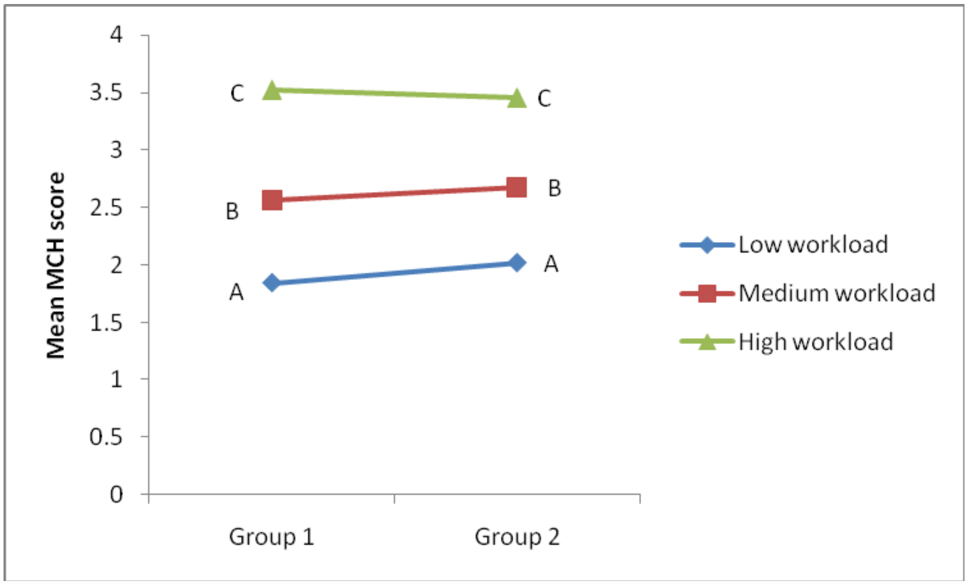


Figure 37. The effect of workload on pilots' mean MCH score for both groups. For each group, different letters represent significant differences due to workload at $p < 0.05$ using Tukey's multiple comparison test.

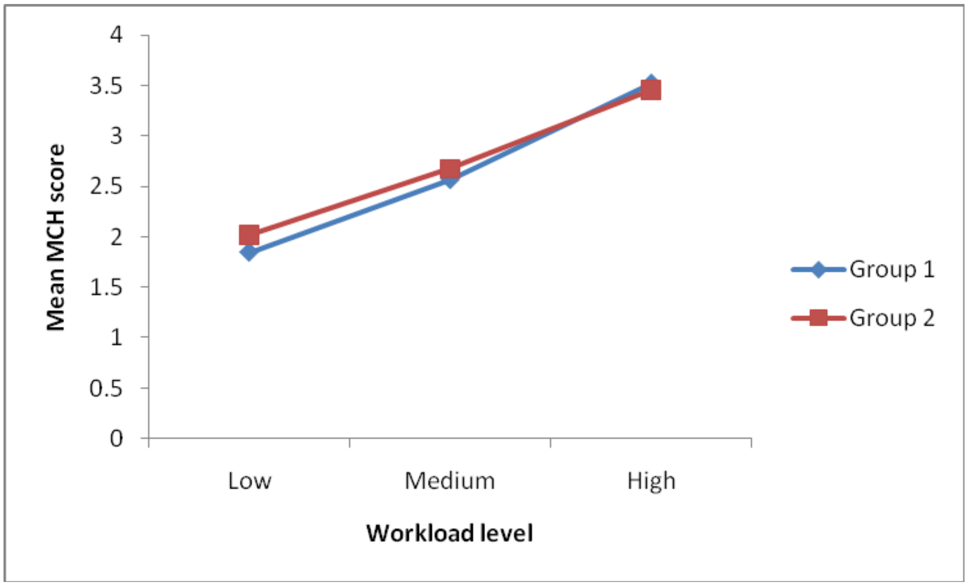


Figure 38. The effect of group on pilots' mean MCH score for all workload levels. For each workload level, the absence of letters represents no significant differences due to the group at $p < 0.05$ using Tukey's multiple comparison test.

Table 22. Mean MCH score and 95% confidence interval values for workload and group.

Workload	Group	Mean MCH score	Lower 95% confidence interval	Upper 95% confidence interval
Low	1	1.8	1.7	2.0
Medium	1	2.6	2.3	2.8
High	1	3.5	3.2	3.8
Low	2	2.0	1.8	2.2
Medium	2	2.7	2.4	2.9
High	2	3.5	3.1	3.8

Discussion

As workload increased, and signal quality decreased, workload ratings increased for both groups. As with flight performance measures, the point at which workload ratings were significantly higher for signal quality was between average and poor (0.6 and 0.4 SII). Pilots' subjective reports of increased workload in high workload-poor signal quality conditions were consistent with the flight performance deviations that occurred in the same situations. Also consistent with objective communications intelligibility results, Group 2 rated workload significantly highest with the DC-FOAM headset configuration, while Group 1 reported no differences in perceived workload due to the headset configuration.

The mean MCH rating for high workload (3.5) indicates that this level of flight workload is between the “acceptable amount of workload” and “high workload, level should be reduced” categories of workload in accordance with the scale’s rating method. This finding raises the question as to whether the high workload level designed for this experiment was practically important and high enough to influence real performance differences.

There are two possible reasons for this finding. Although two research pilots at the USAARL assisted in developing the workload levels, and agreed that the high workload level was indeed high workload in their opinion, it is possible that the pilot participants were conservative in their workload ratings due to a reluctance to admit that a simulator flight task was challenging, particularly to a female experimenter. Another possible answer is that, although they were instructed to base their workload ratings on the flight segment just flown, many of the pilot participants had experience flying in much higher workload environments such as nap-of-the-earth (NOE) flight with night vision goggles (NVG) and even combat, and may have unintentionally rated workload in comparison to their previous flight experiences. Either case would lead to artificially low subjective ratings of workload; therefore it is assumed that the experimental high workload level was indeed a high workload condition.

Situation Awareness

SART

The SART scale was used to assess pilots’ subjective situation awareness following each flight segment. They rated 10 dimensions covering attentional demands (D), attentional supply (S), and understanding (U) on a 1-10 scale. A composite score was calculated to yield a single number rating with the following formula: SA = U – (D-S) (Vidulich, Crabtree, & McCoy, 1993). A larger composite score indicates a subjective opinion of greater situation awareness. Visual inspection of the univariate normality plots of the residuals revealed normal a slope of data values that conformed to the normality values calculated by SAS. The ANOVA summary table for mean SART score is shown in Table 23.

Table 23. ANOVA summary table for mean SART score

Source	df	MS	F	p	
Between					
Group (G)	1	297.780	0.28	0.6038	
Error		1067.010			
Within					
Headset (H)	2	30.867	1.17	0.3098	
Error		26.278			
Workload (W)	2	1232.939	46.92	< 0.0001	*
Error		26.278			
Signal Quality (SQ)	2	189.272	7.20	0.0008	*
Error		26.278			
H × W	4	12.680	0.48	0.7486	
Error		26.278			
W × SQ	4	36.070	1.37	0.2423	
Error		26.278			
H × SQ	4	14.147	0.54	0.7076	
Error		26.278			
G × H	2	131.230	4.99	0.0071	*
Error		26.278			
G × W	2	182.724	6.95	0.0011	*
Error		26.278			
G × SQ	2	5.669	0.22	0.8060	
Error		26.278			
G × H × W	4	56.516	2.15	0.0735	
Error		26.278			
G × W × SQ	4	14.455	0.55	0.6991	
Error		26.278			
G × H × SQ	4	30.177	1.15	0.3330	

*indicates significant result ($p < 0.05$)

Main Effects

An ANOVA revealed a significant main effect of workload ($F = 46.92$, $p < 0.0001$) on SART ratings. As depicted in Figure 39, the mean SART scores for low, medium, and high workload conditions were 36.8, 33.8, and 31.6, respectively. Post-hoc analysis of the mean SART score for each of the three workload conditions using Tukey's multiple comparison test showed a significant difference between all three workload conditions ($p < 0.0001$). Pilots in both groups rated situation awareness lower as workload increased.

ANOVA analysis also revealed a significant main effect of signal quality ($F = 7.20$, $p = 0.0008$) on SART ratings, shown in Figure 40. The mean SART scores in good, average, and poor signal quality conditions were 34.7, 34.6, and 32.9, respectively. Post-hoc analysis of the mean SART score for each of the three signal quality conditions showed a significant lower SART rating when signal quality condition was poor compared to the average and good signal quality conditions. Pilots in both groups rated situation awareness lowest in the poor signal quality condition.

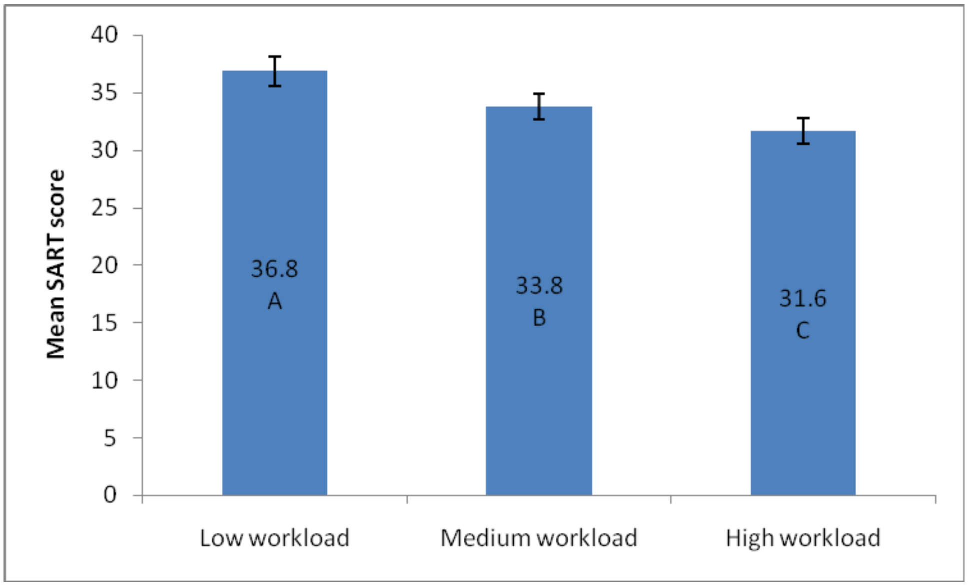


Figure 39. The effect of workload level on pilots' mean SART score. Different letters represent significant differences due to workload at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

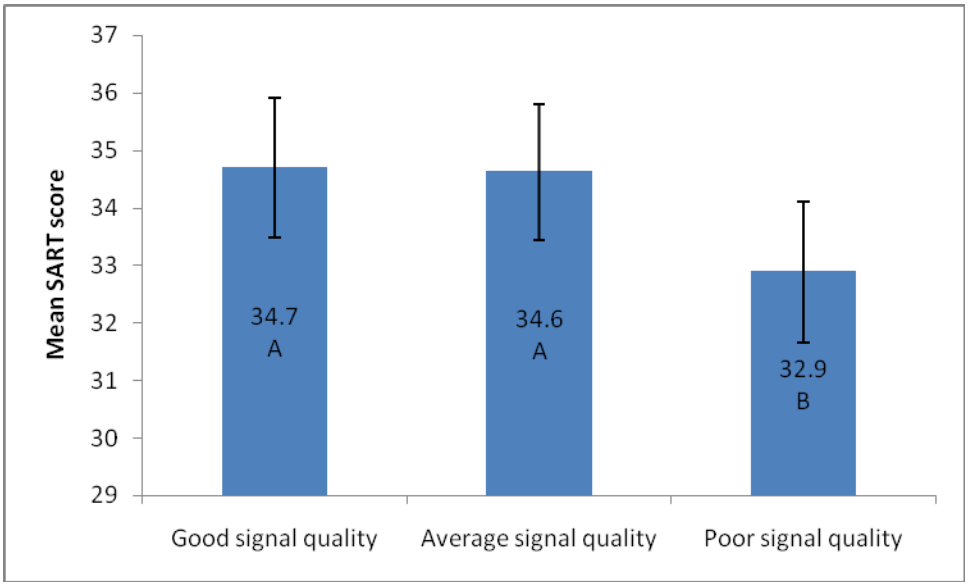


Figure 40. The effect of signal quality on pilots' mean SART score. Different letters represent significant differences due to signal quality at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

Interaction Effects

ANOVA analysis indicated a significant interaction effect between group and headset ($F = 4.99$, $p = 0.0071$), graphically demonstrated in Figure 41 and Figure 42. Post-hoc comparisons were first conducted by isolating the group and performing Tukey's multiple comparison test for each headset on the particular group (Figure 41). This comparison showed that the mean SART score for both groups was significantly different with the DC-FOAM headset configuration. For Group 1; however, the mean SART score was the highest with the DC-FOAM headset configuration, and for Group 2, the DC-FOAM headset resulted in the lowest mean SART score.

The headset variable was then analyzed as a function of the group (Figure 42). The headset configuration was isolated and the effect of the group was evaluated with Tukey's multiple comparison test. There was no main effect of group with this analysis, indicating that the significant difference found in the initial ANOVA was due to the headset rather than the group. The mean SART score values for each group and headset configuration combination and 95% confidence interval values are listed in Table 24.

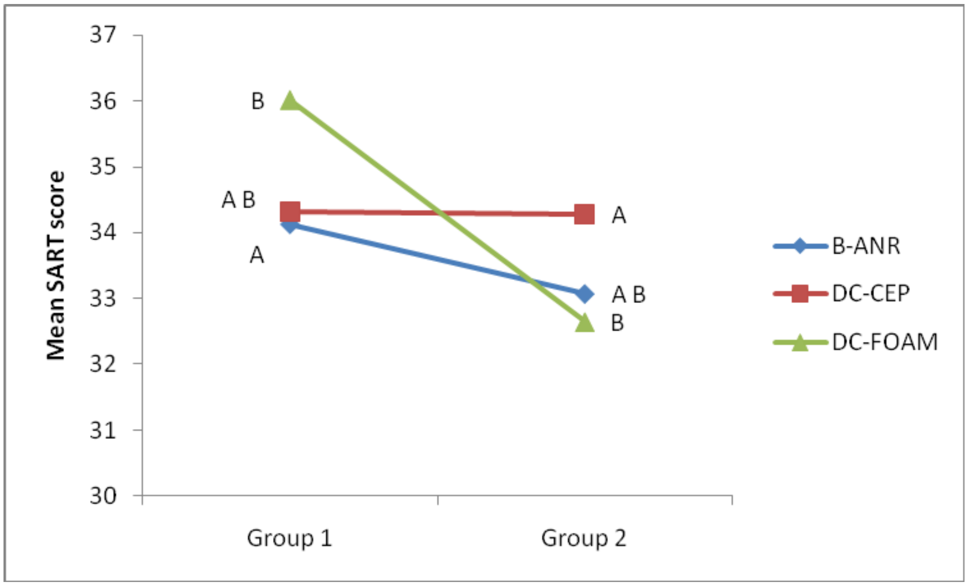


Figure 41. The effect of headset on pilots' mean SART score for both groups. For each group, different letters represent significant differences due to headset at $p < 0.05$ using Tukey's multiple comparison test.

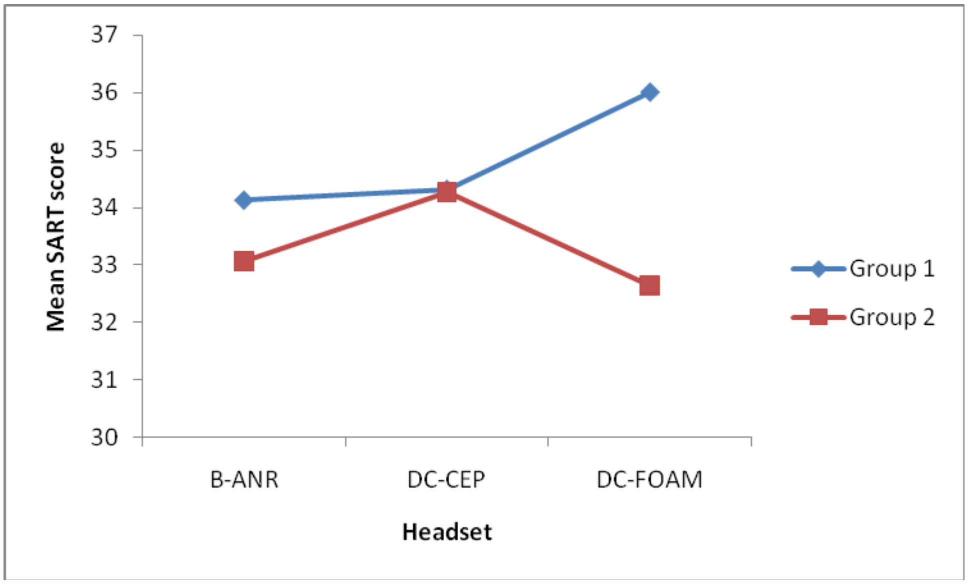


Figure 42. The effect of group on pilots' mean SART score for all headset levels. For each headset, the absence of letters represents no significant differences due to group at $p < 0.05$ using Tukey's multiple comparison test.

