INTRODUCTION

Operationally defined, "command and control (C^2) is the exercise of authority and direction by a properly designated commander over assigned and attached forces in the accomplishment of the mission" (Department of Defense, 1997, p.1-1). More precisely, command is the act of making decisions and ordering actions to be performed, while control is monitoring and influencing the action (Department of Defense, 1997). Without reliable communications, neither command nor control can be exercised effectively.

Command and control tasks often require the commander to make critical decisions in diverse locations and under adverse conditions. Command and control tasks may take place in fixed shelters (e.g., special-purpose hardened bunkers on existing buildings), mobile shelters (e.g., tents), or in purpose-built command and control vehicles, which may include a High Mobility Multipurpose Wheeled Vehicle (HMMWV, often referred to as a "Humvee"), Blackhawk helicopter, or tracked vehicles such as the Bradley Fighting Vehicle. In each of these environments, excessive noise has the potential to disrupt auditory communications, with vehicle-based command and control systems having the greatest potential for disruption due to their higher ambient noise levels.

The ambient noise level in an armored vehicle is typically between 110 and 115 dBA (Whitaker, Peters, and Garinther, 1990). Such intense noise levels are extremely harmful to the crews' hearing and interfere with verbal communication. For this reason, crews in the Bradley Fighting Vehicle wear a Vehicular Intercommunication System (VIS) headset which incorporates active noise reduction (ANR) to reduce the crews' noise exposure and facilitate inter-crew communication. While it is obvious that the noise level in the Bradley Fighting Vehicle is loud enough to cause hearing damage and interfere with speech communication, such intense noise may also interfere with a person's performance on complex cognitive tasks, such as command and control tasks.

When the ANR component of the VIS headset is activated in 114 dBA noise, the sound pressure level (SPL) under the earcup is 83 dBA. When the communication system is turned on but no one is speaking, the noise level under the headset increases to 90 dBA. When the communication system is on and crewmembers are speaking, the noise level under the earcup increases to 94 dBA (M. Vaudrey, personal communication, December 13, 2000). Clearly, excessive noise is being transmitted to the listener via the communication system.

Recognition of this situation has led Adaptive Technologies, Inc. (ATI) to develop a new microphone and speech processing system for use with the Army's VIS headset. The new microphone and communication system (incorporating novel ANR and speech-processing algorithms) developed by ATI is intended to reduce the noise transmitted through the communication system and improve the intelligibility of transmitted speech. In addition to reducing the noise exposure of the affected crew, it is hoped that by improving communications, a measurable improvement in task performance will be achieved in concert with a reduction in workload.

Based on the preceding discussion, it is obvious that the noise level in an armored vehicle is intense enough to cause hearing damage and cause a decrement in speech intelligibility. Such intense noise can also interfere with a person's performance on tasks that are continuous (Davies and Jones, 1982), performed simultaneously (Weinstein, 1974), and impose a high level of information processing (Eschenbrenner, 1971), such as command and control tasks. Therefore, the objective of this dissertation was to determine whether an improved microphone and communication system (incorporating novel ANR and speech-processing algorithms) have any real or subjective benefits for a person engaged in command and control tasks in an environment in which the sound pressure level is equivalent to that of a Bradley Fighting Vehicle.

COMMAND AND CONTROL

Command and control (C^2) can take place in a variety of locations, including buildings, command and control vehicles, and fixed or temporary shelters. Associated command and control tasks may be conducted through face-to-face communication, over a radio or other telecommunication link, or visually through text and graphical (e.g., maps) displays. For example, a commander may initially receive information visually from a computer screen, develop a plan, and then issue orders to subordinates via face-to-face communication, over the radio, or by electronic mail. The method chosen for transmitting the information will depend on the importance of the information, urgency to the situation, and the location of the affected units.

The infrastructure of the United States government, which includes the military, is built on the principle of command and control. The highest-ranking official in government is, of course, the President of United States, who serves as Commander-in-Chief of the Armed Forces. During the *Desert Storm* war, the President stayed abreast of all the events that were happening on the battlefield by receiving information, verbally and visually, from other officials lower in the chain of command. This is an example of how the concept of command and control works. However, the way in which the concept of command and control is exercised is based on the environment and situations in that environment.

The President’s environment may be considered static compared with that of a battlefield commander, who issues commands to subordinates in the battlefield. A battlefield commander in the Army may have to make a decision in an environment filled with noise and toxic chemicals that may exist during wartime, which may influence decision-making. Also, in the case of the President of the United States, the equipment required to perform command and control tasks is not very sophisticated and may not require a lot of cognition as compared with a commander in the battlefield.

However, for situations relative to the battlefield in the military, command and control is exercised differently because the environment is very dynamic. Commanders may use a variety of modes to issue command and control tasks to others (e.g., computers, radio, face-to-face). Regardless of how the commands are issued, the information has to be received quickly in order for the commander to make a decision in a short time-frame.

One way in which information is received and commands are issued is through a computer-automated system. According to Taylor, Charlton, and Canham (1996), such computer-automated systems assist the commander by gathering information about external events and redistributing this information to the commander. In turn, this allows the commander to perform tasks such as directing and assisting soldiers in a more efficient manner.

The displays that are utilized in some military command and control systems include electronic whiteboards, electronic status reports, electronic maps, and electronic mail. The electronic whiteboard display is a computer screen that allows the operator to input information using a keyboard, mouse, or lightpen. An electronic whiteboard allows the commander and others in their unit to communicate synchronously from remote locations, as well as allowing the commander to view vital information needed to make command decisions.

The electronic status report presents logistical information such as tactical status, personnel status, and fuel and ammunition status via a computer screen. While information displayed on the computer screen is received digitally, some information received over the radio is written on paper. If the written information triggers an action, it is entered into the computer and distributed throughout the network. Everyone in the network can then view the information displayed on the computer screen. Individuals responsible for managing particular assets are responsible for updating the status reports (T. W. Davis, personal communication, October 15, 2001).

An electronic map is essential and critical to performing C^2 tasks. The map is updated electronically as different military units (e.g., armor and infantry) move around the battlefield. The map allows the commander to magnify the display to get a more detailed view of the location of the enemy as well as the location of friendly forces. The enemy and friendly forces shown on a map are color coded to allow the commander to distinguish one from the other.

Electronic mail serves as an alternate form of communication and works well for planning and administration. When performing routine operations, such as training and battle simulations, e-mail is used for general messages. However, during actual engagement of an enemy target, verbal communication is used because electronic mail is too slow and cumbersome (S. E. Middlebrooks, personal communication, June 5, 2000).

An example of a computer-automated system, mentioned in a 1996 annual report to Congress, was the Inter-Vehicular Information System (IVIS). Military command and control systems, such as the IVIS, are fundamental for the commander to effectively plan, direct, and control operations for the completion of a mission (Department of Defense, 1997). The Director of Operational Test and Evaluation (DOT&E) stated that the M1A2 tank comes equipped with an IVIS that provides the tank crew with improved command, control, and communication abilities in the battlefield (Department of Defense, 1996). An illustration of this system is shown in Figure 1.

In a military magazine, *Soldiers*, Machamer (1995) told of a story in which the IVIS was used to send digital map overlays and status reports to perform command and control tasks. Machamer's (1995) story described a situation in which a tank crew came under attack of "enemy forces" after completing a final mission at night in a California desert. Once under attack, the commander located the enemy through his binoculars and ordered his subordinates to fire while he looked at his electronic map. Once the enemy was located on the electronic map, the commander ordered the loader to send the location grid to a fire support team to assist them in attacking the enemy. After they defeated the enemy forces, Machamer (1995) stated that the tank crew sent out a status report stating that they were low on fuel and ammunition. Additionally, the commander checked the map screen to locate other soldiers in his company.

In Machamer's (1995) story, command and control tasks were conducted through face-to-face communication and through visual displays. It is important to note that communication was valuable inside the tank, especially when under attack. Also, another point that can be drawn from the story is that electronically sent information aids in defeating the enemy. In this instance, it allowed the crew that was under attack to send information to others in their company concerning the enemy's location. Machamer's (1995) story illustrates how the urgency of command and control tasks

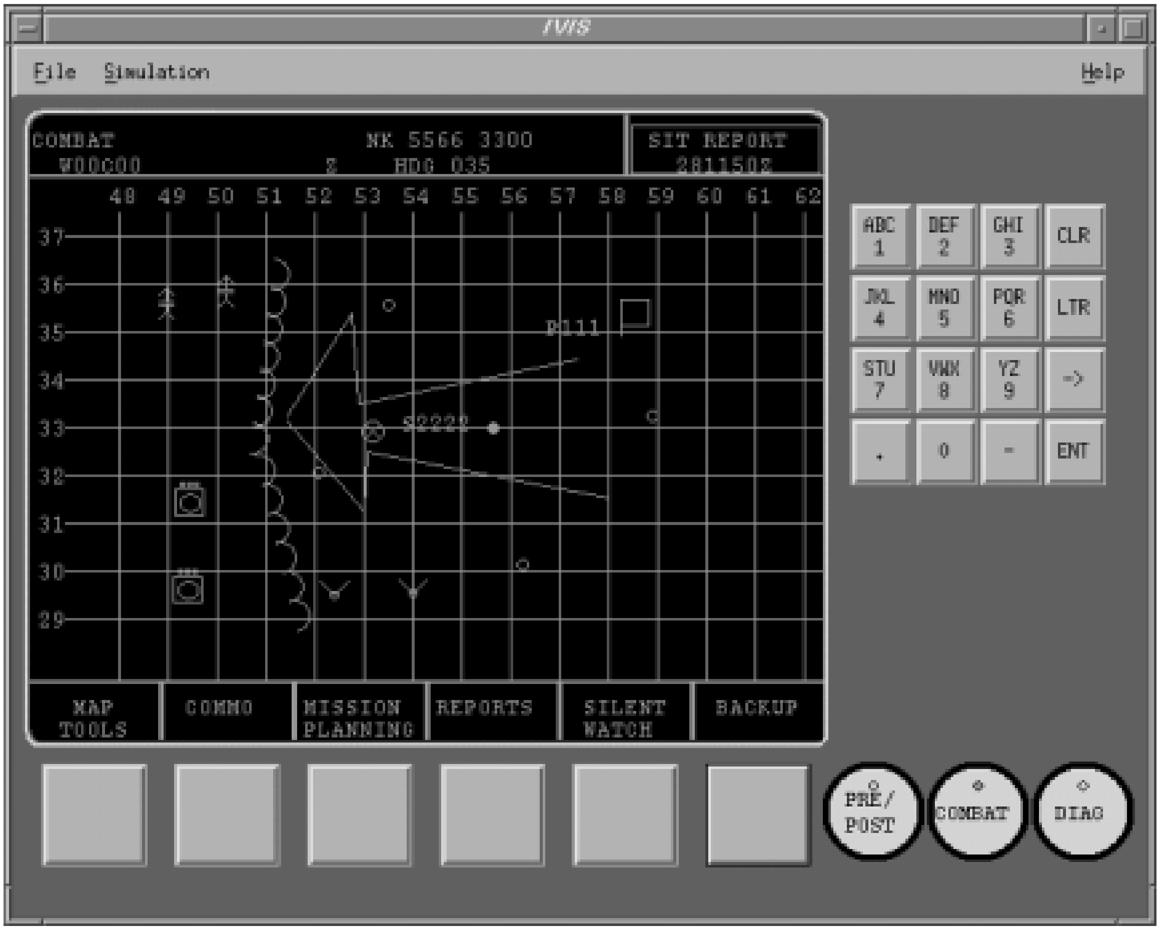


Figure 1. Illustration of an IVIS system
(from http://vesuvius.jsc.nasa.gov/icat/images/ivis.gif).

change following with the outcome of the battle. For instance, when the tank company was engaging the enemy, the most important information was looking at the map, communicating with others in the tank, and sending information to another tank for assistance. However, once they had defeated the enemy, the tank crew sent a status report and checked the location of the other tanks in the company.

Another example of a situation in which the concept of command and control can be exercised is that of a tank platoon. The tank platoon is the smallest unit in a tank company (Department of Defense, 1996). This unit is comprised of four tanks that are arranged in two sections. The platoon consists of a platoon leader, platoon sergeant, and section leaders. The platoon leader and platoon sergeant are the leaders of each section, with the remaining two tanks serving as wingmen to the leading tank.

The platoon leader is assisted by the platoon sergeant to perform command and control tasks during combat. The techniques they utilize to plan operations are issuing orders, communicating, and employing the platoon (Department of Defense, 1996). All the tanks in the tank platoon report to the platoon leader. Whenever an engagement ends, the commanders in the other tanks report to the platoon leader, through voice, what they saw and shot. The platoon leader then selects a spot report (similar to a status report) on a digital system, sends it by electronic mail to the next higher level, and it is forwarded through the system (G. Moore, personal communication, July, 14, 2000). Due to the high noise levels inside the tank and the time constraints, typical of a military combat environment, it is essential that the platoon leader maintain clear dialogue with subordinates in the platoon. This will enable the platoon leader to provide the platoon, as well as other commanders, with the necessary information and instructions that are needed to accomplish the mission.

VEHICULAR INTERCOMMUNICATION SYSTEM

The U.S. Army currently uses the Vehicular Intercommunication System (VIS) to allow intelligible communication in armored vehicles and to protect soldiers against hearing loss. In 1996, the M1A1 was the first armored vehicle in which soldiers used the VIS headset. The Army has since deployed the VIS headset in the Bradley Fighting Vehicle, the M109 self-propelled howitzer, and other vehicles. The VIS headset (shown in Figure 2) consists of a circumaural earmuff with active noise reduction (ANR) electronics, which is known to improve communication in battlefield vehicles (Moore, 1998). Moore (1998) stated that the amount of attenuation afforded by the VIS (incorporating ANR), in 115 dB Bradley Fighting Vehicle noise, is 30 dB, as compared to 15 dB with the conventional passive helmets. It was also reported by Moore (1998) that speech intelligibility scores of 89 percent were achieved with the VIS, using ANR, as compared to 68 percent with the conventional passive helmets. Therefore, since the VIS (incorporating ANR) showed an improved attenuation and speech intelligibility score, it may cause an increase in performance and a reduction in workload. Furthermore, it may reduce the threat of noise-induced hearing loss and allow soldiers to allocate some of their resources to other tasks.

Active Noise Reduction

Hearing protection devices (HPDs) are dichotomized into passive (does not incorporate electronic networks) and active (incorporate electronic networks). As seen in the preceding discussion, the attenuation afforded by conventional passive HPDs does not improve intelligibility better than active HPDs. Passive HPDs are known for attenuating high frequency sound more than low frequency sound. They contain no electronic

Figure 2. Illustration of an active noise reduction as mounted in a VIS helmet.

devices or transducers. They are less expensive than ANR headsets, and they provide less maintenance and are more durable (Casali and Berger, 1996). On the other hand, active HPDs are those that incorporate electronic components and transducers. An active HPD may or may not have a communication component. In comparison to passive HPDs, active HPDs are known for attenuating sounds that are low frequency biased. Due to the advancements in microelectronics and computer technology, active HPDs continue to be developed for hearing protection and speech communication purposes.

The concept of ANR has been around for years, with its name used interchangeably throughout the literature with active noise control and active noise reduction. For the purpose of this research, the term active noise reduction will be used. As reported by Tokhi and Leitch (1992), the birth of the active noise reduction field was due to work by Paul Lueg. Lueg filed a patent in Germany in 1933, and in the United States in 1934, and was awarded U.S. Patent No. 2,043,416 in 1936, entitled “Process of silencing sound oscillation,” that described a technique for active noise cancellation (Nixon, McKinley, and Steuver, 1992). In Lueg’s patent, the principle of superposition to achieve cancellation of unwanted noise was used. His concept consisted of reducing noise by adding anti-noise that was out-of phase with the original noise, which resulted in a wave cancellation that reduced the amplitude of noise.

Today, in addition to the cancellation of noise in spatial volumes or sound fields, the concept of ANR is also applied to reducing the noise levels at the ear and improving communication. Technically, the noise level is reduced by relying on the principle of superposition of two sound waves of equal amplitude, but 180 degrees out-of-phase.

This results in destructive interference, canceling out the undesirable noise, usually under an earmuff-type hearing protection device.

Due to major advancements in technology, ANR-based hearing protection devices and communication headsets have improved throughout the years. Starting in 1957, Willard Meeker developed a paper design and working model of ANR applied to a circumaural earmuff (Meeker, 1957; Nixon, McKinley, and Steuver, 1992). Meeker's ANR headset had an active attenuation bandwidth of approximately 50-500 Hz, with a maximum attenuation of approximately 20 dB (Meeker, 1957). The cancellation of the noise proved that the active noise cancellation concept was practical.

The first active noise reduction system to be used outside the laboratory was developed by P. D. Wheeler (Nixon, McKinley, and Steuver, 1992). This system consisted of an aircrew's helmet incorporating ANR technology to improve speech intelligibility and reduce noise exposure in a cockpit noise environment (Wheeler and Halliday, 1981). The cockpit noise environment was characterized as being low-frequency biased in which the noise levels reached 110-120 dB. The aircrew's helmet (i.e., passive attenuation) was capable of attenuating sound in the higher frequencies (greater than 500 Hz), but poor at attenuating sound at lower frequencies. Therefore, this prompted the researchers to incorporate the helmet with ANR to attenuate the low frequency components of the noise. After testing the ANR system in a Hercules aircraft, the maximum attenuation was reported to be 18 dB over an attenuation bandwidth of 800 to 1000 Hz (Nixon, McKinley, and Steuver, 1992).

Over the years, the bandwidth and attenuation of ANR has changed due to advancements in technology. With analog ANR devices, the maximal attenuation values of 22 dB are usually found from approximately 100 to 250 Hz, with little attenuation above 1000 Hz (Casali and Berger, 1996). Casali and Robinson (1994) discussed digital ANR devices, which are most beneficial for narrow band noises, such as sirens. They stated that attenuation with a digital ANR device may range from 8 to 20 dB at 800 Hz, with approximately 15 dB at 4000 Hz (Casali and Robinson, 1994).

Active noise reduction headsets are designed with a microphone inside the earcup that receives the noise that has penetrated the cushion of the earmuff. The signal is then fed to a feedback filter that reverses the phase of the original noise. After the original noise has been phase-inverted, it is then amplified with the necessary gain, producing an anti-noise that is introduced 180 degrees out-of-phase with the original noise (shown in Figure 3). The anti-noise is outputted to the earphone's loudspeaker, resulting in cancellation inside the earcup (Casali, 1992).

Because the cancellation signal does not reach the earphone, microphone, and listener's eardrum at the same time, a zero time-delay is not possible. Therefore, the resultant amplitude once the two out-of-phase waves are added does not result in complete cancellation, especially in the high frequency region. Thumann and Miller (1990) stated that low frequency noises that emanate from repetitive sources are the easiest to cancel due to their longer wavelengths. Additionally, repetitive sounds are more stable, thus allowing the ANR electronics to invert the phase and match amplitude of the anti-noise to that of the signal. At the higher frequencies, more accurate wave cancellation is necessary because high frequency sounds are characterized by shorter wavelengths, making them more difficult to cancel due to timing and geometric issues.

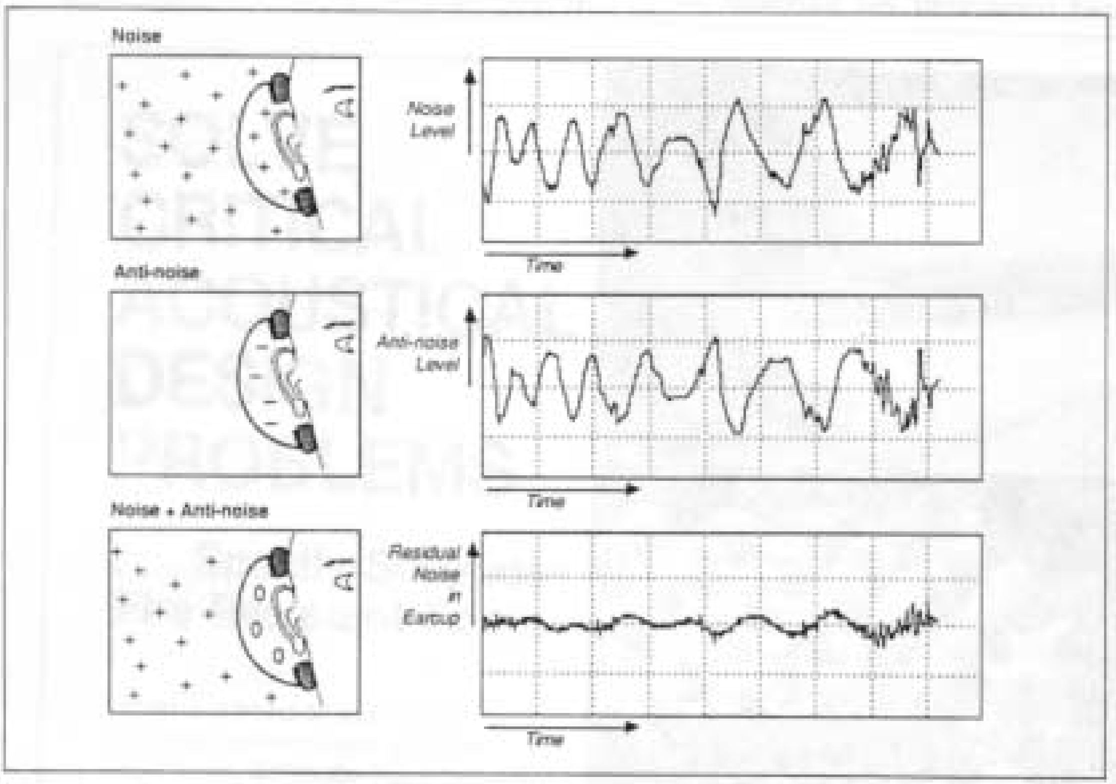


Figure 3. Illustration of how the Bose ANR headset operates (from Gauger and Sapiejewski, 1987).

Active noise reduction headsets can be dichotomized into open-back (supra-aural) or closed-back (circumaural). The open-back system provides only ANR, with little or no passive attenuation. If the ANR system fails, there is no protection afforded by the device. Closed-back devices have earmuff cups that afford the wearer active and passive attenuation. Active attenuation occurs at frequencies below 1000 Hz, and passive attenuation occurs at frequencies greater than 1000 Hz. By attenuating sound below the 1000 Hz region, ANR reduces the effect of the upward spread of masking, which is when lower frequencies mask higher frequencies. The passive attenuation afforded by closed-back devices, along with the active attenuation from active noise reduction electronics, give the wearer protection in the higher and lower frequencies. One of the greatest benefits of closed-back devices is that they provide the wearer protection if the ANR system fails. Closed-back devices are found in the majority of the ANR systems to date.

In many environments, where noise dominates the low frequencies, such as in military and industrial environments, ANR is fortuitous. Gauger and Sapiejewski (1987) reported that pilots who flew the "Voyager," a non-stop experimental flight around the world, used Bose ANR headsets. During a trial flight, which lasted 111 h and 44 min, and the non-stop flight, which lasted 216 h and 3 min, the crew experienced noise levels ranging from 93.0 to 103.3 dBA (Stephenson, Billings, and Jutila, 1990). Gauger and Sapiejewski (1987) reported that significant benefits accrued to the crew using the headset, in that the headsets reduced the noise from the engine and other ambient noise that exists in the cockpit. Additionally, it was reported that the Bose ANR headset allowed radio communication to be intelligible. However, the electronics of the ANR headset failed about halfway through the nine-day flight. The crew did not suffer noise-induced hearing loss due to the passive attenuation afforded by the Bose headset, in addition to the foam earplugs that were worn under the headset (Stephenson, et al., 1990).

Gower and Casali (1994) conducted a study on speech intelligibility and the attenuation of noise using a David Clark H10-76 passive headset and a Bose Aviation ANR headset (operated both with and without the ANR feature), in which they used the Modified Rhyme Test (ANSI, 1989) to quantify communication. The noise levels used in their experiment consisted of 105 dB (linear), 110 dB, and 115 dB of pink noise (flat by octaves) and noise from a Bradley M-2 tank. The noise from the Bradley M-2 tank was low frequency biased with a –3 dB/octave rolloff above 500 Hz. In order to obtain the signal-to-noise ratio for a required level of intelligibility, the noise levels were manipulated, and the speech levels were set by the subjects to obtain an intelligibility score of 70%.

The researchers found that when the ANR system was activated in the Bose headset, it reduced the noise exposure at the ear, thus increasing the allowable exposure time. Additionally, the signal-to-noise ratio for the Bose (active) needed to be 3 dB higher than the David Clark H10-76 in order to achieve the same speech intelligibility score in pink and tank noise. Although the conventional passive hearing protector outperformed the Bose ANR (active) unit, the Bose unit would still be more beneficial in an environment where low frequency noise is a problem.

At-Ear Exposure of the VIS Headset

The VIS headset is equipped with a voice-activated noise canceling lip microphone (Moore, 1998), which is part of the communication system. This microphone picks up noise from the external environment when the communication system is activated. In tank noise, with a sound pressure level of 114 dBA, this microphone picks up considerable noise, as well as speech. As reported earlier, the noise level, when no one is speaking and the VIS headset is activated, is 83 dBA. When the communication system is activated and no one is speaking, the noise level under the headset increases to 90 dBA. Additionally, when the communication system is activated and someone is speaking, the noise level increases to 94 dBA. The noise-canceling microphone serves as an important part of a crews' performance because communication allows the soldiers to exchange important information with each other in order to perform various tasks. However, when the noise level reaching the ear is 94 dBA, it may cause hearing damage and speech intelligibility decrements, both of which may have an impact on performance.

Summary

In summary, ANR has been shown to be effective in headsets, such as the VIS. However, the at-ear exposure of the VIS headset is a problem. Since military soldiers perform continuous missions, such intense noise may cause permanent hearing loss over extended periods of time. For this reason, if the ANR electronics inside the VIS headset could provide a reduction in noise when the communication system is activated, this would reduce the noise level that reaches the eardrum when someone is speaking.

HUMAN AUDITION AND NOISE

Anatomy and Physiology of the Ear

The normal, healthy human ear transforms sound pressure oscillations into sensory information that the brain resolves through nerve impulses. Anatomically, the human ear is divided into three parts (as shown in Figure 4): the outer ear, middle ear, and inner ear, with each part contributing to the hearing process.

Outer ear. The outer ear is composed of the pinna and auditory canal. The auditory canal is the open channel leading to the tympanic membrane, or eardrum. The pinna collects the sound waves, modifies them, and funnels them to the ear canal. As the sound waves travel through the ear canal, they are amplified in the frequency range of 2000 and 4000 Hz by 10 to 15 dB. For this reason, sounds that fall within the 2000 to 4000 Hz frequency range are more hazardous to a person's hearing because the ear is more sensitive to sounds in this region. Once the sound waves reach the end of the ear canal, they vibrate the eardrum at the same frequency as the pressure wave striking it. The canal ends at the eardrum, which separates the outer ear from the middle ear.

Middle ear. The middle ear has two functions: (1) to protect the inner ear from foreign bodies, and (2) to transmit the acoustical energy of the original sound wave to the inner ear (Ward, 1986). The middle ear consists of the tympanic membrane, ossicles (comprised of three small bones called the malleus, incus, and stapes), and the oval window. The stapes is attached to the oval window; the malleus is attached to the tympanic membrane, and the incus is located between the malleus and stapes. Functionally, the malleus moves with the eardrum and transfers the motion to the incus.

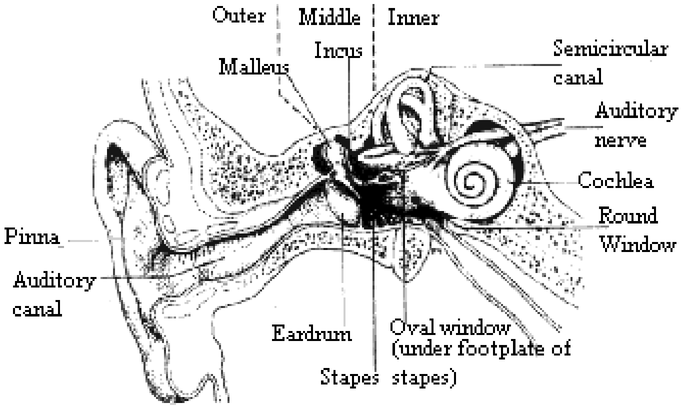


Figure 4. Anatomy of the ear (from Goldstein, 1989).

In turn, the incus transfers the motion to the stapes, which presses the oval window of the inner ear.

Due to the size difference between the tympanic membrane and the oval window (the tympanic membrane is 17 times larger than the oval window) and the orientation of the three bones in the ossicles, the vibration of the tympanic membrane is amplified by a factor between 22 and 100 before the vibrations reach the oval window. This amplification is needed because as sound travels from one medium to another, much of the energy is lost. If the amplification process did not take place, only about 1/1000th of the original sound energy would be transmitted to the inner ear (Ward, 1986).

Additionally, the tensor tympani and the stapedius muscle are two muscles that are attached to the malleus and stapes. These two muscles contract when one or both ears are exposed to a high intensity sound greater than 80 dB (Kryter, 1994; Ward, 1986). This contraction reduces the amount of high intensity noise reaching the inner ear. Richards and Goodman (1977) reported in Kryter (1994), that the intensity of the sound is proportional to the degree of contraction that occurs for short duration bursts of noise. For instance, if the intensity increases, the degree of contraction increases. However, with a steady state noise, the aural reflex is not relevant as a defense mechanism. Kryter (1994) stated that the ear becomes relaxed and adapts to an intense steady state noise after about 15 minutes.

Inner ear. The inner ear consists of the cochlea and the auditory nerve. The inner ear converts the mechanical vibration at the oval window into fluid motion in the cochlea, which causes the basilar membrane to vibrate. Located on top of the basilar membrane is the organ of corti, which consists of hair cells and the tectorial membrane.

The vibration of the basilar membrane causes the hair cells to vibrate. The hair cells cause a set of bristles (or cilia) atop of the hair cells to brush against the tectorial membrane. The hair cells act as tiny transducers that convert the mechanical impulses they pick up from the fluid in the cochlea into electrical energy that the brain interprets as sound.

Noise and Hearing Loss

An extensive amount of research has been conducted about the potential of noise to cause hearing loss. The appropriate frequencies, intensities, and durations of loud noise can result in permanent hearing loss. Jones (1983) stated that there are three type of hearing loss attributed to loud noise exposure: temporary threshold shift, noise-induced permanent threshold shift, and acoustic trauma. Temporary threshold shift (TTS) and noise-induced permanent threshold shift (NIPTS) are two types of conditions that would be experienced by a person in 114 dBA tank noise, because this noise is sufficiently strong and continuous. Acoustic trauma is experienced when a person is exposed to a sudden, very high intensity sound, such as a gunshot, explosion, or blow to the head.

Temporary threshold shift is a short-term elevation of the hearing threshold after exposure to intense noise. Over time, the person will regain their normal threshold of hearing and the hearing loss will disappear. The duration and intensity of a sound determine the magnitude of the threshold shift. For instance, if a person is exposed to a 90 dBA sound for 10 minutes, his/her recovery period will be shorter than a person exposed to a 90 dBA sound for 1 hour.

Miller (1974) stated that sounds that exceed 60-80 dBA will cause a person to experience a temporary threshold shift. Therefore, a person will likely experience some

TTS when the ANR system in the VIS headset is activated in 114 dBA noise because the noise level to which he or she is exposed is 83 dBA. Additionally, when a person is exposed to noise levels of 90 dBA, which is when the communication system is turned on and no one is speaking, and 94 dBA which corresponds to the noise level when the communication system is on and the crew is talking, he or she will also likely experience TTS.

On the other hand, noise-induced permanent threshold (NIPTS) shift is just as the name implies--permanent. The loss of hearing will not be restored over time. A TTS less than 25 dB over 8 hours will disappear after 16 hours of quiet (Jones, 1983). However, continuous exposure can produce a permanent threshold shift.

Extra-Auditory Effects of Noise on Humans

Smith and Miles (1985) conducted an experiment in which subjects performed a serial choice reaction-time task once they had been exposed to 75 dBA noise for two and five hours, respectively. They concluded that there was an increase in the number of errors for subjects who had experienced the noise for five hours. The results from this study are consistent with studies conducted by Broadbent (1979), in which noise levels greater than 95 dBA, with a shorter exposure time, were used.

Broadbent (1979) stated that low to moderate levels of noise might increase a person’s performance in routine tasks. However, when humans are performing tasks that may require concentration, they are more vulnerable to noise disruption than when completing routine tasks (Broadbent, 1979). Therefore, it may be stated that the effect of noise on performance depends on the task to be performed.

In some environments, such as those of the military, tasks that may seem routine during training may not be routine during wartime. For example, intense noise and the realization of the dangers inherent in the environment may interfere with a soldier's ability to make timely and accurate decisions. Therefore, soldiers may struggle to adapt to the environment, especially if it is their first experience of such an environment. Intense noise, such as the 114 dBA tank noise, may affect cognition, which may cause a decrement in performance. Additionally, the noise level of a tank may cause physiological disorders, induce stress, and impair decision-making. The following is a summary of the effects these variables have on a human.

Cognition and performance. When conducting command and control tasks in a noisy environment, the noise levels may be high enough to affect performance. Tank noise is low frequency biased, meaning that the noise dominates the lower frequencies more than the higher frequencies. This noise level has been reported to be between 110 and 115 dBA (Whitaker, Peters, and Garinther, 1990). More specifically, the noise level inside an operational U.S. Army Bradley Fighting Vehicle is 114 dBA.

A vast majority of the research centered around high noise levels (defined here as greater than 100 dB) and performance is not conducted on human subjects, due to the risk of causing hearing damage (Smith, 1989). Smith (1989) stated that results from studies using noise levels greater than 100 dB can be correlated to noise levels between 75 and 85 dB with longer duration. Davies and Jones (1982), cited in Sanders and McCormick (1993), stated that noise would impact performance on tasks that are continuous, such as command and control tasks. Weinstein (1974) stated that performance would decline over time on tasks that are simultaneous. Eschenbrenner (1971), cited in Sanders and

McCormick (1993), stated that tasks that pose high information processing and perceptual skill are also affected by noise. Command and control tasks have characteristics described by these authors previously cited, in that they are continuous, performed simultaneously, and require high information processing (Kane and Kay, 1992; Expanded Cognitive Assessment Battery, 1988).