Table 24. Mean SART score and 95% confidence intervals for group and headset.

Headset	Group	Mean SART score	Lower 95% confidence interval	Upper 95% confidence interval
B-ANR	1	34. 1	32.6	35.7
DC-CEP	1	34.3	32.3	36.3
DC-FOAM	1	36.0	34.1	37.9
B-ANR	2	33.0	31.5	34.6
DC-CEP	2	34.2	32.8	35.8
DC-FOAM	2	32.6	30.9	34.4

There was also a significant interaction effect of group and workload (F = 6.95, p = 0.0081) on the mean SART score, shown in Figure 43 and Figure 44. Post-hoc comparisons were first analyzed by isolating the group variable and conducting Tukey's multiple comparison test for workload on each group (Figure 43). This analysis showed that Group 1 rated all three workload levels differently. Group 2, however, rated the medium and high workload levels the same with the low workload level influencing significantly higher SART ratings. For both groups, the low workload condition resulted in the highest SART ratings.

The data were then analyzed by isolating the workload level and the effect of the group was evaluated with Tukey's multiple comparison test (Figure 44). There was no main effect of group with this analysis, indicating that the interaction was due to the workload levels instead of the group variable. The mean SART score values for each group and workload level combination and 95% confidence interval values are listed in Table 25.

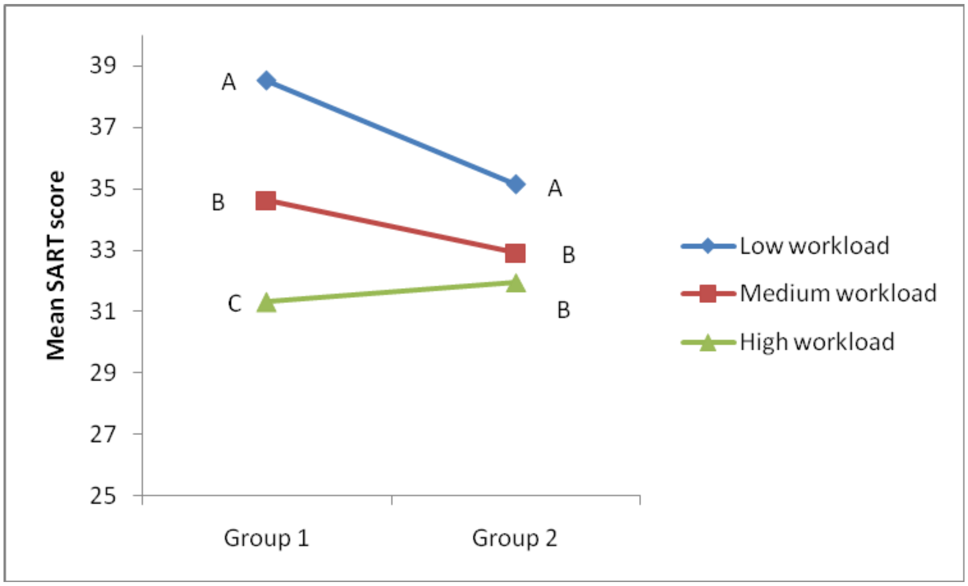


Figure 43. The effect of worklaod on pilots’ mean SART score for both groups. For each group, different letters represent significant differences due to workload at $p < 0.05$ using Tukey’s multiple comparison test.

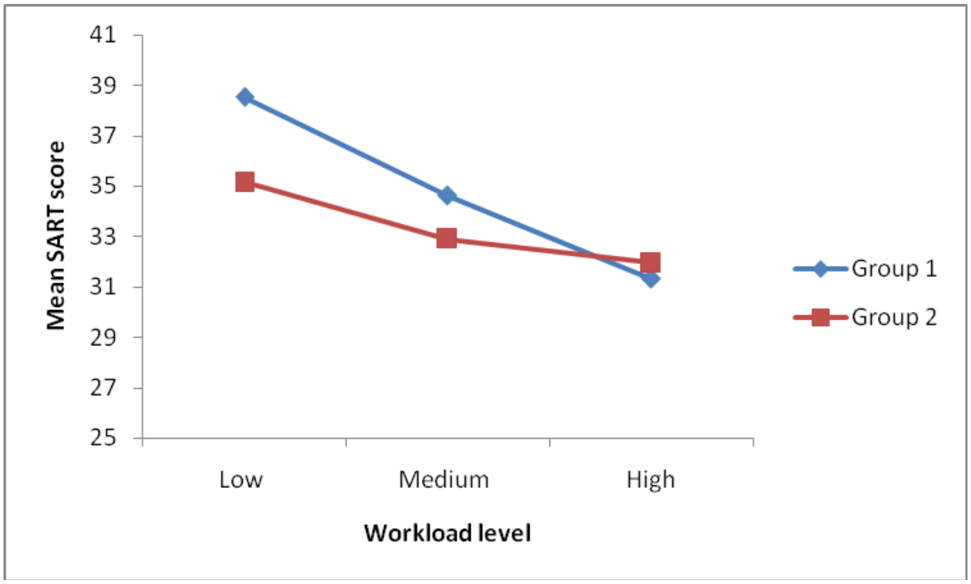


Figure 44. The effect of group on pilots’ mean SART score for all workload levels. For each workload level, the absence of letters represents no significant differences due to group at $p < 0.05$ using Tukey’s multiple comparison test.

Table 25. Mean SART score and 95% confidence intervals for workload and group.

Workload level	Group	Mean SART score	Lower 95% confidence interval	Upper 95% confidence interval
Low	1	38.5	36.7	40.3
Medium	1	34.6	32.9	36.3
High	1	31.3	29.7	33.0
Low	2	35.1	33.5	36.8
Medium	2	32.9	31.4	34.4
High	2	32.0	30.4	33.5

Discussion

All three workload conditions were rated significantly different for situation awareness. As with MCH workload ratings, the results appear to validate the experimental design to the extent that they support the notion that decreased workload leads to improved situation awareness.

Considering the interaction of group and workload (Figure 43 and Figure 44) it is clear that the situation awareness ratings for each workload level were significantly different for Group 1, which in addition to being the better hearing group, was also the least experienced group of pilots. In contrast are the pilots in Group 2, who happen to have more flight experience than Group 1. Group 2 pilots did not rate situation awareness for medium and high workload differently in a statistical sense, and there is less spread of the ratings compared to Group 1. This indicates that the effect of varying workload levels was less pronounced for Group 2 pilots. It could be that pilots with hearing loss do not acquire the same amount of auditory information as a normal-hearing pilot to assist with situation awareness, and therefore do not perceive as much difference across workload levels, or it could be due to the unintended, but unavoidable variable of experience. The latter is suspected; however, most Army pilots with hearing waivers are experienced pilots, so one would expect this effect in a real world situation.

It is interesting to note, again, that the point at which ratings become statistically different for signal quality is between average and poor (0.6 SII and 0.4 SII), just as with some of the previously discussed results. This subjective result supports previous research that has shown performance decrements when communications intelligibility dropped below 50%.

Situation awareness was rated significantly lower when using B-ANR and DC-FOAM for Group 2. This finding partially supports the hypothesis that as flight workload increased and communication signal quality decreased, ratings of situation awareness (SA) would decrease first with the passive headset/earplug combination, then ANR, followed by CEP, regardless of hearing loss category. For Group 1, ratings of situation awareness were actually higher for the DC-FOAM headset.

Headset Comfort/Speech Intelligibility Ratings

The headset comfort/speech intelligibility rating scale was based on the rating scale used in a previous study (Park & Casali, 1991). In that study, response scoring calculated a single composite score; however, in this study, it seemed more appropriate to analyze each bipolar rating scale question separately in order to see effects of individual headsets and how they related to the previous data analysis. Each question response was transformed to numeric values, with a response of 1 on the far left of the scale, and a response of 6 on the far right of the scale (see rating scale in Appendix H). For all of the following questions, a visual inspection of the normal probability plot showed residuals that generally followed the predicted values calculated in SAS. Further, a scatterplot of the residual versus the predicted values was consistent with independently distributed residuals with constant variance, therefore ANOVA was applied. The ANOVA revealed no significant differences for group, headset, or group crossed with headset configurations for the rating items shown in Table 26. Thus, pilots did not rate any differences between the headsets with respect to these metrics.

Table 26. Headset comfort/speech intelligibility rating scale items that did not show statistical significance.

Rating scale item	*F*	*p*
Tight/Loose	1.64	0.1647
Non-restrictive/ Restrictive	1.22	0.3104
No extraneous noise/Extraneous noise	1.64	0.1654
Light/Heavy	1.21	0.3182
Background hum present/No background hum present	0.40	0.8495
Medium	2.00	0.3000
High	2.00	0.8000

Uncomfortable/Comfortable.

ANOVA analysis (Table 27) revealed a main effect of headset on comfort ratings (F = 6.62, p = 0.0036). Post-hoc analysis using Tukey's multiple comparison test indicated that the B-ANR headset was rated as more comfortable than the DC-CEP and the DC-FOAM headset configurations, which did not differ with respect to this metric. Mean values and significant differences for this rating scale item are displayed in Figure 45.

Table 27. ANOVA summary table for bipolar ratings of comfortable/uncomfortable

Source	df	MS	*F*	*p*	
Between					
Group (G)	1	0.000	0.00	1.0000	
Error		1.740			
Within					
Headset (H)	2	13.216	6.62	0.0036	*
Error		1.996			
G × H	2	1.850	0.93	0.4051	
Error		1.996			

*indicates significant result ($p < 0.05$)

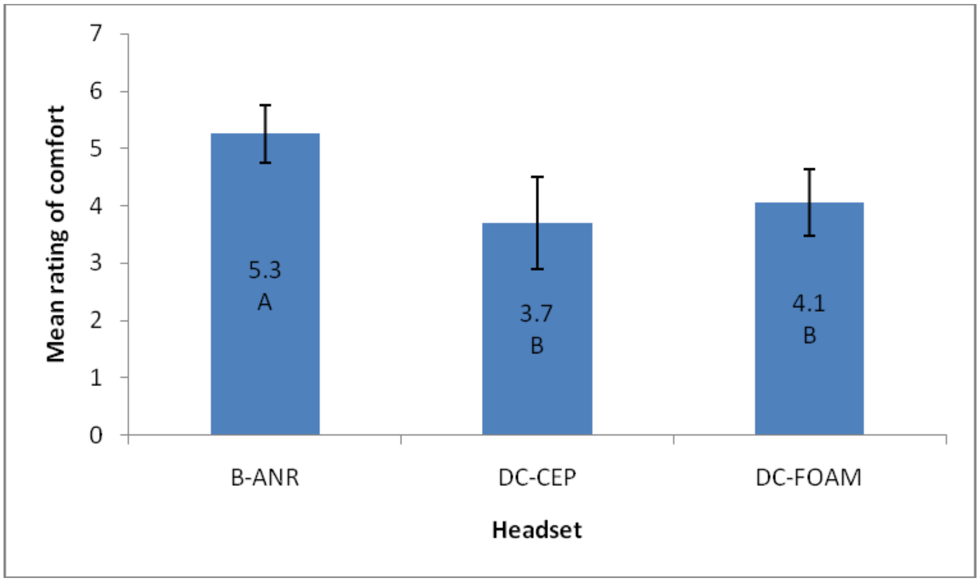


Figure 45. Pilot ratings of *comfort* for each headset (1=uncomfortable, 6=comfortable). Different letters represent significant differences due to headset at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

Uncomfortable Pressure/No Uncomfortable Pressure

ANOVA analysis (Table 28) revealed a significant main effect of headset for pressure ratings (F = 5.69, p = 0.0071). Further post-hoc Tukey's multiple comparison test analysis (Figure 46) showed that the B-ANR headset configuration did not differ from the DC-FOAM headset, and the DC-FOAM and DC-CEP did not differ from each other.

Table 28. ANOVA summary table for bipolar ratings of uncomfortable pressure/no uncomfortable pressure

Source	df	MS	*F*	*p*	
<u>Between</u>					
Group (G)	1	0.066	0.03	0.8661	
Error		2.277			
<u>Within</u>					
Headset (H)	2	11.316	5.69	0.0071	*
Error		1.988			
G × H	2	2.216	1.11	0.3391	
Error		1.988			

*indicates significant result ($p < 0.05$)

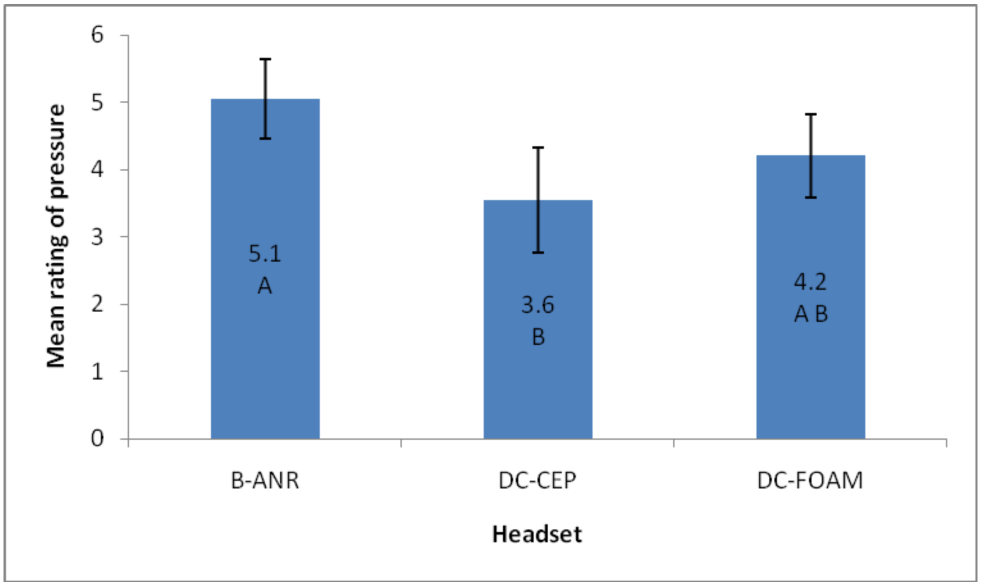


Figure 46. Pilot ratings of *pressure* for each headset (1=uncomfortable pressure, 6=no uncomfortable pressure). Different letters represent significant differences due to headset at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

Painful/Painless

ANOVA showed a significant main effect of headset on pain ratings ($F = 3.32$, $p = 0.0473$), as shown in Table 29. Post-hoc Tukey's multiple comparison test analysis (Figure 47) indicated that pilots rated the B-ANR and DC-CEP headset configurations significantly different, with the B-ANR being the least painful. The B-ANR and DC-FOAM headset configurations were not considered significantly different from each other. The same is true of the DC-CEP and DC-FOAM headset configurations.

Table 29. ANOVA summary table for bipolar ratings of painful/painless

Source	df	MS	*F*	*p*	
Between					
Group (G)	1	0.066	0.03	0.8721	
Error		2.517			
Within					
Headset (H)	2	7.350	3.32	0.0473	*
Error		2.211			
G × H	2	2.516	1.14	0.3317	
Error		2.211			

*indicates significant result ($p < 0.05$)

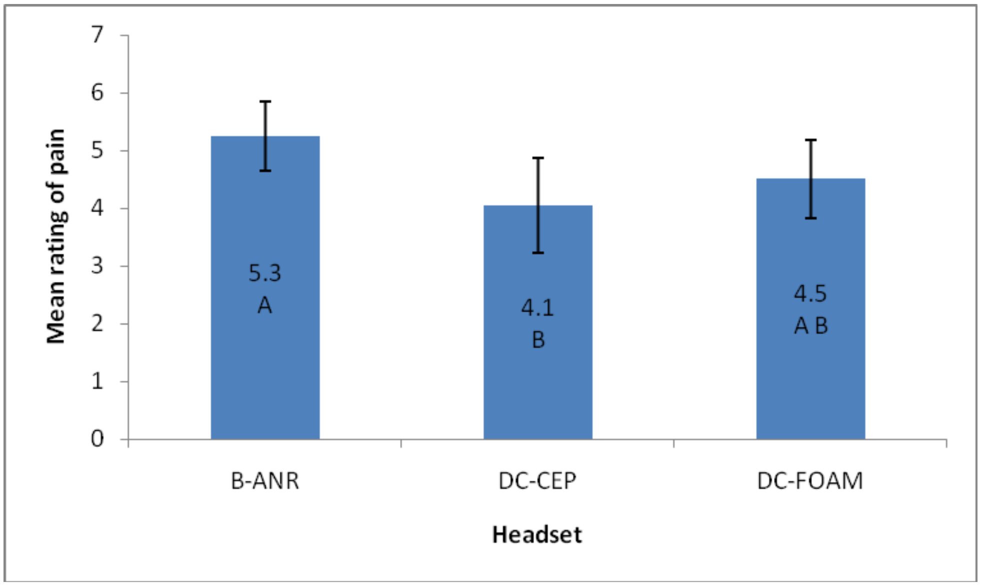


Figure 47. Pilot ratings of *pain* for each headset (1=painful, 6=painless). Different letters represent significant differences due to headset at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

Ear Blocked/Ear Open

Statistical analysis showed a significant main effect of headset on ear blocked ratings ($F = 8.23$, $p = 0.0011$), as displayed in Table 30.

Post-hoc analysis conducted with Tukey's multiple comparison test (Figure 48) revealed that the B-ANR headset was rated the highest concerning the ear feeling open and was significantly different from the DC-CEP and DC-FOAM headsets, which were judged to be the headsets that made the ears feel most blocked.

Table 30. ANOVA summary table for bipolar ratings of ear blocked/ear open

Source	df	MS	*F*	*p*	
Between					
Group (G)	1	2.816	0.55	0.4663	
Error		5.083			
Within					
Headset (H)	2	14.816	8.23	0.0011	*
Error		1.800			
G × H	2	0.116	0.06	0.9374	
Error		1.800			

*indicates significant result ($p < 0.05$)

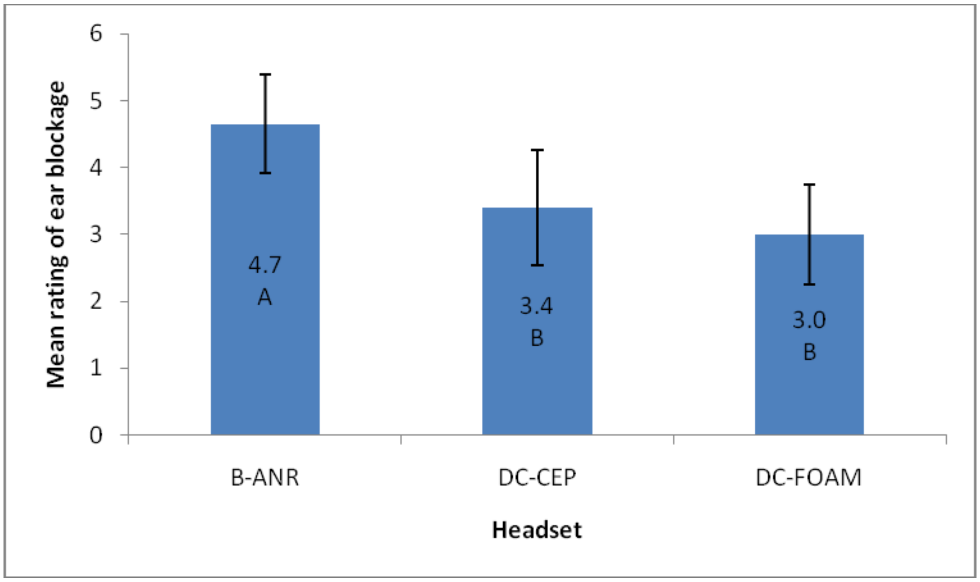


Figure 48. Pilot ratings of *ear blockage* for each headset (1= ear blocked, 6= ear open). Different letters represent significant differences due to headset at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

Cumbersome/Convenient

An ANOVA indicated a significant main effect of headset on convenience ratings (F = 6.44, p = 0.0041), as displayed in Table 31. Post-hoc analysis conducted with Tukey's multiple comparison test showed that the B-ANR headset was rated significantly more convenient than the DC-FOAM headset, as shown in Figure 49. There was no statistical difference with the DC-CEP headset.

Table 31. ANOVA summary table for bipolar ratings of convenient/cumbersome

Source	df	MS	*F*	*p*	
<u>Between</u>					
Group (G)	1	2.400	1.67	0.2126	
Error		1.437			
<u>Within</u>					
Headset (H)	2	10.617	6.44	0.0041	*
Error		1.648			
G × H	2	1.050	0.64	0.5347	
Error		1.648			

*indicates significant result ($p < 0.05$)

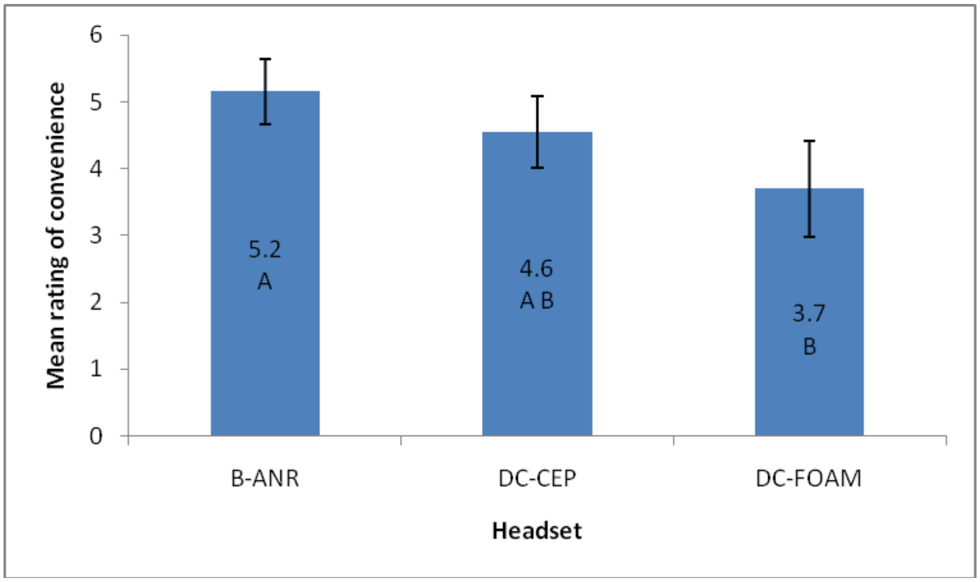


Figure 49. Pilot ratings of *convenience* for each headset (1=cumbersome, 6=convenient). Different letters represent significant differences due to headset at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

Low Fidelity Communications/High Fidelity Communications

As displayed in Table 32, the ANOVA showed a significant main effect of headset on communications fidelity ratings ($F = 9.91$, $p = 0.0004$). Post-hoc analysis conducted with Tukey's multiple comparison test (Figure 50) revealed that the B-ANR headset and DC-CEP headsets ratings were statistically higher, indicating perceptions of higher fidelity communication with those headsets compared to the DC-FOAM headset.

Table 32. ANOVA summary table for bipolar ratings of high fidelity communications/low fidelity communications

Source	df	MS	*F*	*p*	
<u>Between</u>					
Group (G)	1	0.150	0.08	0.7859	
Error		1.972			
<u>Within</u>					
Headset (H)	2	17.017	9.91	0.0004	*
Error		1.717			
G × H	2	3.750	2.18	0.1272	
Error		1.717			

*indicates significant result ($p < 0.05$)

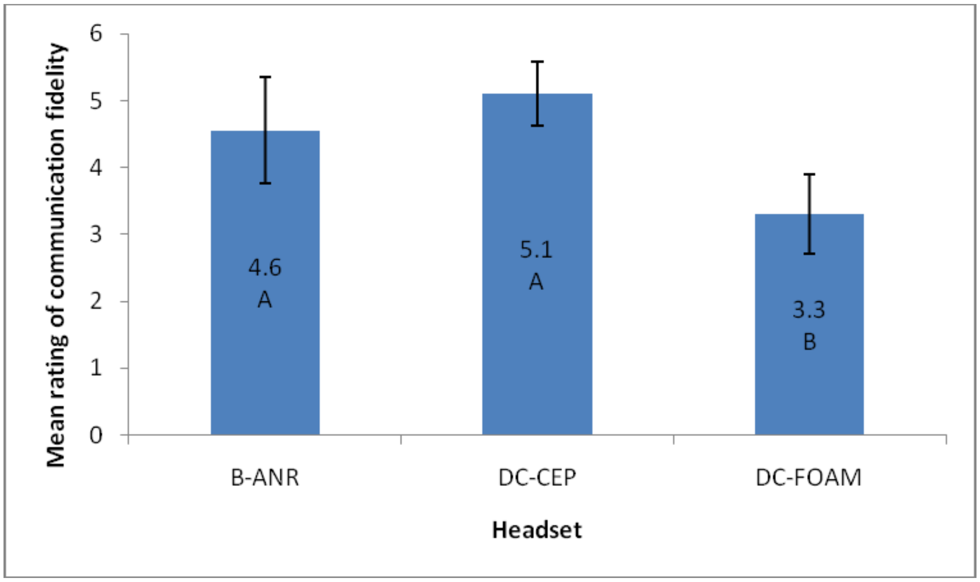


Figure 50. Pilot ratings of *communication fidelity* for each headset (1= low fidelity communications, 6= high fidelity communications). Different letters represent significant differences due to headset at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

Sound Distortion/No Sound Distortion

Statistical analysis showed a significant main effect of headset on sound distortion ratings ($F = 14.0686$, $p < 0.0001$) as shown in Table 33. Post-hoc analysis conducted with Tukey's multiple comparison test (Figure 51) revealed that the B-ANR headset and DC-CEP headsets ratings were statistically higher, indicating perceptions of less sound distortion with those headsets compared to the DC-FOAM headset.

Table 33. ANOVA summary table for bipolar ratings of no sound distortion/ sound distortion

Source	df	MS	*F*	*p*	
<u>Between</u>					
Group (G)	1	0.0166	0.01	0.9362	
Error		2.528			
<u>Within</u>					
Headset (H)	2	21.6500	20.73	<.0001	*
Error		1.044			
G × H	2	1.2166	1.16	0.3234	
Error		1.044			

*indicates significant result ($p < 0.05$)

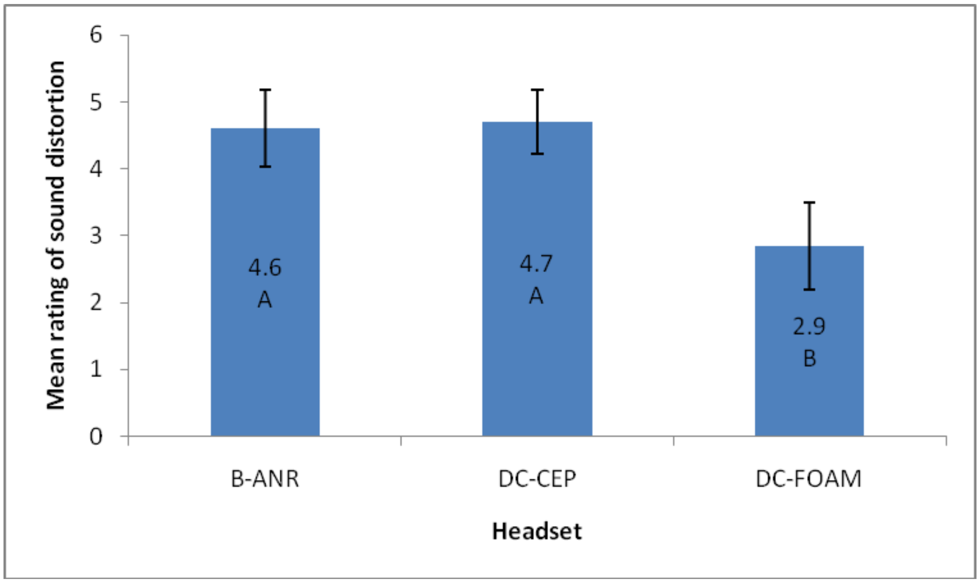


Figure 51. Pilot ratings of *sound distortion* for each headset (1= sound distortion, 6= no sound distortion). Different letters represent significant differences due to headset at $p < 0.05$ using Tukey's multiple comparison test. Vertical bars indicate 95% confidence intervals about the means.

Hard/Soft

ANOVA indicated a main effect of headset on hardness ratings ($F = 4.99$, $p = 0.0122$), shown in Table 34. Post-hoc analysis conducted with Tukey's multiple comparison test showed that the B-ANR headset was rated significantly softer than the DC-FOAM headset (Figure 52). ANOVA also showed an interaction of group and headset ($F = 4.29$, $p = 0.0213$), displayed in Figure 53. Post-hoc analysis with Tukey's multiple comparison test revealed that Group 1 rated the B-ANR headset as significantly softer than the DC-CEP headset and the DC-FOAM headset. Mean response values to this rating scale item for each headset level/group combination and 95% confidence interval values are listed in Table 35.

Table 34. ANOVA summary table for bipolar ratings of hard/soft

Source	df	MS	*F*	*p*	
<u>Between</u>					
Group (G)	1	0.067	0.03	0.8621	
Error		2.148			
<u>Within</u>					
Headset (H)	2	6.650	4.99	0.0122	*
Error		1.331			
G × H	2	5.717	4.29	0.0213	*
Error		1.331			

*indicates significant result ($p < 0.05$)

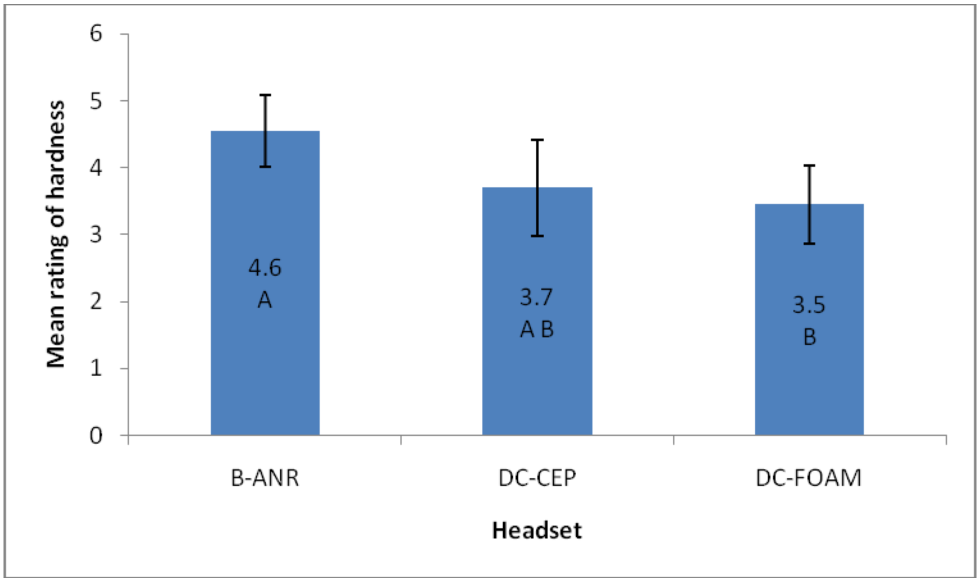


Figure 52. Pilot ratings of *hardness* for each headset (1= hard, 6= soft). Different letters represent significant differences due to headset at $p < 0.05$ using a Tukey pairwise comparison test. Vertical bars indicate 95% confidence intervals about the means.