Weinstein (1974) conducted an experiment on the effects of noise on an intellectually challenging task. He had two groups of subjects proofread text containing contextual errors and noncontextual errors in a quiet and a noise condition. Contextual errors are errors such as misspelled words, and noncontextual errors are considered grammatical errors. A teletype keyboard and a paper reader were used as sources of background noise. The noise level at the participant’s ear was 70 dBA, which was presented for 30 seconds every minute. The noise level in the quiet condition was 39 dBA. Weinstein (1974) reported that subjects in the noise condition did not differ from the subjects in the quiet condition when detecting noncontextual errors. However, the noise group differed significantly in finding contextual errors in the passage that was proofread. Also, the experiment revealed that subjects took longer to proofread the text in the noise condition than in the quiet condition. These conclusions led the researcher to note that when detecting contextual errors, different cognitive processes are used than when detecting noncontextual errors, meaning that contextual errors require more decision- making for them to be detected. Conversely, when detecting noncontextual errors, pattern recognition skills are used, which do not require as much cognitive processing.

Theologus, Wheaton, and Fleishman (1974) conducted an experiment on the effects of intermittent and random noise. The random noise consisted of 85 dB of broadband white noise in bursts of random duration. The intermittent noise consisted of broadband white noise of 85 dB presented in five-second bursts, with two seconds of quiet between each burst of noise. Each task assessed a different cognitive ability: reaction time, rate control, and time-sharing. The reaction time task consisted of the participant pressing a button when a light behind the button came on, and the rate control task consisted of tracking a light with a joystick. The time-sharing task consisted of the participant performing both the reaction time and rate control task simultaneously. The researchers found that the reaction time task was affected by the random noise and the rate control task was not, while the time-sharing task was affected by continuous exposure to the random noise. Intermittent noise showed nonsignificant effects across all three tasks.

The studies discussed briefly in the preceding passages cannot be considered representative of all tasks, because different tasks pose different cognitive demands on a person (Theologus et al., 1974). However, certain aspects of cognition are affected differently. For instance, in the Weinstein (1974) study, he stated that noise affects only tasks that require a person's decision-making skills. Therefore, according to Weinstein (1974), tests that are classified as simple motor tests would not be affected by noise because they do not require a high level of cognitive processing.

Physiological effects. It has been well established in the literature that noise is capable of producing physiological responses in humans (DeJoy, 1984; Suter, 1992). Rosen (1970) stated that loud noise causes blood vessels to constrict, pupils to dilate, and the voluntary and involuntary muscles to tense. Also, increased heart rate and peripheral blood flow was found with increased noise exposure (Griefahn and DiNisi, 1992).

DeJoy (1984) and Suter (1992) stated that noise is considered to be a non-specific biological stressor, which evokes a response that prepares the body for action. The stimulation of the brain’s reticular activating system by noise is thought of as being responsible for this action. The reaction is influenced by an activation of the autonomic nervous system. Anticaglia and Cohen (1970) stated that this system is believed to be responsible for physiological effects due to noise.

The majority of the research centered on physiological responses to noise deals with the cardiovascular system. More specifically, blood pressure has been shown to increase with noise exposure (Anticaglia and Cohen, 1970; DeJoy, 1984). Suter (1992) stated that the Environmental Protection Agency performed a study on the effects of noise on blood pressure in monkeys. The experimenters used typical office noise, 85 to 90 dBA, to which the monkeys were exposed for nine months. The results of the experiment showed a rise in systolic and diastolic blood pressure. The fact that the increase in blood pressure remained long after the exposure stopped, raised major concerns. However, it is important to note that the results from experiments using rodents and other laboratory animals cannot be extrapolated to humans (Suter, 1992). The primary reason why animals are used instead of humans is due to the long exposures that were experienced by the monkeys. It is not safe for humans to be exposed to 85 to 90 dBA noise levels for nine months without adhering to the OSHA regulation. The OSHA regulation states that workers who are exposed to more than a 90 dBA time-weighted average over an eight-hour shift are to be protected by means of administrative and engineering controls. When workers are exposed to more than 85 dBA, over an eight-hour shift, they are to be enrolled in a hearing conservation program (OSHA, 1990).

The effects of exposure to high noise levels have been studied numerous times in the literature, with the outcome being fairly consistent. Cohen (1976) studied the effect of a hearing conservation program on extra-auditory problems experienced by 400 workers working in high noise (95 dBA or higher) areas in a boiler plant. Prior to the introduction of a hearing conservation program, the boiler plant experienced problems such as frequent job injuries, medical problems, and absences. However, after the implementation of the hearing conservation program, in high noise areas, these problems were reduced significantly. Cohen (1976) also found that in low noise (80 dBA or less) areas of the same boiler plant, the problems were not different from problems in the high noise level areas. Therefore, it was concluded that the high noise contributed to extra-auditory problems.

A study conducted by Talbott, Helmkamp, Matthews, Kuller, Cottington, and Redmond (1985) found that there was not a significant difference in blood pressure for workers exposed to high noise levels as compared to those who are exposed to lower noise levels. However, they did state that there is a relationship between severe noise-induced hearing loss and high blood pressure, regardless of the exposure level of the noise.

There have been other physiological disorders identified as being related to excessive noise, such as headaches and reproduction difficulties in females. Ohrstrom and Bjorkman (1983), reported in Jones (1983), stated that workers in textile mills complain frequently about headaches due to periodic noise. In an Australian study, conducted by Fernandez and Sheffield (1996), noise was one of the main causes of headaches, while mental stress was the most important factor.

Stress. Command and control in a military environment is very demanding; lives depend on the efficient and expeditious decisions of the commander or person who is in charge of issuing commands. In this environment, stress is an extremely crucial issue. Sanders and McCormick (1987) defined stress as an “undesirable condition, circumstance, task, or other factor that impinges upon the individual” (p. 197).

The onset of stress can arise from external or internal events or a combination of the two. For instance, an example of external events may be the nature of the working environment (such as noise and heat), or the pressure felt when outnumbered by adversaries during war. An example of an internal event may be a person realizing the consequences of making a critical mistake. A situation in which an external and an internal event could trigger stressors would be a soldier having difficulty understanding his commander in a critical situation due to noise. The soldier may experience a sense of fear while under attack by enemy forces, which may decrease performance.

Finch and Stedmon (1998) developed a taxonomy of stressors, shown in Table 1. Based on Table 1, noise would be classified as a physical stressor. However, it has already been shown in the literature that noise can be classified as both a physiological and a psychological stressor. Therefore, the stressors in Table 1 are not limited to the category to which they are assigned. The effects of the stressors are not dependent only on the task and the environment, but also on the traits of the individual performing the task.

Driskell and Salas (1996), cited in Wickens, Gordon, and Liu (1998), stated that various stressors have been shown to affect task performance, such as crowding, noise, performance pressure, workload, and anticipated threat of shock or dangerous conditions.

Table 1. A taxonomy of stressors (from Finch and Stedmon, 1998).

Classification	**Stressors**	**Example**
Physical	Motion	Soldiers map-reading during patrol marches
	Vibration	Radar operators inside a ship during heavy swells
	Temperature	Tank gunners performing heavy work in NBC units
	Lighting	Aircraft weapon-loaders working on unlit flights
Chemical	Medication	Soldiers who have consumed anti-nerve agent drugs
	Alcohol	Short term and long term effect following abuse
	Caffeine	Anxiety attacks following/during over-use
	Sleep deprivation	SAR pilots during long endurance flights
Physiological	Illness	Performance inhibitor during combat missions
	Exhaustion	Rapier crews during prolonged states of high alert
	Injuries	Initial trauma followed by restricted performance
	Sleep deprivation	Soldiers on long observation exercises
Psychological	Social	Soldiers requested to ‘volunteer’ for dangerous tasks
	Fear	Unknown situations with real possibility of dying
	Shock	Airmen seeing a colleague shot down
	Sleep deprivation	Airborne radar crews during repeated missions

Military operations can be classified as having these stressors because soldiers are exposed to hostile environments in which they must make critical decisions in a short time period. More specifically, during C^2 tasks, vital information is exchanged on a continuous basis. As the information is delivered to other soldiers, they are to execute the task as soon as possible, because other information is steadily being filtered through the system.

Van Gemmert and Van Galen (1997) conducted experiments focused on the effects of cognitive stress and physical stress on human performance when performing an aiming task and a number-writing task. Cognitive stress is encountered when a multi-faceted task is performed, and physical stress is encountered when a person's environment is induced by stressors, such as noise. The results from their study showed how reaction time increased and movement time decreased under high levels of cognitive stress during the number-writing task. In addition, they reported that reaction time and movement time decreased for the writing task under physical stress. Both physical and cognitive stress did not have an affect on the aiming task.

Lehto (1997) stated that stress, along with time pressure, could influence decision-making. He stated that stress caused by time pressure causes poor performance and a shift in the cognitive strategies used to make decisions. The stressors faced in a military environment may cause this shift. For instance, if a person is placed in a life or death situation, they may abandon the information stored in long-term memory (which was gained from training) and rely solely on their intuition to make a decision. Cognitive and physical stress may have significant bearing in a military environment, thus leading to a decrease in performance. Physical stress will greatly be influenced by the noise in the environment, while cognitive stress may be due to frustration from the environment and not fully understanding the commander.

Decision-making. Whenever a decision is made, the decision-maker must consider the risk involved with the decision. By exhausting all the possible choices to solve a problem, the most desirable action will be selected. However, when other factors are present which may interrupt the person's decision-making process, problems may occur. Janis (1982) stated that time stress and attentional resource limitations are two factors that may contribute to problems in decision-making. Wickens et al. (1998) stated that such factors are most commonly found in complex and dynamic operating environments.

In a dynamic environment such as the battlefield, naturalistic decision-making may be used to decide on the best action to reach a goal. Wickens, Gordon, and Lui (1998) and Zsambok (1997) stated that naturalistic decision-making is the way people use their experience to make decisions in field settings. Furthermore, Wickens et al. (1998) cited several researchers as defining characteristics of tasks that are performed in real world environments. These characteristics are:

- Ill-structured problems
- Uncertain, dynamic environments
- Information-rich environments where situational cures may change equally
- Cognitive processing that precedes in iterative action/feedback loops
- Multiple shifting and/or competing individual and organization goals
- Time constraints or time stress
- High risk
- Multiple persons involved in the decision

During C^2 tasks, a soldier may receive information, ask for additional information, make decisions, and evaluate those decisions. The information may be presented as distorted or clear, simple or complex, and complete sentences or sentences filled with gaps. Nevertheless, in instances where information is not easily discernable, the listener interprets the talker's intent to the best of his or her ability and makes a guess based on his or her experience or education concerning their next action.

Kaempf, Klein, Thordsen, and Wolf (1996) conducted a study on the decision-making undertaken by experienced naval officers in a complex, time-pressured command and control setting. They reported that when the officers made decisions, they used recognitional decision-making (RDM), and that situation awareness was of major concern. Recognitional decision-making occurs in simple and complex forms. Simple RDM is when the decision-maker identifies the situation as being familiar or typical. This concept allows the decision-maker to implement a course of action without any hesitation (Kaempf et al., 1996). Complex decision-making is when the decision-maker evaluates the courses of action. The decision-maker closely examines each course of action to determine which one will work by simulating the action mentally. Situational awareness refers to the state of knowledge that includes the perception of things in the environment, the understanding of their meaning, and the understanding of their anticipated status in the future (Kaempf et al., 1996).

Endsley and Smith (1996) conducted a study on decision-making involving fighter pilots engaged in a tactical task. They found that there was a high degree of variability in decision-making among highly trained and experienced pilots. They found that many pilots would plan ahead based on the actions reported to them, thus allowing them to determine an exit from the battle. However, from the two studies cited, it is important to emphasize that their results may not correlate to other tasks, because the problems and level of difficulty encountered are different, thus making decision-making different.

Summary

From the studies cited, it is obvious that noise has effects on humans. In command and control tasks, which involve monitoring, talking, listening, making decisions, and writing/typing, performance may be disrupted. Therefore, it seems best to have some type of engineering control in place to allow a reduction in noise levels. An engineering control, namely a hearing protection device of some type, may reduce the amount of stress due to noise and improve decision-making. The reduction of noise levels may contribute to better task performance. Additionally, since speech is important for command and control tasks, the degradation of speech due to noise may increase the amount of stress and impair decision-making. Therefore, it is better to abate the noise, which may improve performance, while at the same time reducing hearing loss and improving speech intelligibility.

SPEECH INTELLIGIBILITY

Communication is one way that people express their ideas and thoughts to one another. Communication can be conducted by sign language, braille, written words, or speech. Speech is the most common mode of communication, allowing people to exchange information verbally. The quality or naturalness of speech lends to a person's understanding of what has been spoken. Naturalness refers to the clear quality of speech (Crandall, 1917).

Words are formed from combinations of syllables. Syllables are formed by two or more phonemes, and are the basic unit of speech perception. Phonemes are the smallest units of speech, which if changed, will change the meaning of the word. They correspond to the letters or combinations of letters of the alphabet and comprise the consonant and vowel sounds. Vowels are the parts of the speech waveform that are low in frequency and high in sound energy, while consonants are higher in frequency and have less amplitude (Suter, 1992).

The goal of the listener in tasks where speech communication is important is to parse speech sounds into language that is meaningful. By parsing the sounds of speech, the person is able to decipher the meaning of various words, thus making speech intelligible. Speech intelligibility is the degree to which a speech message is correctly identified by the listener. The success of the listener depends on the listener's ear in receiving and interpreting the various components of speech in their initial order and spacing in time (French and Steinberg, 1947). More importantly, French and Steinberg (1947) stated that recognizing and interpreting speech depends on the intensity of the sound at the ear, and the intensity of the other sounds that are present, such as ambient noise, which may cause masking. For example, when the communication system of the VIS headset is activated, the diaphragm of the microphone picks up speech and ambient noise. The ambient noise (tank noise) is considered the masker of the speech.

Speech Intelligibility Tests

Over the years, test materials have been developed to measure speech intelligibility. Hawley (1977) stated that speech intelligibly tests offer the most convincing and trustworthy method of measuring the ability to understand speech. These tests use nonsense syllables, monosyllabic words, single words, and sentences. Since consonants contribute the most to intelligibility, most tests give emphasis to the ability to distinguish them.

ANSI S3.2-1989. ANSI S3.2-1989 is titled the *Method for Measuring the Intelligibility of Speech Over Communication Systems* (ANSI, 1989). This test has three different methods for measuring speech intelligibility: the phonetically balanced word test, the modified rhyme test, and the diagnostic rhyme test. Each of these three methods will be discussed individually.

The *phonetically balanced* (PB) word test is the oldest of the three tests mentioned. It consists of a 20-word list, consisting of 50 monosyllabic English words in each list. The frequency of occurrence of various phonemes is proportional to their frequency in everyday English. The words are presented to the listener in a carrier sentence (“Would you write______now”). The listener responds by writing his or her response in an open format.

When using this test, at least 10 hours of training is needed for the listeners to attain and maintain stable performance (ANSI, 1989). Once the desired performance level has been attained, the test takes approximately 200 seconds to administer each PB word list. Although this test has disadvantages, PB words are still useful because they are sensitive to changes in the speech-to-noise ratio. For instance, a change in the signal-to-noise ratio between –5 dB to 5 dB would produce a large change in intelligibility (ANSI, 1989).

The *modified rhyme* test (MRT) consists of 300 monosyllabic English words grouped into 50 six-word sets. The sets of six-words are arranged according to response ensembles, with each ensemble characterized by one vowel that is the nucleus of every word (ANSI, 1989; House, Williams, Hecker, and Kryter, 1965). While most of the words are in the consonant-vowel-consonant (CVC) format, others are in the form of CV or VC. A carrier sentence (identical to the carrier sentence used for the PB test) may or may not be used to present the participant with one word from each ensemble. The participant responds by circling one of the six words in each ensemble.

It takes approximately 120 to 180 s to administer each 50-word set when carrier sentences are used. When carrier sentences are not used, the test takes approximately 75 s (ANSI, 1989), and since the test has a closed-response set, it is easy to administer and score.

The *diagnostic rhyme* test (DRT) is similar to the MRT, except it consists of 96 pairs of words. The DRT distinguishes differences only in the initial consonants and uses no carrier sentence (ANSI, 1989). The listener in the test has two choices between the pairs of words presented. The words are not presented to the listener via a carrier sentence. Because the word list is relatively small (as compared to the other two tests), the DRT is quicker to administer (ANSI, 1989).

Articulation testing method. Egan (1948) discussed another method to test the intelligibility of speech in which test words are phonographically recorded. This test consists of one person, who has a list of the words recorded on a record. The list of words are covered with an index card or sheet of paper. After the listener hears the word and decides what it is, he or she removes the sheet of paper covering the word to determine if the word was heard correctly. Obviously, this test relies heavily on a person's honesty to evaluate whether the word presented was the same as the word that was heard.

Fairbanks rhyme test. Fairbanks (1958) discussed the *rhyme test*, in which words are drawn from a 250-monosyllable vocabulary. This test consists of a 50-word list that is constructed by drawing one word from a set of 5 rhyming words. Within each set of five rhyming words, the words differ by initial consonant phoneme. The stem of each word is spelled alike, but the initial consonant (a single letter) is different. In this test, the stem of a word is all the letters except the first letter (e.g., -oy, -at, -ip). The participant is given a response sheet that has 50 stems, in which he or she is to write in the letter of the word once it is heard. Fairbanks stated that there are 536 possible words corresponding to the stems, with the responses for each stem being 8 or 9 possible words.

Speech identification test. Hochhaus and Antes (1973) conducted an experiment in which subjects rated their attempt to correctly identify a word masked in white noise. The experimenters used two lists of 50 words. The first list consisted of a context list of monosyllable food-related words, and the second list consisted of a random set of monosyllable words that match words in the first list in their frequency of occurrence in English. Each list was presented three times to each participant. The participant's task was to listen to the word and write the word they heard on a sheet of paper, indicating how certain they were about the word they wrote by circling a number between one and five, located next to the word. On the rating scale, 5 indicated that the participant was very certain, and 1 indicated that the participant was very uncertain (Hochhaus and Antes, 1973). The results showed a significant difference between the context list of monosyllable food-related words and the random list of monosyllable words.

Auditory discrimination test. In an audiology text on the evaluation of appraisal techniques in speech, Locke (1979) mentioned the auditory discrimination test (ADT), developed by Wepman. This test consists of 40 pairs of monosyllabic words. With this test, two words are read at the same time, and the participant's task is to determine if the words are the same or different. Of the 40 pairs of monosyllabic words, 30 are different (e.g. theft-sheaf), with the remaining 10 being identical. If the two words are the same, the participant is instructed to say, "same or yes." If the two words are different, the participant is instructed to say, "no or different." Locke (1979) stated that this test was invalid because of its inability to correlate well to English spoken words.

Summary

Of all the methods discussed for measuring speech intelligibility, the MRT test in ANSI S3.2-1989 appears to have the most promise. A review of the literature revealed that the MRT has been used in past studies involving measuring the intelligibility of speech (Morrison and Casali, 1997; Peters and Garinther, 1990; Urquhart, Robinson, and Casali, 2001). The MRT is better than the PB because it requires less training and is quicker to administer due to the closed response set of the test. Additionally, the MRT responses are indicative of errors in both initial and final consonants heard in single words, while the DRT is indicative of only errors in the initial consonants (ANSI, 1989).

EFFECTS OF NOISE ON SPEECH INTELLIGIBILITY

Hawley (1977) constructed a table of factors that affect speech intelligibility (see Table 2), in which one of those factors was noise. Noise can interfere with speech in various ways. For instance, if a person is talking in 114 dBA tank noise, the listener will not be able to hear what is spoken due to the level of the background noise. Additionally, a person's hearing threshold will be elevated in the presence of loud noise, which will also interfere with that person's ability to hear speech.

Masking. When a signal is presented in the presence of a background noise, and the noise is at a sufficiently higher level than the signal, the noise level masks the signal. Masking is experienced when the threshold of audibility of one sound is raised in the presence of another sound. Much of the research done on masking is based on experiments using broadband noise as the masker. Broadband noise is noise whose sound energy is distributed across all the frequency bands. The frequency range of speech is from 1000 to 4000 Hz, and noises whose frequencies are the strongest in this region will mask speech more seriously.

Arlinger and Gustafsson (1991) conducted an experiment in which they used low redundancy sentences that consisted of five words each, to determine if an amplitude modulated noise will mask human speech to a higher degree than a steady-state noise. Amplitude modulation is the change in the peak of a sound wave, in which the peak may be reduced or amplified. An example of amplitude modulation is shown in Figure 5. The pool of participants consisted of young, normal-hearing listeners and two groups of elderly people: those with normal hearing (54-69 years), and those with light to moderate

Table 2. Factors affecting speech intelligibility (adapted from Hawley, 1977)

- Interference
  - Continuous white, pink, speech spectrum, aircraft, machinery noise
  - Impulsive, interrupted, fading, modulated noise
  - Sine waves, square waves, sawtooth waves
  - Competing messages, backwards speech
- Speech processing
  - High-pass, low-pass, band-pass, specially shaped filtering
  - Compression, expansion in time with or without frequency shift
  - Shifts, transposition, or inversion of frequencies
  - Compression limiting in dynamic range
  - Center or peak clipping
  - Intermodulation or harmonic distortion
  - Interruption with different duty cycle, switching rate, gaps empty or filled with noise, same words, other words
- Speech characteristics
  - Pitch, rate of speaking, vocal effort, inflection, regional or national accent, language, age, sex of speaker
- Listener characteristics
  - Intelligence, motivation, vocabulary size, hearing impairments, speech impairments, native language, age and sex of listener
- Speaking conditions
  - Reverberation time, side tone with or without delay and level change
  - In noise, high or low air pressure, helium, confining equipment
- Listening conditions
  - Free field or on earphone
  - In noise or competing messages from same or different azimuth
  - Reverberation time
  - Listening level, forward or backward masking
  - Monaural or binaural presentation, speech at same or different phase at two ears, noise at same or different phase
  - Signal simultaneous or delayed at two ears, with and without noise
  - Signal switched between ears at various rates, with and without noise
  - Different signals at two ears (dichotic listening)
  - In noise, high or low air pressure, underwater, confining equipment
- Combinations of all the above

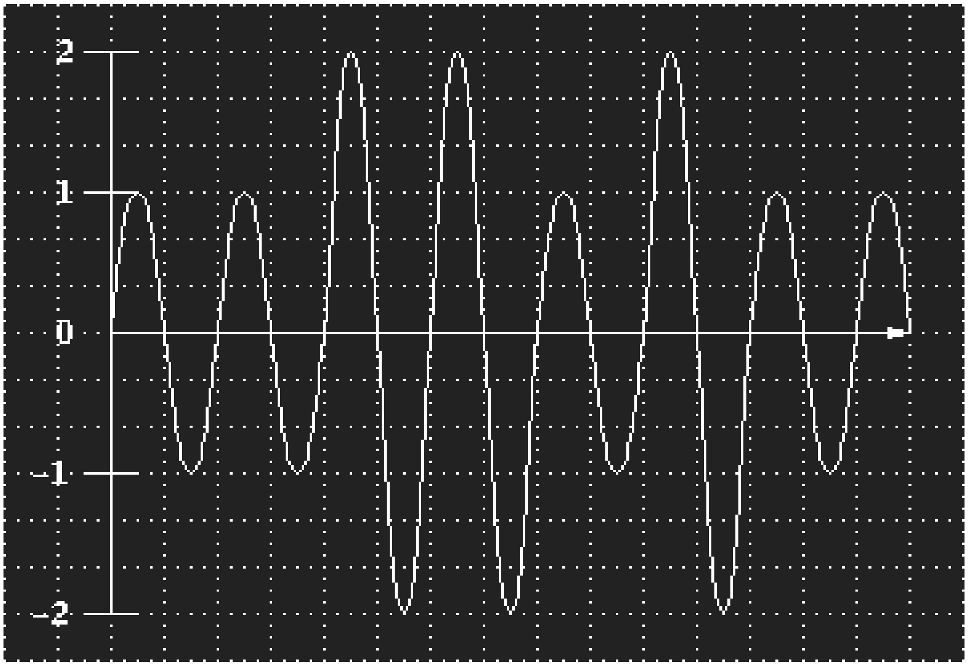


Figure 5. An example of amplitude modulation.

sensorineural hearing loss (60-70 years). In the experiment, subjects were monaurally presented ten sentences at 65 dB (linear). For the hearing impaired group, instead of using 65 dB (linear), the most comfortable sound pressure level was chosen for the subjects. The researchers reported that amplitude modulation improved speech recognition for the young, normal hearing group, as well as for the elderly groups. They concluded that amplitude modulation in the masking noise provided a release of the speech masking, but just the opposite was noticed with unmodulated noise.

Stevens, Miller, and Truscott (1946) investigated how three waveforms (sine waves, square waves, and pulses) can mask speech. They reported that sine waves of low frequency masked speech more than tones of high frequency. This is due to the upward spread of masking to the higher frequencies as the intensity is increased. Furthermore, this upward spread of masking is aided by the distortion in the ear, which produces aural harmonics at high sound levels. The square waves and the pulses were more effective at masking speech than the sine waves. This was due to the wider frequency range that the square wave and pulses covered, especially at lower frequencies.

Noise exposure. Pollack (1958) conducted an experiment on speech intelligibility in high noise levels for a short time period. He successively used eight phonetically balanced (PB) word lists, which were read against background noise levels between 45 and 130 dB (linear) for a 13 minute period. He reported that speech intelligibility decreased because of continuous exposure to noise levels of 115 dB (linear) and higher. In addition, he concluded that at noise levels above 120 dB (linear), the speech intelligibility scores were halved 13 minutes into testing. It may also be concluded that at such high noise levels, a person’s hearing threshold will be raised, which may cause a decrement in the intelligibility scores.

Workload and Speech Intelligibility

There is a consensus in the literature that a decrease in speech intelligibility causes an increase in workload (Whitaker and Peters, 1990; Whitaker, 1991). Whitaker, Peters and Garinther (1989) conducted a study on the effects of intelligibility on crew performance in a tank simulator. They had subjects performing tasks under different levels of speech intelligibility (0, 25, 50, 75, and 100 percent). The subjective workload assessment technique (SWAT) was used to measure the workload due to changes in the speech intelligibility level. The SWAT analysis showed that as speech intelligibility decreased, it interfered with performance and increased mental workload.

Whitaker and Peters (1990) studied the effect of speech intelligibility among Bradley Fighting Vehicle crewmembers. The crewmembers were tested in a Bradley Fighting Vehicle simulator. The primary goal of their task was to execute the instructions from the commander, which were to navigate the tank and shoot. It was concluded that the SWAT ratings were affected, and the navigation performance declined at the first drop in speech intelligibility. Results identical to these two studies have been reported elsewhere (Garinther and Peters, 1990; Garinther, Whitaker, and Peters, 1994; Peters and Garinther, 1990; Whitaker, Peters, and Garinther, 1989; Whitaker and Peters, 1993). Hence, the research conducted on speech intelligibility and workload shows that a reduction in the intelligibility of speech will lead to an increase in workload.

There also exists a relationship between speech intelligibility and C^2 tasks, since verbal communication plays a major role in the execution of C^2 tasks. Verbal communication allows a soldier to instantly exchange information in critical situations that require immediate feedback. For example, if an armored vehicle is in the area of enemy forces, there may not be enough time to type information into an electronic whiteboard; instead, a soldier might rely on verbal communication to relay information. Also, if the status reports, electronic whiteboard, electronic mail, and electronic maps fail to function properly, verbal communication may be relied upon more heavily. If verbal communication is misunderstood, the results may be grave.

WORKLOAD MANIPULATION AND ASSESSMENT

Sophisticated systems, such as those used for C^2 tasks, may impose very high levels of workload on the operator. Tsang and Wilson (1997) stated that the results of high levels of mental workload are errors and system failure. Such high levels may be experienced by a commander performing C^2 tasks. The reason may be due to the need for the commander to process information displayed on a computer screen rapidly and accurately in a short time span, in addition to the other events occurring in his or her environment.

Criteria for Workload Techniques

The majority of the research surrounding mental workload techniques involves determining which workload technique to use. For this reason, the following criteria were developed to assist researchers in assessing mental workload:

(1) *Sensitivity*: As the difficulty level or resource demand of the task changes, the workload technique should detect the change. Sensitivity deals with the degree to which the test chosen can detect changes in the levels of load imposed on the operator.

(2) *Diagnosticity*: Wierwille and Eggemeier (1993) stated "diagnosticity is the ability to discern the type or cause of workload, or the ability to attribute it to an aspect or aspects of the operator's task" (p. 264). The test should not only indicate when the workload of the operator changes, but the cause of the change. Knowing what is causing a change in workload can assist in implementing solutions to a problem.

(3) *Selectivity*: Wickens (1992) stated that the measure should be selectively sensitive to differences in capacity demand and not to changes in such factors as physical load or emotional stress, which may be unrelated to mental workload or information-processing ability.

(4) *Intrusion*: Intrusion occurs when the technique that is responsible for measuring workload interferes with the task that the operator is performing.

(5) *Reliability*: As with any measure, it must be reliable. However, in a time-changing environment such as the battlefield, the workload will have to be tracked from task to task. Therefore, the technique chosen should be reliable enough to estimate changes.

Many techniques exist for assessing workload, with many of them meeting some of the criteria, but with few meeting all the criteria (Wickens, 1992). The techniques are categorized as being objective and subjective. Physiological and task performance are examples of objective measures of workload, and a questionnaire is an example of a subjective measure. Techniques under each of the categories are discussed below.

Objective techniques

Objective techniques used to measure mental workload use the operator's behavior to determine the workload imposed by a specific task. It is expected that an operator's performance will decrease as workload increases on a task. Conversely, if an operator's performance increases, the operator's workload level is low. Three objective techniques that are used to assess mental workload are primary task measures, secondary task measures, and psycho-physiological measures. Each of these methods are discussed individually.

Primary task measures. Primary task measures are task-specific, meaning that it is important that the tasks be closely related to the strategy used by the operator to perform tasks (Sanders and McCormick, 1993). With this method, the performance of the operator and system are monitored to determine the operator's changes as the task varies. For instance, when studying a military tank, one can evaluate a person's performance while operating the tank, with the tank being the system, and determining how well the operator is operating the tank across different terrain being the primary task.

Meshkati, Hancock, Rahimi, and Dawes (1995) stated that primary task measurement is the most apparent method for measuring mental workload. Workload can be measured as the number of errors made and/or time taken to perform the primary task. If many errors are made in a timed task, it may be stated that the task imposed a high workload. On the other hand, if the task was completed with 100% accuracy in the least amount of time, it may be stated that the workload level was low. Sometimes the primary task measurement is used to assess other workload measures, such as an operator's performance on a secondary task (Williges and Wierwille, 1979).

Secondary task measures. A secondary task is a task a person performs concurrently with the primary task of interest. An example of a secondary task would be a person driving a car and adjusting the radio station. Driving is the primary task, and adjusting the radio station is the secondary task. Since the primary task and the secondary task are competing for limited resources, changes in one measure would show a change in another measure. For instance, if the operator is able to perform well on the secondary task, then the primary task is believed to be easy. Conversely, if the secondary task suffers, the primary task is believed to be very demanding. The workload is measured as the difference between the two conditions, with and without the primary task (Meshkati et al., 1995).

O'Donnell and Eggemeier (1986) stated that the secondary task changes the nature of the primary task, and contaminates the workload measure. Wierwille and Eggemeier (1993) stated that an alternative way to reduce the chances of the data being contaminated would be to identify a task that already exists in the procedures and use it as the secondary task. For example, an operator performing command and control tasks could be a primary task, while responding to communications could be used as a secondary task. In the previous example, the secondary task is in the realm of the overall task, and would not seem artificial to the operator.

Psycho-physiological measures. Psycho-physiological methods measure physiological responses to determine workload. Charlton (1996) stated that the U. S. Army has used physiological measures extensively to determine a soldier's physical workload in the battlefield. They assessed the soldier's heart rate, ventilation rate, skin temperature, and core temperature. Charlton (1996) stated that physiological tests are not used as much for measuring mental workload due to the cost and complexity of some of the instrumentation.

Subjective Techniques

A great deal of research has been devoted to determining a way to measure mental workload subjectively. Subjective methods of measuring mental workload are the most common methods in the human factors community (Charlton, 1996). These methods rely on the operator to rate the level of mental effort imposed by a task. The methods range from simple, unidimensional scales to multidimensional scales. Unidimensional scales provide the experimenter with a single workload measurement, while multidimensional scales measure several components of workload. The most commonly used subjective methods are: the Cooper-Harper, Subjective Workload Assessment Technique, and the National Aeronautical Space Administration Task Load Index. Charlton (1996) stated that these measures are “sensitive to changes in workload levels, minimally intrusive, diagnostic, convenient, relevant to a wide variety of tasks and posses a high degree of operator acceptance” (p. 188).

Cooper-Harper Scales. The Cooper-Harper scales were developed for the purpose of assessing aircraft handling characteristics. The major advantage of using the Cooper-Harper Scale is that it is easy to administer; it is popular in the scientific community, and the resulting ratings correlate highly with other workload scales (Charlton, 1996; Lysaght, Hill, Dick, Plamondon, Linton, Wierwille, Zaklad, Bittner, and Wherry, 1989). The major drawback of the Cooper-Harper scale is that it can only be used to measure workload associated with aircraft control.

Wierwille and Casali (1983) modified the Cooper-Harper Scale, calling it the Modified Cooper-Harper (MCH) Scale. The MCH is shown in Figure 6. The MCH can be used for applications beyond that of aircraft control. Additionally, the MCH allows the experimenter to measure workload associated with tasks that require the participants perceptual, cognitive, and communication skills.