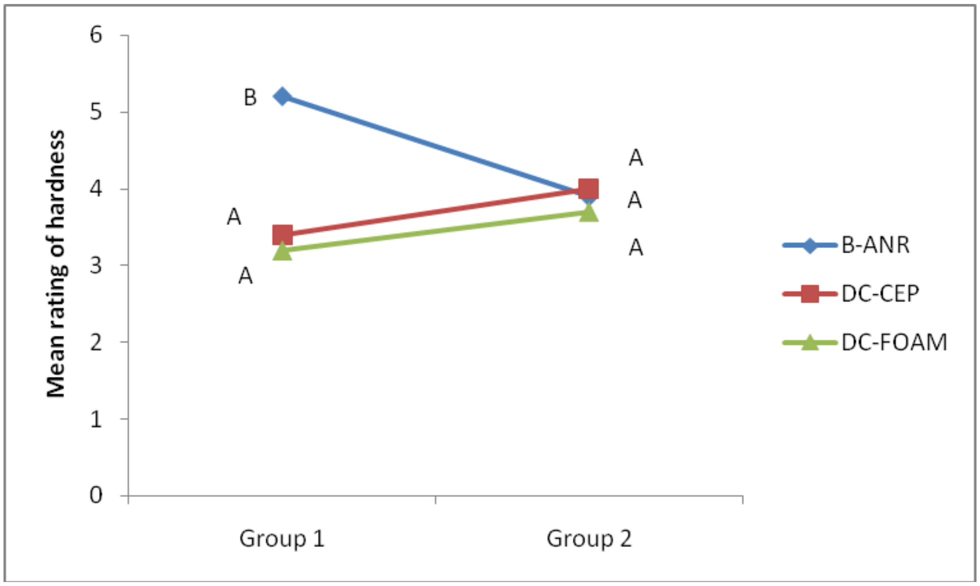


Figure 53. Pilot ratings of *hardness* (1=hard, 6= soft). For each group, different letters represent significant differences due to headset at $p < 0.05$ using Tukey's multiple comparison test.

Table 35. Mean response values of Item 12 (hard/soft) for each group/headset combination and 95% confidence interval values.

Headset	Group	Mean response question 12	Lower 95% confidence interval	Upper 95% confidence interval
B-ANR	1	5.2	4.8	5.7
DC-CEP	1	3.4	2.1	4.7
DC-FOAM	1	3.2	2.4	4.0
B-ANR	2	3.9	3.0	4.8
DC-CEP	2	4.0	3.2	4.8
DC-FOAM	2	3.7	2.7	4.7

Intolerable/Tolerable

Statistical analysis for the intolerable/tolerable item on the rating scale showed no significant main effects, but did reveal a significant group/headset interaction (*F*= 4.62, *p* = 0.0164, displayed in Table 36. Post-hoc Tukey's multiple comparison test showed that Group 1 perceived the B-ANR headset as significantly more tolerable than the DC-CEP headset (Figure

54). Mean response values to this rating scale item for each headset level/group combination and 95% confidence interval values are listed in Table 37.

Table 36. ANOVA summary table for bipolar ratings of tolerable/intolerable

Source	df	MS	*F*	*p*	
<u>Between</u>					
Group (G)	1	0.067	0.03	0.8649	
Error		2.237			
<u>Within</u>					
Headset (H)	2	0.450	0.31	0.7357	
Error		1.454			
G × H	2	6.717	4.62	0.0164	*
Error		1.454			

*indicates significant result ($p < 0.05$)

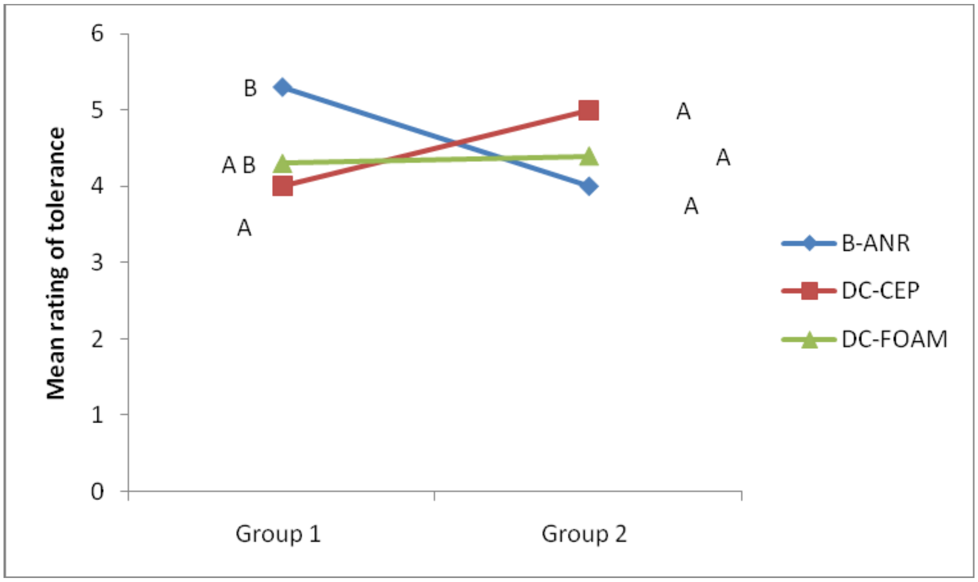


Figure 54. Pilot ratings of *tolerance* (1=intolerable, 6= tolerable). For each group, different letters represent significant difference due to headset at $p < 0.05$ using Tukey's multiple comparison test.

Table 37. Mean response values of Item 14 (hard/soft) for each group/headset combination and 95% confidence interval values.

Headset	Group	Mean response question 14	Lower 95% confidence interval	Upper 95% confidence interval
B-ANR	1	5.3	4.8	5.8
DC-CEP	1	4.0	2.8	5.2
DC-FOAM	1	4.3	4.0	4.6
B-ANR	2	4.0	2.7	5.4
DC-CEP	2	5.0	4.3	5.7
DC-FOAM	2	4.4	3.3	5.5

Discussion

Based on the previous analysis, it is clear that the B-ANR was favored by both groups of pilots for general comfort when compared to the DC-CEP and DC-FOAM headset. However, when asked specifically about uncomfortable pressure, pain, and headset tolerance, both the B-ANR and DC-FOAM headsets were rated higher (better) than the DC-CEP headset. The B-ANR and DC-CEP headsets were favored over the DC-FOAM headset for items concerning communication quality. The DC-FOAM headset configuration was rated the lowest (worse) in regards to sound distortion and low fidelity communications. This supports the flight performance finding that airspeed deviations and ATC readbacks occurred more often with the DC-FOAM headset for Group 2. However, it is unclear why there was not a group effect for these two items on the rating scale since Group 1 required fewer ATC readbacks and rated situation awareness higher for the DC-FOAM headset configuration.

CONCLUSIONS

Objective Flight Performance

The overall objective of this study was to determine if flight performance was affected by the combination of hearing loss, flight workload, ATC signal quality, and the type of headset used. As discussed previously, the current criteria for a continued-flight hearing waiver in the Army is good binaural speech recognition ability in quiet ("Standards of Medical Fitness," 2007). The results of this study suggest that flight workload is also an important factor in predicting pilot performance; all three workload levels led to significant differences in the measured flight performance parameters, regardless of hearing waiver category. Similarly, degraded signal quality decreased flight performance, as evidenced in the altitude deviation measurements. The combination of increased workload and decreased signal quality was especially problematic for certain flight performance parameters. Choice of headset was important for the group of hearing-impaired pilots, evident in the communications intelligibility data. This group of pilots required significantly more ATC command readbacks when using the passive headset/passive earplug combination compared to the other headsets. This is not surprising; however, it is an important finding with regard to the experimental objective of providing communication headset selection guidance considering pilots' hearing abilities.

The heading, altitude, and airspeed deviations in this experiment, though statistically significant, may appear to be small in magnitude. In many cases, such as when a pilot is flying under instrument flight rules (IFR) at altitude and under ATC-controlled conditions, a small deviation may be inconsequential; however, under more common terrain flight training or combat flight scenarios, deviations due to increased workload and degraded signal quality are much more critical.

Terrain flight can be divided into three categories - low level, contour, and nap-of-the-earth (NOE) flight. The mission and threat environment in which Army helicopters operate in a tactical environment will dictate which type of terrain flight a pilot will conduct. Low level flight occurs at the highest altitude of the three categories, and is conducted at a constant altitude and constant airspeed. Contour flight more closely follows the terrain features, and is conducted at varying altitude and a constant airspeed. NOE flight, in which a helicopter flies as close to the earth as obstacles will permit, is the most demanding type of terrain flight and is conducted at varied altitude and airspeed. It is critical that a pilot who is flying NOE maintain a high level of situation awareness to detect and avoid obstacles while maintaining maximum concealment and desired flight path. Additionally, the pilot (and co-pilot) must monitor instruments and maintain adequate speech communications, both by radio and within the helicopter in a fast-moving perceptual environment with limited room for error (Dura, 2009).

In this study, workload-influenced heading deviations ranged from 0.3 to 1.1 degrees. This amount of heading deviation could affect low level and contour terrain flight not controlled and monitored by ATC. For example, the final destination of a flight may be determined by flying segments of pre-planned distance based on time, airspeed, and heading. Depending on the distance of several flight segments, a heading deviation of one degree for each segment could result in a substantial deviation from the intended objective.

Statistically significant altitude deviations were up to 40 feet in this study. Obviously, when conducting terrain flight to avoid detection or enemy fire, an altitude deviation of 40 feet is more critical than when flying at higher altitudes because the aircraft is much closer to objects in the environment. The terrain features dictate altitude for contour and NOE flight, so it follows that adequate SA is critical for maintaining altitude in this situation. The results of this study showed that high workload and poor signal quality conditions resulted in significantly lower subjective SA ratings. Of particular concern is the hearing-impaired aviator in this situation who is not equipped with an assistive communication device such as ANR or CEP.

Airspeed deviations in this experiment of 0.4 to 1.3 knots were significantly different. As with heading, this amount of airspeed deviation can make a difference if a pilot is flying mission segments that depend on timing, as would be the case when there are no external reference points to gauge position or distance. For example, if a pilot is supposed to fly for 30 minutes at 120 knots and then change heading and fly an additional 30 minutes at 120 knots, over several hours, an airspeed deviation of 1.3 knots can alter the final destination of the helicopter enough to compromise mission success. Airspeed deviations at maximum cruising speed (up to 159 knots at 2,000 feet altitude) could result in even greater errors, although flights at this altitude are most likely ATC-controlled.

Objective flight performance data (heading and altitude deviation) showed that performance differences due to workload occurred between low and medium workload, whereas performance differences due to signal quality (altitude deviation) occurred between poor and average signal quality. These findings indicate that a key consideration for pilots is ensuring manageable flight workload and adequate communications signal quality. A study conducted at the USAARL in the mid 1970's reported that pilots and crew members were spending 30% of their time in communication concerning navigation during NOE flight (Sanders, Hofmann, Harden, & Frezell, 1975). The amount of time spent on communication concerning navigation for NOE flight may be reduced now due to technology such as global positioning systems, but the fact remains that NOE flight is a high workload environment that is less tolerant of compromised communication brought on by decreased speech signal quality due to hearing loss or inadequate communication headsets.

Although not statistically significant at $p < 0.05$, when the pilots with hearing loss, as specified by the Army's current hearing waiver criteria, used the passive headset/passive earplug combination, the result was a trend of increased airspeed deviation. This finding is supported by findings on the number of readbacks required for correct ATC comprehension. Again, the number of readbacks that resulted in statistical significance in this study may appear small in magnitude; however, the potential consequences of missing one-half of an ATC message can be enormous. Imagine a scenario in which a helicopter is flying at 120 knots and told to change heading immediately. If a pilot does not hear or understand the ATC command and spends an additional fifteen seconds requesting and receiving ATC repetition, he or she could have travelled approximately a half mile during that time period. Again, this may not be a critical amount of time, especially with a co-pilot to assist in message reception and navigation, but in a worst-case scenario where there is not room for such error, an additional fifteen seconds could be critical for the safety of the pilot and crewmembers.

The statistical results from the number of readbacks required for ATC command comprehension are similar to the flight performance data. Again, the number of readbacks increased once speech intelligibility dropped below about 50% and when workload increased beyond the medium level in this experiment, for both groups. Of course, neither of these parameters usually occurs in isolation, and the readback data once again showed that flight performance suffers when a pilot has to cope with a combination of high workload and poor signal quality.

The group of pilots without a hearing waiver did not show significant differences with any of the headset configurations. However, as with airspeed deviation, the hearing-impaired pilots required more ATC command readbacks with the passive headset/passive earplug combination compared to the other headsets. This is another important finding regarding the previously described experimental objective of providing data on which to base communication headset selection for Army pilots.

Subjective Reports on Workload and Situation Awareness

The fact that pilots in both hearing level groups rated the three workload levels used in the experiment differently on the MCH workload rating scale validates the experimental design and supports the objective findings of significant differences in flight performance due to workload levels. Just as with the altitude deviation, the subjective workload ratings indicate that once speech intelligibility is below about 50%, workload increases significantly. An interesting finding in the MCH data is that the group of non-hearing waivered pilots rated workload no differently with any of the three headset configurations, while the group of pilots with a hearing waiver reported higher workload with the passive headset/passive earplug combination, and this finding supports the airspeed deviation finding. As with objective flight performance, the results of the MCH rating scale indicate that the combination of high workload and poor signal quality is not favored by either group of pilots. These results suggest that a tool such as the MCH could be used to assess pilots’ perceptions of workload in an effort to balance all workload factors (perceptual, psychomotor, and communication) and maximize pilot efficiency and safety.

Supporting the notion that workload is interrelated to situation awareness, pilots rated SA differently for all three workload levels. The fact that the normal-hearing pilot group reported more difference between each workload level and the hearing-impaired pilot group did not report further degradation between medium and high workload could be an indication of experience. The groups naturally partitioned into a less experienced (as measured by flight hours), better hearing group, and a more experienced, poorer hearing group. It is reasonable to assume that the normal-hearing pilot group would notice more spread in workload as it relates to situation awareness than the hearing-impaired pilot group did, due solely to flying experience.

Consistent with the objective flight performance measures and the subjective ratings, pilots reported decreased SA when speech intelligibility dropped below about 50%. Also consistent with other findings is that the hearing-impaired group rated SA significantly lower with the passive headset/passive earplug combination, while the normal-hearing group reported significantly *higher* situation awareness with the same headset. This result is surprising, but it could be that pilots with less flight experience perceive that they are receiving more information that contributes to their SA with the passive headset/passive earplug combination because speech was louder and possibly background aircraft noise was louder. Pilots with hearing loss (and more experience) may have accommodated to their hearing loss and as a result, rely more on the feel of the aircraft rather than the sound.

The headset comfort/speech intelligibility results regarding the passive headset/passive earplug combination supports the results found with airspeed deviation, ATC command readbacks, MCH ratings, and SART ratings, particularly with the hearing-impaired group of pilots. Based on these data, it is apparent that the headset of choice for hearing-impaired pilots is not a passive headset/passive earplug combination, but rather a headset with some sort of assistive technology, such as ANR or CEPs.

CEPs are readily available and have obtained approval for use in Army helicopters, but this technology is not without limitations. A common saying about hearing protection is that the best hearing protector is the one that is worn. As such, audiologists and flight surgeons should consider headset comfort when recommending a particular device for pilots. While the CEP-equipped headset contributed to good flight performance for hearing-impaired pilots, and was rated high regarding communications, it was not rated favorably compared to the other headsets regarding comfort. For long flights, this could be a concern because pilots may remove the CEP due to discomfort and risk further hearing damage.

Review of Experimental Objectives, Goals, and Hypotheses

As discussed previously, the research objectives of this experiment are listed below with a brief discussion of how they relate to the experimental results.

1. *To refine audiometric waiver criteria, in part based on flight-related performance data to ensure the maintenance of a readily mobile, effective fighting force of Army aviators*.

The results of this study did not show adequate group effects on flight performance data to warrant recommendations on audiometric waiver criteria refinement. Due to the limited hearing ability groups that spanned a large range of audiometric thresholds in each group, it was not possible to infer a point at which hearing loss affected flight performance.

2. *To yield data on which to base flight recommendations and headsets for hearing-impaired pilots.*

CEPs are commonly used by Army helicopter pilots, and are commonly recommended by Army audiologists for pilots with hearing waivers, but use of CEPs or other assistive communications device is not required by Army regulation. Objective findings of airspeed deviation and ATC command readbacks and subjective findings as indicated by MCH, SART, and speech intelligibility ratings strongly support a recommendation that pilots with hearing loss sufficient to warrant a hearing waiver be required by Army regulation to use assistive communication technology with their flight helmet, and not be permitted to fly with a passive headset/passive earplug combination.

3. *To provide selection guidance for communication headsets, considering aviators' hearing abilities, that will improve combat effectiveness as realized through optimized performance.*

As with the first objective, this study did not yield data sufficient to predict pilot performance based on a continuum of hearing sensitivity levels. There were fewer group/headset interaction effects with the normal hearing group compared to the hearing-impaired group; therefore, conclusions on headset recommendation is not as clear for the normal hearing group, other than to recommend a headset with adequate hearing protection properties to prevent permanent hearing loss. As discussed, it is clear that hearing-impaired pilots, as defined by current Army regulation, did not perform as well with a passive headset/passive earplug combination as they did with assistive communication devices; however, the same effect was not seen with pilots who did not have a hearing waiver.

The experimental goals of this project are as follows with discussion on how the experimental results achieved or failed to achieve them.

1. *To determine the level of flight mission-induced workload at which performance decreased with a CEP-equipped headset, ANR-equipped headset, and a passive headset/ passive earplug combination between aviators with normal hearing and aviators with varying degrees of hearing loss*.

A common finding for the effect of flight workload on flight control performance (heading, altitude, and airspeed deviation) was that as flight workload increased from low to medium, performance suffered, with all headsets and with both groups of pilots. Recall that the difference in low and medium workload was decreased visibility (6 SM to 1.75 SM), more flight maneuvers per ATC command (one to two), and longer ATC commands reflecting the flight maneuvers, and that both groups' MCH ratings validated the workload level differences used in the experiment. There were no interactions of workload with group or headset on objective flight performance, so a conclusion cannot be drawn on the performance differences between groups or between headsets due to workload.

2. *To determine the level of communications signal quality at which performance decreased with a CEP-equipped headset, ANR-equipped headset, and a passive headset/passive earplug combination between aviators with normal hearing and aviators with varying degrees of hearing loss.*

As with workload data, there was a common finding regarding the effect of communications signal quality on flight control performance. The point at which altitude deviations increased and ATC command readbacks increased was between average and poor signal quality, which generally corresponds to about 50% speech intelligibility. There were no interactions of signal quality with group or headset on objective flight performance, so conclusions cannot be drawn on performance differences between groups or between headsets due to signal quality.

3. To determine the level of hearing loss at which performance decreased with a CEP-equipped headset, ANR-equipped headset, and passive headset/passive earplug combination beyond acceptable levels as a function of workload and communication signal quality.

As discussed previously, the findings from this study show that flight performance for both groups repeatedly decreased as flight workload increased and signal quality decreased. There were no interactions of headset with workload or signal quality on objective flight performance, so it is not possible to draw conclusions regarding performance differences between groups or between headsets due to workload/signal quality combinations.

Lastly, formal hypotheses related to the research presented previously are listed below. A brief discussion of each hypothesis as it relates to experimental results follows.

H1: As flight workload increased and communication signal quality decreased, normal-hearing aviators would perform better than hearing-impaired aviators, regardless of the communications system used.

The experimental data did not support this hypothesis entirely. In many instances, flight performance decreased as flight workload increased and signal quality decreased for all headset configurations, but that effect was seen regardless of hearing ability.

H2: For hearing-impaired aviators, as flight workload increased and communication signal quality decreased, performance would decrease first with passive headset/passive earplug combination, then the ANR-equipped headset, followed by the CEP-equipped headset.

There were no interactions of workload or signal quality with either headset or group; therefore, the data do not completely support this hypothesis. Again, several flight performance measures showed degradation as flight workload increased and signal quality decreased for both groups, regardless of headset configuration. And, as mentioned previously, hearing-impaired pilots experienced degradation of certain flight performance parameters with the passive headset/passive earplug combination; however, this occurred regardless of workload or signal quality level.

H3: For normal-hearing aviators, as flight workload increased and communication signal quality decreased, performance would decrease first with the passive headset/passive *earplug combination, then the ANR-equipped headset, followed by the CEP-equipped headset.*

Again, this hypothesis is not supported entirely. High workload and poor signal quality clearly affected pilot flight performance for both groups of pilots, but there were no significant interaction effects of workload with hearing level group.

H4: As flight workload increased and communication signal quality decreased, ratings of SA would decrease first with the passive headset/passive earplug combination, then the ANR-equipped headset, followed by the CEP-equipped headset, regardless of hearing loss category.

This hypothesis was partially supported by the experimental results that revealed, for both groups, ratings of SA decreased as flight workload increased and as signal quality decreased. For the hearing-impaired group of pilots, SA ratings decreased first with the passive headset/passive earplug combination, followed by the CEP-equipped headset and the ANR-equipped headset. In contrast, ratings of SA decreased with the ANR-equipped and CEP-equipped headsets first, followed by the passive headset/passive earplug combination for the normal-hearing group.

Limitations of this Study

Several limitations to this study were realized during the course of data collection and data analysis. The experiment was designed to be as realistic as possible, and factors such as a full motion simulator, realistic aircraft noise, and actual Army helicopter pilots as study participants greatly contributed to the realism and the overall external validity of the study. Some aspects of the study; however, were put in place for ease of data collection, but limited the external validity of the study. The first aspect is that there was no co-pilot to aid the pilot as is usually found in helicopters. The main advantage of having two pilots, which was lost in this experiment, is that when two pilots are listening to ATC commands, there is twice the opportunity to hear, understand, and comply with the command. As a result, an available co-pilot reduces the amount of communication workload on the primary pilot. In a similar manner, the psychomotor flight workload on the primary pilot is reduced if there is another pilot available to share in the tasks required to fly the helicopter.

Another factor that was not implemented in this study was that of multiple radios. Under actual flight conditions, a military pilot must attend to more than one radio and filter out a great deal of auditory ‘clutter’, thereby increasing the amount of communication workload. These factors were not added to the experiment in an effort to reduce confusion in data collection. For that reason, communication workload may have been artificially reduced in this experiment. Future studies could add a co-pilot and multiple radios, but considerable thought would have to go into isolating pilots and ATC communications for data collection purposes.

Lastly, the number of hearing loss categories was changed from three to two due to the difficulty in recruiting pilots with a slight hearing loss that did not require a hearing waiver. This modification in participant group stratification led to problems meeting some of the experimental goals and objectives, particularly the ones concerned with refining the hearing waiver criteria, and determining the level of hearing loss at which performance decreased between different headsets.

RECOMMENDATIONS FOR FUTURE RESEARCH

The results of this study identified the relative effects of workload, signal quality, and headset choice on helicopter flight performance in an effort to include factors other than binaural speech recognition on pilot performance. Because this experiment was conducted in a simulator, there are factors inherent in flying an actual helicopter in the air that could not be accurately captured in a simulator. A similar study conducted in a research helicopter in actual air traffic would provide even more realism that could identify workload, signal quality, and headset effects more accurately. Because there was no co-pilot or additional crew in this experiment, the effects of degraded cockpit crew coordination was not explored. Future studies could include measures of crew communications and how the variables presented in this study affect intra-cockpit coordination.

The results of this experiment did not warrant refinement of the current hearing waiver criteria, in part because of the limited number of hearing level groups, but also because all of the communication aspects of flight workload could not be isolated for analysis. In other words, due to the design of the experiment, each workload level was the same for perceptual, psychomotor, and communication workload. There was no way to measure the interaction of high communication workload with low psychomotor workload and low perceptual workload, for example. Future studies should be designed to do so by including more narrowly defined categories of hearing levels, and by designing the experiment in such a way that the effects of the communication workload on pilot performance could be assessed individually.

The results of this study consistently revealed that a high workload/poor signal quality flight condition resulted in decreased pilot performance supported by subjective pilot ratings. There were also repeated findings of decreased performance, again supported by subjective ratings, with the passive headset/passive earplug combination worn by hearing-impaired pilots. These results should be sufficient to justify modifications to current Army regulation that would *require* pilots with hearing waivers to use a headset or helmet modified with assistive communication technology, and not be permitted to fly with a passive headset/passive earplug combination. CEPs are readily available and approved for Army aviation; however, ANR technology is currently available in earplug form that is lighter (and perhaps safer in the event of an aircraft accident) than the headset mounted ANR components, which may change the airworthiness status of ANR. For this reason, further research on hearing-impaired pilot performance with ANR technology is warranted.

One research objective of this experiment was to refine the current audiometric waiver criteria. This objective was not obtained due, in part, to limited hearing ability groups. Another way to refine the waiver criteria would be to conduct a functional hearing evaluation for Army helicopter pilots, and relate the findings to a more realistic (and accurate) audiometric flight evaluation.

In order to determine the functional hearing requirements of an Army pilot, an in-cockpit communication task analysis, conducted in a realistic flight environment, would be necessary. This type of evaluation, which would identify the relative importance of hearing-related job tasks, could more clearly define what a pilot needs to hear and what combination of workload, signal quality, headset, and hearing sensitivity is adequate for the hearing requirements. An evaluation of this type could lead to significant changes in future flight physical hearing evaluations.

Ideally, a functional hearing evaluation to determine pilot fitness for duty could consist of two-way ATC-type communications between the audiologist and the pilot, conducted in simulated noise of the pilot’s primary aircraft, with simultaneous workload tasks that mimic helicopter operation; however, an evaluation of this type may not be realistic, especially if a desktop simulator device is required. A more straightforward hearing-in-noise evaluation for all pilots, regardless of audiometric thresholds may be appropriate, with the background noise, speech/noise spatial separation, and signal-to-noise ratio closely replicating that of the pilots’ primary aircraft, and with the speech signal mimicking ATC or crew communication heard by the pilot while in flight. The same type of criteria for allowing continued flight status could be implemented (i.e. a minimum requirement of 84% speech intelligibility), but a more realistic hearing-in-noise test as described would give better assurance that a pilot has the functional hearing requirements to fly safely.

REFERENCES

Abel, S. M., Alberti, P. W., Haythornthwaite, C., & Riko, K. (1982). Speech intelligibility in noise: effects of fluency and hearing protector type. *Journal of the Acoustical Society of America, 71*, 708-715.

Abel, S. M., & Spencer, D. L. (1997). Active noise reduction versus conventional hearing protection: Relative benefits for normal-hearing and impaired listeners. *Scandinavian Audiology, 26*(3), 155-167.

Ahroon, W. A., Gordon, E., Mozo, B. T., & Katz, L. C. (2000). Communications Effectiveness When Using the Communications Earplug or Expandable-Foam Earplug With the HGU-56/P Aviator Helmet. *USAARL Report No. 2000-27*.

Ahroon, W. A., Hamernik, R. P., & Davis, R. I. (1993). Complex noise exposures: An energy analysis. *The Journal of the Acoustical Society of America, 93*, 997-1006.

Anderson, B. W., & Garinther, G. R. (1997). *Effects of active noise reduction in armor crew headsets.* Paper presented at the AGARD, Neuilly-Sur-Seine, France.

ANSI. (1997). *ANSI S3.5-1997, American national standard methods for calculation of the speech intelligibility index*: American National Standards Institute, New York.

Army Hearing Program. (2008). *Special Text 4-02.501*. Fort Sam Houston, TX: United States Army Medical Department Center and School

Battiste, V., & Bortolussi, M. (1988). *Transport Pilot Workload: A Comparison of Two Subjective Techniques.* Paper presented at the Human Factors Society, Santa Monica, CA.

Bell, H. H., & Waag, W. L. (1995). Using observer ratings to assess situational awareness in tactical air environments. In D. J. Garland & M. R. Endsley (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 93-99). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Berger, E. H. (2000). Hearing Protective Devices. In E. H. Berger, L. H. Royster, J. D. Royster, D. P. Driscoll & M. Layne (Eds.), *The Noise Manual* (5th ed., pp. 379-454). Fairfax, VA: American Industrial Hygiene Association.

Bhattacharya, S. K., Roy, A., Tripathi, S. R., & Chatterjee, S. K. (1985). Behavioural measurements in textile weavers wearing ear protectors. *Indian Journal of Medical Research, 82,* 56-64.

Brungart, D. S., Kordik, A. J., Simpson, B. D., & McKinley, R. L. (2003). Auditory Localization in the Horizontal Plane with Single and Double Hearing Protection. *Aviation, Space, and Environmental Medicine, 74*, 937-946.

Buck, K., & Zimpfer-Jost, V. (2005). *Active Hearing Protection Systems and Their Performance*: French-German Research Institute

Casali, J. G. (2006). Sound and Noise. In G. Salvendy (Ed.), *Handbook or Human Factors and Ergonomics* (Third ed., pp. 612-642). Hoboken, NJ: John Wiley & Sons, Inc.

Casali, J. G., & Berger, E. H. (1996). Technology advancements in hearing protection circa 1995: Active noise reduction, frequency/amplitude-sensitivity and uniform attenuation. *American Industrial Hygiene Association Journal, 57*(2), 175-185.

Casali, J. G., & Gerges, S. (2006). Protection and enhancement of hearing in noise. In R. C. Williges (Ed.), *Reviews of Human Factors and Ergonomics, Volume 2* (pp. 195-240). Santa Monica, CA.

Casali, J. G., & Grenell, J. F. (1990). Noise-attenuating earmuff comfort: A brief review and investigation of band-force, cushion, and wearing-time effects. *Applied Acoustics, 29*, 117-138.

Casali, J. G., Lam, S. T., & Epps, B. W. (1987). Rating and Ranking Methods for Hearing Protector Wearability. *Sound and Vibration, 21*(12), 10-18.

Casali, J. G., Lancaster, J. A., Valimont, R. B., & Gauger, D. (2007). Headsets in the light aircraft cockpit: Speech intelligibility. *Proceedings of the 2007 International Congress of Noise Control Engineering*.

Casali, J. G., Mauney, D. W., & Burks, J. A. (1995). Physical vs. Psychophysical Measurement of Hearing Protector Attenuation - a.k.a. MIRE vs. REAT. *Sound and Vibration*(July), 20-27.

Casali, J. G., & Robinson, G. S. (1994). Narrow-Band Digital Active Noise Reduction in a Siren-Cancelling Headset: Real-Ear and Acoustical Manikin Insertion Loss. *Noise Control Engineering Journal, 42*(3), 101-115.

Casali, J. G., & Wierwille, W. W. (1983). A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors, 25*(6), 623-641.

Casali, J. G., & Wierwille, W. W. (1984). On the measurement of pilot perceptual workload: A comparison of assessment techniques addressing sensitivity and intrusion issues. *Ergonomics, 27*(10), 1033-1050.

Casto, K. L., Ostler, D. A., & Ahroon, W. A. (2001, Aug 26-30). *Communications Earplug Use by Hearing-impaired Army Aviators.* Paper presented at the Military Audiology Short Course, Albuquerque, NM.

Chan, J. W., & Simpson, C. A. (1990). *Comparison of Speech Intelligibility in Cockpit Noise Using SPH-4 Flight Helmet with and without Active Noise Reduction* (No. 90-G-1). Moffett Field, CA: NASA Ames Research Center.

Chung, D. Y., & Gannon, R. P. (1979). The effect of ear protectors on word discrimination in subjects with normal hearing and subjects with noise-induced hearing loss. *Journal of the Acoustical Society of America, 5*, 11-16.

Conover, W. J. (1999). *Practical Nonparametric Statistics, Third Edition*. Hoboken, New Jersey: John Wiley and Sons, Inc.

Cooper, G. E., & Harper, R. P. (1969). *The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities* (No. NASA TN D-5153). Washington, DC: National Aeronautics and Space Administration.

Corwin, W. H., Sandry-Garza, D. L., Biferno, M. H., Boucek, G. P., Logan, A. L., Jonsson, J. E., et al. (1989). *Assessment of Crew Workload Measurement Methods, Techniques, and Procedures, Volume 1: Process, Methods, and Results*. Retrieved. from http://stinet.dtic.mil/cgi-bin/GetTRDoc?AD=A217699&Location=U2&doc=GetTRDoc.pdf.

Damos, D. L. (Ed.). (1991). *Multiple Task Performance*. Bristol, PA: Taylor & Francis.