The use of the MCH for C^2 task would be very beneficial to measuring mental workload on C^2 tasks because it taps into constructs that are important for the operator in

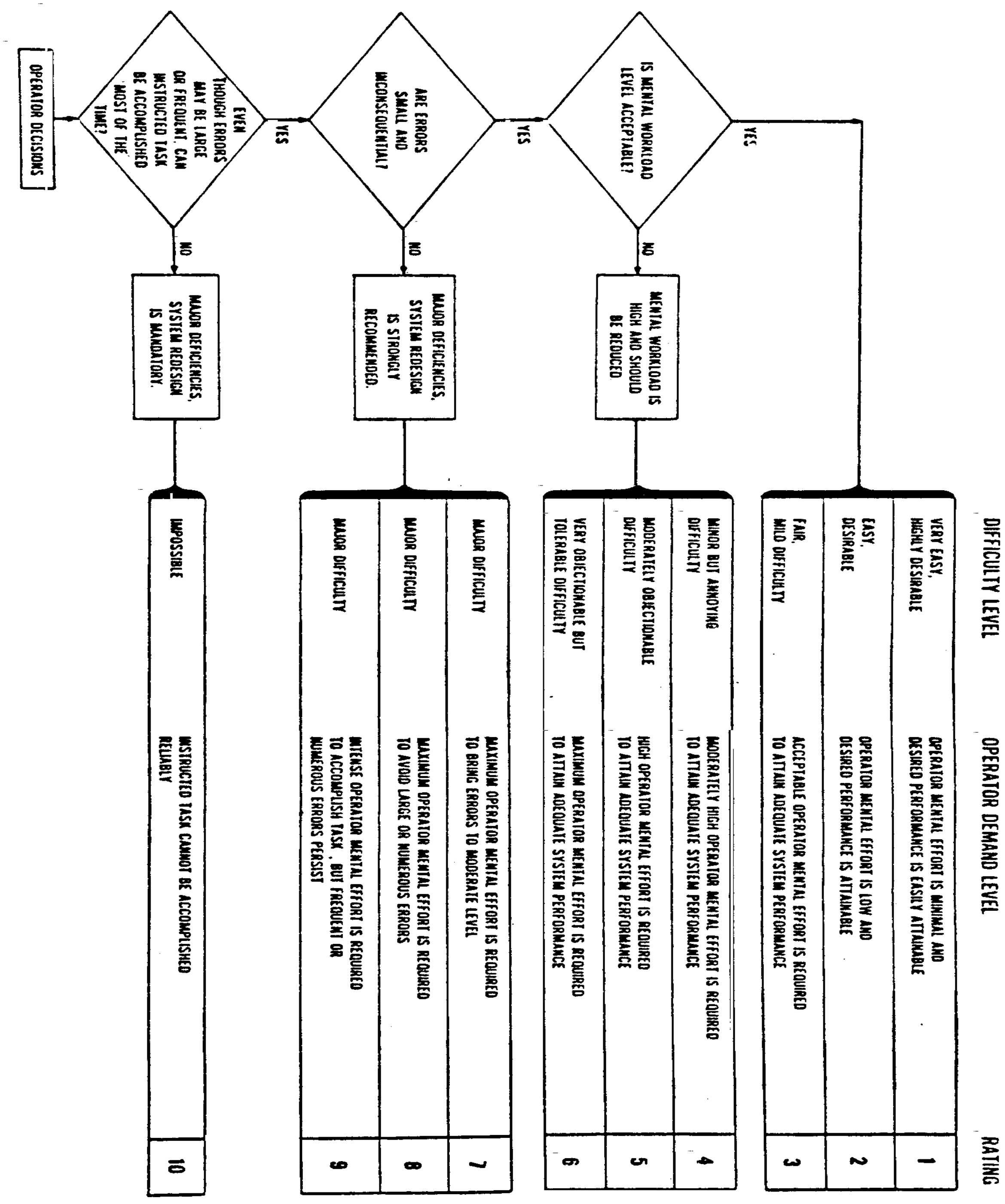


Figure 6. The Modified Cooper-Harper Scale (from Wierwille and Casali, 1983).

performing C^2 tasks. For instance, during C^2 tasks, the operators rely on their perceptual skills to listen and interpret information from different modalities. They use cognition to assist in helping them to make decisions, while communication is used to transfer information verbally once it has been received.

This test is usually administered to participants after the completion of the task. The quality of the data gathered from the participant depends heavily on how well he or she remembers the task. One may argue that a delay in time between the test completion and the administering of the MCH would lead to contaminated data. However, Lysaght et al. (1989) stated that delays in the time of a test and the administering of the MCH do not lead to contamination of the MCH.

Subjective Workload Assessment Technique. The subjective workload assessment technique (SWAT) has a multidimensional view comprised of three dimensions: time load, mental effort load, and psychological stress load (Reid, Shingledecker, and Eggemeier, 1981). Boyd (1983) stated that this technique "carries with it the assumption that people can accurately predict the amount of workload they would experience under various levels of the three dimensions" (p.125). The contribution of time constraints and interruptions to the measurement of workload corresponds to *time load*. This corresponds well to military C^2 tasks because sometimes a commander is faced with a situation that requires an instant response. The difficulty and uncertainty of the task corresponds to the *mental load* and the frustration, confusion, or uneasiness of the task corresponds to *psychological stress*. Mental load relates to C^2 tasks as well, because C^2 tasks do not consist of one single task, and they may demand a person to dedicate attention to several displays to make a decision, which may impose a mental strain on the decision-maker. Psychological stress can easily be experienced during C^2 tasks, because frustration and confusion may arise when communication that is necessary to complete a mission is not understood.

There has been an extensive amount of background research conducted on SWAT. The research has shown that this technique is reliable, well-developed, and is known as a valid metric of workload (Charlton, 1996). A disadvantage of SWAT is that it requires participants to be pretrained before the test is administered, and it requires a serious amount of groundwork to set-up the materials. Additionally, when the SWAT is compared to the NASA Task Load Index, the SWAT technique is considered to be less sensitive (Nygren, 1991). Charlton (1996) stated that the time and effort required to administer the test might cause the sample size to be small, due to subjects' aversion to participation, which is another drawback of using this method.

Whitaker, Peters, and Garinther (1990) used the SWAT in a study to determine how performance is affected by degradation in communication. The experiment consisted of thirty-six Bradley Fighting Vehicle soldiers who were tested performing a route exercise. The thirty-six participants were divided into twelve groups that consisted of three soldiers in each group. Five levels of intelligibility were set using the MRT. The levels of intelligibility were 1.9%, 27.8%, 51.1%, 72.7%, and 96%. After each test, the subjects rated how difficult it was to hear the words from the MRT, using the SWAT. The SWAT rating showed that as intelligibility improved, the level of difficulty of the route exercise decreased. The authors concluded that the effectiveness of military personnel is dependant on communication, thus if communication is poor, performance will suffer.

National Aeronautical Space Administration Task Load Index. The National Aeronautical Space Administration (NASA) Task Load Index (TLX) is similar to SWAT, in that it is based on a multidimensional approach. Instead of using three dimensions, the NASA-TLX uses six, and divides each dimension subscale into 20 levels. These six dimensions are mental demand, physical demand, temporal demand, performance, effort, and frustration. Two of the six levels within the NASA-TLX, mental demand and temporal demand, correspond to the mental and time load measures of SWAT, respectively. Also, in NASA-TLX, effort and frustration are different scales, whereas in SWAT they are combined into one scale--psychological stress. The performance scale allows the subjects to rate how well they performed on the task. The physical demand scale relates to how much physical activity was required to perform the task.

This technique employs a different weighting procedure than SWAT, in that participants are instructed to make a paired comparison on each subscale. If one subscale is used more frequently than another, then the number of times that subscale is chosen is used as the weighting for that subscale. A weighting of zero is assigned to a subscale that is never chosen. The workload metric is computed by taking the weighting of each subscale, multiplying it by its task-weighting, adding up the subscale scores, and dividing them by the total number of paired comparisons (Charlton, 1996).

Hill, Zaklad, Bittner, Byers, and Christ (1988) compared the NASA-TLX with four other scales: SWAT, MCH and Overall Workload (OW). The scales were applied to six soldier-operators of a Line of Sight-Forward (Heavy) air defense missile system several weeks after the participant had completed two missions. The participant's task was to view a videotape of two missions in which they had been participants and then complete each subjective rating technique. Once all the subjective rating techniques were completed, the researchers compared the validity of the results and the relationships between performance and workload.

The results of the study conducted by Hill et al. (1988) showed that the NASA-TLX has the highest validity. A principal component analysis revealed a single component that explained 79.6% of the total variance. A second analysis was conducted, Jackknife principal component analysis, which revealed an ordering of the mean factor loading, in which the NASA-TLX (.935) and OW (.927) were significantly greater than the MCH (.862) and SWAT (.860). Hill et al. (1988) compared the results of this experiment with the results of an earlier test of a remotely piloted vehicle. There was a significant interaction of the four workload scales. From the comparison of the two tests, the researchers stated that the NASA-TLX and MCH had the highest and lowest loadings. There was also a significant multiple correlation between performance and the NASA-TLX, with $r = 0.66$. They noticed that a decrease in performance was due to an increase in workload.

Workload and Command and Control Tasks

It has already been established that a decrease in communication will cause an increase in workload (Whitaker, Peters, and Garinther, 1990; Whitaker, 1991; Whitaker and Peters, 1993). When performing C^2 tasks, workload may be affected by communication, temperature, noise, location of the enemy, urgency of the response, or a combination of all these factors. These factors may cause workload to be variable and unpredictable over time. This problem is confounded even more when the operator has more than one task to perform. Stark, Scerbo, Freeman, and Mikulka (2000) stated that an operator's mental capacity may be exceeded in environments where multiple task demands are placed on a individual. Additionally, when an operator is experiencing high workload, the different tasks will be prioritized to give the most important task more attention (Taylor, Charlton, and Canham, 1996), thus causing the other tasks to suffer. Therefore, when C^2 tasks are performed in an environment where communication is difficult, it will cause a decrease in performance and an increase in workload.

Summary

As reported by Hendy, Hamilton, and Landry (1993) subjective measures have emerged as the primary source for measuring workload. With these measures, the researcher is provided with information on the relationship between a task and the operator's performance on a task. The workload method chosen to assess this relationship may reveal that performance will suffer if the workload is high, or too low. Objective methods of measuring mental workload do not require the participant's opinion, since the data are collected while performing the task. On the other hand, subjective methods rely heavily on the participant's opinion to measure the workload imposed by a task. Although, several experiments conducted on speech intelligibility in tanks (Whitaker, Peters, and Garinther, 1990; Whitaker, 1991; Whitaker and Peters, 1993) have used the SWAT technique to determine how speech intelligibility affects performance, the SWAT technique does not glean enough detailed information about the task from the participant. For this reason, a more sensitive measure, such as the NASA-TLX, probably should have been used to tap into other dimensions that are not addressed in SWAT. Lastly, it is believed that using two measures of mental workload would reduce the chance of important information going unnoticed. Therefore, in addition to the NASA-TLX, the MCH is also a good measure since it places an emphasis on task with a communication component.

THEORETICAL BASIS FOR COGNITION

When performing C^2 tasks, a soldier must sense the information in the environment, make a decision based on the information, and execute a response. Proctor and Zandt (1994) stated that humans process information in three stages: the perceptual, cognitive, and action stages. An illustration of each one of these stages is shown in Figure 7. In the perceptual stage, the objects are identified through a person's sensory organs. Once the stimulus has been identified, the cognitive stage determines the most feasible response or action to be executed. After the perception and cognitive stages, a response is selected and action is taken to execute the response.

The most important part of the human information processing stage that will be discussed in this dissertation is cognition. Cognition involves the ability of the brain to transform, reduce, elaborate, store, recover, and use information that is important to performance (Bailey, 1996). It can be affected by different stimuli, both externally and internally. An example of an internal stimulus would be high-blood pressure or a headache. An example of an external stimulus would be noise, such as that found in a battlefield. More specifically, the noise from a tank, which has been reported to be 114 dBA, may affect cognition. In turn, this may cause performance to decrease when performing C^2 tasks.

Gomes, Pimenta, and Branco (1999) conducted a study to determine if prolonged exposure to low frequency noise had an effect on cognition. The sound pressure level of the noise was greater than 90 dB (linear), with the low frequency range less than 500 Hz. The study was conducted on 70 male engine technicians who worked at Oficinas Gerais

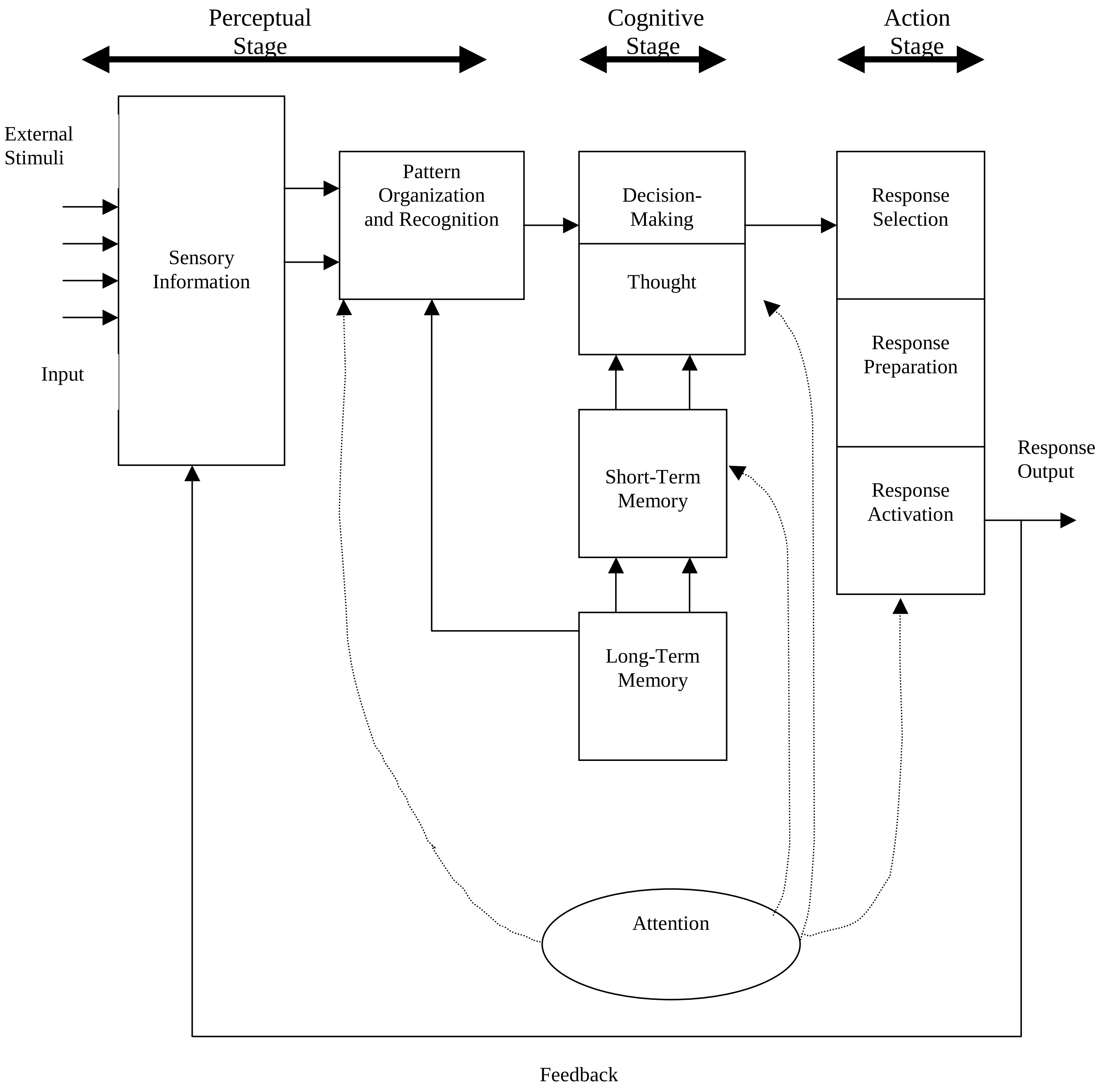


Figure 7. An elaborated model of human information processing (from Kantowitz, 2000).

de Material Aeronautic in Portugal. The experiment consisted of two groups of subjects, the exposed group and the control group. The exposed group was presented with noise levels between 90 and 110 dB (linear), and the control group was not exposed to the noise. Both groups performed engine run-up procedures that were identical to their regular job routines.

Several tests were run on these subjects after completion of the task. The experimenters administered the Wechsler Memory Scale, Toulouse-Pieron Test, and the P300 ERP to determine if the low frequency noise had an effect on cognition. The results from the Toulouse-Pieron test revealed that there was not a significant difference between the two groups. Significance was found when the Wechsler Memory Scale data was analyzed (shown in Table 3). As can be seen from Table 3, there was a significant difference in “immediate verbal memory” and the “memory quotient.” Additionally, when the results of the P300 ERP were analyzed, they revealed significant differences in "Psychic Control," "Digit Span," "Visual Memory" and "Word Association." Gomes, Pimenta, and Branco (1999) stated that the "Visual Memory" test revealed that 14.6% of the exposed group showed disturbances when performing the P300 ERP.

Again, since C^2 tasks are performed by using complex systems, and with multiple people transferring information among one another, cognition may be affected by the noise. The reason why these systems are complex is that they consist of sophisticated technology that process and tune information to assist the commander in achieving the mission (Essens, Post, and Rasker, 2000). The noise (from machinery, guns, vehicles, helicopters, etc.) in a military environment, coupled with performing multiple tasks with and without the technology, may cause cognition to deteriorate.

Table 3. Data from the Wechsler Memory Scale (from Gomes, Pimenta, and Branco, 1999).

	Exposed Group (n=40)		Control Group (n=30)		Exposed vs. Control	
	Average	SD	Average	SD	F	Deg. Signif.
Personal and General Information	6.00	0.00	6.00	0.00	-	-
Immediate Orientation	5.00	0.00	5.00	0.00	-	-
Psychic Control	7.41	1.55	7.86	1.43	1.5074	0.2238
Immediate Verbal Memory	10.00	2.23	11.31	2.33	5.6468	0.0203*
Digit Span	9.80	1.93	10.52	1.94	2.3105	0.1331
Visual Memory	10.20	2.52	11.90	1.86	1.6184	0.2076
Word Association	16.07	3.28	17.84	4.48	3.6589	0.0600
Memory Quotient	101.78	8.66	108.31	8.18	10.1102	0.0022*

* denotes significance at $p < 0.05$.

The decision-making construct associated with command centers can also be correlated to C^2 tasks, since issuing commands is the first phase of C^2 tasks. More simply stated, before a task can be executed, a command must be issued to inform the person of the task and allow the commander to maintain control based on the situation.

Noyes, Barber, and Leggatt (2000) stated that some tasks performed in armored vehicles require the use of a keyboard. The keyboard allows the commander in an armored vehicle to type information concerning a mission. Next, this information may be transferred to subordinates or someone in a higher level of the military. This process is continuous, with everyone staying abreast of the events in the battlefield. Therefore, certain constructs of cognition are important for performing C^2 tasks. These constructs are memory retrieval, planning, situation assessment, problem solving, decision-making, and attention to detail.

Planning involves making a decision while considering several other factors that must be taken into account in selecting an optimal course of action. During C^2 operations, the commander has to conceptually plan how to coordinate the different activities in the battlefield. Planning can be done prior to the mission or during the mission. Regardless of when it is performed, planning may enhance performance.

O'Hare (1997) stated that situation awareness is fortuitous in an environment where a human interacts with a dynamic and complex system. Situation awareness relates to C^2 tasks because before a person can issue C^2 tasks to soldiers, he or she must have some knowledge of the situation. This is very fundamental to situations that are unpredictable, such as a battlefield environment. In the battlefield, the commander is presented with multiple tasks that must be prioritized, monitored, and managed to successfully complete the mission. This leads to the commanders' need to assess the situations or events that are occurring in their environment.

Bailey (1996) stated that problem-solving is one of the most complex forms of human information processing. Problem-solving involves a person using his or her existing knowledge to derive new solutions to problems that arise. Additionally, this person may rely on his or her short-term memory to recall events. During C^2 tasks, a commander may encounter a situation in which he or she may not be able to comprehend information presented verbally or visually. When such situations arise, the commander must rely on his or her own knowledge to develop a solution and continue to C^2 his or her subordinates. When thinking of a solution to the problem, the commander must carefully consider all the alternatives.

Attention to detail entails carefully looking at aspects of a situation that would be overlooked by a novice. For instance, it has already been established that C^2 tasks may be conducted through visual displays. The visual displays may show overlays of maps, emails, or other type of textual information. When reporting from maps or emails, the richness of all the information is important when making critical decisions because it may determine life or death.

Decision-making involves weighing all the possible solutions to a problem and deciding on one solution that will result in the most favorable outcome. In C^2 tasks, the desired outcome is to listen, monitor, and transmit information as quickly as possible to effectively accomplish the mission. However, factors in a person’s environment may influence the decision-making process. For instance, when making a decision in a static environment, the number of alternative solutions will be greater than if the person was in a dynamic environment in which he or she has a limited amount of time to arrive at a solution.

The constructs discussed in the preceding section are all used in performing C^2 tasks. The number of times each construct occurs is unknown, due to the unpredictable in a military environment, which will cause different types of C^2 tasks to be performed. However, a test battery has been developed and validated to assess how performance is affected by such cognitively demanding tasks.

COMPLEX COGNITIVE ASSESSMENT BATTERY

In this research, the Complex Cognitive Assessment Battery (CCAB) was used to quantify performance. It is believed that the tests within the CCAB pose the same cognitive constraints as C^2 tasks. The CCAB was developed by the U.S. Army Research Institute, the Army Medical Research and Development Command and the Tri-Service Joint Working Group for Drug Dependent Degradation on Military Performance (Expanded Complex Cognitive Assessment Battery, 1988). The purpose of this battery is to evaluate performance on tasks that are classified as being high-level and requiring complex cognitive skills (Kane and Kay, 1992). Additionally, this battery was developed to measure the complex cognitive abilities that are required to perform Army C^2 and Operational tasks (Expanded Complex Cognitive Assessment Battery, 1988).

The tests in the CCAB were selected based on fourteen cognitive constructs, with each construct being categorized as four different types of information processing: (1) responding to data, (2) going beyond data, (3) taking action based on data, and (4) creating data or solutions. A taxonomy representing the four types of information processing is shown in Figure 8. The fourteen cognitive constructs are presented below, along with examples of how they may relate to C^2 tasks:

(1) *Attention to detail*: Attention to detail is concerned with the detection of details in data that provide some type of insight to events in context. An

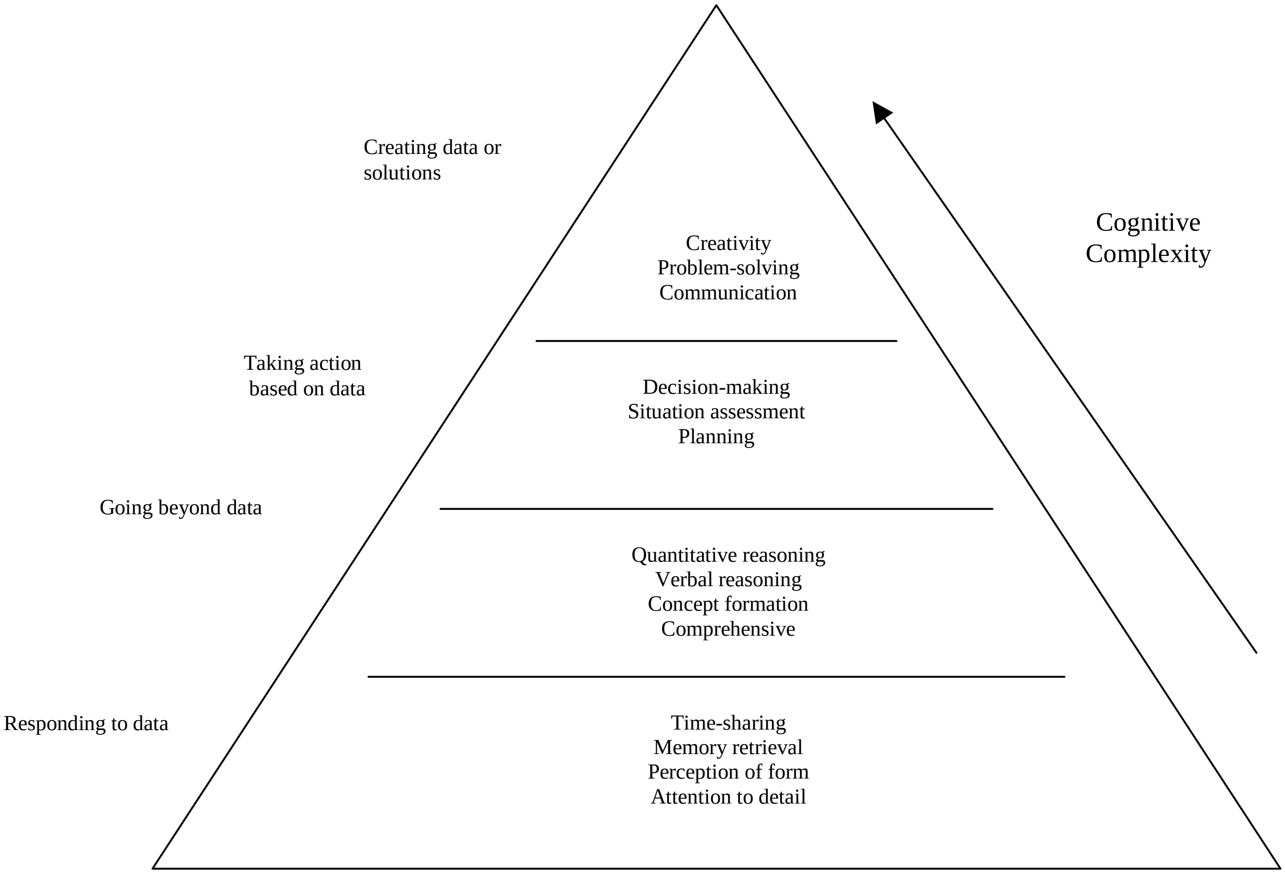


Figure 8. CCAB taxonomy (from Expanded Complex Cognitive Assessment Battery, 1988).

example would be a commander monitoring an electronic map to locate enemy forces.

(2) *Perception of form*: Perception of form deals with noticing a pattern in a field of data that is unusual. For example, a commander scanning an electronic map may notice paths that adversaries are taking, and update his soldiers on this pattern.

(3) *Memory retrieval*: Memory retrieval involves a person searching through his or her long-term memory to locate the knowledge required to perform a task. For example, in order to command soldiers effectively, a commander will have to remember events, as well as communications, that are occurring in the environment.

(4) *Time-sharing*: Time-sharing involves the execution of two or more cognitive tasks that require a person to shift resources among tasks. An example of a time-sharing task would be a commander receiving information via the auditory channel while trying to look at information displayed on an electronic map.

(5) *Comprehension*: Comprehension involves understanding written and verbal information. For example, a commander may need to transform verbal communication into written communication when completing and sending a status report to someone at a central location.

(6) *Concept formation*: Concept formation involves creating an idea that can be applied to other situations. An example of a situation in which concept formation would be used is when a commander has to decide how to

communicate with soldiers in another area, when the events occurring in his or her environment are known.

(7) *Verbal reasoning*: Verbal reasoning involves the ability to apply basic rules to verbal problems for determining a solution. For example, given a verbal command, the soldier must perform the action based on a set of principles that was taught in training.

(8) *Quantitative analysis*: Quantitative analysis involves using general mathematical methods to find solutions. An example of a situation in which quantitative analysis would be used is when a tank company is reporting the distance they are from the enemy or the rate at which the enemy is moving.

(9) *Planning*: Planning involves using cognitive skills to plan and prepare for military operations. For example, before the military engage in war, they devise a plan, outlining the strengths and weaknesses of the enemy, to schedule the activities for a mission.

(10) *Situational assessment*: Situational assessment involves the ability to know, describe, and stay abreast of the events that are occurring in the environment. For example, a commander must stay aware of all the events occurring in the battlefield, to assist in delegating tasks to other individuals.

(11) *Decision-making*: Decision-making involves scanning the possible solutions to a problem or situation and selecting the best solution. For example, a commander may have to choose one of several tactical decisions and weigh the outcome of each.

(12) *Communication*: Communication involves expressing ideas and knowledge to others. During C^2 operations, the commander must have a clear line of communication with other soldiers and commanders to stay abreast of events occurring in the environment.

(13) *Problem-solving*: Problem-solving requires the person to use his/her knowledge and experience to apply reasoning to solve a problem. For example, a soldier may have to rely on his/her experience from previous combat operations to devise a solution to defeat the enemy.

(14) *Creativity*: Creativity involves a person creating a new idea or solution to a problem. For example, a soldier only having basic training experience may be drafted to fight in a war. This soldier will have to use his/her creative wisdom to eliminate or decrease his or her chances of being captured by enemy forces.

The actual tests chosen to measure these constructs are: (1) Tower Puzzle, (2) Following Directions, (3) Word Anagrams, (4) Grammatical/Logical Reasoning, (5) Mark Numbers, (6) Numbers and Words, (7) Information Purchase, (8) Route Planning, and (9) Missing Items. However, only three tests will be used for this dissertation.

Cognitive Tests

Of the nine computerized complex cognitive tests, three have been chosen for this research: *Tower Puzzle, Numbers and Words*, and *Logical Relations*. These specific tests were chosen because they measure a particular set of cognitive constructs more effectively than the others tests. A synopsis of the cognitive constructs measured by each test and the level of association with each construct is shown in Table 4. The most

Table 4. Level of Association Between CCAB Constructs and Tests (adapted from Expanded Complex Cognitive Assessment Battery, 1988).

[1 = Low, 2 = Medium, 3 = High]

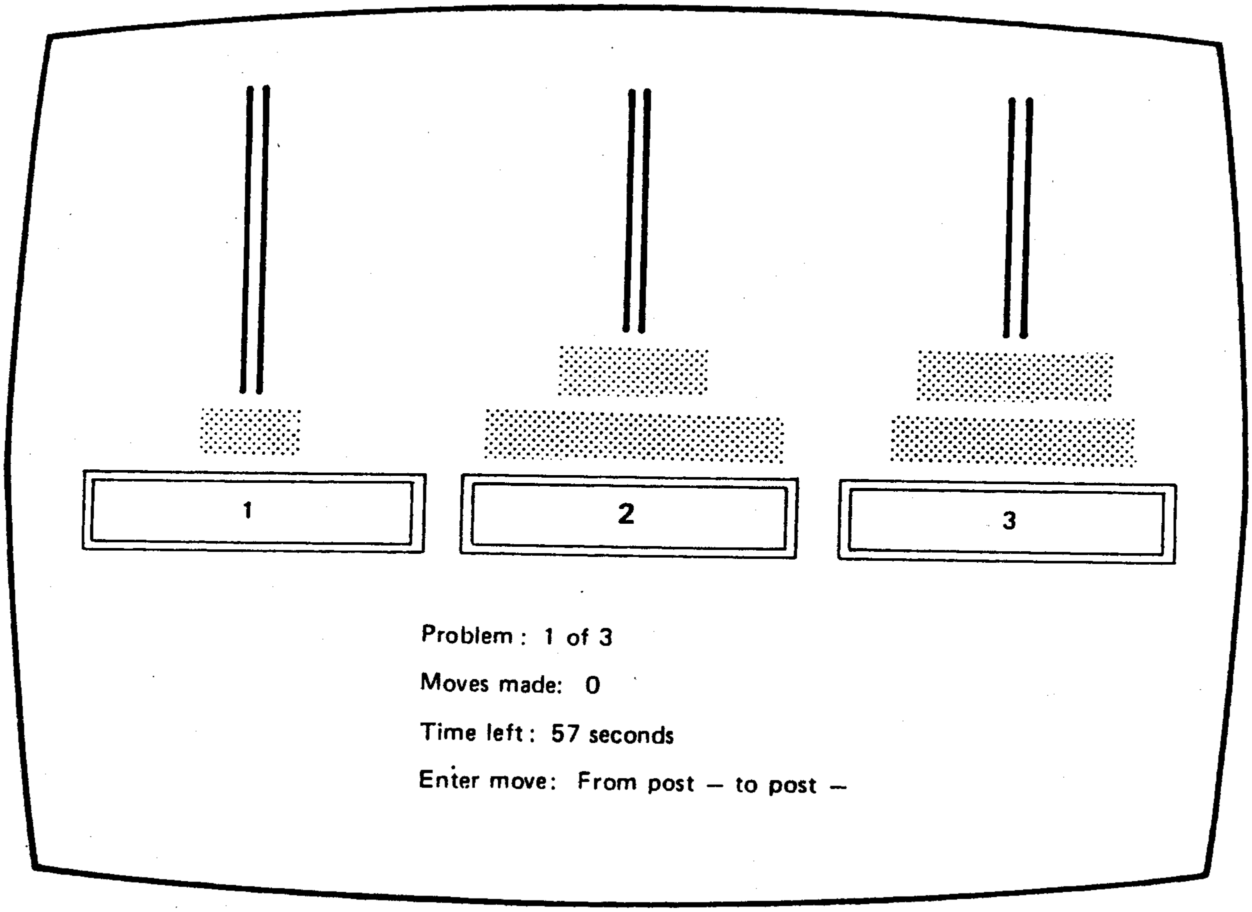
Cognitive Complexity Categories	Cognitive Construct Measured	CCAB Tests*		
		TP	NW	LR
Responding to Data	Attention to Detail		**2**	1
	Perception of Form	1	**3**	
	Memory Retrieval		**2**	1
	Time-Sharing		**3**	
Going beyond Data	Comprehension			**3**
	Concept Formation	1	2	**1**
	Verbal Reasoning			**3**
	Quantitative Reasoning	1		**2**
Taking Action Based on Data	Planning	**3**		
	Situation Assessment	**3**		
	Decision-Making	**2**	2	
Creating Data or Solutions	Communication	**1**		
	Problem-Solving	**3**		1
	Creativity	**1**		

*Codes used in the table for tests are as follows: *Tower Puzzle* (TP), *Numbers and Words* (NW), and *Logical Relations* (LR).

dominant set of constructs that each test measures is emphasized with bold letters in the table.

Tower Puzzle. The *Tower Puzzle* test (shown in Figure 9) is intended to assess the cognitive functions of planning, situation assessment, problem-solving, decision-making, communication, and creativity (Analytical Assessment Corporation, 1988). In the test, the participant is presented with five blocks of different sizes distributed across three stacks. The participant is required to arrange all five blocks in descending order (largest at the bottom, smallest at the top) on the middle stack. Only one block at a time can be moved and a larger block cannot be placed on a smaller block. Blocks are moved using the numeric keypad on a computer keyboard. (For example, to move the top block from stack 3 to stack 1, the participant would press "3" followed by "1.")

Numbers and Words. The *Numbers and Words* (shown in Figure 10) test is intended to evaluate a person's ability to perform two tasks simultaneously. This test assesses the cognitive functions of attention to detail, perception of form, memory retrieval, and time-sharing. In the *Numbers* portion of the test, the computer sequentially presents a series of numbers on the computer screen. The participant's task is to recall the previous number presented by pressing the corresponding key on the keyboard. In the *Words* portion of the test, a white box appears above a list of candidate words (each candidate word is associated with a specific letter, e.g., A – CAT, B – HAT, C – BAT, etc.). The white box gradually fades to become a three-letter word. The participant's task is to press the letter corresponding to the word that appears as the white box fades. The two parts of this test are presented to the participant simultaneously.



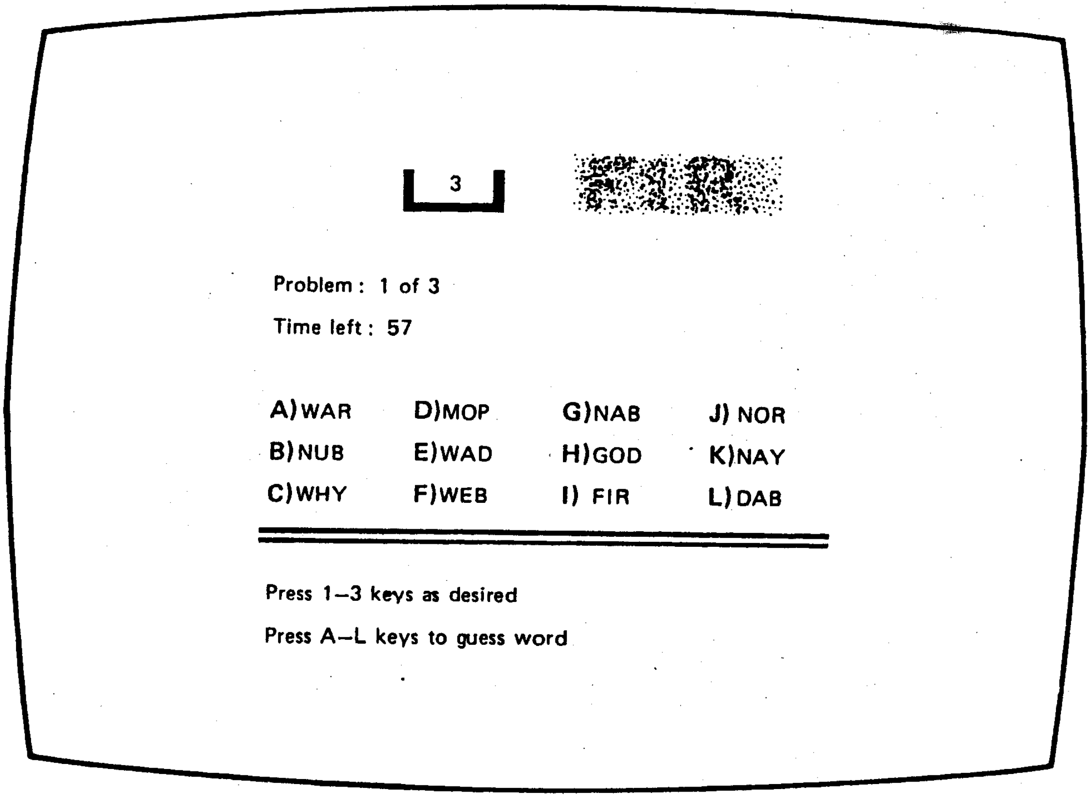
Problem : 1 of 3

Moves made: 0

Time left: 57 seconds

Enter move: From post – to post –

Figure 9. Sample illustration of the *Tower Puzzle* screen.



Problem : 1 of 3

Time left : 57

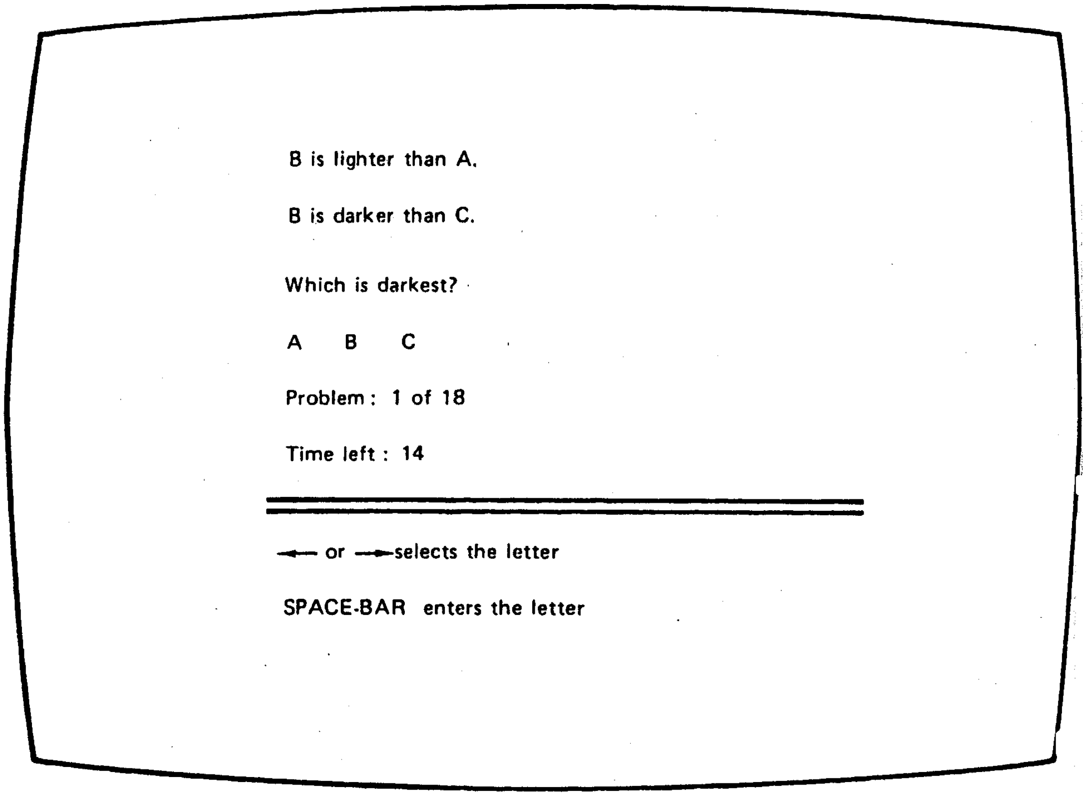
A) WAR
B) NUB
C) WHY
D) MOP
E) WAD
F) WEB
G) NAB
H) GOD
I) FIR
J) NOR
K) NAY
L) DAB

Press 1–3 keys as desired

Press A–L keys to guess word

Figure 10. Sample illustration of the *Numbers and Words* test screen.

Logical Relations. The *Logical Relations* (shown in Figure 11) test is intended to assess comprehension, verbal reasoning, concept formation, and quantitative reasoning by evaluating an individual's ability to judge the logical relationships among multiple items. In the test, the participant is presented with several statements describing the relationship among three letters (e.g., “A is worse than B, C is better than B, B is worse than A") and asked a question regarding the relationship (e.g., “Which is best?”). The participant's task is to answer the question by pressing the appropriate key on the keyboard.



B is lighter than A.

B is darker than C.

Which is darkest?

A B C

Problem : 1 of 18

Time left : 14

← or → selects the letter

SPACE-BAR enters the letter

Figure 11. Sample illustration of the *Logical Relations* test screen.

RESEARCH OBJECTIVE AND HYPOTHESES

From the preceding discussion, it has been shown that noise has an effect on speech intelligibility, performance and workload. To recapitulate these issues, Figure 12 shows the relationships among these factors. Furthermore, several studies conducted by Whitaker (1991), Whitaker and Peters (1993, 1990), Peters and Garinther (1990) and Whitaker, Peters and Garinther (1990, 1989) have demonstrated the importance of speech information and how performance is impacted when speech intelligibility decreases, thus causing an increase in workload. These same results may be generalizable to C^2 tasks, and it is to this question that this dissertation was directed.

The microphone/communication system in the VIS headset was one of the major points of concern in this research, because it is the component of the VIS headset responsible for transmitting speech. A major reason for this concern is that the sound pressure level under the earcup, in 114 dBA tank noise, when the communication system is activated and someone is speaking, is reported to be 94 dBA. The sound pressure level is reported to be 90 dBA when the communication system is activated and no one is speaking. Lastly, when the ANR component of the VIS headset is activated in 114 dBA noise, the sound pressure level (SPL) under the earcup is 83 dBA (M. Vaudrey, personal communication, December 13, 2000). Obviously, a problem exists when the communication system is turned on and someone is speaking.

In accordance with OSHA regulation, when employees are exposed to an 8-hour time–weighted average (TWA) that meet or exceed 85 dBA, the employer must make hearing protection devices (HPDs) available to all employees. Also, when employees are

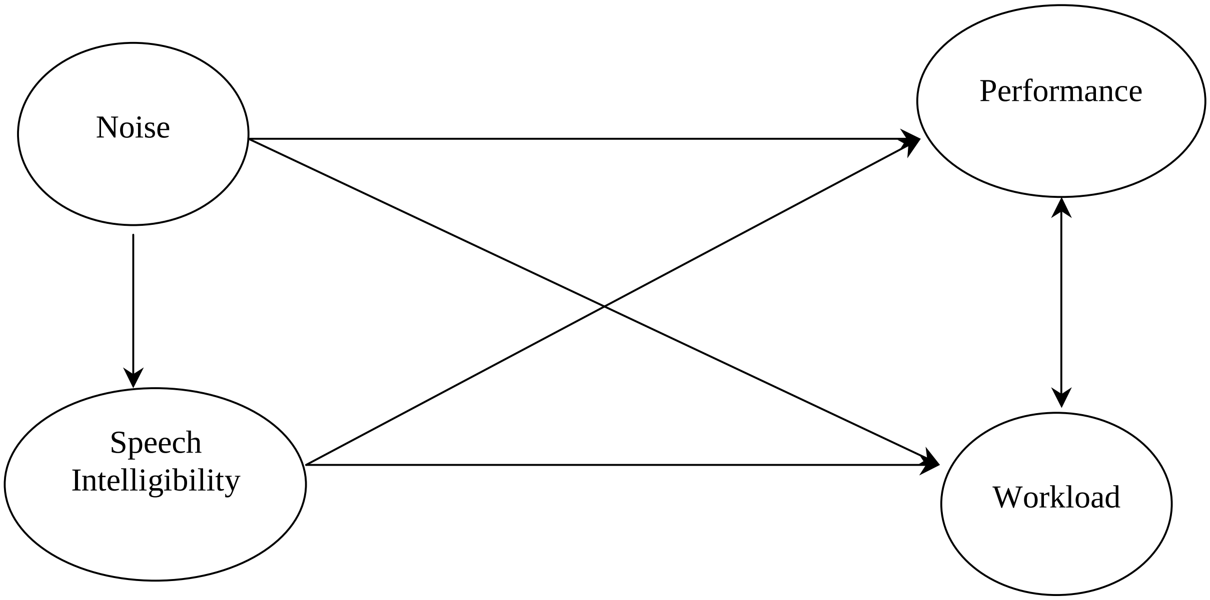


Figure 12. Block diagram showing the relationships among speech intelligibility, noise, performance, and workload.

exposed to an 8-hour TWA of 90 dBA, it is mandatory that employees wear HPDs (OSHA, 1990). Additionally, according to OSHA, workers are only to be exposed to 85 dBA and 90 dBA for 16 and 8 hours a day, respectively. For that reason, it is important to reduce the sound pressure level that reaches the ear in order to increase the exposure time.

Therefore, the objectives of this dissertation were to: 1) determine which of two communication microphones provide the greatest noise reduction at the ear, 2) determine if a reduction in noise at the ear will cause an improvement in speech intelligibility, and 3) determine if a reduction in noise reaching the ear results in any improvement in cognitive performance or a reduction in subjective workload. The two communication microphones used in the study were the Gentex Model 1453 microphone currently used by the US Army in their VIS headset, and a prototype communication microphone developed by ATI. The Bradley Fighting Vehicle noise sample used in the experiment was provided to the Auditory Systems Laboratory by ATI.

The hypotheses examined during this research study were as follows:

- In 114 dBA noise, the prototype microphone/communication system developed by ATI reduces the noise transmitted to the listener more than the current Gentex microphone.
- The reduction in noise reaching the ear results in an improvement in speech intelligibility.
- Reduced noise and improved speech intelligibility results in an improvement in cognitive performance.
- Subjective workload is lower when subjects perform the tasks using the ATI prototype system.

EXPERIMENTAL DESIGN AND METHOD

Experimental Design

The overall experiment will be discussed as two separate analyses because the data consisted of two types: objective and subjective. The experimental design employed for the first analysis (objective data) consisted of a three factor repeated measure design (two communication microphones used at two speech levels for the Cognitive Assessment Tests). A schematic representation of the first experimental design is shown in Figure 13. The objective data consisted of scores from the Modified Rhyme Test (MRT; ANSI, 1989) and the Complex Cognitive Assessment Battery (CCAB). A two-factor repeated measure design was employed for the second analysis (subjective data), in which speech level and communication microphone were the two independent variables. Subjective data consisted of scores from the Modified Cooper-Harper (MCH) and National Aeronautical Space Administration Task Load Index (NASA-TLX). These two subjective workload instruments were used to measure perceived mental workload as a result of using the two communication microphones at each speech level. This experimental design is shown in Figure 14.

Due to presbycusis, hearing loss due to aging, gender was not considered to be a factor in the experiment. This effect first is first noticed in the high frequency range, causing hearing threshold levels to become elevated. Also, since consonants, which are the most important component of a word that lends to intelligibility, are found in the high frequency range, intelligibility is profoundly affected. However, males tend to experience hearing loss before females due to the nature of tasks (e.g. skeet shooting, racing cars, etc.) in which males participant, which for the most part are noisy.

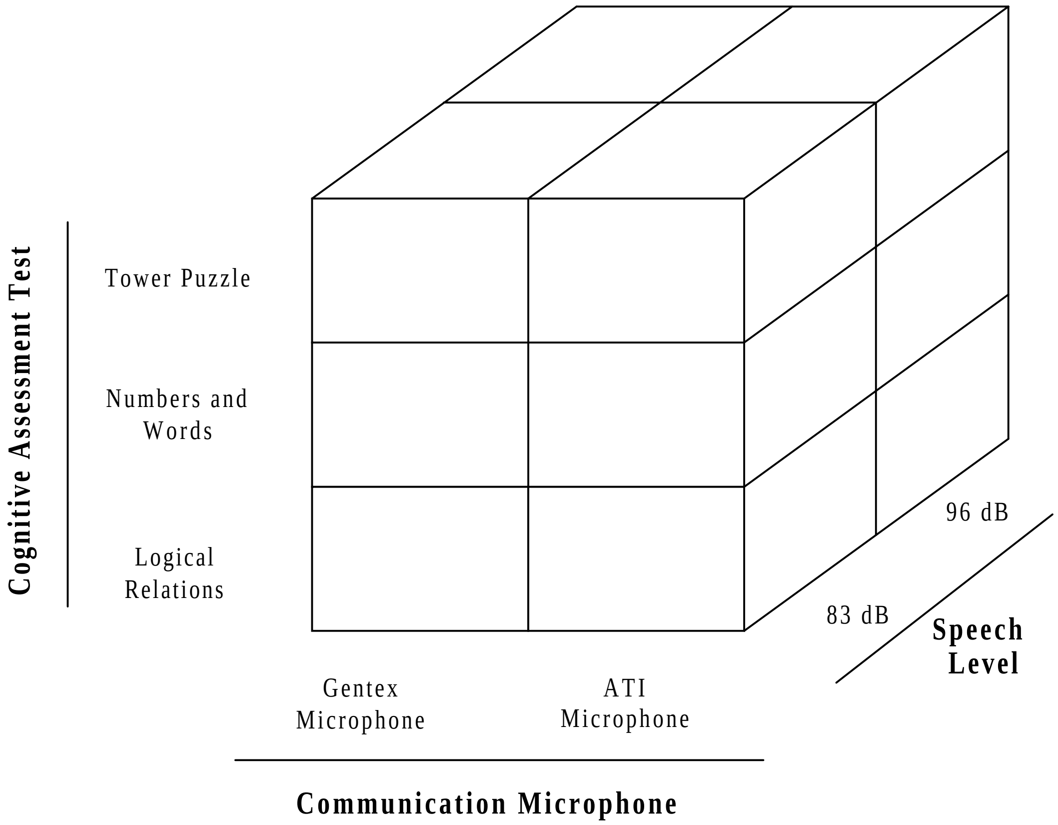


Figure 13. Schematic diagram of the experimental design used for yielding the objective data from the Complex Cognitive Assessment Battery and the Modified Rhyme Test.

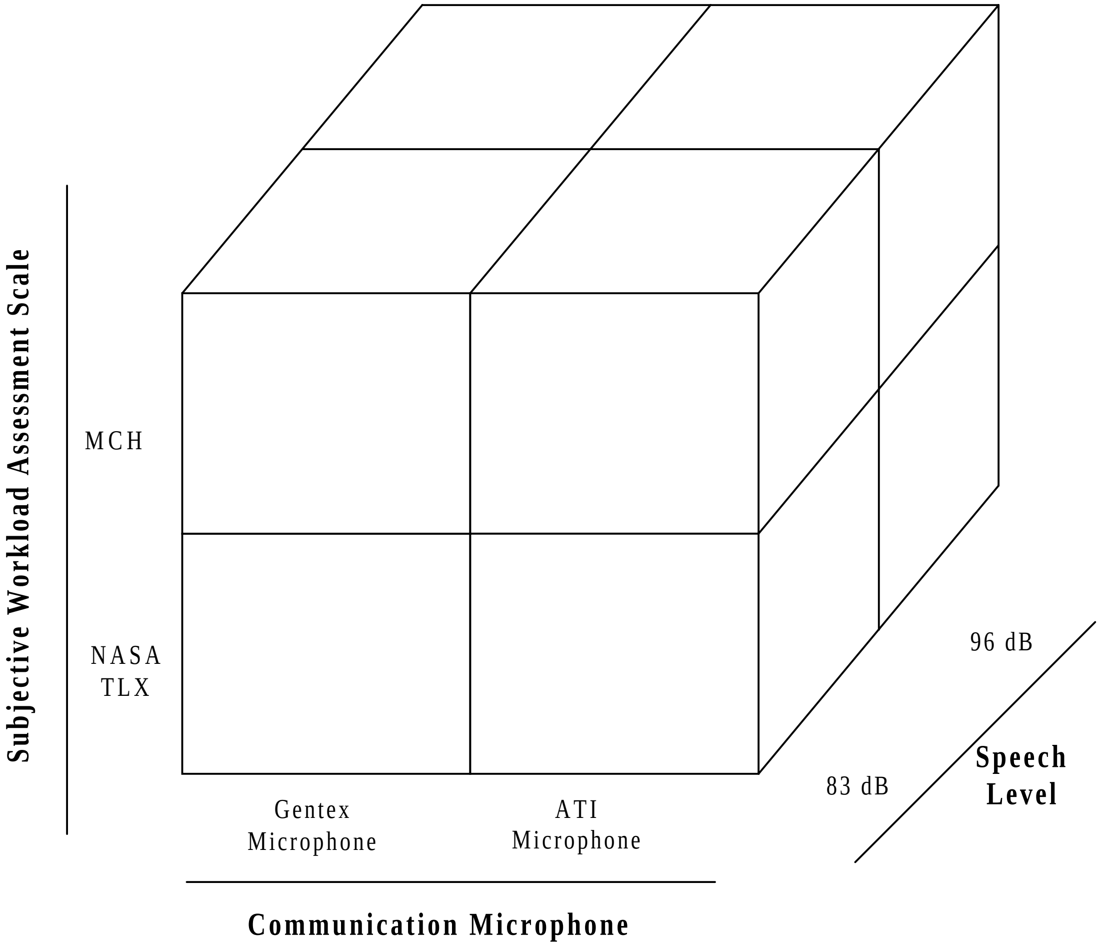


Figure 14. Schematic diagram of the experimental design used for yielding the subjective data from the two Subjective Workload Assessment Scales.

Therefore, unless severe hearing loss occurred in a person's childhood, it was expected that young participants (in this case, less than 30 years of age) with hearing threshold levels in accordance with ANSI S3.2 (1989), would be appropriate for this dissertation. Besides, audiometric testing was conducted to assure this speculation was true. A study was conducted by Ellis, Reynolds, Fucci, and Benjamine (1996), in which thirty participants (aged 19-24) listened to audiotaped male and female speech samples, in an effort to determine whether gender had an effect on their judgment of speech intelligibility. The researchers used magnitude-estimation scaling responses to obtain responses from the participants. The results revealed that there was not a significant difference between the male and female magnitude-estimation responses, thus lending credence to the previous notion that a difference would not be expected in young participants.

Independent variables. In the first analysis, in which objective performance was measured, there were three independent variables: *Communication Microphone*, *Speech Level*, and *Cognitive Assessment Test*. The first factor, *Communication Microphone,* had two levels: the Gentex communication microphone and the prototype communication microphone developed by ATI. The second factor, *Speech Level*, also had two levels: 83 dB and 96 dB. These two speech levels had been shown in pilot tests to produce speech intelligibility levels of approximately 50% and 90%, respectively. The third factor, *Cognitive Assessment Test*, had three levels representing the three individual tests from the CCAB (Expanded Complex Cognitive Assessment Battery, 1988). These tests were the *Tower Puzzle*, *Numbers and Words,* and *Logical Relations*. Each of these tests is designed to measure a specific set of cognitive functions.

The second analysis was used to assess subjective mental workload using two different workload scales: NASA-TLX and MCH. This design included two independent variables: *Speech Level and Communication Microphone*. Both of these variables were the same as those described for the previous design.

These two workload assessment scales were selected due to their ability to provide information with respect to communication and complex cognitive performance. Wickens (1992) stated that the NASA-TLX conveys more information, since it contains a greater number of scales. For instance, if a researcher was interested in determining the amount of physical demand that a lifting task places on an operator, the NASA-TLX takes this construct into account by having a scale for which a person rates his or her physical demand. The MCH is appropriate for tasks with an emphasis on communication (Casali and Wierwille, 1983). The results from the MCH have also been shown to be correlated with two other workload scales, Subjective Workload Assessment Technique and Overall Workload (Charlton, 1996; Lysaght, Hill, Dick, Plamondon, Linton, Wierwille, Zaklad, Bittner, and Wherry, 1989). After completing the entire experiment, participants used these two subjective workload instruments to rate their perceived mental workload.

Dependent variables. For the first experimental design, from which objective performance was measured, the dependent variables were the MRT and the scores for each of the specific CCAB tests. The MRT consists of 300 monosyllabic English words grouped into 50 six-word sets (See Appendix A; ANSI, 1989). The six-word sets are arranged according to response ensembles, with each ensemble characterized by one vowel that is the nucleus of every word (ANSI, 1989; House, Williams, Hecker, and

Kryter, 1965). While most of the words are in the consonant-vowel-consonant (CVC) format, others are in the form of CV or VC. One word from each set was randomly selected to generate the six distinct 50-word lists used in the experiment. This test was used to quantify the intelligibility of speech presented aurally in the presence of the Bradley Fighting Vehicle noise.

In addition to the MRT, computer generated scores were obtained as a measure of performance on each CCAB test (*Tower Puzzle, Logical Relations,* and *Number and Words*). This computer-generated score differed for each test because the CCAB used a different scoring scheme and criteria for each test.

For the second experimental design, in which subjective performance was measured, subjective mental workload scores from the NASA-TLX and MCH were obtained. The NASA-TLX quantifies mental workload based on six subscales: mental demand, temporal demand, physical demand, frustration, effort, and performance. The participant selected one of two dimensions (e.g., mental demand versus physical demand) that contributed most to his or her experience. Each of the six dimensions were then rated on a scale of "zero to ten", in which a score of "zero" indicated "no" workload and a score of "ten" indicated "heavy" workload imposed on the participant, to obtain the overall workload score.

The MCH assesses subjective mental workload associated with tasks that require a participant's perceptual, cognitive, and communication skills. The scale is based on a decision tree that combines several factors into a unidimensional ten-point scale. After progressing through the decision tree, the participant selected a more refined description of his or her experience, resulting in a single number rating (ranging from one to ten) describing his or her perceived mental workload (Wierwille & Casali, 1983). The workload scores for the MCH may range from “one to ten”. A workload rating of “one” indicates that the participant’s mental effort is minimal and desired performance is easily attainable”, while a rating of “ten” indicates that the “instructed task cannot be accomplished reliably.”

Participants

Originally, it was envisioned to have sixteen participants, but the experiment concluded at eight due to problems caused by the intense noise level throughout Whittemore Hall and other safety concerns on behalf of the experimenter. Nevertheless, the eight paid volunteers (five males and three females) were from the Virginia Tech/Blacksburg, Virginia community. Participants were required to read a description of the experiment and provide their informed consent in writing prior to participation. All participants were required to be native English speakers, be at least 18 years of age, have no obvious injury or infection of the ears, and possess pure-tone hearing thresholds, as measured on a screening audiogram, no higher than 20 dBHL and no lower than –10 dBHL from 125 Hz to 8000 Hz. Participants were paid at a rate of $8 per hour for their time.

Facilities and Instrumentation

The experimental apparatus (shown in Figure 15) was located in Room 519N, Whittemore Hall on the Virginia Tech campus. An audio CD-based recording of noise inside of a Bradley Fighting Vehicle was used as the noise source for the experiment. The output of a Sony compact disc player (CDP-XE400) was directed through an AudioControl (C-131) one-third octave band equalizer, in conjunction with a BSS

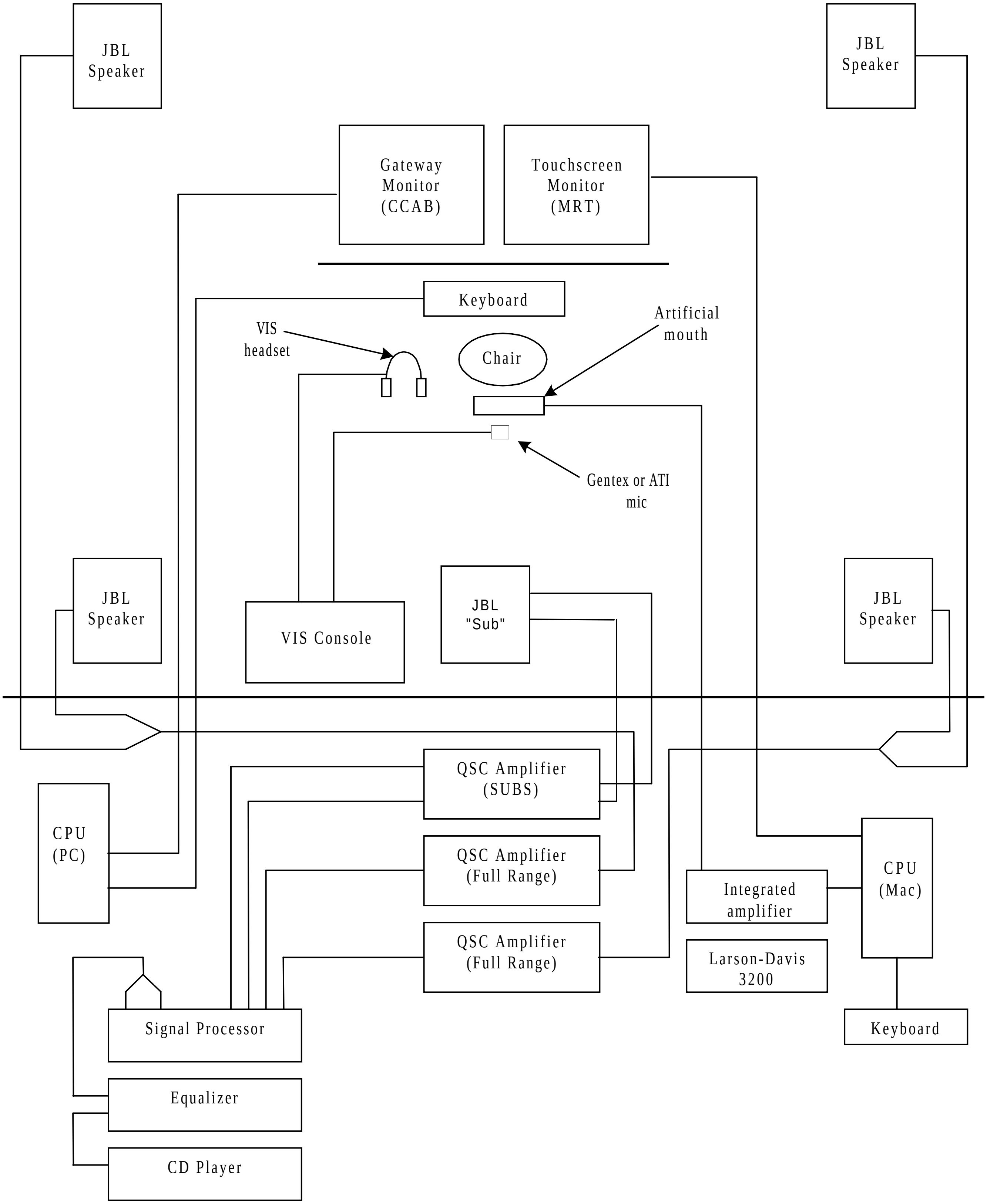


Figure 15. Functional block diagram of the experimental apparatus (not to scale).

Minidrive (FDS-334) signal processor to shape the spectrum of the noise. Three QSC (CX1102) amplifiers were driven by the outputs of the BSS Minidrive to present the noise at a level of 114 dBA through four JBL full range speakers (SP215-6) and a JBL dual subwoofer (SP128-6).

Recording the speech stimuli. The speech stimulus sentences were digitally recorded by the Auditory Systems Laboratory staff using Sound Edit Pro software running on a PowerMac 8500/120. The male talker spoke into an AKG C414-B-ULS microphone connected to a TASCAM Mulitracker 260 mixer (for level control). A Larson-Davis 3200 real-time spectrum analyzer was connected in parallel with the computer so the talker could monitor his speech levels in real-time. Recordings were made in an anechoic room to prevent ambient noise and sound reflections from affecting the recording. Figure 16 illustrates how the equipment was arranged for recording the MRT stimulus sentences.

Each MRT word was recorded within the carrier sentence: “Mark the word_____now.” After all 300 sentences were recorded, the peak A-weighted levels (using a “slow” 1s meter time constant) for each word in each sentence was normalized. To do so, the audio output of the computer was connected to the analog input of the spectrum analyzer (see Figure 17) and the amplification of each word in each sentence was digitally adjusted so that the level of all utterances were equal. A 2-minute segment of speech-weighted noise, generated by a Beltone 2000 audiometer, was also digitally recorded and its unweighted Leq equated to that used to equalize the individual MRT words. [The peak A-weighted level of individual words measured using a slow response is equal to the long-term unweighted Leq of continuous speech (Kryter, 1985).] This

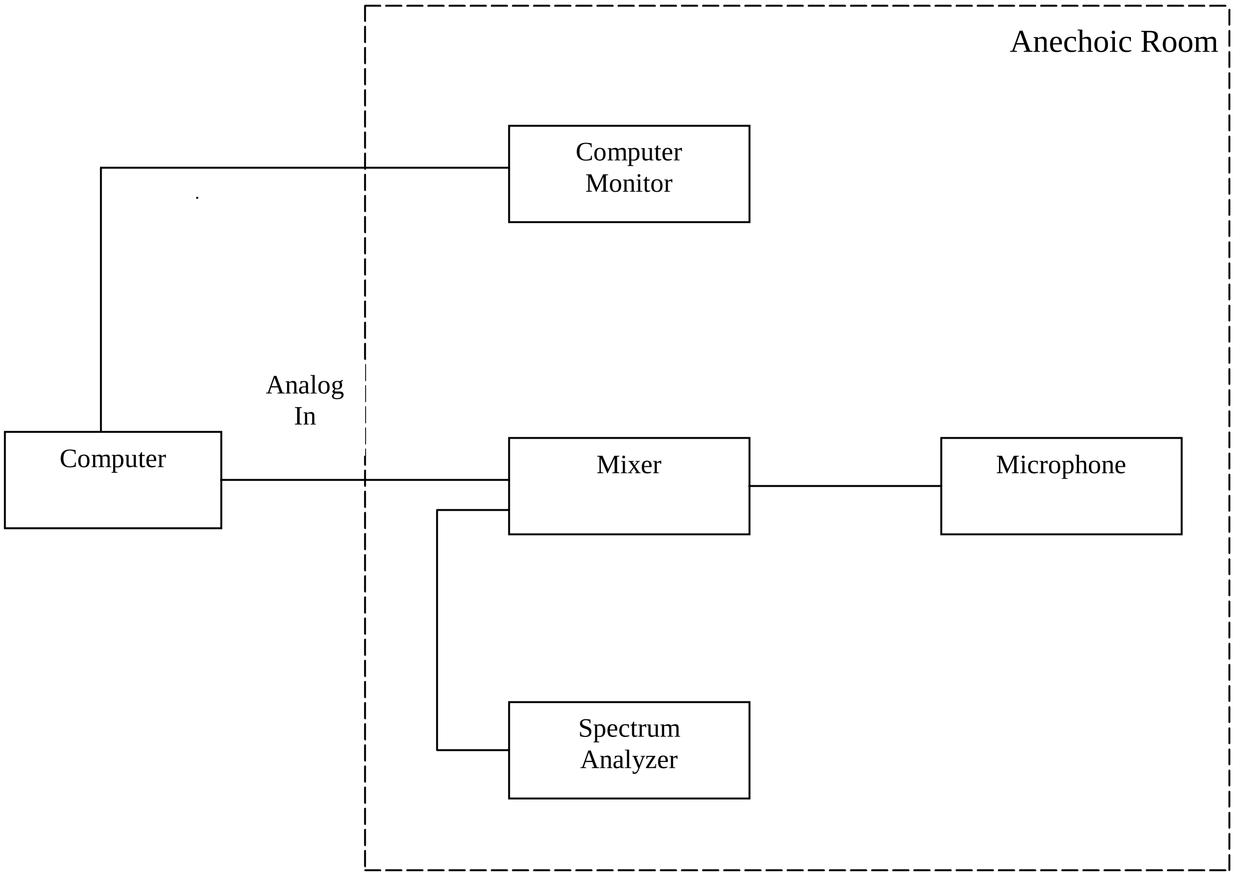


Figure 16. Instrument configuration used to record MRT stimulus sentences.