Dura, D. (2009). Terrain Flight.   Retrieved 20 June, 2009, from http://www.ianrp.org/Safety/terrain_flight.htm

Edwards, D. E. (1990). *Pilot: Mental and Physical Performance*. Ames: Iowa State University Press.

Eggemeier, F. T., & Wilson, G. F. (1991). Performance-based and subjective assessment of workload. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 217-278). Bristol, PA: Taylor & Francis.

Eggemeier, F. T., Wilson, G. F., Kramer, A. F., & Damos, D. L. (1991). Workload assessment in multi-task environments. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 207-216). Bristol, PA: Taylor & Francis.

Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). *Proceedings of the National Aerospace and Electronics Conference (NAECOM)*, 789-795.

Endsley, M. R. (1995). A taxonomy of situation awareness errors. In R. Fuller, N. Johnston & N. McDonald (Eds.), *Human factors in aviation operations* (pp. 287-292). Aldershot, England: Avebury Aviation, Ashgate Publishing.

Endsley, M. R. (1996). Situation awareness measurement in test and evaluation. In T. G. O'Brien & S. G. Charlton (Eds.), *Handbook of Human Factors Testing and Evaluation* (pp. 159-180). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R. (1999). Situation Awareness in Aviation Systems. In D. J. Garland, J. A. Wise & V. D. Hopkin (Eds.), *Handbook of Aviation Human Factors* (pp. 257-276). Mawah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R. (2000). Theoretical Underpinnings of Situation Awareness: A Critical Review. In M. R. Endsley & D. J. Garland (Eds.), *Situation Awareness: Analysis and Measurement* (pp. 3-32). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R., & Garland, D. J. (2000). *Situation Awareness: Analysis and Measurement*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Fitzpatrick, D. T. (1988). An Analysis of Noise-Induced Hearing Loss in Army Helicopter Pilots. *Aviation, Space, and Environmental Medicine, 59*, 937-941.

Gates, G. A., Schmid, P., Kujawa, S. G., Nam, B.-h., & D'Agostino, R. (2000). Longitudinal threshold changes in older men with audiometric notches. *Hearing Research, 141*(1-2), 220-228.

Gower, D. W., & Casali, J. G. (1994). Speech intelligibility and protective effectiveness of selected active noise reduction and conventional communications headsets. *Human Factors, 36*(2), 350-367.

Hansen, C. H. (2001). *Understanding Active Noise Cancellation*. London: Spon Press.

Harley, T. R., Shields, F. D., & Hendrix, J. E. (1994). *Digital Active Noise Reduction Ear Plugs*: United States Air Force Research Laboratory.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload.* (pp. 139-183). Oxford, England: North-Holland.

Hartel, C. E., Smith, K., & Prince, C. (1991). *Defining aircrew coordination: Searching mishaps for meaning.* . Paper presented at the Sixth International Symposium on Aviation Psychology.

Hawkins, D. B., Montgomery, A. A., Mueller, H. G., & Sedge, R. K. (1988). Assessment of speech intelligibility by hearing-impaired listeners. In B. Berglund, U. Berglund, J. Karlsson & L. T. (Eds.), *Noise as a public health problem* (Vol. 2, pp. 241-246). Stockholm: Swedish Council for Building Research.

Hearing Conservation Program. (1998). *Department of the Army Pamphlet 40-501*. Washington, DC: Headquarters, Department of the Army.

Henderson, D., & Hamernik, R. P. (1982). Asymptotic threshold shift from impulse noise. In R. P. Hamernik, D. Henderson & R. Salvi (Eds.), *New Perspectives on Noise-Induced Hearing Loss* (pp. 265-281). New York: Raven Press.

Henderson, D., & Hamernik, R. P. (1986). Impulse noise: Critical review. *Journal of the Acoustical Society of America, 80*(2), 569-584.

High, R. (2001). Should You Be Using PROC MIXED to Analyze Continuous Data? Retrieved June 9, 2009, from http://cc.uoregon.edu/cnews/summer2001/procmixed.html

Hightower, J. A. (1999). *Guide to Aeromedical Triservice Terms*. Pensacoloa, FL: Naval Operational Medicine Institute.

Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner, A. C., Zaklad, A. L., & Christ, R. E. (1992). Comparison of Four Subjective Workload Rating Scales. *Human Factors, 34*(4), 429-439.

Howell, K., & Martin, A. M. (1975). An investigation of the effects of hearing protectors on vocal communication in noise. *Journal of Sound and Vibration, 41*(2), 181-196.

Humes, L. E., Joellenbeck, L. M., & Durch, J. S. (Eds.). (2006). *Noise and Military Service: Implications for Hearing Loss and Tinnitus*. Washington, DC: The National Academies Press.

Itoh, Y., Hayashi, Y., Tsukui, I., & Saito, S. (1989). Heart rate variability and subjective mental workload in flight task validity of mental workload measurement using H.R.V. method In M. J. Smith & G. Salvendy (Eds.), *Work with Computers: Organizational, Management, Stress and Health Aspects* (pp. 209-216). Amsterdam: Elsevier Science.

James, S. H. (2005). *Defining the Cockpit Noise Hazard, Aircrew Hearing Damage Risk and the Benefits Active Noise Reduction Headsets Can Provide*: Future Systems & Technology Division QinetiQ, Farnborough, Hants, GI14 OLX, UK.

Jones, D. G. (2000). Subjective Measures of Situation Awareness. In M. R. Endsley & D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 113-127). Mahwah, NJ: Lawrence Erlbaum Associates.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

King, R. B., Saliba, A. J., & Brock, J. R. (1999). A comprehensive noise survey of the S-70A-9 Black Hawk helicopter. *Aviation, Space, and Environmental Medicine, 70*(2), 107-116.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In D. L. Damos (Ed.), *Multiple-Task Performance* (pp. 279-317). Washington, DC: Taylor & Francis.

Kryter, K. D. (1946). Effects of ear protective devices on the intelligibility of speech in noise. *Journal of the Acoustical Society of America, 18*, 413-417.

Kujawa, S. G., & Liberman, M. C. (2006). Acceleration of Age-Related Hearing Loss by Early Noise Exposure: Evidence of a Misspent Youth. *Journal of Neuroscience, 26*(7), 2115-2123.

Laroche, C., Soli, S., Giguere, C., Lagace, J., Vaillancourt, V., & Fortin, M. (2003). An approach to the development of hearing standards for hearing-critical jobs. *Noise and Health, 6*(21).

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS System for Mixed Models*. Cary, North Carolina: SAS Institute Inc. .

Mason, K. T., & Mozo, B. T. (1995). Communication Earplug and Active Noise Reduction: Hearing Protection Technologies for Air Warrior. *USAARL Report No. 95-26*.

McGrath, B. J., Estrada, A., Braithwaite, M. G., Raj, A. K., & Rupert, A. H. (2004). *Tactile Situation Awareness System Flight Demonstration Final Report* (No. 2004-10). Ft. Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Middelweerd, M. J., Festen, J. M., & Plomp, R. (1990). Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. *Audiology, 29*(1), 1-7.

Milliken, G. A., & Johnson, D. E. (1984). *Analysis of Messy Data Volume I: Designed Experiments*. New York: Van Nostrand Reinhold.

Mills, J. H., Gengel, R. W., Watson, C. S., & Miller, J. D. (1970). Temporary Changes of the Auditory System due to Exposure to Noise for One or Two Days. *Journal of the Acoustical Society of America, 48*(2 (Part 2)), 524-530.

Moore, T. J., & von Gierke, H. E. (1991). Military performance in acoustic noise environments. In R. Gal & A. D. Mangelsdorff (Eds.), *Handbook of military psychology*. (pp. 295-311). Oxford, England: John Wiley & Sons.

Mozo, B. T. (2004, August 31, 2004). CEP Introduction. Retrieved November 15, 2007, from http://www.cep-usa.com/index.htm

Mozo, B. T., & Murphy, B. A. (1997a). *The Assessment of Sound Attenuation and Speech Intelligibility of Selected Active Noise Reduction Devices and the Communications Earplug when used with the HGU-56/P Aviator Helmet* (No. 97-08). Ft. Rucker, AL: U.S. Army Aeromedical Laboratory.

Mozo, B. T., & Murphy, B. A. (1997b). *Evaluation of the Communications Earplug in the H-53 and CH-46 Helicopter Environments* (No. 97-36). Ft. Rucker, AL: U.S. Army Aeromedical Laboratory.

Mozo, B. T., Murphy, B. A., & Ribera, J. E. (1995). *User Acceptability and Comfort of the Communications Earplug (CEP) When Used in the UH-1 Helicopter* (No. 95-17). Ft. Rucker, AL: U.S. Army Aeromedical Laboratory.

NIH. (1990). Noise and Hearing Loss. National Institutes of Health Consensus Development Panel. *Journal of the American Medical Association, 263,* 3185-3190.

Nixon, C. W., McKinlely, R. L., & Steuver, J. W. (1992). Performance of actice noise reduction headsets. In A. L. Dancer, Henderson, D., Salvi, R.J., Hamernik, R.P. (Ed.), *Noise-induced hearing loss* (pp. 389-400). St. Louis, MO: Mosby-Year Book.

Noble, W., Byrne, D., & Lepage, B. (1994). Effects on sound localization of configuration and type of hearing impairment. *Journal of the Acoustical Society of America, 95*(2), 992-1005.

Peters, L. J., & Garinther, G. R. (1990). The effects of speech intelligibility on crew performance in an M1A1 tank simulator. *Technical memo 11-90, Aberdeen Proving Ground, MD: Human Engineering Laboratory*.

Pew, R. W., Mavor, Anne S. (Ed.). (1998). *Modeling Human and Organizational Behavior: Application to Military Simulation*: The National Academy Press

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech & Hearing Research, 29*(2), 146-154.

Poulton, E. C. (1978). A new look at the effects of noise: A rejoinder. *Psychological Bulletin, 85*(5), 1068-1079.

Pritchett, A. R., & Hansman, R. J. (2000). Use of Testable Responses for Performance-Based Measurement of Situation Awareness. In M. R. Endsley & D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 189-209). Mahwah, NJ: Lawrence Erlbaum Associates.

Pyykko, I., Toppila, E., Zou, J., & Kentala, E. (2007). Individual susceptibility to noise-induced hearing loss. *Audiological Medicine, 5*(1), 41-53.

Raby, M., & Wickens, C. D. (1994). Strategic workload management and decision biases in aviation. *International Journal of Aviation Psychology, 4*(3), 211-240.

Raynal, M., Kossowski, M., & Job, A. (2006). Hearing in Military Pilots: One-Time Audiometry in Pilots of Fighters, Transports, and Helicopters. *Aviation, Space, and Environmental Medicine, 77*, 57-61.

Reid, G. B., & Nygren, T. E. (1988). The Subjective Workload Assessment Technique: A scaling procedure for measuring mental workload. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload.* (pp. 185-218). Oxford, England: North-Holland.

Ribera, J. E., Mozo, B. T., & Murphy, B. A. (1996). *Subjective Evaluation of the Communications Earplug with Flexible Harness (CEP/FH) among CH-47D crewmembers* (No. 96-29). Ft. Rucker, AL: U.S. Aeromedical Research Laboratory.

Ribera, J. E., Mozo, B. T., & Murphy, B. A. (2004). Speech Intelligibility With Helicopter Noise: Tests of Three Helmet-Mounted Communication Systems. *Aviation, Space, and Environmental Medicine, 75*(2), 132-137.

Rink, T. L. (1979). Hearing protection and speech discrimination in hearing-impaired persons. *Sound and Vibration, 13*, 22-25.

Robinette, M. B., Ahroon, W. A., & Ostler, D. A. (2003). *Noise attenuation of the HGU-56/P aircrew integrated helmet system when worn with the combat vehicle crewman hood (balaclava)*. Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Robinson, G. S., & Casali, J. G. (2000). Speech Communications and Signal Detection in Noise. In E. H. Berger, Royster, L.H., Royster, J.D., Driscoll, D.P., Layne, M. (Ed.), *The Noise Manual* (5th ed., pp. 567-600). Fairfax, VA: American Industrial Hygiene Association.

Roscoe, A. H. (Ed.). (1987). *The Practical Assessment of Pilot Workload, AGARDograh No. 282*. Neuilly Sur Seine, France: AGARD.

Sanders, M. G., Hofmann, M. A., Harden, D. F., & Frezell, T. L. (1975). *Communication during terrain flight*. Fort Rucker, Alabama: Army Aeromedical Research Laboratory.

Sanders, M. S., & McCormick, E. J. (1993). *Human Factors in Engineering and Design* (7th ed.). New York: McGraw-Hill, Inc.

Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *International Journal of Aviation Psychology, 1*(1), 45-57.

Schuknecht, H. F., & Gacek, M. R. (1993). Cochlear pathology in presbycusis. *Ann Otol Rhinol Laryngol, 102,* 1-16.

Selcon, S. J., & Taylor, R. M. (1990). *Evaluation of the situational awareness rating technique (SART) as a tool for aircrew systems design*. Neuilly Sur Seine, France: NATO-AGARD.

Shaw, E. A. (1974). Transformation of sound pressure level from the free field to the eardrumin the horizontal plane. *Journal of the Acoustical Society of America, 56*(6), 1848-1861.

SII: Speech Intelligibility Index. Retrieved March 1, 2009, from http://www.sii.to/index.html

Skipper, J. H., Rieger, C. A., & Wierwille, W. W. (1986). Evaluation of decision-tree rating scales for mental workload estimation. *Ergonomics, 29*(4), 585-599.

Smith, A. T. (1991). A review of the non-auditory effects of noise on health. *Work & Stress, 5*(1), 49-62.

Soli, S. D. (2003). *Hearing and Job Performance*. Los Angeles: House Ear Institute.

Standards of Medical Fitness. (2007). *Army Regulation 40-501*. Washington, DC: Headquarters, Department of the Army.

Staples, N. (1981). Hearing conservation - Is management shortchanging those at risk? *Noise and Vibration Control Worldwide, 12*(6), 236-238.

Staton, R. N., Mozo, B. T., & Murphy, B. A. (1997). *Operational Test to Evaluate the Effectiveness of the Communication Earplug and Active Noise Reduction Devices when used with the HGU-56/P Aviator Helmet* (No. 97-09). Ft. Rucker, AL: U.S. Army Aeromedical Laboratory.

Suter, A. H. (1985). Speech recognition in noise by individuals with mild hearing impairments. *Journal of the Acoustical Society of America, 78*(3), 887-900.

Suter, A. H. (2000). Standards and Regulations. In E. H. Berger, Royster, L.H., Royster, J.D., Driscoll, D.P., Layne, M. (Ed.), *The Noise Manual* (5th ed., pp. 639-668). Fairfax, VA: American Industrial Hygiene Association.

Suter, A. H. (2002). *Hearing Conservation Manual* (Fourth ed.). Milwaukee, WI: Council for Accreditation in Occupational Hearing Conservation (CAOHC) Executive Office.

Svensson, E., Angelborg-Thanderz, M., Sjoberg, L., & Olsson, S. (1997). Information complexity-mental workload and performance in combat aircraft. *Ergonomics, 40*, 362-380.

Taylor, R. M. (1990). *Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design*. Neuilly-sur-Seine, France: Advisory Group for Aerospace Research and Development.

Tenney, Y. J., Adams, M. J., Pew, R. W., Huggins, A. W. F., & Rodgers, W. H. (1992). *A Principles Approach to the Measurement of Situation Awareness in Commercial Aviation*. Hampton, VA: Langley Research Center, National Aeronautics and Space Administration.

Tsang, P. S., & Vidulich, M. A. (1994). The Roles of Immediacy and Redundancy in Relative Subjective Workload Assessment. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 36*, 503-513.

Tsang, P. S., & Vidulich, M. A. (2006). Mental Workload and Situation Awareness. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 243-268). Hoboken, NJ: John Wiley & Sons, Inc.

Valimont, R. B., Casali, J. G., & Lancaster, J. A. (2006). *ANR vs. Passive Communications Headsets: Investigation of Speech Intelligibility, Pilot Workload, and Flight Performance in an Aircraft Simulator.* Paper presented at the Human Factors and Ergonomics Society 50th Annual Meeting, San Francisco, CA.