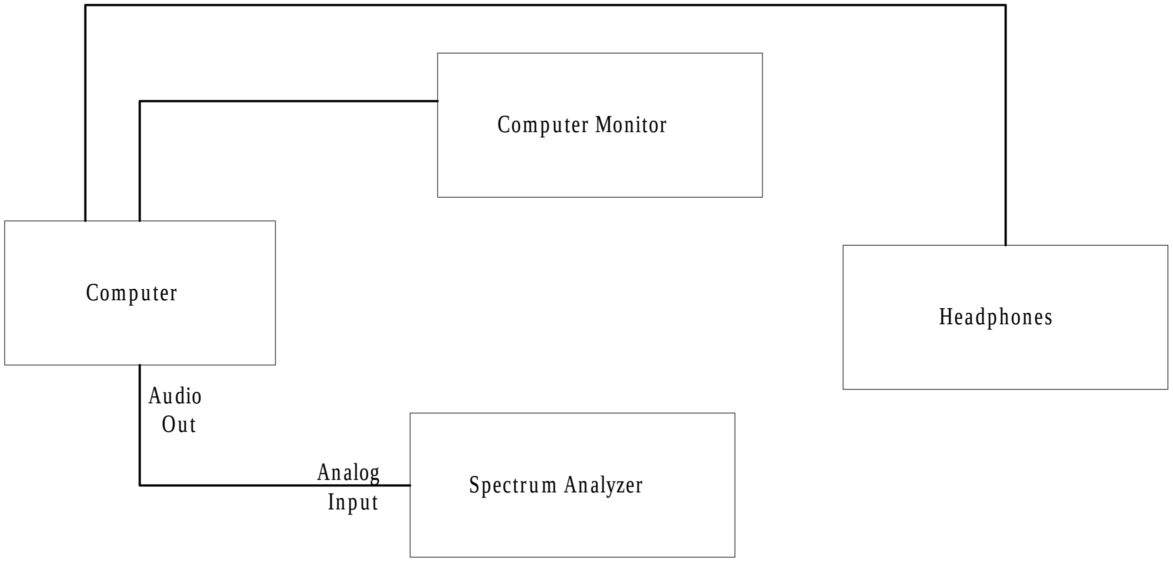


Figure 17. Instrument configuration used to normalize the levels of the recorded MRT stimuli.

relationship and the 2-minute segment of speech-weighted noise allowed the speech levels to be quantified.

The speech was then embedded in a HyperCard Stack, which was ran on a Power Macintosh (8500/120) computer. A Sony receiver (STR-DE135) was used to direct the audio output from the computer to a Brüel and Kjaer (Model 4227) mouth simulator located directly behind the participant's head, while the word choices (the visual component of the MRT) were presented on a Sony Trinitron (Multiscan 15sf) touchscreen monitor. In an effort to maintain consistency with the placement of the communication microphone, the Brüel and Kjaer lip ring (SO 0005) was used to place the communication microphone at the center of the artificial mouth (shown in Figure 18) at a distance of 25 mm. Figure 19 shows the VIS console that controlled the Gentex communication microphone and the VIS headset, and a separate controller that controlled the ATI microphone. The CCAB tests were presented via a Gateway (GP6-300) computer and a Sony Trinitron (Multiscan 100ES) monitor.

Reverberation time. The reverberation time (RT(60)) is the time it takes a sound to decay 60 dB below its initial level. Prior to performing the tests of the acoustics of the test room, it was necessary to adjust the absorptive characteristics of the room to more closely represent the conditions inside of a Bradley Fighting Vehicle. Therefore, the walls of the room were covered with two-inch thick Sonex^{TM} sound-absorbing foam to reduce the reverberation time in the room. Next, the room was excited through four JBL Sound Power loudspeakers and a JBL dual subwoofer by a steady pink noise signal generated by the LD 3200 analyzer. When the noise was on, data collection was initiated; when all of the analyzer's filters were producing data, the noise was terminated

Figure 18. The Gentex communication microphone in front of the artificial mouth.

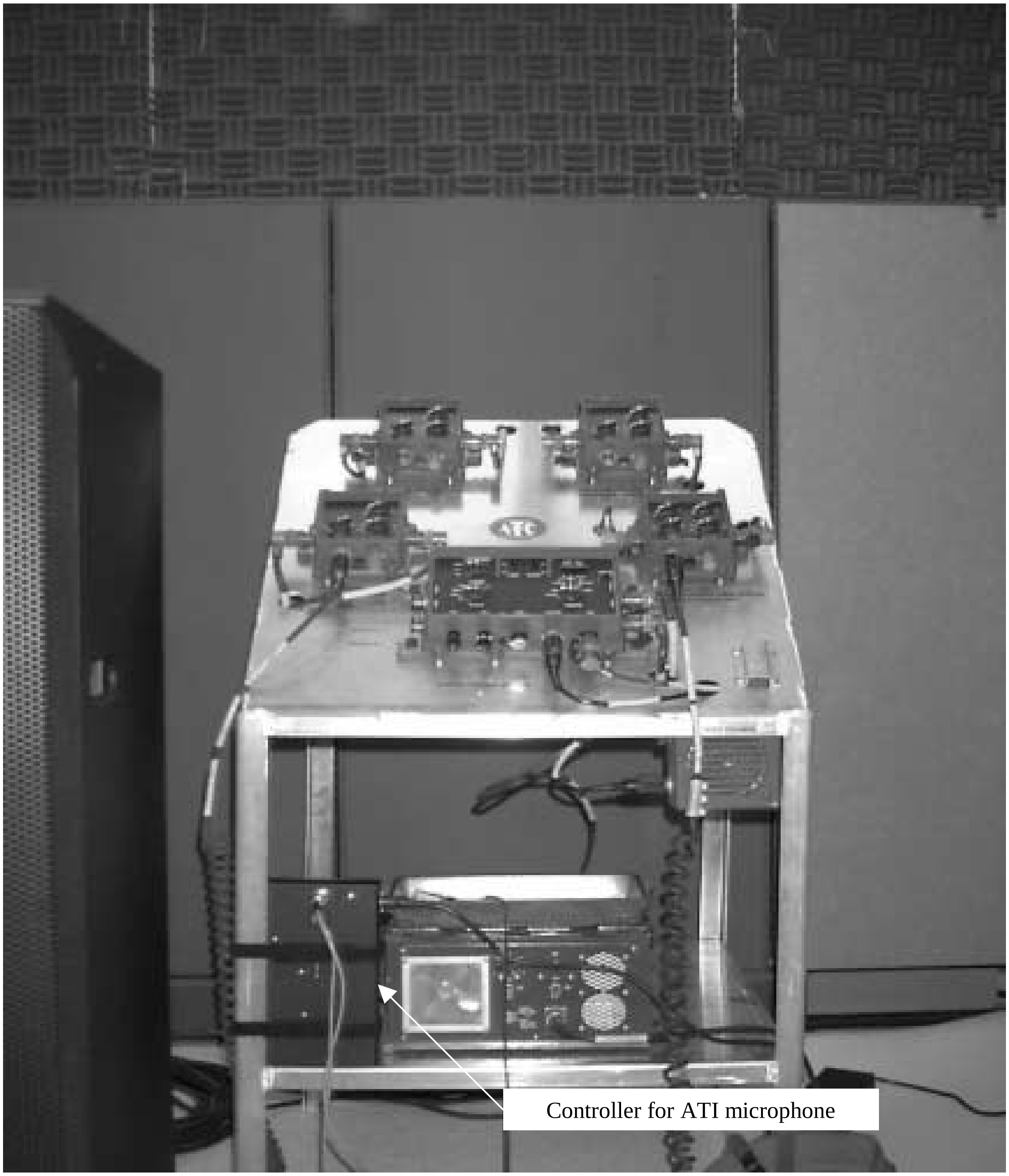


Figure 19. The VIS control console.

abruptly and the analyzer allowed to run for several seconds until the noise had decayed completely. Decay curves (SPL vs. time) were displayed on the LCD screen of the analyzer and printed. The RT(60) was calculated by the LD 3200 using the linear portion of the decay curve. (The analyzer performs a linear regression on the linear portion of the decay curve between two cursors positioned by the user to calculate a slope from which the RT60 is estimated.) Data obtained in this manner is shown in Table 5.

SPL at six positions about the participant's head. The sound pressure levels at 6 positions (± 6 in.) about the participant's head center position, an indication of the uniformity of the sound field, appear in Table 6. The six positions were: front, back, left, right, up, and down. A total of six equivalent continuous sound level (Leq) measurements were obtained over 60 s periods using the LD 3200 analyzer, the LD 2559 microphone, and the LD 900B preamplifier. Earshen (1996) stated that Leq equals the continuous sound level which, integrated over a specific time, would result in the same energy as a variable sound integrated over the same time. Four of the six measurements (left, right, front, and back) were made with the diaphragm of the microphone placed 6 in. from the center of the room corresponding to the desired position. For the other two measurements (up and down), the microphone was placed at the center of the room and adjusted ± 6 in. to obtain the up or down position. The room was configured exactly as for testing, but without the participant, chair, or desk present. The 1/3 OB measurement was obtained using a steady (i.e., not pulsed) pink noise (flat by octaves) test signal at a broadband level of 114 dBA.

Experimental Procedure

Participants attended a combined screening and practice session and four experimental sessions. For each practice and experimental session, the CCAB tests, MRT

Table 5. Reverberation time of the test room, RT(60)

Center Frequency (Hz)	RT60 (sec)
250	0.520
500	0.362
1000	0.213
2000	0.177
4000	0.133
8000	0.146

Table 6. SPL Variation at Six Positions About Head Center Position.

1/3 OB Center (Hz)	Right (dB)	Left (dB)	**Diff.* R-L**	Up (dB)	Down (dB)	Front (dB)	Back (dB)	**Diff.** 6-Pos.**
63	107.0	107.5	**-0.5**	107.0	107.3	106.4	107.9	**1.5**
80	107.6	107.4	**0.2**	107.2	107.7	107.5	107.6	**0.5**
100	121.2	121.2	**0.0**	119.3	122.6	120.4	121.9	**3.3**
125	108.1	108.3	**-0.2**	107.2	108.6	107.5	108.3	**1.4**
160	110.0	109.6	**0.4**	109.3	110.1	110.2	109.2	**1.0**
200	114.7	112.6	**2.1**	114.2	114.4	116.4	110.2	**6.2**
250	106.1	107.1	**-1.0**	109.4	106.5	107.9	107.0	**2.9**
315	106.0	106.8	**-0.8**	109.2	105.3	106.4	106.9	**3.9**
400	103.1	100.8	**2.3**	107.1	105.7	102.8	103.4	**2.3**
500	104.0	101.9	**2.1**	103.3	107.4	103.2	102.4	**4.1**
630	100.0	99.6	**0.4**	103.6	101.4	98.1	99.7	**2.2**
800	95.4	92.9	**2.5**	102.0	102.1	88.9	94.0	**5.1**
1000	91.8	96.5	**-4.7**	100.0	101.3	94.4	91.4	**3.0**
1250	94.3	97.2	**-2.9**	98.3	97.0	96.3	93.6	**2.7**
1600	101.9	102.4	**-0.5**	101.3	103.7	102.3	102.6	**2.4**
2000	99.0	96.0	**3.0**	97.2	98.4	96.9	98.4	**3.0**
2500	90.1	95.0	**-4.9**	95.8	94.2	93.6	89.4	**4.2**
3150	89.7	89.2	**0.5**	89.0	89.4	90.2	88.5	**1.7**
4000	86.3	87.0	**-0.7**	85.4	87.0	85.4	86.6	**1.6**
5000	83.2	83.5	**-0.3**	82.8	83.3	83.6	81.8	**1.8**
6300	82.5	83.3	**-0.8**	85.0	84.1	84.7	85.9	**1.2**
8000	79.1	79.0	**0.1**	81.2	80.4	79.2	81.3	**2.1**

* Absolute value dB difference between right and left positions.

** Maximum absolute value dB difference between all pairs of the six microphone positions.

test words, and the subjective workload scales were balanced across the conditions so that each was used an equal number of times. During their first session, prospective participants were screened for the experiment, and if they qualified, they were trained in the experimental procedures. This screening and practice session lasted approximately 1 hr. 30 min. Each of the four experimental sessions was conducted on four separate days and lasted about 1hr 45 min.

Screening and practice session. Participants were screened in the Auditory Systems Laboratory at Virginia Tech, Room 538 Whittemore Hall (screening form appears in Appendix F). After reading and signing the informed consent form (Appendix E), participants were asked general questions about their hearing health. An otoscopic examination was also performed to detect any visible outer- and middle-ear abnormalities. The *method of limits*, a psychophysical technique, was employed to quantify the hearing threshold of each participant by using a Beltone Model 114 clinical pure-tone audiometer at the following frequencies: 125, 250, 500, 1000, 2000, 3000, 4000, 6000, and 8000 Hz.

(Prior to the beginning of the practice and experimental sessions, the experimenter personally fit the VIS headset on each participant to ensure that a good fit was obtained.) Prospective participants who did not meet the screening criteria (hearing threshold levels no higher than + 20 dBHL and no lower than – 10 dBHL in each ear) were thanked and compensated for their time. Prospective participants who qualified for the experiment and chose to continue, next practiced the procedures used in the MRT and the CCAB.

Practice sessions were conducted in Room 519N, Whittemore Hall. Before each practice session, the Larson Davis 3200 spectrum analyzer and ½ -inch microphone were used to set the sound pressure levels of the Bradley Fighting Vehicle noise to 85 dBA and the speech-weighted noise to 61 dB (linear), with both measurements being a 60 s Leq. For the background noise, the microphone was placed at the participant's head-center position. Also, the microphone was placed in front of the mouth simulator with the aid of a calibration jig (UA 0901) supplied by Brüel and Kjaer to set the desired speech level.

The MRT and each individual CCAB test were first practiced individually. For the MRT practice, the participant was presented with six possible word choices (Appendix A) on the touchscreen monitor while the target word, contained within the carrier sentence "Mark the word _____ now," was presented through the artificial mouth located behind the participant. (The participant heard the auditory prompts through the VIS headset since the Gentex or ATI microphone was placed in front of the mouth simulator.) The participant was instructed to select the word heard by touching it on the monitor. Once a word was selected, the computer automatically presented the next stimulus sentence and word choices. This process was repeated until all 50 words were presented. (If the participant did not touch a word within 10 s, the computer automatically presented the next set of words.) The participant performed at least two practice trials of the MRT. Although the purpose of the practice trial was to familiarize the participant with the task, the participant's answers were examined to ensure that the participant was paying attention.

Once the speech intelligibility practice trials were completed, the individual CCAB tests were practiced (*Tower Puzzle, Numbers and Words,* and *Logical Relations*). For each CCAB test, participants read the instructions aloud to assure the experimenter that they were following directions. After each practice test, the computer automatically presented the participant with a series of TRUE/FALSE questions, followed by another practice test. The intention of the TRUE/FALSE questions was to determine if the participant understood the task. Immediate feedback was provided for incorrect answers. The CCAB tests were scored automatically by the computer for both accuracy and time. After practicing each test individually, the participant practiced the CCAB in conjunction with the MRT test twice. The practice session ended when the experimenter was satisfied with the participant's performance on the combined tasks (CCAB plus MRT). Satisfactory performance consisted of a score between the range of 25% and 35% on the MRT and a score of at least 700 on *Logical Relations*, 800 on the *Tower Puzzle*, and 400 on *Numbers and Words* test. Next, each participant was provided with a set of instructions corresponding to the NASA-TLX and the MCH. Prior to practicing each workload scale individually, each participant was allowed to ask any questions after reading the instructions. The protocol used for the screening and practice sessions is shown in Table 7.

Experimental session. Four experimental sessions were conducted on separate days to minimize each participant's exposure to the noise. The same calibration procedures used for the practice session were followed before each experimental session. A pre-experimental audiogram was conducted in Whittemore Hall, Room 538 prior to each experimental session at the following selected frequencies: 125, 500, 1000, 2000, and 4000 Hz. The pre-experimental audiogram was conducted to determine the participant's hearing threshold prior to the experiment to compare to their post-audiogram. Next, the experimenter and the participant went to Room 519N to begin the pre-experimental practice trial. The pre-experimental practice trial was identical to the

Table 7. Protocol for Screening and Practice Session.

1. The experimenter greeted the prospective participant.
2. The experimenter gave the prospective participant a brief overview of the experiment and the screening procedures.
3. The experimenter asked the prospective participant to read and sign the Informed Consent.
4. The experimenter asked the prospective participant about his or her auditory health and visually examined the participant's ear with an otoscope.
5. The experimenter performed an audiogram on the prospective participant.
6. The experimenter and the prospective participant went to Room 519N, Whittemore Hall and while the participant sat outside the room, the experimenter set the speech-weighted noise to 61 dB (linear) and the tank noise to 85 dBA for the practice session.
7. The experimenter asked the participant to enter the room and the experimenter placed the VIS headset on the participant.
8. The experimenter allowed the participant to practice the MRT twice at the speech and noise levels mentioned in Step 4.
9. The experimenter allowed the participant to practice the CCAB twice in 85 dBA noise.
10. The experimenter allowed the participant to practice the MRT and CCAB twice simultaneously in the speech and noise levels mentioned in Step 4.
11. The experimenter allowed the participant to practice the NASA-TLX and MCH.
12. The experimenter paid the participant for the time they participated in the screening and practice session and scheduled him/her for the experimental session.

earlier practice trial except that the experimenter read the instructions to the participant and the True/False tests were not presented. After practicing both the MRT and CCAB tests separately, the two tests were practiced simultaneously. <u>During the practice trials, participants were monitored by the experimenter to observe any unusual behavior and their scores were examined based on the same criteria that was used for the screening/practice session.</u> At the completion of the practice trials, the participant doffed the VIS headset and was asked to leave the room so the experimenter could set the appropriate noise and speech levels. The Bradley Fighting Vehicle noise was set to 114 dBA (using a 60 s Leq) and the speech level was set to either 83 or 96 dB (linear), using a 30 s Leq for both measurements. Once the noise and speech levels were set, the experimenter asked the participant to re-enter the room. The experimenter explained that the noise level would be higher than it had been in the practice trials and that the participant would have to listen carefully to hear the words. Additionally, the experimenter told each participant to weigh both tasks equally. The experimenter then fit the VIS headset on the participant's head, ensured that all cables were properly connected and switches in their appropriate positions, and, after alerting the participant, started the noise presentation and the test. An outline of the procedures followed in administering the experiment is shown in Table 8.

At the completion of the test session, the noise was terminated and the participant was allowed to complete the MCH and NASA-TLX workload scales. Lastly, the experimenter performed a post-audiogram at the same frequencies mentioned earlier, which were compared against the pre-test audiogram, to ensure that the participant did not experience a temporary threshold shift as a result of his or her participation in the

Table 8. Protocol used for the Experimental Session.

1. The experimenter greeted the prospective participant in Room 538, Whittemore Hall and performed a pre-experimental audiogram at the following selected frequencies: 125, 500, 1000, 2000, and 4000 Hz.

2. The experimenter and the participant went to 519N, Whittemore Hall and while the participant sat outside the room, the experimenter set the speech weighted noise to 61 dB (linear) and the tank noise to 85 dBA for the practice session.

3. The experimenter asked the participant to enter the room and the experimenter placed the VIS headset on the participant.

4. The experimenter allowed the participant to practice the MRT twice at the speech and noise levels mentioned in Step 4.

5. The experimenter allowed the participant to practice the CCAB twice in 85 dBA noise.

6. The experimenter allowed the participant to practice the MRT and CCAB twice simultaneously in the speech and noise levels mentioned in Step 4.

7. The experimenter asked the participant to leave the room.

8. The experimenter set the noise level to 114 dBA and the speech-weighted noise level was set to either 83 or 96 dB (linear).

9. The experimenter asked the participant to re-enter the room and explained to the participant that the noise levels would be high; they would have to listen intently to hear the words, and to weigh each task equally.

10. The experimenter placed the VIS headset on the participant and made sure that the cables were properly connected.

11. The experimenter alerted the participant and began the test session.

12. The experimenter allowed the participant to complete the NASA-TLX and MCH.

13. The experimenter performed a post-audiogram on the participant in Room 538, Whittemore Hall at the frequencies that were used in Step 2.

14. The experimenter paid the participant for the time he or she participated in the experimental session.

experiment. If there existed a 10 or 15 dB difference between the pre- and post-audiogram, the experimenter would not have allowed the participant to participate in the other trials. An outline of the procedures used in administering the NASA-TLX and MCH are shown in Tables 9 and 10, respectively.

Insertion Loss Measurements

Another objective of the research was to determine if the prototype ATI microphone/communication system actually transmitted less noise through the communication system than did the Gentex microphone. To determine this, an acoustical test fixture (containing a 1-inch microphone) was placed at the participant's head-center position.

A series of insertion loss measurements were made using both communication microphones, Gentex and ATI. (All measurements were conducted in the same background noise level (114 dBA) used in the other experiment; the speech level was set to 96 dB (linear), using a 60 s Leq.) These measurements consisted of an unoccluded measurement and three occluded measurements using each microphone. The occluded measurements were conducted with the active noise reduction electronics turned on (no speech), active noise reduction electronics turned on with speech-weighted noise, and active noise reduction electronics turned on with real speech (using the MRT stimulus words).

Casali and Robinson (1994) conducted insertion loss measurements to determine the attenuation characteristics of a supra-aural siren canceling headset (SCH) that incorporated active noise reduction electronics (supplied by Noise Cancellation Technologies, Inc.). Their approach was different than the method previously discussed,

Table 9. Protocol used to obtain the workload score for the NASA-TLX.

1. The experimenter gave the participant the instructions to read concerning the NASA-TLX.

2. The experimenter gave the participant a sheet of paper describing the different weighting scales and a separate sheet of paper with the workload scales shown (see Appendix A).

3. The experimenter gave the participant the weighting cards and instructed them to circle the weighing scale title that corresponds to his or her experience when performing the task.

4. The experimenter allowed the participant to complete the workload scales that were provided in Step 2.

Table 10. Protocol used to obtain the workload score for the MCH.

---

1. The experimenter gave the participant instructions for the MCH scale along with a copy of the MCH scale on an 8.5 x 11 in sheet of white paper.

2. The experimenter informed the participant that he or she was considered the "operator".

3. The experimenter informed the participant that he or she was to start from the lower left hand side of the scale and follow the path until he or she decided on a rating.

4. The experimenter instructed the participant to circle only one number on the sheet and the number would be arrived at by following the logic on the sheet.

5. The experimenter instructed the participant not to skip any steps in the logic.

---

in that they used two test methods (KEMAR and MIRE), at three siren levels (90, 95, and 100 dB), and three siren modes (“Wail”, “Yelp”, and “Hi-Lo”). The results of their experiment indicated that at sounds below 98 dB, the SCH was successful at accomplishing its goal to reduce the noise energy in the frequency where it was most dominant, but some distortion occurred above 98 dB. It is noteworthy to mention that this experiment was conducted using siren stimuli obtained from the cab of an ambulance with the windows rolled up without additional noise from external sources.

RESULTS

Data Reduction

Insertion loss measurements. For the insertion loss measurements, the insertion loss was computed by subtracting the occluded measurement from the unoccluded measurement for each microphone. Additionally, the insertion loss for the two microphones were subtracted from each other to determine the attenuation afforded by the ATI microphone in each condition.

Speech intelligibility. The raw scores for the MRT were recorded in a separate text file and copied into Microsoft Excel. Next, the raw score was entered into the following formula, from ANSI S3.2-1989, that adjusted the raw score for chance or guessing:

$$R_a = R - [W/(n - 1)]$$

where R_a is the number of items correct adjusted for chance/guessing, R is the number of items correct, W is the number of items incorrect, and n is the number of alternative choices per item.

The scores for the MRT were expressed in percentages and can range between 0 and 100%. However, a score of “0” would indicate that the participant was not paying attention and such a participant would have been eliminated from the experiment. Once the R_A number was calculated, it was divided by 50 (the total number of words presented to the participant) and then multiplied by 100 to obtain the percentage of correct words.

Cognitive performance. The raw scores for each cognitive assessment test were automatically computed by the CCAB and stored in a file on the computer on which the test was administered. At the end of each test session, the scores for each test were printed. To obtain the arithmetic mean score, the score for the trials that were completed were entered into Microsoft Excel.

For the CCAB tests, only the minimum score on each test could be obtained because each of the scores were computed by a multiplicative formula that consists of several factors. For instance, the *Tower Puzzle* and *Logical Relations* test scores are computed based on accuracy, speed, problem difficulty, and a range constant. Whereas, for the *Number and Words* test, the factors that determine the score are accuracy, speed, and range constant. Neverthless, the minimum score for each test was obtained by allowing time to expire on each test without touching the keys on the keyboard. For the *Tower Puzzle* test this score was 122, for the *Number and Words* test it was 334, and for the *Logical Relations* test it was 212. Nevertheless, the actual range for the scores in this experiment were between 847 and 2146 for the *Tower Puzzle* test; 744 and 1435 for the *Logical Relations* test; and 405 and 728 for the *Numbers and Words* test. Once again, this is due to the scoring schema for which the CCAB was programmed to compute.

Modified Cooper-Harper. It was not necessary to perform secondary procedures to obtain the workload scores for the Modified Cooper-Harper scale. The MCH is a one-dimensional scale in which each participant selected a single number based on their perceived mental workload. This number represented the overall workload for that particular task, along with a description of the workload experienced.

NASA-TLX. To recapitulate, the NASA-TLX is based on six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. To score the NASA-TLX, two of the six dimensions were paired on 15 cards for the participant to select one. Next, each participant rated his or her perceived mental workload using a scale for each dimension. Based on these results, the scale selected on each card was summed to determine the total number of times a particular dimension was selected. The results were recorded on a scoring sheet supplied by the NASA-TLX under the heading “Tally.” The tally marks were then counted and entered into the “Weight” column. Afterwards, the numerical value corresponding with the hashmarks on each scale were entered into the “Raw Weighting” column that corresponded to that particular scale. Next, the “Weight” column was multiplied by the “Raw Weighting” column to obtain the adjusted weighting. The “Adjusted Rating” was summed and divided by 15 to obtain the overall workload score, which was entered into the spreadsheet. This scoring procedure aligns with that recommended by Hart and Staveland (1988).

Insertion Loss Measurements

The measurements conducted to determine the reduction in noise at the ear are shown in Table 11. Also, the difference between the insertion loss measurements for each microphone is shown in Table 12. Based on Table 12, the ATI communication microphone reduces the noise transmitted through the communication system when compared to the existing Gentex microphone.

Objective and Subjective Performance Measures

Based on the insertion loss measurements, it would appear that the ATI microphone would improve speech intelligibility. However, without further exploration, this would be only speculation. Therefore, objective and subjective data were analyzed in an effort to determine whether the communication microphone system developed by ATI produced any significant improvements in speech intelligibility over the current Gentex

Table 11. Insertion loss (IL) measurements for the two communication microphones.

Condition	dBA	IL (dB)
Unoccluded	113.9	
Gentex Microphone		
ANR on - no speech	86.0	27.9
ANR on - speech-weighted noise	93.1	20.8
ANR on - speech (MRT words)	89.7	24.2
ATI Microphone		
ANR on - no speech	83.9	30.0
ANR on - speech-weighted noise	90.8	23.1
ANR on - speech (MRT words)	87.9	26.0

Table 12. Difference in insertion loss between the two microphones.

Gentex Microphone	dBA	ATI Microphone	dBA	Diff. (dB)*
ANR on – no speech	86.0	ANR on - no speech	83.9	2.1
ANR on - speech-weighted noise	93.1	ANR on - speech-weighted noise	90.8	2.3
ANR on - speech (MRT words)	89.7	ANR on - speech (MRT words)	87.9	1.8

*Positive values indicate an advantage in insertion loss for the ATI microphone.

microphone, and if there were any concomitant improvements in cognitive performance or a reduction in perceived mental workload.

A summary of the standard deviation and mean CCAB test scores for each test and the corresponding adjusted MRT score, at each speech level, are shown in Tables 13 through 15 and Figures 20 through 22.

First Analysis

The first experimental design was used to examine the main effects of *Communication Microphone*, *Speech Level*, and *Cognitive Assessment Test*, and all the associated interactions. For this design, a repeated-measures multivariate analysis of variance (MANOVA) was performed using Statistical Analysis Software (SAS; 1999), Version 8. The Wilks' lambda (Λ) criterion was used as the test statistic in both analyses. Where appropriate, the MANOVA was followed by an analysis of variance (ANOVA) on each dependent variable. In the ANOVA, Geisser-Greenhouse corrections were used to guard against any violation of the assumption of homogeneity of covariance where there were more than two levels of the independent variable (Winer, Brown, and Michels, 1991). Post hoc analyses of significant main effects were conducted using a Student Newman-Keuls test. For significant interactions, simple effects *F*-tests were conducted, followed by a Student Newman-Keuls test, when appropriate.

The MANOVA summary table for the analysis using the cognitive assessment and speech intelligibility scores appears in Table 16. Results revealed a significant two-way and three-way interaction: *Speech Level by Communication Microphone* (Λ = 0.192;

Table 13. Raw CCAB and MRT scores for the *Tower Puzzle* (TP) test (units and ranges discussed in text).

	Gentex Microphone				**ATI Microphone**			
	83 dB		**96 dB**		**83 dB**		**96 dB**	
Subjects	**TP Score**	**MRT**	**TP Score**	**MRT**	**TP Score**	**MRT**	**TP Score**	**MRT**
1	2009	28	2146	77	2119	25	2023	72
2	1172	30	1839	74	1284	25	1732	63
3	847	44	1818	84	1284	46	1732	86
4	1525	35	1969	77	1771	21	1904	53
5	1146	21	1947	77	1896	28	1698	65
6	1221	42	1919	77	1716	29	1649	70
7	1199	23	1254	42	1532	25	1199	23
8	1385	35	1299	77	1536	51	873	46
Means	**1313**	**32**	**1774**	**73**	**1642**	**33**	**1601**	**60**
***Std. Dev.**	**320**	**8**	**302**	**12**	**272**	**10**	**355**	**18**

*Std. Dev. = standard deviation

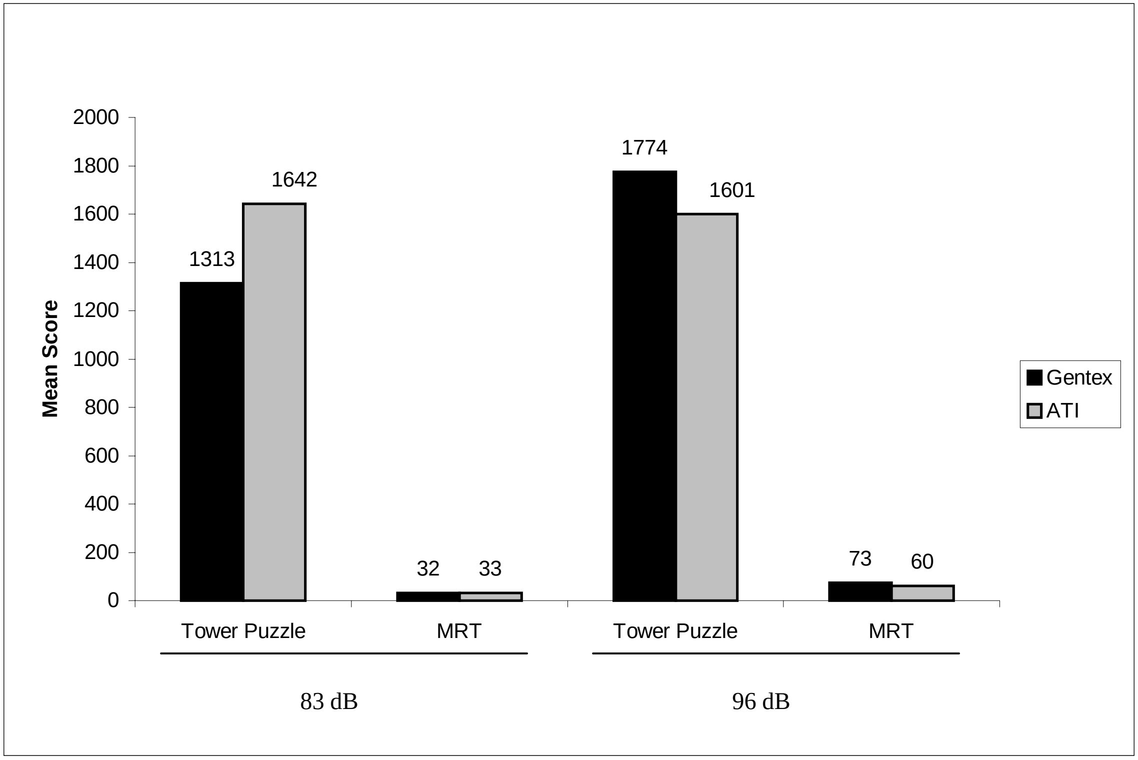


Figure 20. Mean cognitive assessment scores and speech intelligibility scores for the *Tower Puzzle* test.

Table 14. Raw CCAB and MRT scores for the *Numbers and Words* (NW) test (units and ranges discussed in text).