Vidulich, M. A. (2003). Mental Workload and Situation Awareness: Essential Concepts for Aviation Psychology Practice. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and Practice of Aviation Psychology* (pp. 115-145). Mahwah, NJ: Lawrence Erlbaum Associates

Vidulich, M. A., Crabtree, R. B., & McCoy, A. L. (1993). Developing subjective and objective metrics of pilot situation awareness. *Proceedings of the Seventh International Symposium on Aviation Psychology*(2), 896-900.

Vidulich, M. A., & Hughes, E. R. (1991). Testing a Subjective Metric of Situation Awareness. *Proceedings of the Human Factors Society 35th Annual Meeting*, 1307-1311.

Vidulich, M. A., & McMillan, G. (2000). The Global Implicit Measure: Evaluation of Metrics for Cockpit Adaptation. In P. T. McCabe, M. A. Hanson & S. A. Robertson (Eds.), *Contemporary Ergonomics 2000* (pp. 75-80). London: Taylor & Francis.

Vidulich, M. A., & Tsang, P. S. (1987). *Absolute Magnitude Estimation and Relative Judgment Approaches to Subjective Workload Assessment.* Paper presented at the Human Factors Society 31st Annual Meeting, Santa Monica, CA.

Waag, W. L., & Houck, M. R. (1994). Tools for assessing situational awareness in an operational fighter environment. *Aviation, Space, and Environmental Medicine, 65*(5 Supplement), A13-A19.

Ward, W. D., Royster, L. H., & Royster, J. D. (2000). Anatomy and Physiology of the Ear: Normal and Damaged Hearing. In E. H. Berger, L. H. Royster, J. D. Royster, D. P. Driscoll & M. Layne (Eds.), *The Noise Manual* (5th ed., pp. 101-122). Fairfax, VA: American Industrial Hygiene Association.

Ward, W. D., Royster, J.D., Royster, L.H. (2000). Auditory and Nonauditory Effects of Noise. In E. H. Berger, Royster, L.H., Royster, J.D., Driscoll, D.P., Layne, M. (Ed.), *The Noise Manual* (Fifth ed., pp. 123-147). Fairfax, VA: American Industrial Hygiene Association.

Weston, H. C., & Adams, S. (1935). *The performance of weavers under varying conditions of noise*. London: Medical Research Council Industrial Health Research Board.

Whitaker, L. A., Peters, L. J., & Garinther, G. R. (1989, June, 1989). *Tank Crew Performance: Effects of Speech Intelligibility on Target Acquisition and Subjective Workload Assessment.* Paper presented at the Proceedings of the Human Factors Society - 33rd Annual Meeting.

Wickens, C. D. (2001). Workload and Situation Awareness. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, Workload, and Fatigue* (pp. 443-450). Mahwah, NJ: Lawrence Erlbaum Associates

Wickens, C. D., & Hollands, J. G. (2000). *Engineering Psychology and Human Performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Wierwille, W. W., & Casali, J. G. (1983, October 10-14). *A validated rating scale for global mental workload measurement applications.* Paper presented at the Proceedings of the 1983 Human Factors Society 27th Annual Conference, Norfolk, VA.

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for Mental Workload Measurement in a Test and Evaluation Environment. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 35*, 263-281.

Wierwille, W. W., Rahimi, M., & Casali, J. G. (1985). Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. *Human Factors, 27*(5), 489-502.

Wildzunas, R. M. (1999). *Notes on the Calculation of Power in Determining Sample Size and the Construction of Digram-Balanced Designs* (No. 99-14). Fort Rucker, Alabama: U.S. Army Aeromedical Research Laboratory.

Wilson, G. F., & Eggemeier, F. T. (1991). Psychophysiological assessment of workload in multi-task environments. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 329-360). Bristol, PA: Taylor & Francis.

Wolfinger, R. D., & Chang, M. (1998). Comparing the SAS GLM and MIXED Procedures for Repeated Measures. Retrieved June 8, 2009, from http://www.ats.ucla.edu/stat/sas/library/mixedglm.pdf

Woodson, J. (2006). *Flight Surgeon's Aeromedical Checklists*. Retrieved October 17, 2006. from https://aamaweb.usaama.rucker.amedd.army.mil/AAMAWeb/policyltrs/Army_APLs_Mar06_v3.pdf.

Yost, W. A., Nielson, D. W. (1985). *Fundamentals of Hearing. An Introduction* (2nd ed.). New York: CBS College Publishing.

APPENDIX A – List of Abbreviations

ACAP	Aeromedical consultants advisory panel
AEDR	Aviation Epidemiology Data Register
AERO	Aeromedical electronic resource online
AI	Articulation Index
AL	Action level
ANR	Active noise reduction
ANOVA	Analysis of variance
APL	Aeromedical policy letter
AR	Army regulation
ATC	Air traffic control
B-ANR	Bose Aviation X ANR headset
CAOCH	Council for the Accreditation of Occupational Hearing Conservationists
CEP	Communications earplug
CFR	Code of federal regulations
CHPPM	Center for Health Promotion and Preventive Medicine
CIRCLE	Coordination index rating for crew linguistic events
DA PAM	Department of the Army pamphlet
DAC	Department of the Army civilian
dB	Decibel
dBA	Decibel (A weighting network)
dBC	Decibel (C weighting network)
dBHL	Decibel (hearing level)
dBP	Decibel (peak)
DC-CEP	David Clark model H10-66 headset with Communications Earplug
DC-FOAM	David Clark model H10-76 headset with foam earplug
DoD	Department of defense
DoDI	Department of defense instruction
EEG	Electroencephalogram
EP	Evoked potential
FFD	Full flying duties
FS	Flight surgeon
GIM	Global implicit measure
GLM	Generalized linear model
HCP	Hearing conservation program

HPD	Hearing protective device
HRC	Human resources command
HTL	Hearing threshold level
Hz	Hertz
ICS	Intercommunications system
IFR	Instrument flight rules
IMC	Instrument meteorological conditions
IO	Instructor operator
MANOVA	Multivariate analysis of variance
MCH	Modified Cooper-Harper
MCL	Most comfortable listening level
NASA-TLX	National Aeronautics & Space Administration – Task load index
NIHL	Noise-induced hearing loss
NOE	Nap of the earth
OSHA	Occupational Safety and Health Administration
PEL	Permissible exposure limit
PTS	Permanent threshold shift
RAS	Reticular activating system
REML	Restricted Maximum Likelihood
SA	Situation awareness
SA-SWORD	Situation awareness-subjective workload assessment technique
SAGAT	Situation awareness global assessment technique
SART	Situation awareness rating technique
SARS	Situational awareness rating scale
SAS	Statistical Analysis Software
SII	Speech Intelligibility Index
SL	Sensation level
SM	Statute miles
SRT	Speech reception threshold
STI	Speech Transmission Index
STS	Significant threshold shift
SWAT	Subjective Workload Assessment Technique
SQ	Signal quality
TFOE	Transfer function of the open ear
TTS	Temporary threshold shift
TWA	Time weighted average

USAAMA	United States Army Aeromedical Activity
USAAMC	United States Army Aeromedical Center
USAARL	United States Army Aeromedical Research Laboratory
USAMRMC	United States Army Medical Research and Materiel Command
VFR	Visual flight rules
VMC	Visual meteorological conditions
WL	Workload

APPENDIX B - Informed Consent Form

CONSENT TO PARTICIPATE IN RESEARCH

Title of Project: *Experimental Evaluation of Army Pilot Hearing Requirements, Flight Workload, Communications Headsets, and Speech Signal Quality in a Black Hawk Helicopter Flight Simulator*

You are asked to participate in a research study conducted at the United States Army Aeromedical Research Laboratory (USAARL) by Dr. William A. Ahroon, Ph.D. (Primary Investigator (PI)) and MAJ Kristen L. Casto (Associate Investigator (AI)). Your participation in this study is voluntary. You should read the information below and ask questions about anything you do not understand before deciding whether or not to participate.

- **PURPOSE OF THE STUDY**

The purpose of this study is to assess how flying performance in a UH-60 flight simulator is affected by different headsets, different levels of flight workload, different levels of communication signal clarity, and hearing ability. In this study, you will be asked to fly three one-hour (approximate) missions in a UH-60 research flight simulator located at the USAARL, Ft. Rucker, Alabama. The simulator is a motion-based system that includes an operational crew station, computer-generated visual display, temperature control system, realistic sound system, and a multichannel data acquisition system. The simulator controls have the look and feel of those in a real helicopter, and the simulator response is derived from accurate flight aerodynamics.

- **PROCEDURES**

If you volunteer to participate in this study, we would ask you to do the following things:

You must have a current (within the past year) flight physical that allows full flying duties and you must have flown a military helicopter or military flight simulator in the past year to participate in this study.

You will report to the Acoustics Branch at the USAARL for the study. The study session consists of one introductory session that lasts about an hour, one simulator familiarization session that will last approximately 30 minutes to an hour, and three simulated flight missions that each last approximately one hour. The study also consists of filling out rating forms about your perceptions of flight workload and situation awareness nine times during each flight mission, and headset comfort/communications rating once at the end of each flight mission. Each rating form takes approximately one minute to complete, resulting in approximately 20 minutes per mission, or a total of approximately one hour for all three flight missions combined for rating form completion.

The first session is an introductory session in which you will undergo a brief hearing evaluation. The evaluation consists of answering questions about your history of noise exposure, hearing loss, and any ear surgeries; an examination of your ear canal and eardrum with a lighted scope that allows the examiner to see into your ear; a hearing test to measure your hearing ability; and a measurement of how your eardrum and the space behind your eardrum is functioning. These are all routine tests used by audiologists, and there is no risk to your hearing posed by them. You will be assigned to a study group based on your hearing threshold levels. Including you, 27 instrument rated helicopter pilots (active duty, DA civilian, and contract) of at least 19 years of age will be participating in the experiment.

The second session is familiarization session. The purpose of this daylight flight mission session is to ensure that you are comfortable with the simulator controls.

The last session consists of three, one-hour (approximate) simulated flight missions (ideally, all three missions will be flown during one session; however, the missions may need to be flown during separate sessions). During each flight mission the visibility, required flight maneuvers, and quality of communication signals will vary. You will fly each mission with one of the following headset configuration: 1) a headset worn over foam earplugs, 2) a headset equipped with active noise reduction (ANR) and 3) a headset equipped with the Communications Earplug (CEP). For each mission, your flight control performance, such as altitude, heading, and airspeed deviation will be measured by the simulator's computer. The number of times you request a read-back from the radio communications or make an error because of communication difficulty will also be measured. You will also be asked to complete forms on which you rate your perception of flight workload, situation awareness, headset comfort/speech intelligibility.

- **POTENTIAL RISKS AND DISCOMFORTS**

The risks of harm anticipated in the proposed research are not greater, considering the probability and magnitude, than those encountered in daily life. Sound levels within the simulator will mirror those of an actual UH-60 helicopter, but the flight time will be fairly short, about one hour per flight. The headsets and hearing protection that you will wear will adequately protect you from the helicopter noise exposure.

- **ANTICIPATED BENEFITS TO SUBJECTS**

Your participation in this experiment will provide information that may be helpful in understanding pilot performance with different headsets and different hearing levels in situations that vary in flight workload and communication demand. The results of this research may help refine the audiometric flight waiver criteria for Army pilots so that the waiver decision is not based only on hearing levels, but also on performance data. The results of this experiment may also provide information that could help flight surgeons and audiologists recommend appropriate communication system for individual pilots. No guarantee of direct benefits has been made to encourage you to participate.

- **ALTERNATIVES TO PARTICIPATION**

There are no alternative therapeutic, diagnostic, or preventive procedures that should be considered before deciding whether or not to participate in the study. The alternative to participation in this study is to not participate in the study.

- **PAYMENT FOR PARTICIPATION**

You will not be monetarily, or otherwise compensated, for your participation in this experiment.

- **MEDICAL CARE FOR RESEARCH RELATED INJURY**

Active Duty Military:

If you are a Department of Defense healthcare beneficiary (i.e., active duty military), you are entitled to medical care for research-related injuries, to the same extent that such medical care would be provided if the injuries were not research-related. If any medical expenses for research-related injuries are incurred that are not covered with the DoD healthcare system, the individual will contact the U.S. Army Medical Research and Materiel Command Office of the Staff Judge Advocate to assist in pursuing reimbursement for research-related medical expenses (not otherwise provided or reimbursed). Reimbursement is not guaranteed.

Civilian:

If you are hurt or get sick because of this research study, you can receive medical care at an Army hospital or clinic free of charge. You will only be treated for injuries that are directly caused by the research study. The Army will not pay for your transportation to and from the hospital or clinic. If you have questions about this medical care, talk to the PI for this study, (Dr. William A. Ahroon at (334) 225-6828 or AI, MAJ Kristen L. Casto at (540) 230-5379). If you pay out-of-pocket for medical care elsewhere for injuries caused by this research study, contact the PI. If the issue cannot be resolved, contact the U.S. Army Medical Research and Materiel Command (USAMRMC) Office of the Staff Judge Advocate (legal office) at (301) 619-7663/2221.

- **CONFIDENTIALITY**

When the results of the research are published or discussed in conferences, no information will be included that would reveal your identity. If photographs, videos, or audiotape recordings of you will be used for educational purposes, your identity will be protected or disguised. Authorized representatives of the U.S. Army Medical Research and Materiel Command, FDA, and the manufacturer of the device being tested, Bose, and/or Communications and Ear Protection, Inc. may need to review records of individual subjects. As a result, they may see your name; but they are bound by rules of confidentiality not to reveal your identity to others.

Complete confidentiality cannot be guaranteed because information bearing on a soldier's health may be required to be reported to appropriate medical or command authorities. Representatives of the USAARL and USAMRMC are authorized to review research records as part of their responsibility to protect human research volunteers. However, the information and data that you provide will have your name removed and only a subject number will identify you during analyses and any written reports of the research. All subject numbers will be secure and stored on the experimenter's password protected personal computer. If you would like to receive a summary of this research when it is completed, please provide a self-addressed envelope and/or email address.

- **PARTICIPATION AND WITHDRAWAL**

Your participation in this research is voluntary. If you choose not to participate, that will not affect your relationship with USAARL or your right to health care or other services to which you are otherwise entitled. If you decide to participate, you are free to withdraw your consent and discontinue participation at any time without prejudice. To avoid biasing other potential subjects, please do not discuss the study with anyone until six months from the date of signing this form.

- **WITHDRAWAL OF PARTICIPATION BY THE INVESTIGATOR**

The investigator may withdraw you from participating in this research if circumstances arise which warrant doing so. If you experience motion sickness or if you become ill during the research, you may have to drop out, even if you would like to continue. The investigator will make the decision and let you know if it is not possible for you to continue. The decision may be made either to protect your health and safety, or because it is part of the research plan that people who develop certain conditions may not continue to participate.

- **NEW FINDINGS**

During the course of the study, you will be informed of any significant new findings (either good or bad), such as changes in the risks or benefits resulting from participation in the research or new alternatives to participation, that might cause you to change your mind about continuing in the study. If new information is provided to you, your consent to continue participating in this study will be re-obtained.

- **IDENTIFICATION OF INVESTIGATORS**

In the event of a research related injury or if you experience an adverse reaction, please immediately contact one of the investigators listed below. If you have any questions about the research, please feel free to contact:

Dr. William A. Ahroon, (334) 255-6828, U.S. Army Aeromedical Research Laboratory, Bldg 6901, Ft. Rucker, AL

MAJ Kristen L. Casto (540) 230-5379, Virginia Polytechnic & State University, Blacksburg, VA (with temporary duty at USAARL, Ft. Rucker, AL)

- **RIGHTS OF RESEARCH SUBJECTS**

You may withdraw your consent at any time and discontinue participation without penalty. You are not waiving any legal claims, rights or remedies because of your participation in this research study. **Also, if you have any questions or concerns about this study or your rights as a research volunteer participant you may contact the USAMRMC Human Subject Research Review Board (HSRRB) at 301-619-2165, or email at HSRRB@amedd.army.mil, or mail at Commanding General, U.S. Army Medical Research and Materiel Command, ATTN: MCMR-RPH, 504 Scott Street, Fort Detrick, Fredrick, MD 21702.**

IF THERE IS ANY PORTION OF THIS VOLUNTEER AGREEMENT AFFIDAVIT THAT YOU DO NOT UNDERSTAND, ASK THE INVESTIGATOR BEFORE SIGNING THIS FORM.