	Gentex Microphone				**ATI Microphone**			
	83 dB		**96 dB**		**83 dB**		**96 dB**	
Subjects	**NW Score**	**MRT**	**NW Score**	**MRT**	**NW Score**	**MRT**	**NW Score**	**MRT**
1	649	23	728	65	661	28	618	79
2	512	25	553	77	523	39	541	70
3	646	35	720	86	523	49	541	86
4	527	28	574	77	479	21	502	58
5	483	37	618	77	582	35	514	74
6	495	35	597	70	580	37	601	70
7	417	21	405	60	468	18	417	39
8	517	46	581	74	542	63	523	35
Means	**531**	**31**	**597**	**73**	**545**	**36**	**532**	**64**
***Std. Dev.**	**74**	**8**	**95**	**8**	**58**	**14**	**58**	**17**

*Std. Dev. = standard deviation

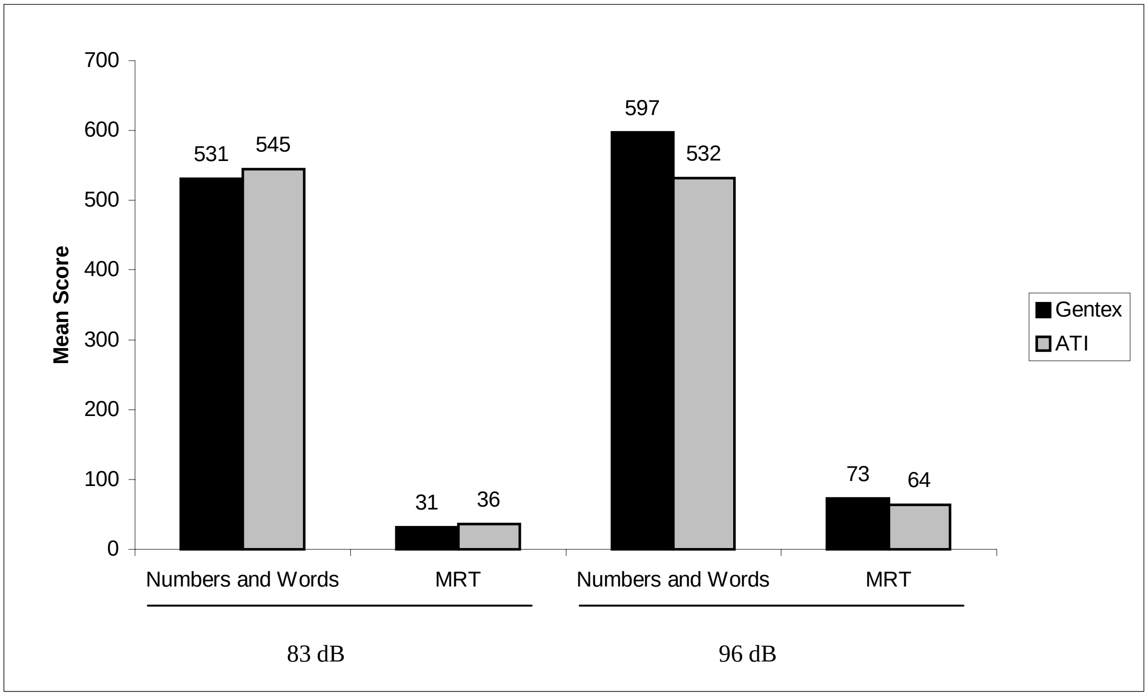


Figure 21. Mean cognitive assessment scores and speech intelligibility scores for the *Number and Words* test.

Table 15. Raw CCAB and MRT scores for the *Logical Relations* (LR) test (units and ranges discussed in text).

	Gentex Microphone				**ATI Microphone**			
	83 dB		**96 dB**		**83 dB**		**96 dB**	
Subjects	**LR Score**	**MRT**	**LR Score**	**MRT**	**LR Score**	**MRT**	**LR Score**	**MRT**
1	1460	37	1417	86	1435	25	1408	74
2	1036	35	1050	74	1024	35	1046	67
3	1073	46	1215	79	1024	30	1046	70
4	970	25	1239	77	1145	14	1185	58
5	820	23	896	77	819	39	755	74
6	1114	32	1262	72	1155	46	1115	67
7	756	23	809	51	744	32	756	18
8	1169	42	1289	84	1225	65	1086	53
Means	**1050**	**33**	**1147**	**75**	**1071**	**36**	**1050**	**60**
***Std. Dev.**	**204**	**8**	**196**	**10**	**207**	**14**	**202**	**17**

*Std. Dev. = standard deviation

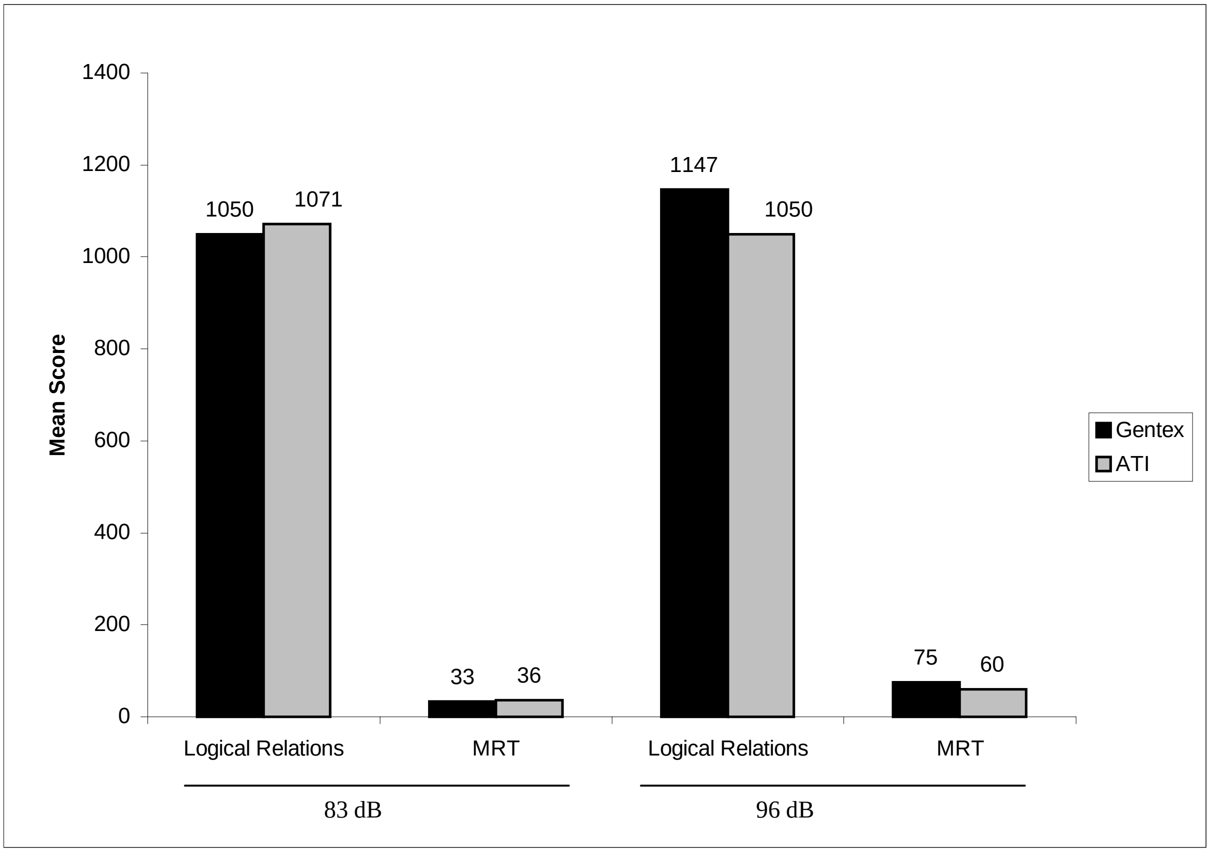


Figure 22. Mean cognitive assessment scores and speech intelligibility scores for the *Logical Relations* test.

Table 16. MANOVA Summary Table for Dependent Variables *Cognitive Assessment Score* and *Speech Intelligibility Score*.

Source	**Num DF**	**Den DF**	Λ	**F**	p
Between-Subjects					
Within-Subjects					
Speech Level (Splv)	2	6	0.098	27.61	0.0009*
Subj. x Splv.					
Test (Tst)	4	26	0.061	19.84	0.0001*
Subj. x Tst					
Comm. Mic. (Mic*)*	2	6	0.551	2.44	0.1674
Subj. x Mic					
Splv x Tst	4	26	0.750	1.01	0.4227
Subj x Splv x Test					
Splv x Mic	2	6	0.192	12.60	0.0071*
Subj. x Splv x Mic					
Tst x Mic	4	26	0.548	2.28	0.0876
Subj. x Tst x Mic					
Tst x Mic x Splv	4	26	0.241	6.73	0.0007*
Subj x Tst x Mic x Splv					

* Denotes significance at ($p < 0.05$).

$p = 0.0071$) and *Speech Level by Communication Microphone by Test* ($\Lambda = 0.241$; $p = 0.0007$). The significant main effects were *Test* ($\Lambda = 0.061$, $p < 0.0001$) and *Speech Level* ($\Lambda = 0.098$, $p = 0.0009$). The *Communication Microphone* main effect was not significant ($\Lambda = 0.551$, $p = 0.1674$). Only the significant interactions and main effects from the MANOVA can be explored in the subsequent ANOVAs (Johnson, 1998).

Speech intelligibility. The ANOVA summary table for the dependent variable, *speech intelligibility score*, is shown in Table 17. There was a significant *Speech Level by Communication Microphone* ($p = 0.0410$) interaction and a *Speech Level* main effect ($p = 0.0002$). Simple effects F-tests showed no simple main effect of *Communication Microphone* at a *Speech Level* of 83 dB ($p = 0.3942$). However, at 96 dB, there was a significant main effect of *Communication Microphone* ($p = 0.0219$), in that speech intelligibility scores were significantly higher for the Gentex microphone (73.8) at 96 dB than for the ATI microphone (61.3). The graph of this interaction is shown in Figure 23.

The simple effects F-test examining the simple main effects of *Speech Level* for each communication microphone was also significant ($p < 0.0001$ for the Gentex and $p = 0.0101$ for the ATI microphone). The interaction is shown in Figure 24. The main effect of *Speech Level* ($p = 0.0002$), shown in Figure 25, was also significant.

Cognitive performance. The ANOVA summary table for the dependent variable, *cognitive assessment score*, is shown in Table 18. The ANOVA revealed that the three-way interaction of *Speech Level by Communication Microphone by Test* was not significant ($p = 0.2980$), and was therefore not analyzed further.

The *Speech Level by Communication Microphone* ($p = 0.0012$) interaction was significant. This interaction was explored in two ways. First, the simple main effect of

Table 17. ANOVA Summary Table for the Dependent Variable *Speech Intelligibility Score*.

Source	**df**	**SS**	**MS**	**F**	***p***	**G-G**
Between-Subjects						
Subject (Subj.)	7	6910.8	987.3			
Within-Subjects						
Speech Level (Splv)	1	27812.0	27812.0	49.86	0.0002*	
Subj. x Splv.	7	3904.5	557.8			
Test (Tst)	2	58.2	29.1	0.57	0.5785	0.3598
Subj. x Tst	14	715.5	51.1			
Comm. Mic. (Mic)	1	580.2	580.2	5.66	0.0489	
Subj. x Mic	7	717.3	102.5			
Splv x Tst	2	9.8	4.9	0.10	0.9024	0.9501
Subj x Splv x Tst	14	661.2	47.2			
Splv x Mic	1	1395.4	1395.4	6.25	0.0410*	
Subj. x Splv x Mic	7	1563.8	223.4			
Tst x Mic	2	90.6	45.3	1.11	0.3555	0.1709
Subj. x Tst x Mic	14	569.4	40.7			
Tst x Mic x Splv	2	19.3	9.7	0.47	0.6329	0.8056
Subj x Tst x Mic x Splv	14	286.0	20.4			
Total	95					

* Denotes significance at ($p < 0.05$). (only main effects and interactions that were significant in the MANOVA can be analyzed in the subsequent ANOVA.)

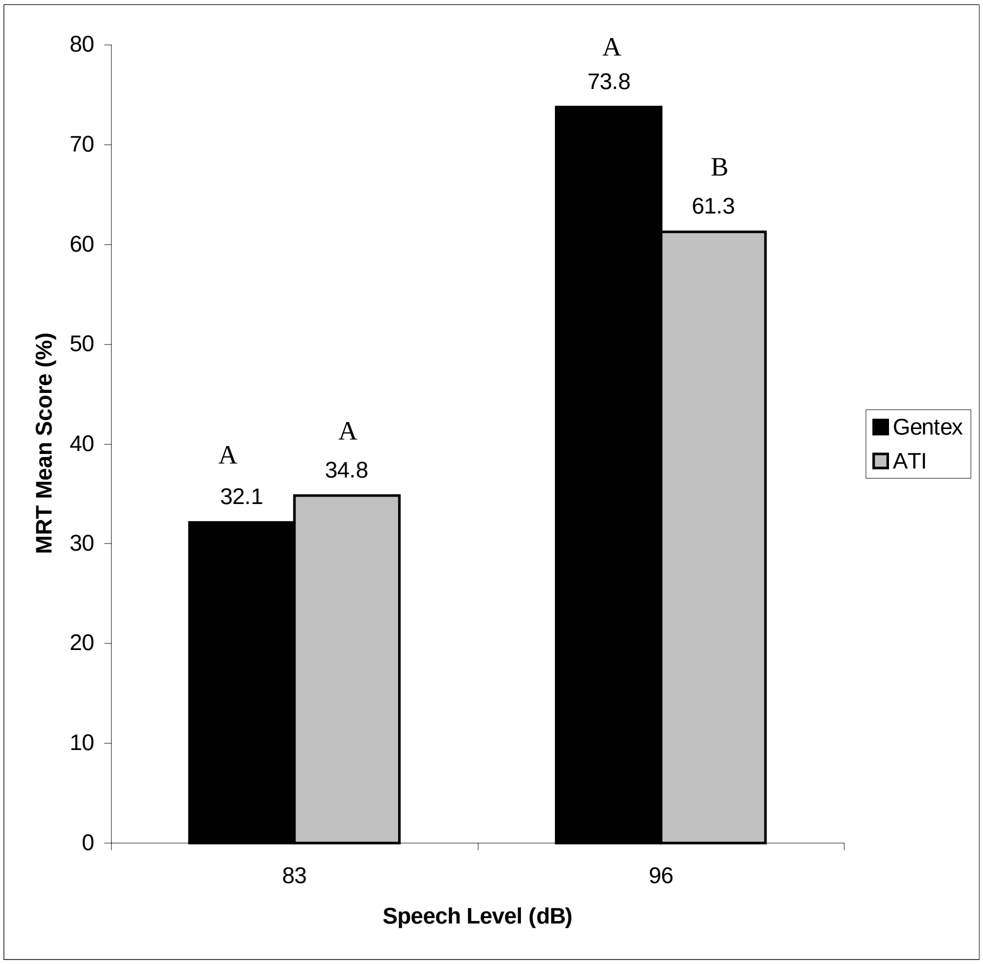


Figure 23. Simple main effect of *Communication Microphone* at each speech level for the *speech intelligibility score* dependent variable. (Means with the same letter are not significantly different from each other at $p < 0.05$.)

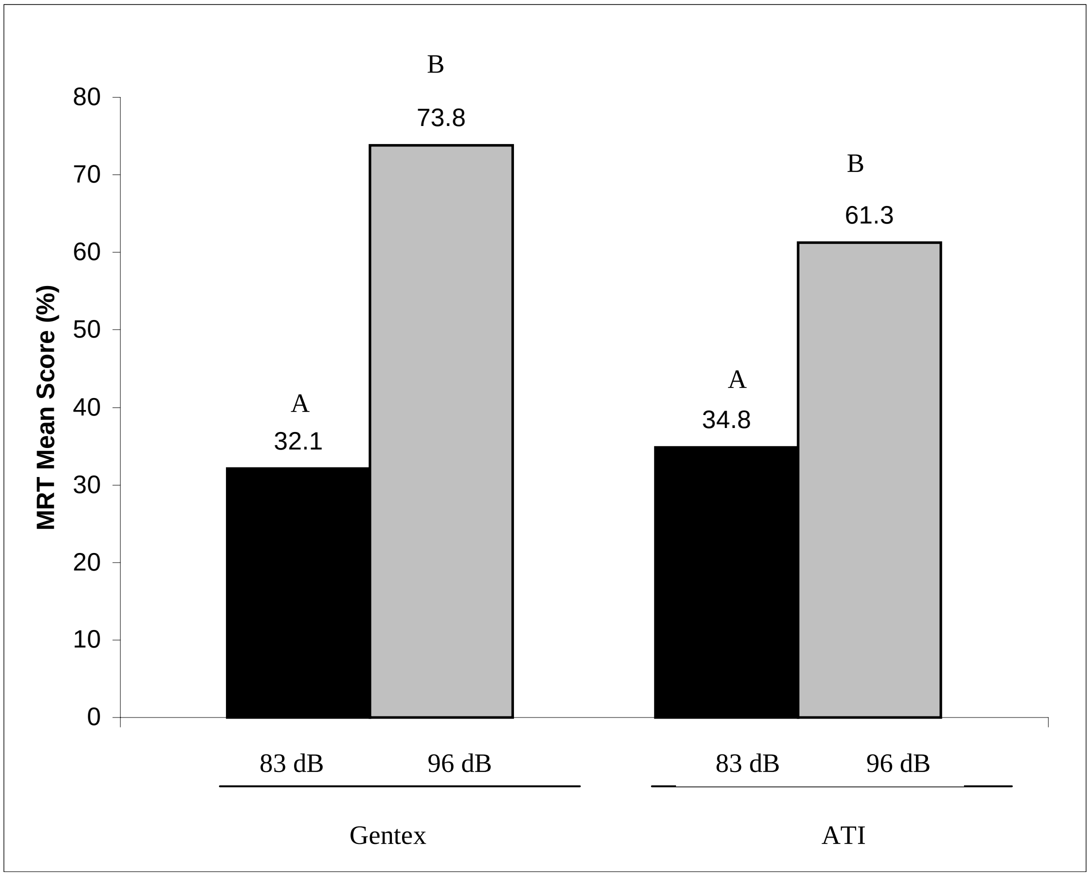


Figure 24. Simple main effect of *Speech Level* for each communication microphone for the *speech intelligibility score* dependent variable. (Means with the same letter are not significantly different from each other at $p < 0.05$.)

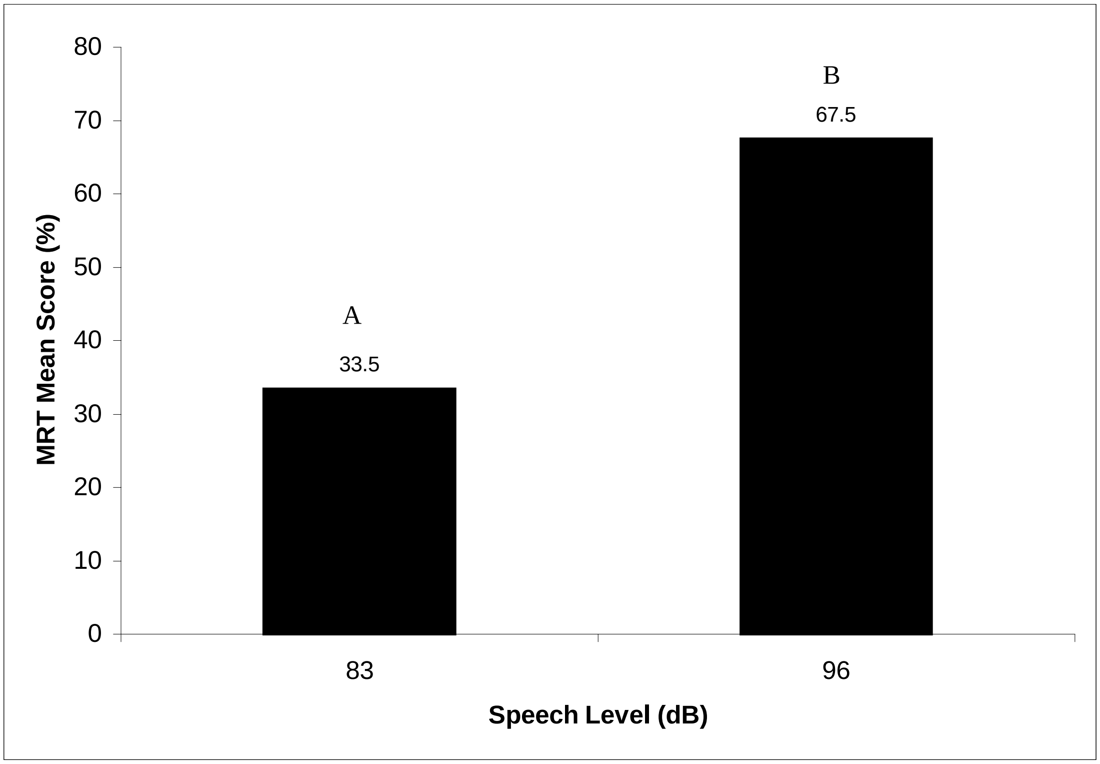


Figure 25. Main effect of *Speech Level* for the *speech intelligibility score* dependent variable. (Means with the same letter are not significantly different from each other at $p < 0.05$.)

Table 18. ANOVA Summary Table for the Dependent Variable *Cognitive Assessment Score*.

Source	df	SS	MS	F	*p*	G-G
Between-Subjects						
Subject (Subj.)	7	2147981.7	306854.5			
Within-Subjects						
Speech Level (Splv)	1	201025.5	201025.5	3.69	0.0963	
Subj. x Splv.	7	381539.7	54505.7			
Test (Tst)	2	17025196.6	8512598.3	95.24	0.0001	< 0.0001*
Subj. x Tst	14	1251275.9	89376.9			
Comm. Mic. (Mic)	1	595.0	595.0	0.09	0.7672	
Subj. x Mic	7	43957.9	6279.7			
Splv x Tst	2	168754.1	84377.0	2.17	0.1507	0.0016
Subj x Splv x Tst	14	543723.4	38837.4			
Splv x Mic	1	326550.0	326550.0	27.67	0.0012*	
Subj. x Splv x Mic	7	82623.2	11803.3			
Tst x Mic	2	65158.3	32579.2	4.33	0.0343	0.0259
Subj. x Tst x Mic	14	105218	7515.5			
Tst x Mic x Splv	2	218031.1	109015.5	20.59	0.0001	0.2980
Subj x Tst x Mic x Splv	14	74138.4	5295.6			
Total	95					

* Denotes significance at ($p < 0.05$). (only main effects and interactions that were significant in the MANOVA can be analyzed in the subsequent ANOVA.)

Communication Microphone at each speech level was examined. The simple effects *F*-test revealed that the simple main effect of *Communication Microphone* was significant at both speech levels, (p =0.0051 at 83 dB and p = 0.0025 at 96 dB) at $p < 0.05$. This interaction is illustrated graphically in Figure 26.

The simple main effect of *Speech Level* for each communication microphone was significant only for the Gentex microphone (p = 0.0051), but not for the ATI microphone (p = 0.6510). The graph of this interaction is shown in Figure 27. Also, the main effect of *Test* was also significant ($p < 0.0001$), while the main effect of *Speech Level* was not significant (p = 0.0963). A Student Newman-Keuls post- hoc test was performed on the significant main effect of *Test* revealing that scores on each of the three tests were significantly different from each other at $p < 0.05$. The results are shown graphically in Figure 28.

Second Analysis

The second analysis was conducted based on the second experimental design in which there were two levels for each independent variable, *Speech Level* and *Communication Microphone*. The purpose of this design was to assess the perceived mental workload imposed by the main effects, *Speech Level* and *Communication Microphone*. Again, the statistical procedure employed to explore this design was a MANOVA, which used the Wilks' lambda (Λ) criterion as the test statistic. Results of the MANOVA (shown in Table 19) revealed no significant main effects or interactions. Though no significant differences were uncovered, the means for the two workload measures are presented in Table 20.

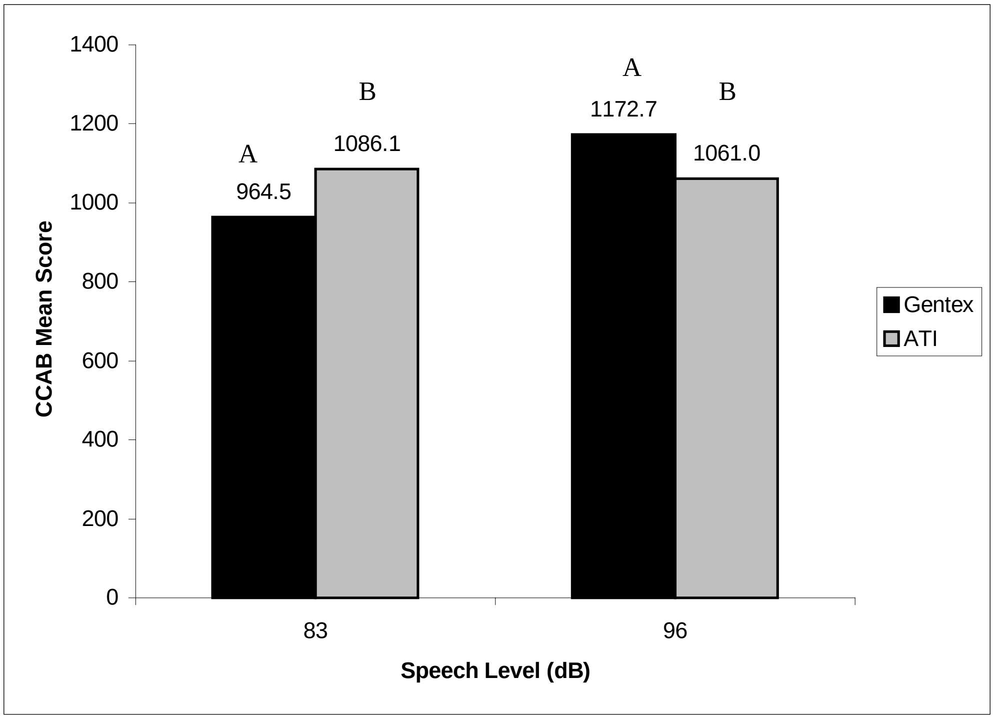


Figure 26. Simple main effect of *Communication Microphone* for each speech level for the *cognitive assessment score* dependent variable. (Means with the same letter are not significantly different from each other at $p < 0.05$.)

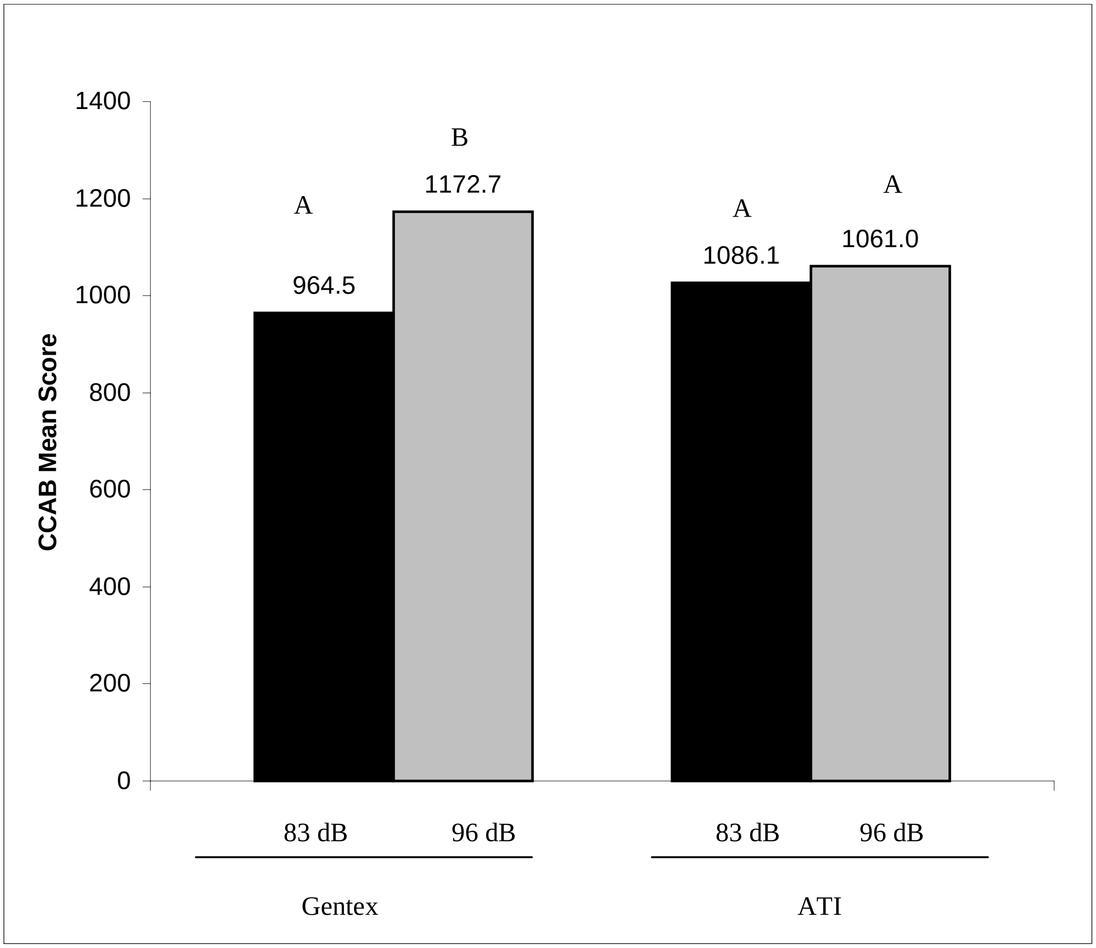


Figure 27. Simple main effect of *Speech Level* for each communication microphone for the *cognitive assessment score* dependent variable. (Means with the same letter are not significantly different from each other at $p < 0.05$.)

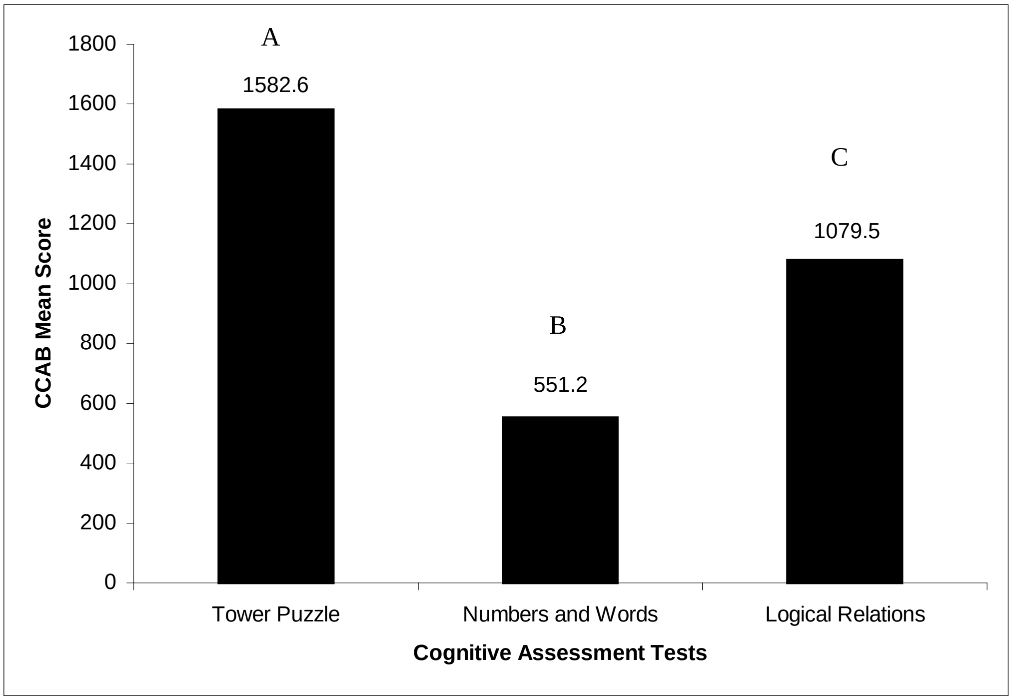


Figure 28. Main Effect of *Cognitive Assessment Test* for the *cognitive assessment score* dependent variable. (Means with the same letter are not significantly different from each other at $p < 0.05$. Data are only presented to show the magnitude on each scale).

Table 19. MANOVA Summary Table for Dependent Variables *NASA-TLX* and *Modified Cooper-Harper*.

Source	**Num DF**	**Den DF**	Λ	**F**	p
Subject (Subj.)					
Speech Level (Splv)	2	6	0.542	2.54	0.1589
Subj. x Splv.					
Comm. Mic. (Mic)	2	6	0.834	0.60	0.5796
Subj. x Mic					
Splv x Mic	2	6	0.513	2.85	0.1347
Subj. x Splv x Mic					

Table 20. Mean scores and standard deviation* for the two workload assessment scales across both speech levels and communication microphones.

	Speech Level			
	83 dB		**96 dB**	
Comm. Mic	**NASA-TLX**	**MCH**	**NASA-TLX**	**MCH**
Gentex	6.4 (1.0)	5.6 (1.5)	5.5 (1.8)	6.2 (1.5)
ATI	5.5 (1.7)	5.3 (1.9)	4.8 (1.5)	5.1 (1.5)

*****Standard deviation in parenthesis

Correlation. In addition to the statistical analyses discussed above, a correlation was performed to determine the relationship between the two workload scales and the relationship between the CCAB and the two workload scales. In this case, the null hypothesis was that no association exists between the NASA-TLX and MCH, nor does there exist a relationship between the CCAB and the two workload scales. Although the Pearson product-moment correlation and Spearman rank-order correlation are used most frequently in the literature, it is important to distinguish between these two correlations to determine which is appropriate for this research.

The Pearson product-moment correlation is appropriate when the dependent variable consists of data that is continuous, while the Spearman rank-order correlation should be used when ordered data is the dependent variable (Graziano and Raulin, 1997). Furthermore, Hatcher and Stepanski (1994) stated that when one variable is ordinal and the other is interval, the Spearman-rank order correlation may be appropriate. Therefore, since the two workload assessment scales used to measure mental workload are classified as having interval (NASA-TLX) and ordinal (MCH) scales, the Spearman-rank order correlation coefficient (r_s) was used to determine the relationship between these two scales.

Based on the results, there was a fairly moderate, significant positive ($p = 0.0210$; $r_s = 0.40629$) correlation between the two workload scales, NASA-TLX and MCH. The graph of this result is shown in Figure 29. Also, there was a weak, non-significant negative correlation between the CCAB and NASA, ($p = 0.6403$; $r_s = -0.04830$) and CCAB and MCH, ($p = 0.8981$;$r_s = -0.01324$). The graph of the results for these two

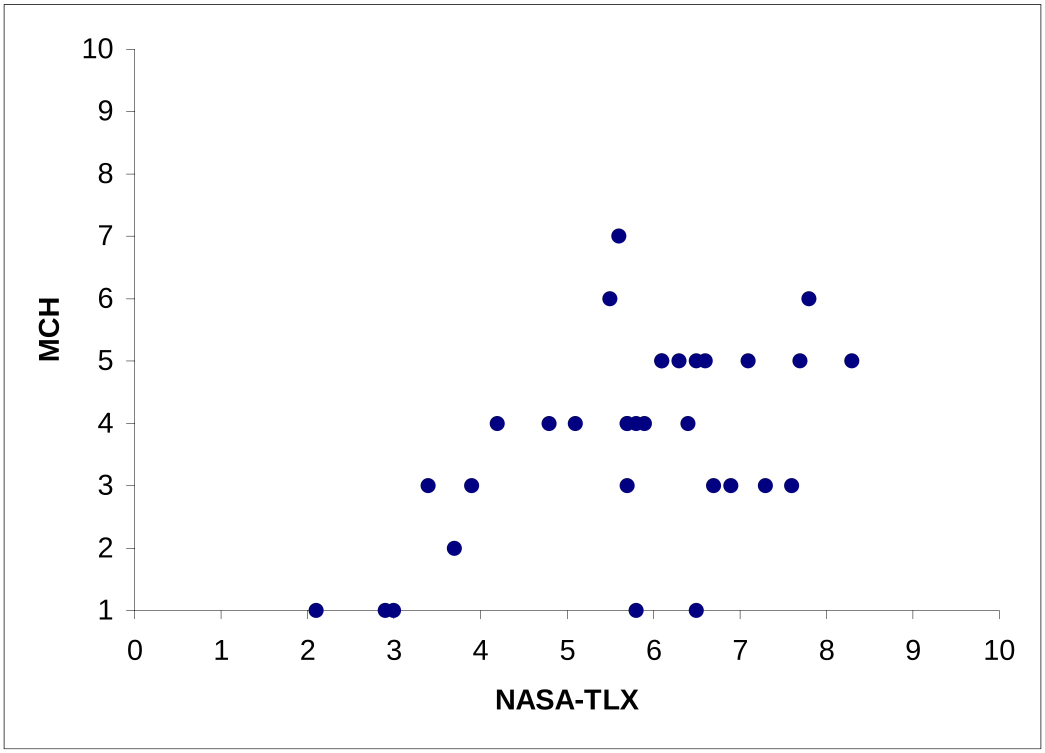


Figure 29. Graph of the data from the NASA-TLX and MCH scores.

correlations are shown in Figures 30 and 31, which also justifies using the Spearman rank-order correlation, in that the Pearson correlation is appropriate when a linear relationship exists (Hatcher and Stepanski, 1994), which is not the case based on each of the three previously mentioned graphs.

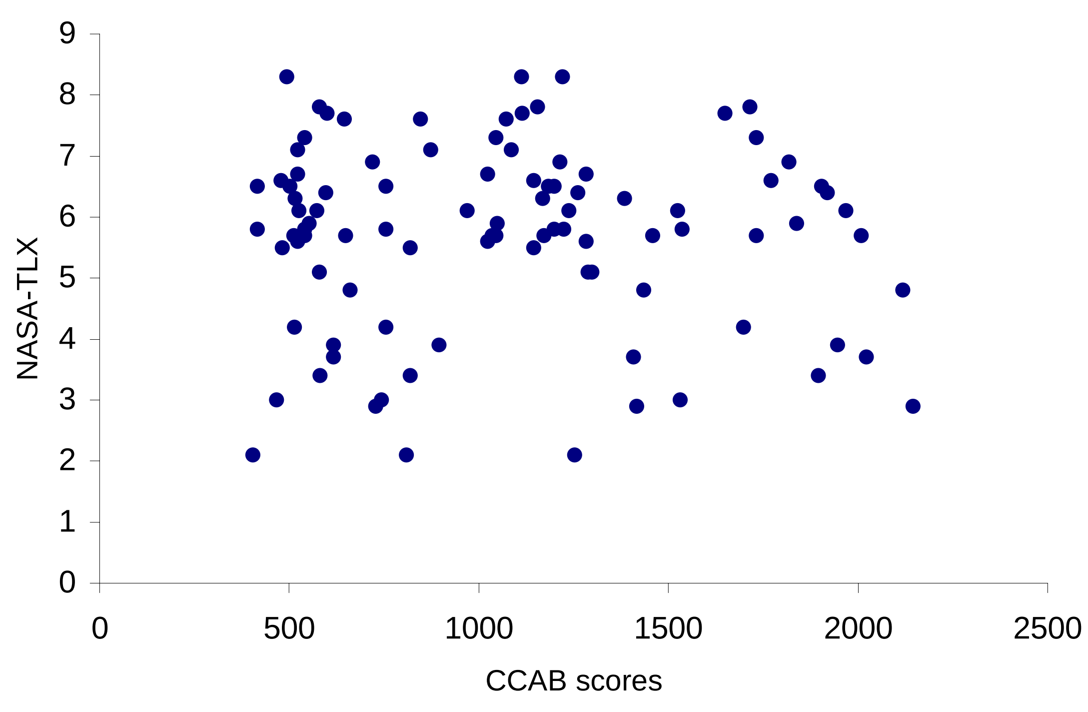


Figure 30. Graph of the data from the NASA-TLX and CCAB scores.

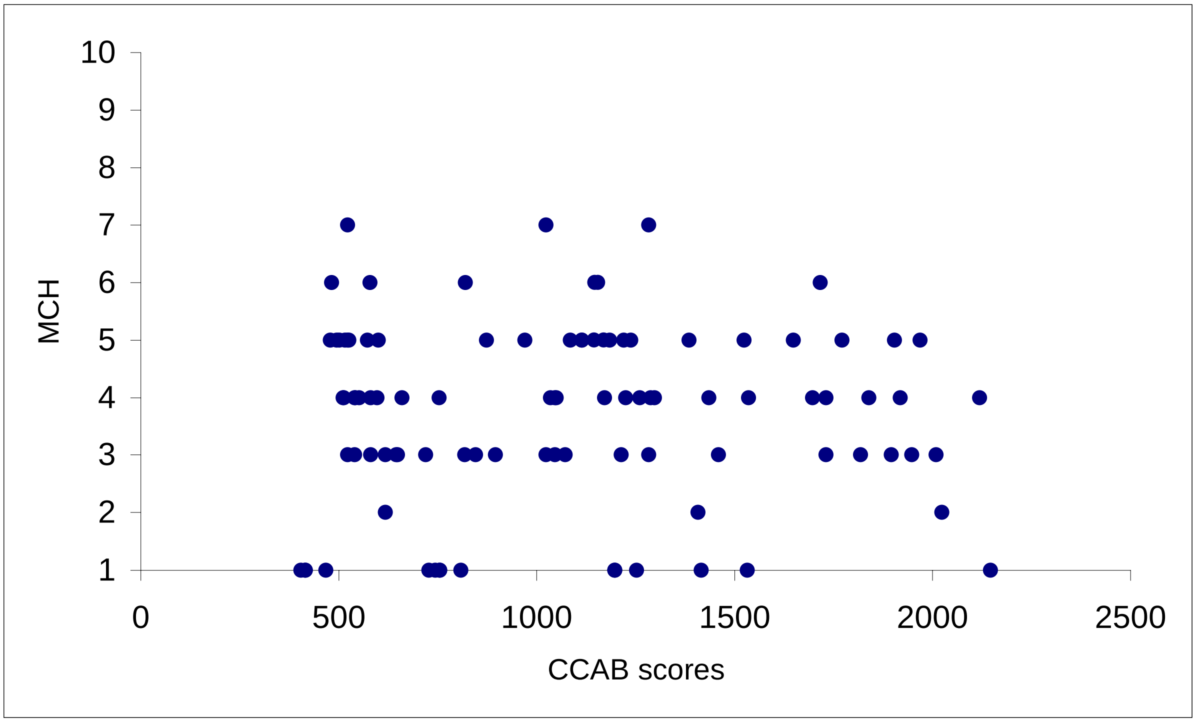


Figure 31. Graph of the data from the MCH and CCAB scores.

DISCUSSION

The primary purpose of this dissertation was to test hypotheses concerning whether a reduction in noise at the ear will cause an improvement in speech intelligibility, an improvement in cognitive performance, or a reduction in subjective mental workload. Speech intelligibility was measured using the MRT, while cognitive performance was measured using specific tests chosen from the CCAB. Subjective mental workload was assessed using two subjective workload scales that presently dominate the literature, the NASA-TLX and the MCH (Casali, Robinson, Dabney, and Gauger, 2000; Draper, Ruff, and LaFleur, 2001; Matthews, Campbell, and Falconer, 2001; Matthews and Shattuck, 2001; Park and Woldstad, 2000).

This discussion begins by briefly summarizing the results of the insertion loss measurements that were conducted to determine which microphone provided the greatest attenuation. Next, the results of the experiment are discussed with respect to the speech intelligibility and the cognitive performance scores. Then, the discussion concludes with the subjective mental workload scores.

Insertion Loss Measurements

Based on the results shown in Table 12, it is apparent that both microphones, ATI and Gentex, were capable of reducing the 114 dBA noise reaching the listeners' ear for each condition (ANR on –no speech, ANR on – speech weighted noise, and ANR on-speech. However, the ATI microphone reduced the noise transmitted through the communication microphone system more than the existing Gentex microphone. The largest difference (2.3 dB) between the two microphones was in the "ANR on – speech weighted noise condition", in which the sound pressure level measured using the ATI and

Gentex communication microphones were 90.8 dBA and 93.1 dBA, respectively. Although the difference in the insertion loss between the two microphones was not very large numerically, a 2.3 dB is a considerable amount of attenuation, in that a 3 dB reduction is equivalent to a halving of the sound intensity. In addition to increasing the allowable exposure time, the attenuation afforded by the ATI microphone improves the signal-to-noise (S/N) ratio and increases the exposure time under the VIS headset.

The Occupational Safety and Health Act (OSHA; 1989) states that some type of engineering or administrative control be in place for workers exposed to a 90 dBA noise level for an 8 hour period (OSHA, 1990). The protection afforded by the Gentex and ATI communication microphones (in the noise plus speech condition) reduced the noise to 89.7 dBA and 87.9 dBA, respectively. The corresponding time (in one 24-hour period), under which OSHA allows workers to be exposed to these noise levels is 8.3 for the Gentex microphone and 10.7 for the ATI microphone. Clearly, the ATI communication microphone has the potential to reduce the noise level at the listeners' ear and increase the amount of time that a person is allowed to be exposed to this noise level without being in violation of OSHA guidelines.

Since the technology discussed in this dissertation is intended for the US Army, it is pertinent to mention the military's current noise regulation. The US Army's hearing conservation program (HCP) states that for soldiers in training, combat or nonindustrial scenarios, a solider must wear single hearing protection (regardless of duration) when exposed to noise levels between 85 dBA time-weighted average (TWA) and 103 dBA TWA (Department of Defense, 1998). For sounds between 103 dBA TWA and 108 dBA TWA, double hearing protection must be worn (Department of Defense, 1998). Also, the

HCP states that when the TWA of the sound is below 85 dBA (e.g. 81 dBA TWA), the hearing protection requirement can be waived. There are not any implications that can be drawn from the Army's HCP since duration time is not a factor in the Army's HCP. This is understandable since a soldier inside of a Bradley Fighting Vehicle wearing the VIS headset may be engaged in war for several hours, and in such a condition, human life takes precedence over exposure time. However, if exposure time was an issue and it was different from OSHA, implications may be drawn based on the amount of attenuation afforded by each microphone.

Speech Intelligibility

Speech level main effect. There was a significant difference in the speech intelligibility scores at both speech levels for each microphone. At the high speech level (96 dB), the mean speech intelligibility score (67.5) was significantly higher than at the lower speech level (33.4). This result was expected since one would expect intelligibility to improve as speech level increases.

Speech level by communication microphone interaction. The dependent variable, *speech intelligibility score*, showed a significant difference at the 96 dB speech level, with the Gentex microphone showing an advantage of 12.5. Therefore, the Gentex microphone was able to reduce the masking of the noise, thus decreasing the difficulty of the speech. Also, this difference of 12.5 may indicate that at the high speech level, communication will be understood in most circumstances. However, at 83 dB, there was not a significant difference between the two microphones. This result is similar to a previous result reported by Urquhart, Robinson, and Casali (2001), in which a single factor (communication microphone) speech intelligibility experiment was conducted using a tank noise level of 107.2 dBA. The researchers stated that there was not a significant difference between the Gentex and ATI microphone ($p = 0.08$). Without a detailed knowledge of the internal operation of the VIS system and the prototype ATI microphone, it is not possible to say why this result was obtained. One possibility is that at elevated speech levels, the noise reduction built into the ATI system may have been trying to cancel the speech. Another possibility is that the system was adding its own noise or distortion to speech, making it less intelligible. However, this is only a supposition.

An interesting added fact about this result is that during wartime missions, soldiers rely on communication to transmit information to the various units in their battlespace. For those soldiers riding in tracked vehicles, such as the Bradley Fighting Vehicles, their voices may become elevated in an effort to speak over the noise of the vehicle. In audition, this phenomenon is termed the Lombard Reflex, in which people tend to adjust their speech to prevent the masking effect of the background noise (Robinson and Casali, 2000). Therefore, the low speech level (83 dB) is not of interest for such operations due to the typical manner in which people communicate in a military environment.

Cognitive Performance

Test main effect. As expected, the main effect *Test* was significant. Further examination of this main effect showed that all three tests, *Tower Puzzle* (1582.6), *Number and Words* (551.2), and *Logical Relations* (1079.5), differed significantly from each other. This was expected because each of the measurement scales are different with respect to the cognitive function that each test is expected to measure. Therefore, the CCAB scores each test differently based on a set of programmed algorithms running on the computer which administered and computed the scores for each test (Expanded Complex Cognitive Assessment Battery, 1988). Therefore, the discussion on these three tests will be limited to the results obtained from the analysis.

Speech level by communication microphone interaction. This interaction was analyzed in two ways. First, when the communication microphones were examined at both speech levels (see Figure 26), the results showed that at the lower speech level (83 dB), participants' performance on the cognitive assessment tests were better by 121.6 points when the ATI communication microphone was used. However, at the high speech level (96 dB) participants performed better on the cognitive assessment tests when the Gentex microphone was used for a difference of 111.7. This pattern is similar in direction with that shown earlier for the speech intelligibility dependent variable, in which the ATI microphone had the advantage with respect to the low speech level condition, and the Gentex microphone had the advantage with respect to the high speech level condition (see Figure 23). Secondly, the effect of speech level on the Gentex communication microphone revealed that at the high speech level (96 dB), participants scored significantly higher by 208.2 points than at the lower speech level (83 dB). However, for the ATI communication microphone, this difference between the two scores was not significantly different.

Based on the manner in which this interaction was analyzed, the results may be explained using the context of the multiple resource theory proposed by Wickens (1992). This theory, shown in Figure 32, states that a mentally demanding task requires a larger pool of resources than a simple task, thus limiting the resources that are to be shared by

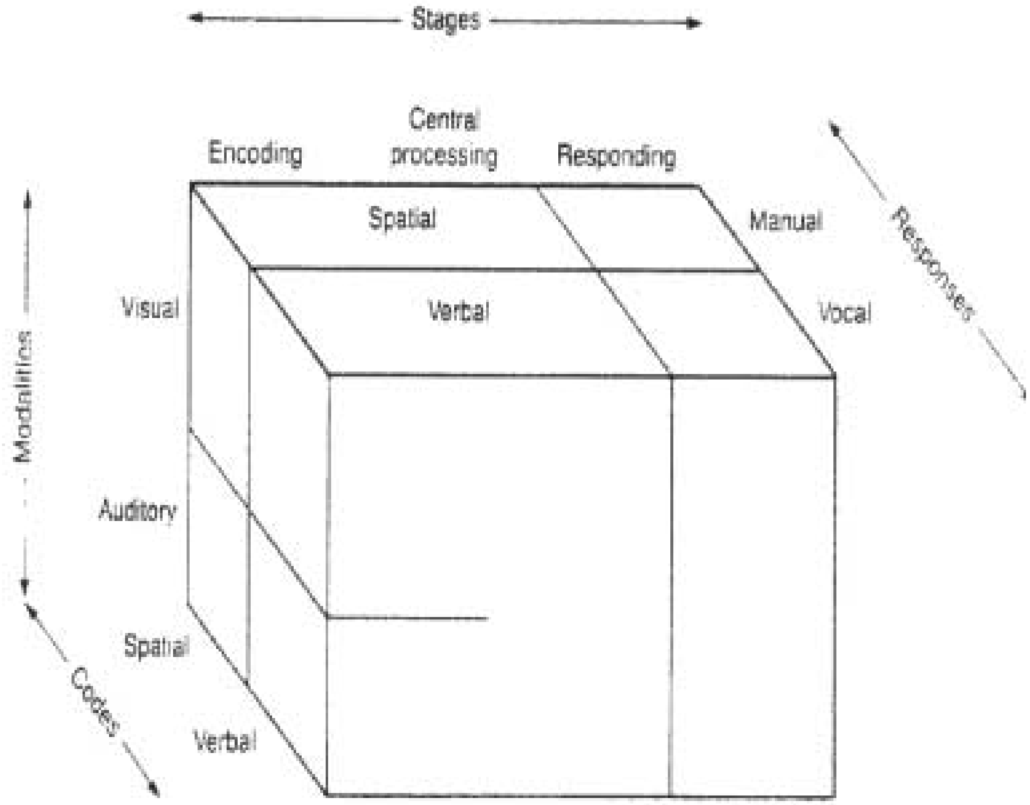


Figure 32. A model of multiple resources (adapted from Wickens, 1992).

other tasks that are performed simultaneously. In addition, when tasks are successfully time-shared, it is presumed that they compete less for the same resources, thus allowing them to be performed concurrently. For instance, at the high speech level (96 dB) with the Gentex microphone, the results showed that participants performed better on the cognitive tests. Additionally, at this same speech level, speech intelligibility scores were higher when the Gentex microphone was used. Thus, when speech is more intelligible, participants were able to devote more attention to the cognitive assessment test, resulting in higher scores.

Additionally, this result was expected, given previous research findings, which stated that performance improves as speech intelligibility improves. For example, Peters and Garinther (1990) conducted a study in which they had tank crews perform missions at nominal levels of intelligibility in an armor simulator. The result of their study is outlined in Table 21. Clearly, the results show that there was an inverse relationship between speech intelligibility level and mission errors, in that as speech intelligibility decreased, mission errors increased. Another interesting fact is that the number of enemy targets which were identified decreased from 98% to 68% as speech intelligibility decreased. In turn, the number of enemy targets “killed” was reduced from 94% to 41%.

Subjective Mental Workload

Since the statistical test failed to detect any differences for the microphones or the speech levels on the workload measures, the discussion on these measures cannot be elaborated upon any further. The possible range of scores for the MCH and the NASA-TLX were 1 to 9 and 0 to 10, respectively. But, for this dissertation, the actual scores for

Table 21. Speech intelligibility levels and mission errors results based on a study the NASA conducted by Peters and Garinther (1990).

Speech intelligibility (%) level	Mission Errors (%)
93.5	7
73.6	13
52.1	14
26.3	23
7.1	28

the NASA-TLX were between 2.1 and 8.3. The actual range of scores for the MCH were between 3.0 and 7.0._ Therefore, it is important to note that there was some mental workload imposed by the combined CCAB and MRT tasks (as was shown in Table 20), ranging from “fair, mild difficulty” to “major difficulty”.

Based on the results from the correlation analysis, there was a positive relationship between the two workload scales. This indicates that when participants rated their perceived mental workload as “high” using the NASA-TLX, they made the same rating for the MCH. Since there exists a positive relationship between the two workload scales, there seems to be evidence that in the future, one scale can be used to glean workload information for the specific tasks that were described herein instead of two workload scales. For this purpose, it is desirable to determine which one may be deemed useful in future research.

Based on the results of this research, the MCH seems to be promising to achieve the intended goal for the reasons discussed herein. The MCH was devised from the Cooper-Harper scale, which was intended to measure tasks in the aviation domain. Because of the Cooper-Harper scale was designed specifically for aircraft handling qualities, Wierwille and Casali (1983) made the MCH applicable to a wider range of tasks, including those associated with aviation emphasis. Also, these authors stated that the MCH is applicable to tasks with a communication emphasis, which is a major facet of this research.

Presently, no research exists that has effectively utilized the MCH and NASA-TLX to assess perceived mental workload. However, there has been research solely comparing the NASA-TLX and SWAT (Colle and Reid, 1998) and comparing the MCH, NASA-TLX, SWAT, and Overall Worklaod (OW) (Hill, Iavecchia, Byers, Bittner, Zaklad, and Christ, 1992). Since the latter study encompasses the two scales of interest, these research results will be briefly discussed. Hill et al. (1992) conducted five experiments comparing these five subjective workload tools. They found that the NASA-TLX was more sensitive and was accepted more by the operators, followed by the OW. Additionally, the NASA took more time to administer and reduce the data because of the dimension which are used to obtain the workload score. The MCH is quicker to administer and takes less data reduction time, because the scale is one-dimensional. The authors suggested that the MCH and OW may be beneficial when it is used as a screening tool to identify potential bottlenecks in workload. To this end, the authors emphasized that the mental workload instrument selected depends on the needs of the study.

Therefore, if the researcher is interested in determining what is causing the mental workload to be high (or low), the NASA-TLX serves as an excellent diagnostic tool in that it gives the researcher a more detailed view of the various components that comprise the overall workload score. Additionally, it allows the researcher to pinpoint what is causing the workload based on the rating on each dimension. However, if the researcher is only interested in determining the overall workload imposed on a person, as was the goal in this dissertation, the researcher should employ the MCH as its workload tool.

Lastly, this point is continuously arguable in the scientific community, but among the most important factors in comparing workload scales are the clarity of the instructions on how to use the scale and ensuring that training was rendered to reduce uncertainty. For this research, each participant received adequate training and the directions were clear and concise in order to reduce ambiguity. Additionally, participants were encouraged to ask questions. The instructions for the NASA-TLX are relatively standard, while those for the MCH may be different for various researchers. The instructions for the MCH used in this research were adapted from those of one of the researchers who devised the scale (Casali, 1982), and it is believed that since each participant was trained in the use of the scale with these instructions, it was administered in a reasonable manner.

CONCLUSION

The results of this dissertation showed that the active noise reduction electronics incorporated in the ATI communication microphone system allowed for a greater reduction in noise at the listener’s ear. However, using the ATI microphone did not offer any advantages in improving speech intelligibility, in that the Gentex microphone produced significantly higher scores (at the high speech level). But, the ATI microphone did show a significant improvement in cognitive performance at the lower speech level. This effect reversed with the higher speech level, in which the Gentex microphone showed a significant improvement in cognitive performance.

Based on the history of ANR, it is clear that low frequency noise can be attenuated with a properly designed ANR system. Therefore, it may be beneficial for ATI to further develop its communication microphone with the intention of increasing the functional bandwidth of attenuation inside their communication microphone to better match the most dominant frequencies of the low-frequency noise. This will provide for greater attenuation of the low-frequency noise, thus causing a greater reduction in noise at the ear and allowing speech to be heard more clearly.

Finally, given the above results, when people need to perform complex tasks and communicate in an environment where intense noise is a concern, the Gentex microphone has the advantage. The reasoning behind this assessment is that people tend to elevate their voices in the presence of noise, thus causing speech to be higher, which was a condition in this research where the Gentex microphone proved to be significantly better. The only real benefit the ATI microphone seems to offer is a moderate reduction of noise at the listener’s ear and an improvement in cognitive performance at the lower speech level. However, this reduction does not provide an improvement in speech intelligibility, as manifested in the MRT scores.

SUGGESTIONS FOR FUTURE RESEARCH

In general, this research is considered to have been successful. Although it was envisioned to have sixteen participants, the experiment concluded at eight. Therefore, future studies should incorporate more participants to reduce variability. In addition, the following are suggestions for future research.

- Conduct a task analysis with experienced military soldiers who operate the Bradley Fighting Vehicle to determine the tasks that are performed and try to implement those tasks in a noise environment similar to that used in this dissertation.
- A simulation study using the Bradley Fighting Vehicle with experienced military tank operators in 114 dBA noise performing the cognitive assessment tests. The Bradley Fighting Vehicle would be stationary, but placed on a platform that simulates the vehicle being driven on terrain similar to that during wartime.
- It would be beneficial to have the speech intelligibility portion of the experiment include commands that elicit a response on the cognitive task. For instance, the speech intelligibility task could give the participant a cognitive job to perform. Thereby, this would allow the two tasks to be more dependent on each other. The result of such an experiment would probably yield the similar results, but it would be more representative of command and control tasks performed in the military.
- Conduct an experiment with and without a noise condition to discover the extent to which noise contributes to mental workload.

- Determine what is the male to female ratio in the military and recruit a subject pool that is identical to that population. Next, conduct a study similar to the one in this dissertation, with the effect of gender being investigated to determine if there are differences in how males and females respond to speech and complex cognitive tasks.
- Conduct an experiment in which “speech rate” is a factor to determine how “slow or fast” speech may affect intelligibility and performance on complex cognitive tasks.
- Lastly, future research should be conducted to further develop and test ANR headsets.

REFERENCES

American National Standards Institute (ANSI). (1989). Method for measuring the intelligibility of speech over communication systems, ANSI S3.2-1989. New York: American National Standards Institute, Inc.

Anticaglia, J. R. and Cohen, A. (1970). Extra-auditory effects of noise as a health hazard. American Industrial Hygiene Association Journal, 31(3), 277-281.

Arlinger, S. and Gustafsson, H. A. (1991). Masking of speech by amplitude-modulated noise. Journal of the Acoustical Society of America, 151(3), 441-445.

Bailey, R.W. (1996). Human performance engineering, 3rd Ed., Upper Saddle River, NJ: Prentice Hall.

Boyd, S. P. (1983). Assessing the validity of SWAT as a workload measurement instrument. In Proceedings of the Human Factors Society Annual Meeting (pp. 124-128). Santa Monica, CA: Human Factors Society.

Broadbent, D. M. (1979). Human performance and noise. In C.S. Harris (Ed.). Handbook of noise control, pp. 2066-2085. New York, NY: McGraw Hill.

Casali, J. G. (1982). A sensitivity/intrusion comparison of mental workload estimation techniques using a simulated flight task emphasizing perceptual piloting behaviors. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Casali, J. G. (1992). Technology advancements in hearing protection: Active noise reduction, frequency/amplitude-sensitivity, and uniform attenuation. In Proceedings of the Human Factors Society 36th Annual Meeting (pp. 258-262). Santa Monica, CA: Human Factors Society.

Casali, J. G. and 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. and Robinson, G. S. (1994). Narrow-band digital active noise reduction in a siren-canceling headset: real-ear and acoustical manikin insertion loss. Noise Control Engineering Journal, 42(3), 101-115.

Casali, J.G., Robinson, G.S., Dabney, E.C., and Gauger, D. (2000). Effect of active noise reduction and passive hearing protection on detectability of a vehicle backup alarm in noise. Noise-Con 2000, pp. 1-6. Newport Beach, CA.

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

Charlton, S. G. (1996). Mental workload test and evaluation. In T. G. O'Brien and S. G. Charlton, Handbook of human factors testing and evaluation. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cohen, A. (1976). The influence of a company hearing conservation program on extra-auditory problems in workers. Journal of Safety Research, 8,4,146-162.

Colle, H.A. and Reid, G.B. (1998). Context effects in subjective mental workload ratings. Human Factors, 40(4), 591-600.

Crandall, I. B. (1917). The composition of speech. Physics Review, 10(2), 74-76.

Davies, D. and Jones, D. (1982). Hearing and noise. In W. Singleton (Ed.), The body at work. New York: Cambridge University Press.

DeJoy, D. M. (1984). The nonauditory effects of noise: Review and perspectives for research. Journal of Auditory Research, 24, 123-150.

Department of Defense (1998). Hearing conservation program, PAM 40-501, Department of the Army, Washington, DC.

Department of Defense (1997). Staff organizations and operations, FM 101-5, Department of the Army, Washington, DC.

Department of Defense (1996). FY1995 Annual Report. Department of Defense, Washington, DC.

Draper, M. H., Ruff, H. A., and LaFleur, T. (2001). The effects of camera control and display configuration on teleoperated target search tasks. In Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting (pp. 1872-1876). Santa Monica, CA: Human Factors Society.

Driskell, J. E. and Salas, E. (1996). Stress and human performance. Mahwah, NJ: Lawrence Erlbaum.

Earshen, J.J. (1986). Instrumentation and noise descriptors. In E.H. Berger, D. Ward, J. Morrill, and L. Royster (Eds.), Noise and hearing conservation manual (pp. 37-95). Fairfax, VA: American Industrial Hygiene Association.

Egan, J. P. (1948). Articulation testing methods. Laryngoscope, 58(9), 955-991.

Ellis, L., Reynolds, L., Fucci, D., Benjamin, B. (1996). Effects of gender on listerners' judgments of speech intelligibility. Perceptual and Motor Skills, 83(3), 771-775.

Endsley, M. R., and Smith, R. P. (1996). Attention distribution and decision-making in tactical air combat. Human Factors, 38(2), 232-249.

Eschenbrenner, A. J., Jr. (1971). Effects of intermittent noise on the performance of a complex psychomotor task. Human Factors, 13(1), 59-63.

Essens, P., Post, W. M., and Rasker, P.C. (2000). Modeling a command center. In J.M. Schraagen, S.F. Chipman, and V.L. Shalin (Eds.) Task Analysis. Mahwah, N.J.: Lawrence Erlbaum Associates.

Expanded Complex Cognitive Assessment Battery (1988). Alexandria, VA: U.S. Army Research Institute.

Fairbanks, G. (1958). Test of phonemic differentiation: The rhyme test. Journal of the Acoustical Society of America, 30(7), 596-600.

Fernandez, E. and Sheffield, J. (1996). Descriptive features and causal attributions of headache in an Australian community. Headache, 36(4), 246-250.

Finch, M. I. and Stedmon, A. W. (1998). The complexities of stress in the operational military environment. In M.A. Hanson (Ed.), Contemporary Ergonomics (pp. 388-392). Bristol, PA: Taylor and Francis.

French, N. R. and Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. Journal of the Acoustical Society of America, 19(1), 90-119.

Garinther, G. and Peters, L. (1990). Impact of communications on armor crew performance. Army Research, Development and Acquisition Bulletin, 1-5.

Garinther, G. R., Whitaker, L. A. and Peters, L. J. (1994). The effects of speech intelligibility on military performance. In Proceedings of Symposium on speech intelligibility (pp.1-12). Washington, DC: National Academy of Science.

Gauger, D. and Sapiejewski, R. (1987). Voyager pilots avoid hearing loss on historic flight. Journal of Sound and Vibration, 21(5), 10-12.

Goldstein, E. B. (1989). Sensation and perception (3rd ed.). Belmont, CA: Wadsworth.

Gomes, L.P., Pimenta, A.F., Branco, N.A. (1999). Effects of occupational exposure to low frequency noise on cognition. Aviation, Space, and Environmental Medicine, 70(3), A115-A118.

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

Graziano, A. M. and Raulin, M. L. (1997). Research methods, 3rd Ed. New York, NY: Longman.

Griefahn, B. and DiNisi, J. (1992). Mood and cardiovascular functions during noise, related to sensitivity, type of noise, and sound pressure level. Journal of Sound and Vibration, 155(1), 111-123.

Hart, S. and Staveland, L. (1988). Development of NASA-TLX: Results of empirical and theoretical research. In P. Hancock and N. Meshkati (Eds.), Human mental workload. Amsterdam: North-Holland.

Hatcher, L. and Stepanski, E.J. (1994). A step-by-step approach to using the SAS system for univariate and multivariate statistics. Cary, NC: SAS Institute.

Hawley, M. E. (1977). Speech intelligibility and speaker recognition (ed.). Stroudsburg, PA: Dowden, Hutchinson & Ross, Inc.

Hendy, K. C., Hamilton, K. M., and Landry, L. N. (1993). Measuring subjective workload: When is one scale better than many? Human Factors, 35(4), 579-601.

Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Zaklad,A.L. and Christ, R.E. (1992). Comparison of four subjective workload rating scales. Human Factors, 34(4), 429-439.

Hill, S.G., Zaklad, A.L., Bittner, A.C., Byers, J.C., and Christ, R.E. (1988). Workload assessment of a mobile air defense missile system. In Proceedings of the Human Factors Society (pp. 1068-1072). Santa Monica, CA: Human Factors Society.

Hochhaus, L. and Antes, J. R. (1973). Speech identification and “knowing that you know.” Journal of Perception and Psychophysiology, 13, 131-132.

House, A. S., Williams, C. E., Hecker, M. H., and Kryter, K. D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. Journal of the Acoustical Society of America, 37(1), 158-166.

Janis, I. L. (1982). Decision-making under stress. In L. Goldberger and S. Breznitz (Eds.), Handbook of stress: Theoretical and clinical aspects (pp. 69-87). New York: Free Press.

Johnson, D.E. (1998). Applied multivariate methods for data analysis. Pacific Grove, CA: Duxbury Press.

Jones, D. M. (1983). Noise. In G.R.J. Hockey (Ed.), Stress and Fatigue in Human Performance (pp. 61-95). Chichester, UK: Wiley and Sons.

Kaempf, G. L., Klein, G., Thordsen, M. L., and Wolf, S. (1996). Decision-making in complex naval command and control environments. Human Factors, 38(2), 220-231.

Kane, R. L. and Kay, G. G. (1992). Computerized assessment in neuropsychology: A review of tests and test batteries. Neuropsychology Review, 3(1), 1-115.

Kantowitz, B. (2000). Attention and mental workload. In Proceedings of the International Ergonomics Association/Human Factors and Ergonomics Society (p. 456-459). Santa Monica, CA: Human Factors Society.

Kryter, K. D. (1994). The handbook of hearing and the effects of noise. San Diego, CA: Academic Press.

Kryter, K. D. (1985). The effects of noise on man (2nd Ed.). Orlando, Fl: Academic Press.

Lehto, M. (1997). Decision making. In G. Salvendy (Ed.), Handbook of human factors (pp. 1201-1248). New York, New York: John Wiley and Sons.

Locke, J. L. (1979). Auditory discrimination test. In Darley (Ed.), Evaluation of Appraisal Techniques in Speech and Language Pathology (pp. 124-127). Reading, MA: Addison-Wesley.