SIGNATURE OF RESEARCH VOLUNTEER PARTICIPANT

I have read the information provided above. I have been given an opportunity to ask questions and all of my questions have been answered to my satisfaction. I have been given a copy of this form.

____ _______________

Name of Subject SSN

____ ______________

Signature of Subject Date

Physical and electronic address ____

____

Name of person administering consent: ____________________________

Signature of person administering consent: _________________________ Date: ______

Time ______

APPENDIX C - Modified Cooper-Harper Rating Scale Instructions

Modified Cooper-Harper Instructions

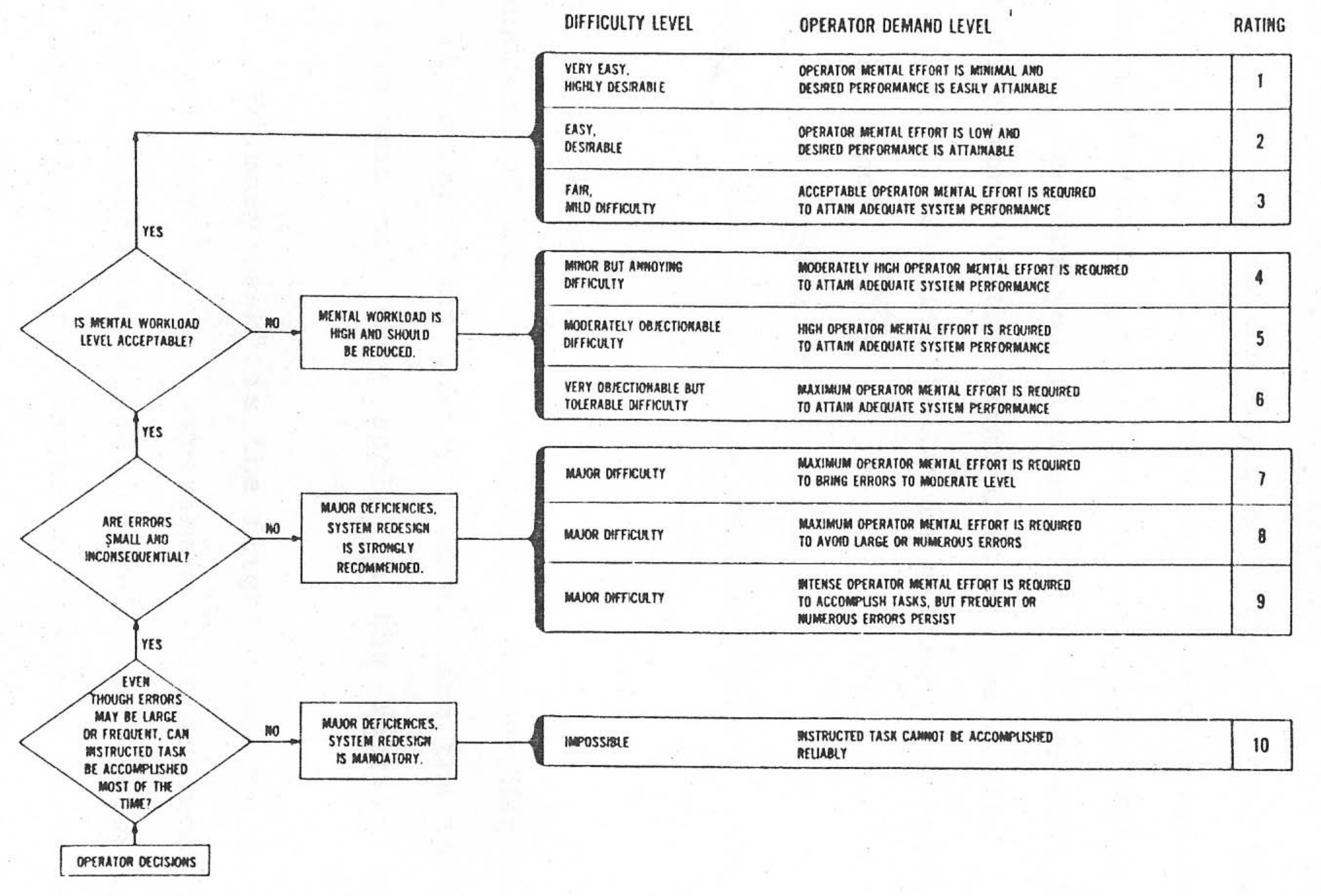
Key Definitions of Terms

Mental Workload: For purposes of the survey, mental workload includes all processes which are purely mental such as (but not limited to) attention, visualization, spatial orientation, decision-making, and memory. Mental workload also includes the combination of mental processes such as perception and the resulting physical actions. For example, the perception of information displayed on the flight instruments and the resulting physical control inputs are considered psychomotor workload, a type of mental workload. The same is true for the perception of the weather's effects on the aircraft and the pilot's resulting actions.

Rating Scale Steps

On the Modified Cooper-Harper scale you will notice that there is a series of decisions that follow a predetermined logical sequence. This logic sequence is designed to help you make more consistent and accurate ratings. Thus, you should follow the logic sequence on the scale for each of your ratings in this experiment. Remember – you are to circle only one number, and the number must be arrived at by following the logic of the scale. You should always begin at the lower level and follow the logic path until you have decided on a rating. In particular, do not skip any steps in the logic. Otherwise, your rating may not be valid or reliable.

APPENDIX D - Modified Cooper-Harper Rating Scale


DIFFICULTY LEVEL
OPERATOR DEMAND LEVEL
RATING
VERY EASY, HIGHLY DESIRABLE
OPERATOR MENTAL EFFORT IS MINIMAL AND DESIRED PERFORMANCE IS EASILY ATTAINABLE
1
EASY, DESIRABLE
OPERATOR MENTAL EFFORT IS LOW AND DESIRED PERFORMANCE IS ATTAINABLE
2
FAIR, MILD DIFFICULTY
ACCEPTABLE OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE
3
MINOR BUT ANNOYING DIFFICULTY
MODERATELY HIGH OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE
4
MODERATELY OBJECTIONABLE DIFFICULTY
HIGH OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE
5
VERY OBJECTIONABLE BUT TOLERABLE DIFFICULTY
MAXIMUM OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE
6
MAJOR DIFFICULTY
MAXIMUM OPERATOR MENTAL EFFORT IS REQUIRED TO BRING ERRORS TO MODERATE LEVEL
7
MAJOR DIFFICULTY
MAXIMUM OPERATOR MENTAL EFFORT IS REQUIRED TO AVOID LARGE OR NUMEROUS ERRORS
8
MAJOR DIFFICULTY
INTENSE OPERATOR MENTAL EFFORT IS REQUIRED TO ACCOMPLISH TASKS, BUT FREQUENT OR NUMEROUS ERRORS PERSIST
9
IMPOSSIBLE
INSTRUCTED TASK CANNOT BE ACCOMPLISHED RELIABLY
10
YES
IS MENTAL WORKLOAD LEVEL ACCEPTABLE?
NO
MENTAL WORKLOAD IS HIGH AND SHOULD BE REDUCED.
YES
ARE ERRORS SMALL AND INCONSEQUENTIAL?
NO
MAJOR DEFICIENCIES, SYSTEM REDESIGN IS STRONGLY RECOMMENDED.
YES
EVEN THOUGH ERRORS MAY BE LARGE OR FREQUENT, CAN INSTRUCTED TASK BE ACCOMPLISHED MOST OF THE TIME?
NO
MAJOR DEFICIENCIES, SYSTEM REDESIGN IS MANDATORY.
OPERATOR DECISIONS

APPENDIX E - SART Scale

Instability of Situation
How changeable is the situation? Is the situation highly unstable and likely to change suddenly (high), or is it very stable and straightforward (low)?

1 2 3 4 5 6 7 8 9 10
LOW HIGH

Complexity of Situation
How complicated is the situation? Is it complex with many interrelated components (high) or is it simple and straightforward (low)?

1 2 3 4 5 6 7 8 9 10
LOW HIGH

Variability of Situation
How many variables are changing in the situation? Are there a large number of factors varying (high) or are there very few variables changing (low)

1 2 3 4 5 6 7 8 9 10
LOW HIGH

Arousal
How aroused are you in the situation? Are you alert and ready for activity (high) or do you have a low degree of alertness (low)?

1 2 3 4 5 6 7 8 9 10
LOW HIGH

Concentration of Attention
How much are you concentrating on the situation? Are you bringing all your thoughts to bear (high) or is your attention elsewhere (low)?

1 2 3 4 5 6 7 8 9 10
LOW HIGH

Division of Attention
How much of your attention is divided in the situation? Are you concentrating on many aspects of the situation (high) or focused on only one (low)?

1 2 3 4 5 6 7 8 9 10
LOW HIGH

Spare Mental Capacity
How much mental capacity do you have to spare in the situation? Do you have sufficient mental capacity to attend to many variables (high) or nothing to spare at all (low)?

1 2 3 4 5 6 7 8 9 10
LOW HIGH

Information Quantity
How much information have you gained about the situation? Have you received and understood a great deal of knowledge (high) or very little (low)?

1 2 3 4 5 6 7 8 9 10
LOW HIGH

Information Quality
How good is the information you have gained about the situation? Is the knowledge communicated very useful (high) or is it not helpful (low)?

1	2	3	4	5	6	7	8	9	10
LOW									HIGH

Familiarity with Situation
How familiar are you with the situation? Do you have a great deal of relevant experience (high) or is it a new situation (low)?

1	2	3	4	5	6	7	8	9	10
LOW									HIGH

APPENDIX F - Audiometric Screening Form

AUDIOMETRIC SCREENING FORM

Participant Number: ________Screening Date: __________________ Qualify: ________

Age:_______ Gender:_______ Number of flight hours:___________

Case history:

History of noise exposure:

History of ear disease:

Hearing aid use:

Known hearing loss:

Otoscopic exam	**Tympanometry**
AD:	AD:
AS:	AS:

Pure Tone Audiometry

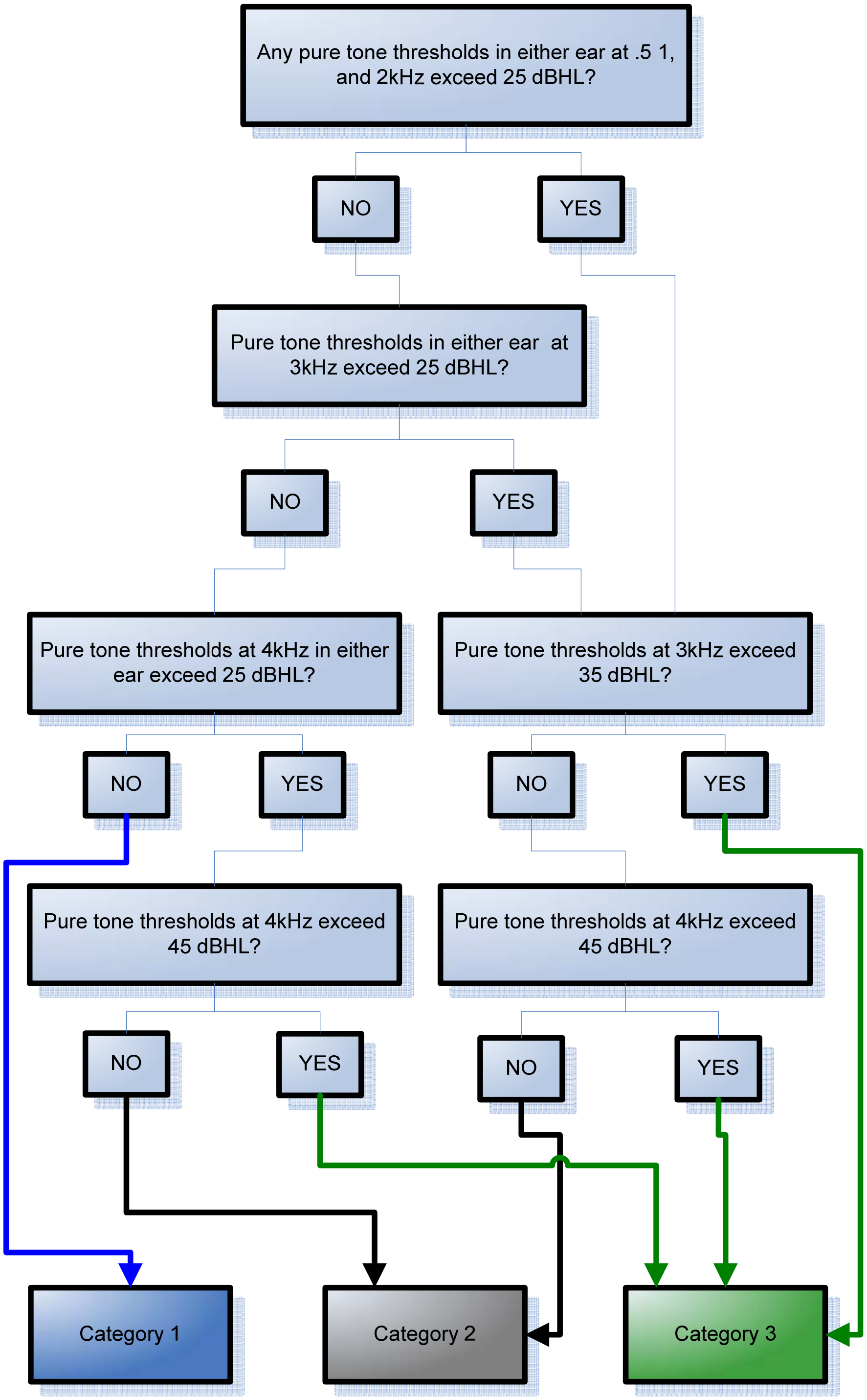
Right ear

250 Hz	500 Hz	1000 Hz	1500 Hz	2000 Hz	3000 Hz	4000 Hz	6000 Hz	8000 Hz

Left ear

250 Hz	500 Hz	1000 Hz	1500 Hz	2000 Hz	3000 Hz	4000 Hz	6000 Hz	8000 Hz

APPENDIX G - Hearing Loss Category Flowchart


Any pure tone thresholds in either ear at .5 1, and 2kHz exceed 25 dBHL?
NO
YES
Pure tone thresholds in either ear at 3kHz exceed 25 dBHL?
NO
YES
Pure tone thresholds at 4kHz in either ear exceed 25 dBHL?
Pure tone thresholds at 3kHz exceed 35 dBHL?
NO
YES
NO
YES
Pure tone thresholds at 4kHz exceed 45 dBHL?
Pure tone thresholds at 4kHz exceed 45 dBHL?
NO
YES
NO
YES
Category 1
Category 2
Category 3

APPENDIX H - Headset Comfort/Communications Rating Scale

The following rating scale is designed for you to rate several aspects related to the comfort of the headset configuration you just wore (headset with ANR, headset with CEP, or headset with foam earplug). The rating scale also includes some items about the communication characteristics of the headset configurations. Please place an X in one of the seven spaces provided between each pair of descriptors that best reflects your opinion. Please provide honest ratings that are as accurate as possible.

Comfortable	____ : ____ : ____ : ____ : ____ : ____	**Uncomfortable**
Uncomfortable pressure	____ : ____ : ____ : ____ : ____ : ____	No uncomfortable pressure
Painful	____ : ____ : ____ : ____ : ____ : ____	Painless
Tight	____ : ____ : ____ : ____ : ____ : ____	Loose
Ear blocked	____ : ____ : ____ : ____ : ____ : ____	Ear open
Convenient	____ : ____ : ____ : ____ : ____ : ____	Cumbersome
Non-restrictive	____ : ____ : ____ : ____ : ____ : ____	Restrictive
High fidelity communications	____ : ____ : ____ : ____ : ____ : ____	Low fidelity communications
No sound distortion	____ : ____ : ____ : ____ : ____ : ____	Sound distortion
No extraneous noise	____ : ____ : ____ : ____ : ____ : ____	Extraneous noise
Restrictive	____ : ____ : ____ : ____ : ____ : ____	Non-restrictive
Hard	____ : ____ : ____ : ____ : ____ : ____	Soft
Light	____ : ____ : ____ : ____ : ____ : ____	Heavy
Tolerable	____ : ____ : ____ : ____ : ____ : ____	Intolerable
Background hum present	____ : ____ : ____ : ____ : ____ : ____	No background hum present