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., Linton, P. M., Wierwille, W. W., Zaklad, A. L., Bittner, A. C., and Wherry, R. J. (1989). Operator workload: Comprehensive review and evaluation of operator workload methodologies (Tech. Rep. 851). Alexandria, VA: U.S. Army Research Institute.

Machamer, R. (1995, April). FORCE XXI- Welcome to the 21st Century. Soldiers. p.36.

Matthews, G. Campbell, S. E.. and Falconer, S. (2001). Assessment of motivational states in performance environments. In In Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting (pp. 906-910). Santa Monica, CA: Human Factors Society.

Matthews, M.D. and Shattuck, L.G. (2001). Situation awareness for military ground forces: Current issues and perspectives. In Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting (pp. 351-355). Santa Monica, CA: Human Factors Society.

Meeker, W. F. (1957). Active ear defender systems, development of a laboratory model, part II (WADC TR 57-368). Dayton, OH: Wright-Patterson Air Force Base.

Meshkati, N., Hancock, P. A., Rahimi, M. and Dawes, S. M. (1995). Techniques in mental workload assessment. In J. R. Wilson and E. N. Corlett (Eds.), Evaluation of human work, 2nd Ed. Bristol, PA: Taylor and Francis.

Miller, J. D. (1974). Effects of noise on people. Journal of the Acoustical Society of America, 56(3), 729-764.

Moore, A. (1998, August). VIS helmet reduces noise. Soldiers. p. 10.

Morrison, H. B. and Casali, J. G. (1997). Intelligibility of synthesized voice messages in commercial truck cab noise for normal-hearing and hearing impaired listeners. International Journal of Speech Technology, 2, 33-44.

Nixon, C. W., McKinley, R. L., and Steuver, J. W. (1992). Performance of active noise reduction headsets. In A. L. Dancer, D. Henderson, R. J. Salvi, and R. P. Hamernik (Eds.), Noise-induced hearing loss (pp. 389-400). St. Louis: Mosby-Year Book.

Noyes, J.M., Baber, C., Leggatt, A.P. (2000). Automatic speech recognition, noise and workload. . In Proceedings of the International Ergonomics Association/Human Factors and Ergonomics Society (p. 762-765). Santa Monica, CA: Human Factors Society.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. Human Factors, 33(4), 17-33.

Occupational Safety and Health Act (1989). Occupational noise exposure. Code of Federal Regulations, 29 Part 1910.95 (pp. 177-194). Washington, D.C.: U.S. Government Printing Office.

Ohrstrom, E. and Bjorkman, M. (1978). Medical symptoms in noisy industries. Journal of Sound and Vibration, 59, 115-118.

O'Donnell, R. D. and Eggemeier, F. T. (1986). Workload assessment methodology. In K.R. Boff, L. Kaufman, and J. Thomas (Eds.), Handbook of perception and performance, Vol. 2: Cognitive processes and performance. (pp. 42.1-42.49). New York: Wiley.

O'Hare, D. (1997). Cognitive ability determinants of elite pilot performance. Human Factors, 39(4), 540-552.

OSHA (1990). Occupational safety and health standards: Occupational noise exposure (29 CFR, Part 1910.95, pp. 204-218). Washinton, D.C.: Office of the Federal Register.

Park, S. H.and Woldstad, J.C. (2000). Multiple two-dimensional displays as an alternative to three-dimensional displays in telerobotic tasks. Human Factors, 42(4), 592-603.

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

Pollack, I. (1958). Speech intelligibility at high noise levels: Effect of short-term exposure. Journal of the Acoustical Society of America, 30(6), 282-285.

Proctor, R.W. and Zandt, R. (1994). Human factors in simple and complex systems. Needham Heights, MA: Allyn and Bacon.

Reid, G. B., Shingledecker, C. A. and Effemeier, F. T. (1981). Application of conjoint measurement to workload scale development. In Proceedings of the Human Factors Society Annual Meeting (pp. 552-526). Santa Monica, CA: Human Factors Society.

Richards, A. M. and Goodman, A. C. (1977). Threshold of the human acoustic stapedius reflex for short-duration burst of noise. Journal of Auditory Research, 13, 183-189.

Robinson, G. and Casali, J. (2000). Speech communications and signal detection in noise. In E.H. Berger, L.H. Royster, J.D. Royster, D.P. Driscoll, and M. Layne (Eds.), The noise manual, 5th Ed. Fairfax, VA: American Industrial Hygiene Association.

Rosen, S. (1970). Noise, hearing, and cardiovascular function. In B.L. Welch and A.S. Welch (Eds.), Physiological effects of noise. New York, NY: Plenum Press.

Sanders, M. S. and McCormick, E. J. (1987). Human factors in engineering and design, 6th Ed. New York, NY: McGraw-Hill Publishing Company.

Sanders, M. S. and McCormick, E. J. (1993). Human factors in engineering and design, 7th Ed. New York, NY: McGraw-Hill Publishing Company.

Smith, A. (1989). A review of the effects of noise on human performance. Scandinavian Journal of Psychology, 30, 185-206.

Smith, A. P. and Miles, C. (1985). The combined effects of noise and nightwork on human function. In D. Oborne (Ed.), Contemporary Ergonomics, pp. 33-41. London: Taylor and Francis

Smith-Jackson, T. L. (1998). Irrelevant speech, verbal task performance, and focused attention : a laboratory examination of the performance of open-plan offices. Unpublished doctoral dissertation, North Carolina State University, Raleigh, NC.

Stark, J.M., Scerbo, M.W., Freeman, F.G., Mikulka, P.J. (2000). Mental fatigue and workload: Effort allocation during multiple task performance. In Proceedings of the Human Factors Society Annual Meeting (pp. 863-866). Santa Monica, CA: Human Factors Society.

Statistical Analysis Software (1999). Cary, NC: SAS Institute

Stephenson, M. R., Billings, B. L., and Jutila, G. A. (1990). Voyager flight crew hearing threshold levels resulting from 5- and 9-day continuous in-flight noise exposure. Journal of the Acoustical Society of America, 87(2), 901-904.

Stevens, S. S., Miller, J., and Truscott, I. (1946). The masking of speech by sine waves, square waves, and regular and modulated pulses. Journal of the Acoustical Society of America, 18(2), 418-424.

Suter, A. H. (1992). Noise sources and effects-A new look. Sound and Vibration, 26(1), 18-38.

Talbott, E., Helmkamp, J., Matthews, K., Kuller, L., Cottington, C., and Redmond, G. (1985). Occupational noise exposure, noise-induced hearing loss, and the epidemiology of high blood pressure. American Journal of Epidemiology, 121(4), 501-514.

Taylor, B. H, Charlton, S. G. and Canham, L. S. (1996). Operability testing in command and control systems. In T.G. O'Brien and S.G. Charlton, Handbook of human factors testing and evaluation. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Theologus, G. C., Wheaton, G. R. and Fleishman, E. A. (1974). Effects of intermittent, moderate intensity noise stress on human performance. Journal of Applied Psychology, 59(5), 539-547.

Thumann, A., and Miller, R. K. (1990). Fundamentals of noise control engineering, 2nd Ed. Englewood Cliffs, N.J.: Prentice Hall.

Tokhi, M. O. and Leitch, R. R. (1992). Active noise control. New York, NY: Oxford Science Publications.

Tsang, P. and Wilson, G. F. (1997). Mental workload. In G. Salvendy (Ed.). Handbook of human factors, 2nd Ed. (pp. 417-449). New York, NY: John Wiley and Sons.

Urquhart, R. L., Robinson, G. S. and Casali, J. G. (2001). The measurement of speech intelligibility and noise transmission for two microphones. (ISE Technical Report No. 20004; Audio Lab No. 5/15/00-4-HP). Blacksburg, VA: Virginia Polytechnic Institute and State University, Industrial and Systems Engineering Department.

Van Gemmert and Van Galen (1997). Stress, neuromotor noise, and human performance: A theoretical perspective. Journal of Experimental Psychology, 23(5), 1299-1313.

Ward, W. D. (1986). Anatomy and physiology. In E. H. Berger, W.D. Ward, J.C. Morrill, and L.H. Royster(Eds.). Noise and hearing conservation manual, 4th Ed. Akron, Ohio: American Industrial Hygiene Association.

Weinstein, N. D. (1974). Effect of noise on intellectual performance. Journal of Applied Psychology, 59(5), 548-554.

Wheeler, P. D. and Halliday, S. G. (1981). An active noise reduction system for aircrew helmets. In Proceedings of NATO-AGARD CP-311 (pp. 22-1 to 22-8). Neuilly sur Sein, France: AGARD.

Whitaker, L. A., Peters, L. J., and Garinther, G. A. (1990). Effects of communication degradation on military crew task performance. In Proceedings of Symposium on Psychology in the Department of Defense (pp. 274-278). Colorado Springs, Colo. U.S. Air Force Academy.

Whitaker, L. A (1991). Performance as a function of communication in military vehicle simulators. In Proceedings of the Human Factors Society Annual Meeting (pp. 614-617). Santa Monica, CA: Human Factors Society.

Whitaker, L. A. and Peters, L. J. (1993). Communication between crews: The effects of speech intelligibility on team performance. In Proceedings of the Human Factors Society Annual Meeting (pp. 630-634). Santa Monica, CA: Human Factors Society.

Whitaker, L. A. and Peters, L. J. (1990). Effects of speech intelligibility among Bradley fighting vehicle crew members: SIMNET performance and subjective workload. In Proceedings of the Human Factors Society Annual Meeting (pp. 186-188). Santa Monica, CA: Human Factors Society.

Whitaker, L. A., Peters, L. J., and Garinther, G. A. (1990). Effects of communication degradation on military crew task performance. In Proceedings of Symposium on Psychology in the Department of Defense (pp. 274-278). Colorado Springs, Colo. U.S. Air Force Academy.

Whitaker, L., Peters, L. and Garinther, G. (1989). Tank crew performance: Effects of speech intelligibility on target acquisition and subjective workload assessment. In Proceedings of the Human Factors Society Annual Meeting (pp. 1411-1413). Santa Monica, CA: Human Factors Society.

Wickens, C. D. (1992). Engineering psychology and human performance, 2nd Ed. New York, NY: Harper Collins.

Wickens, C. D., Gordon, S. E. and Liu, Y. (1998). An introduction to human factors engineering. New York, NY: Addison-Wesley.

Wierwille, W. W. and Casali, J. G. (1983). A validated rating scale for global mental workload measurement applications. In Proceedings of the Human Factors Society 27th Annual Conference, (pp. 129-133). Santa Monica, CA: Human Factors Society.

Wierwille, W. W. and Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. Human Factors, 35(4), 263-281.

Williges, R. C. and Wierwille, W.W. (1979). Behavioral measures of aircrew mental workload. Human Factors, 21(5), 549-574.

Winer, B.J., Brown, D.R., and Michels, K.M. (1991). Statistical principles in experimental design, New York, NY: McGraw-Hill.

Zsambok, C. E. (1997). Naturalistic decision-making: Where are we now? In C.E. Asambok and G. Klein (Eds.), Naturalistic decision making (pp. 3-16). Mahwah, NJ: Erlbaum.

Appendix A

Instructions for completing the NASA-TLX
(from Hart and Staveland, 1988; Smith-Jackson, 1998)

Instructions for the NASA-TLX (rating scales)

The evaluation you are about to perform is a technique that has been developed by NASA to assess the relative importance of six factors in determining how much workload you experienced Six rating scales were developed for you to use in evaluating your experiences during the different conditions. Before you complete the six rating scales, please read the descriptions of the scales carefully. If you have a question about any of the scales in the table, please ask me about it. It is extremely important that they be clear to you. You may keep the descriptions with you for reference during the experiment.

You will evaluate the task by putting an "X" on each of the six scales at the point that matches your experience. Each line has two endpoint descriptors that describe the scale. Please consider your response carefully and try to use the scales to distinguish among the different task conditions. Consider each scale individually. The ratings you will give us will play an important role in the evaluation being conducted.

Thank you.

Instructions for the NASA-TLX (sources of workload evaluation)

You will be presented with a series of pairs of rating scale titles (for example, EFFORT vs MENTAL DEMANDS). Please choose which of the items within each pair was more important to your experience of workload in the task that you just performed. Each pair of scale titles will appear on a separate card. Please circle your choice within each pair on the card using the pencil provided.

Please consider your choices carefully and make them consistent with how you used the rating scales during the particular task you were asked to evaluate. Don’t think that there is any correct pattern; we are only interested in your opinion. If you have any questions, please ask them now. Otherwise, you may begin.

NASA TLX (Continued)

Please consider your choices carefully and make them consistent with how you used the rating scales during the particular task you were asked to evaluate. Don’t think that there is any correct pattern. We are only interested in opinions. If you have any questions, please ask them now. Otherwise, you may begin.

SCALE TITLES FOR THE NASA-TLX

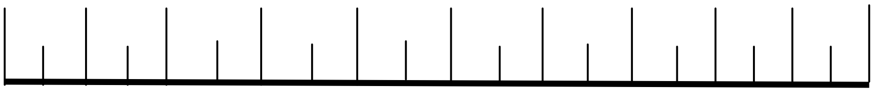
TITLE	ENDPOINTS	DESCRIPTIONS
Mental Demand	Low/High	How much mental and perceptual activity is required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low/High	How much physical activity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

NASA TLX (Continued)

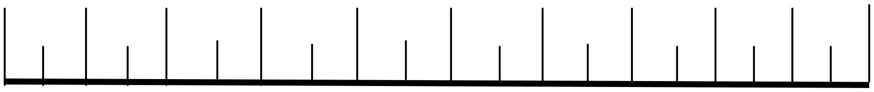
NASA TLX RATING SHEET

Subject # _________ Condition # ________

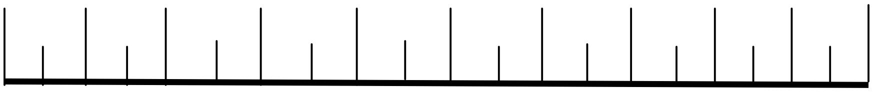
MENTAL DEMAND



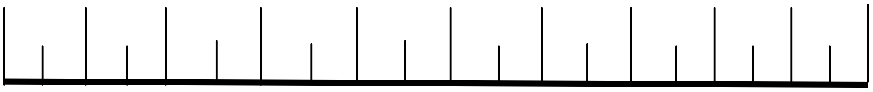
PHYSICAL DEMAND



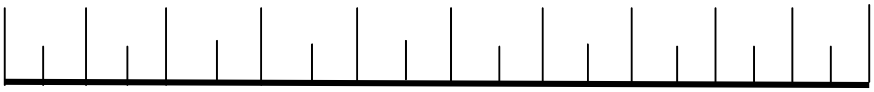
TEMPORAL DEMAND



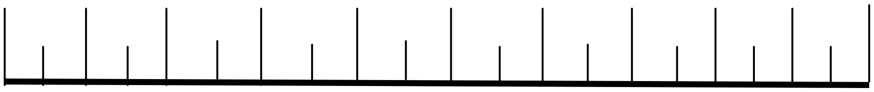
PERFORMANCE



EFFORT



FRUSTRATION



NASA TLX (continued)

Subject #: _______________

Sources of Workload

Task #: ______ Document: ______ Condition: _______

Scale title	Tally	Weight	Raw rating	Adjusted Rating (Weight X Raw)
Mental Demand				
Physical Demand				
Temporal Demand				
Performance				
Effort				
Frustration				

Total Count = ____ (no weight can be greater than 5; total count can’t be > 15).

Write the sum of the adjusted rating column _______. Weighted rating = Sum of adjusted ratings/15

Appendix B

Instructions for completing the Modified Cooper-Harper
(from Wierwille and Casali, 1983; Casali, 1982)

Overview

After each experimental session, you will be asked to give a rating on the Modified Cooper-Harper scale for workload. This rating scale and important definitions for using the scale are shown on the sample, which I have given you. Before you begin, we will review:

1. The definition of the terms used in the scale,
2. The steps you should follow in making your ratings on the scale, and
3. How you should think of the ratings.

If you have any questions as we review these points, please ask me.

Important Definitions

To understand and use the Modified Cooper-Harper scale properly, it is important that you understand the terms used on the scale and how they apply in the context of this experiment.

First, the primary task is the Complex Cognitive Assessment Battery and the Speech Intelligibility test you have been assigned to perform in this experiment. It includes performing both of these tasks simultaneously. Second, the operator in this experiment is you. Because the scale can be used in different situations, the person performing the ratings is called an operator. You will be operating the system and then completing the rating scale to quantify your experience. Third, the system is the complete group of equipment you will be using in performing the primary task. Together you and the system make up the operator/system. Fourth, errors include mistakes and incorrect actions or responses. Finally, mental workload is the integrated mental effort required to perform the primary task. It includes such factors as level of attention, depth of thinking, and level of concentration required by the primary task.

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.

The steps that you will follow in using the rating scale logic are as follows:

1. First, you will decide if the primary task can be accomplished most of the time; if not, then your rating is a 10 and you should circle the 10 on the rating scale.
2. Second, you will decide if adequate performance is attainable. Adequate performance means that errors are small and inconsequential in performing the primary task. By reading the descriptors associated with the numbers 7, 8, and 9, you should be able to select the one that best describes the situation you have experienced. You would then circle the most appropriate number.
3. If adequate performance <u>is</u> attainable, your next decision is whether or not your mental workload for the primary task is tolerable. If it is not tolerable, you should select a rating of 4, 5, or 6. One of these three ratings should describe the situation you have experienced, and you would circle the most appropriate number.
4. If mental workload is tolerable, you should then move to one of the top three descriptors on the scale. You would read and carefully select the rating 1, 2, or 3

based on the corresponding description that best describes the situation you have experienced. You would circle the most appropriate number.

Remember you are to circle only one number, and the number should 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 and reliable.

How You Should Think of the Ratings

Before you begin making ratings, there are several points that need to be emphasized. First, be sure to try and perform the primary task as instructed and make all your evaluations within the context of the primary task. Try to maintain adequate performance as specified for your task.

Second, the rating scale is not a test of your personal skill. On all of your ratings, you will be evaluating the system for a general user population, not yourself. You may assume you are an experienced member of that population. You should make the assumption that problems you encounter are not problems you created. They are problems created by this system and the instructed primary task. In other words, don’t blame yourself if the system is deficient; blame the system.

Finally, always try to “tell it like it is” in making your ratings. If you have any questions, please ask me at this time.

APPENDIX C

Modified Rhyme Test Practice Session Word List

Track 1	Track 2	Track 3	Track 4	Track 5	Track 6
sent	bent	bent	tent	rent	rent
told	hold	hold	cold	gold	hold
pass	pat	pad	pat	pat	path
lake	lay	lame	lame	late	lay
bit	wit	sit	wit	hit	hit
rust	rust	just	just	just	rust
teal	tease	teal	tear	tease	teach
dip	dill	did	dill	did	dim
shed	shed	led	shed	led	led
pin	fin	tin	din	din	sin
duck	dub	dug	duck	dub	dud
sud	sun	sud	sun	sum	sum
seen	seep	seek	seep	seek	seen
lot	not	tot	lot	not	not
west	test	west	west	best	best
pin	pit	pick	pill	pip	pin
back	bat	bat	back	back	bath
day	may	day	say	gay	may
pig	fig	pig	big	rig	big
pale	pace	pay	pave	pay	pave
case	cape	came	cane	came	cane
top	shop	hop	shop	pop	hop
toil	foil	oil	oil	toil	foil
tab	tap	tap	tack	tam	tap
fit	fill	fill	fizz	fill	fib
same	fame	fame	tame	game	fame
eel	peel	keel	keel	keel	eel
bark	mark	park	dark	dark	lark
heap	hear	heath	heat	heave	heal
cuss	cuss	cuss	cuss	cut	cut
paw	thaw	law	paw	saw	thaw
hen	men	den	then	hen	men
pup	puck	pun	pus	pun	puff
beak	bead	beam	bead	beak	beam
beat	feat	seat	beat	neat	seat
rip	rip	sip	hip	sip	hip
kit	kin	kit	king	kick	king
gang	rang	fang	rang	sang	gang
hook	book	hook	cook	took	hook
mad	man	map	man	mass	mad
raze	race	raze	rate	rate	ray
safe	same	sale	save	save	safe
bill	till	hill	fill	till	kill
sit	sip	sip	sit	sick	sip
gale	pale	gale	pale	tale	pale
sick	tick	tick	tick	sick	tick
peat	peach	peas	peach	peat	peat
but	bus	bug	but	buck	but
sat	sass	sack	sass	sack	sad
bun	fun	bun	gun	run	gun

APPENDIX D

Modified Rhyme Test
Experimental Session Word List

Track 1	Track 2	Track 3	Track 4	Track 5	Track 6
tent	rent	dent	sent	bent	bent
hold	sold	gold	cold	cold	told
pat	path	pass	pan	pad	pat
late	lay	lame	lay	lame	lake
wit	bit	hit	wit	kit	sit
dust	rust	just	rust	gust	bust
tease	teach	teal	tear	teak	teal
did	dip	dill	dim	dill	did
wed	led	shed	wed	shed	red
fin	din	tin	sin	pin	fin
dub	dung	duck	dud	duck	dug
sun	sum	sud	sum	sun	sud
seek	seen	seed	seen	seep	seep
not	tot	lot	tot	lot	lot
best	west	test	nest	vest	vest
pill	pin	pit	pip	pill	pick
bath	bat	back	bath	bad	back
may	day	gay	pay	say	gay
rig	big	dig	rig	fig	pig
pave	pay	page	pace	pale	pave
came	cape	cane	came	case	cane
shop	mop	hop	top	shop	pop
oil	foil	toil	coil	oil	coil
tang	tam	tab	tap	tang	tack
fizz	fig	fit	fib	fill	fizz
name	same	fame	name	tame	game
peel	keel	heel	eel	feel	heel
dark	lark	bark	park	dark	mark
hear	heap	heath	heave	heal	heat
cut	cud	cuss	cut	cuff	cud
law	jaw	paw	thaw	paw	saw
pen	den	men	hen	then	pen
pus	pup	puck	pun	puff	pus
beak	bead	beam	beat	bead	beak
beat	neat	seat	feat	neat	seat
sip	lip	rip	hip	dip	lip
kin	king	kit	kick	kid	king
sang	fang	bang	gang	rang	fang
took	cook	book	hook	book	shook
math	mass	mad	map	man	math
rate	raze	race	rate	rave	ray
save	same	safe	save	same	sale
kill	till	hill	bill	fill	hill
sick	sip	sin	sip	sit	sick
gale	pale	tale	pale	tale	gale
pick	wick	tick	sick	pick	pick
peach	peat	peas	peal	peach	peal
bus	but	bug	bus	buff	bus
sass	sat	sap	sack	sad	sap
run	gun	bun	run	fun	run

APPENDIX E

Participant’s Informed Consent

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
GRADO DEPARTMENT OF INDUSTRIAL AND SYSTEMS ENGINEERING
AUDITORY SYSTEMS LABORATORY

Informed Consent for Participants of Investigative Projects

Title of Project: **Evaluation of an Improved Active Noise Reduction Microphone using Speech Intelligibility and Performance-Based Testing**

Principal Investigators: Dr. J. G. Casali, Grado Professor and Head, ISE
Dr. G. S. Robinson, Research Associate Professor, ISE
Ryan Urquhart, M.S., Graduate Research Assistant

Faculty Advisor: Dr. J. G. Casali, Grado Professor and Head, ISE

I. THE PURPOSE OF THIS RESEARCH

The purpose of the tests described herein is to determine what effect two different noise-canceling microphones have on speech intelligibility and cognitive performance at high noise levels at different speech-to-noise ratios. The noise used in this experiment was recorded in a U.S. Army Bradley Fighting Vehicle.

II. PROCEDURES

The procedures for the experiment are as follows. If you wish to become a participant after reading the description of the study, then sign this form. If you have any questions about the study or this form, please feel free to ask them at any time.

First, you will be screened to determine if you qualify for the experiment. Screening will consist of a hearing test and several assessment tests. You will be asked several questions to assess the general health and condition of your ears. You will then be given an examination in which the experimenter will look into your ears using an otoscope. Next, the hearing in your right and left ears will be tested with very quiet tones played through headphones. You will have to be very attentive and listen carefully for these tones. **Depress the button on the hand-held switch and hold it down whenever you hear the pulsed tones and release it when you do not hear the tones.** The tones will be very faint and you will have to listen carefully to hear them.

If you qualify and choose to participate in the study, you will be instructed in and allowed to practice the Modified Rhyme Test (MRT) and the Complex Cognitive Assessment Battery (CCAB). These two tests will be practiced separately and then simultaneously. You will wear a U.S. Army VIS headset during these tests. The combined screening and practice session will last approximately 90 minutes.

For the MRT test, six words will be displayed on a touch screen monitor, with each word enclosed in a rectangular box. One of the six words on the screen will also be presented aurally over the headset within the carrier sentence, "Mark the word ____ now." Your task is to indicate which of the six words was spoken by selecting it on the touch screen. A total of 50 such sentences will be presented during the test. Background noise will be presented through separate loudspeakers in the room. The presence of the background noise will make the words difficult to hear; therefore, you should concentrate and listen carefully to the words. At least two such practice trials will be conducted. After the practice trials have been completed, the experimenter will examine your data and determine if additional practice trials are necessary.

For the CCAB, you will perform three separate tests (Tower Puzzle, Numbers and Words, and Logical Relations). Before practicing each of the three tests, the experimenter will first explain the purpose of the test to you. Next, a set of instructions pertaining to the specific test to be practiced will be presented on the computer screen in front of you. You will be asked to read the instructions aloud. After reading the instructions, a computer-generated practice trial will be initiated. After this trial is completed, a series of 10 TRUE/FALSE questions will appear on the computer asking about the test. You will practice each of the three tests at least twice.

Finally, you will practice performing the MRT and CCAB tests simultaneously. The practice trials will end when both you and the experimenter are satisfied with your ability to perform both tests at the same time. At the completion of practicing the MRT and CCAB, you will be asked to complete two subjective workload scales: the Modified Cooper-Harper (MCH) and the National Aeronautical Space Administration Task Load Index (NASA TLX). The experimenter will explain both of the workload measurement tools to you and then you will be given a set of instructions about each scale before you complete them. The combined screening/practice session should last about 90 minutes.

In the four experimental sessions (conducted on different days), you will again practice both the MRT and CCAB tests individually and in combination. After the practice trials, you will perform the experimental trials just as you did the practice trials, with three exceptions: the background noise will be louder, there will be no computer-based instructions presented immediately prior to the tests, and there will be no TRUE/FALSE questions at the end of the individual trials. It is important that you perform these tests to the best of your ability. Prior to the beginning of each experimental session, a pre-test audiogram will be conducted at the following selected frequencies: 125, 500, 1000, 2000, and 4000 Hz. After completing the experiment and the two workload assessment forms, a post-test audiogram will be conducted to ensure that you did not experience a temporary threshold shift as a result of your participation.

III. RISK

The background noise level in the room during the experimental sessions will be 114 dBA (124 dB). When used in a 114 dBA noise environment, the VIS headset you will be wearing has been shown to reduce the at-ear noise levels to 83 dBA, with the communication system turned off. With the communication system turned on (as it will be in this experiment), the highest at-ear noise level to which you will be exposed is about 94 dBA (in one of the four experimental conditions). The at-ear noise levels in the other three experimental conditions are expected to be much lower (about 90 dBA). Because the background noise level will be 85-90 dBA in the practice session and pre-experimental practice trials, the at-ear exposures in these situations are expected to be about 70 dBA or less.

The Occupational Safety and Health Administration (OSHA) currently allows workers in the United States to be exposed to 95 dBA time-weighted average noise for 4 hours/day. The total length of the noise exposures in this experiment will be about one hour (6 test sequences for which the noise will be "on" for less than 10 minutes each) in each of four experimental sessions. (Each experimental session will be conducted on different days to limit your total daily noise exposure.) Given the short exposure times, it is felt that there is little or no potential for doing any harm to your health or hearing. Furthermore, to ensure that a temporary threshold shift does not occur, an audiogram will be performed before and after each experimental session. (Stimulus levels presented during the experiment will be checked and adjusted before every experimental session.) <u>If, at anytime during this experiment, you should experience discomfort, please inform the experimenter</u>.

During the hearing test, you will be in a sound-proof booth with the experimenter sitting outside. The door to the booth will be shut but <u>not</u> locked; either you may open it from the inside or the experimenter may open it from the outside. There is also an intercom system through which you may communicate with the experimenter by simply talking (there are no buttons to push). If you are or think you may be claustrophobic or if you are uncomfortable in confined spaces, please tell the experimenter at this time. He/she will show you the rooms and let you enter them to see if they make you uncomfortable. The speech intelligibility test will be conducted in a much larger room and the experimenter will be in the room with you.

Since a U.S. Army VIS headset will be worn during all practice and experimental sessions, you may experience some minor discomfort due to the tight fit of the headset. However, the headset will not harm you in any way.

IV. BENEFITS OF THIS RESEARCH

Your participation in this experiment will provide information that will be used to quantify the performance of a prototype noise-canceling microphone that might be used by the U.S. Army in the future. No guarantee of benefits has been made to encourage you to participate. You may receive a summary of the results of this research when completed if you so desire. Please leave or send a self-addressed envelope if you are interested in receiving such a summary. To avoid biasing other potential participants, you are requested not to discuss the study with anyone until six months from now.

V. EXTENT OF CONFIDENTIALITY/ANONYMITY

The results of this study will be kept strictly confidential. At no time will the researchers release the results of the study to anyone other than the individuals working on the project without your written consent. The information you provide will have your name removed and only a participant number will identify you during analyses and any written reports of the research.

VI. COMPENSATION

For participation in this experiment, you will be compensated at a rate of $8.00/hr for each hour that you participate.

VII. FREEDOM TO WITHDRAW

You are free to withdraw from this study at any time without penalty. If you choose to withdraw, you will be compensated for the portion of time you have spent in the study. There may also be certain circumstances under which the investigator may determine that you should not continue as a participant of this project. These include, but are not limited to, unforeseen health-related difficulties, inability to perform the task, and unforeseen danger to you, the experimenter, or the equipment.

VIII. APPROVAL OF THIS RESEARCH

This research project has been approved, as required, by the Institutional Review Board for projects involving human participants at Virginia Polytechnic Institute and State University, and by the Grado Department of Industrial and Systems Engineering.

IX. PARTICIPANT'S RESPONSIBILITIES

I know of no reason why I cannot participate in this study. I have the following responsibilities:

- To listen attentively to the stimulus sounds presented during the tests, to respond appropriately and accurately, and to follow all instructions to the best of my ability.
- To notify the experimenter at any time about discomfort or a desire to discontinue participation.

____
Signature of Participant

X. PARTICIPANT'S PERMISSION

Before you sign the signature page of this form, please make sure that you understand, to your complete satisfaction, the nature of the study and your rights as a participant. If you have any questions, please ask the experimenter at this time. If you decide to participate, please sign your name on this page and the preceding page.

I have read a description of this study and understand the nature of the research and my rights as a participant. I hereby consent to participate, with the understanding that I may discontinue participation at any time if I choose to do so, being paid only for the portion of the time that I spend in the study.

Signature: ______________________________
Printed Name: ______________________________
Date: ______________________________

The research team for this experiment includes Dr. John G. Casali, Director of the Auditory Systems Laboratory, Dr. Gary S. Robinson, Research Associate Professor, and Ryan Urquhart, Graduate Research Assistant. They may be contacted at the following address and phone numbers:

Auditory Systems Laboratory	Dr. Casali: (540) 231-9081

Room 538 Whittemore Hall
Virginia Tech
Blacksburg, VA 24061

Dr. Robinson: (540) 231-2680
Ryan Urquhart: (540) 231-9086

In addition, if you have detailed questions regarding your rights as a participant in University research, you may contact the following individual:

Dr. David Moore
CVM Phase II (0442)
Virginia Tech
Blacksburg, VA 24061
(540) 231-4991

APPENDIX F

Participant's Screening Form

Pure-Tone Audiometric Tests for Normal Hearing

Participant: ______________________________ Age:_____ Sex:_____

Phone:___________________ Screening Date:_________ Qualify?____

Right Ear

Frequency Hz	t-1	t-2	t-3	t-4	t-5	t-6	final threshold
125	______	______	______	______	______	______	______
250	______	______	______	______	______	______	______
500	______	______	______	______	______	______	______
1000	______	______	______	______	______	______	______
2000	______	______	______	______	______	______	______
3000	______	______	______	______	______	______	______
4000	______	______	______	______	______	______	______
6000	______	______	______	______	______	______	______
8000	______	______	______	______	______	______	______

Left Ear

Frequency Hz	t-1	t-2	t-3	t-4	t-5	t-6	final threshold
125	______	______	______	______	______	______	______
250	______	______	______	______	______	______	______
500	______	______	______	______	______	______	______
1000	______	______	______	______	______	______	______
2000	______	______	______	______	______	______	______
3000	______	______	______	______	______	______	______
4000	______	______	______	______	______	______	______
6000	______	______	______	______	______	______	______
8000	______	______	______	______	______	______	______

SCREENING FORM

Otoscopic Data

Occluding wax?: ______________________________

Ear canal irritation?: ______________________________

Unusual canal characteristics: ______________________________

Eardrum perforations?: ______________________________

Eardrum scar tissue? ______________________________

Foreign matter?: ______________________________

Self-Report Data

Tinnitus or head noises: ______________________________

Otopathological history: ______________________________

Occupation: ______________________________

Noisy hobbies: ______________________________

HPD experience: ______________________________

Other: ______________________________

VITA

Ryan L. Urquhart was born and raised in Wakefield, Virginia. He earned his B.S. and M.S. in Industrial and Systems Engineering from North Carolina Agricultural and Technical State University in 1996 and 1998, respectively. For his Master's thesis, he studied the relationship between euphemisms that mention human body parts and visual-based hand signal used by the military for the placement of a tactile sensor device on the human body. Upon completion of his Master's, he continued his studies at Virginia Polytechnic Institute and State University with a concentration in the area of Human Factors Engineering. During this time, he was a Graduate Research Assistant in the Auditory Systems Laboratory conducting the research outlined in this dissertation. Ryan is an active member of the Human Factors and Ergonomics Society, American Society of Safety Engineering, and the Institute of Industrial Engineers.