3-24-2016

# Unmanned Aerial Vehicle (UAV) Operators' Workload Reduction: The Effect of 3D Audio on Operators' Workload and Performance during Multi-Aircraft Control

Sungbin Kim

**UNMANNED AERIAL VEHICLE (UAV) OPERATORS' WORKLOAD REDUCTION:**

**THE EFFECT OF 3D AUDIO ON OPERATORS' WORKLOAD AND PERFORMANCE DURING MULTI-AIRCRAFT CONTROL**

THESIS

Sungbin Kim, Major, Republic of Korea Air Force

AFIT-ENV-MS-16-M-163

**DEPARTMENT OF THE AIR FORCE**
**AIR UNIVERSITY**

***AIR FORCE INSTITUTE OF TECHNOLOGY***

**Wright-Patterson Air Force Base, Ohio**

AFIT-ENV-MS-16-M-163

# UNMANNED AERIAL VEHICLE (UAV) OPERATORS' WORKLOAD REDUCTION: THE EFFECT OF 3D AUDIO ON OPERATORS' WORKLOAD AND PERFORMANCE DURING MULTI-AIRCRAFT CONTROL

THESIS

Presented to the Faculty

Department of Systems Engineering and Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Systems Engineering

Sungbin Kim

Major, Republic of Korea Air Force

March 2016

AFIT-ENV-MS-16-M-163

# UNMANNED AERIAL VEHICLE (UAV) OPERATORS' WORKLOAD REDUCTION: THE EFFECT OF 3D AUDIO ON OPERATORS' WORKLOAD AND PERFORMANCE DURING MULTI-AIRCRAFT CONTROL


Sungbin Kim

Major, Republic of Korea Air Force


Committee Membership:

Michael E. Miller, PhD
Chair

Christina F. Rusnock, PhD
Member

John Elshaw, PhD
Member

AFIT-ENV-MS-16-M-163

## Abstract

The importance and number of Unmanned Aerial Vehicle (UAV) operations are rapidly growing in both military and civilian applications. This growth has produced significant manpower issues, producing a desire to invert the ratio of vehicles to operators such that multiple aircraft are controlled by a single operator as opposed to the current model where one aircraft sortie may require multiple operators. A potential issue with the revised concept of operations is the need for an operator to monitor radio traffic for the call signs of multiple aircraft. As a result, an investigation of the use of 3D sound was undertaken to investigate whether an automatic parser, which preselected the spatial location of relevant versus irrelevant call signs, could aid UAV operators in increasing performance with reduced workload. Furthermore, because the 3D audio system may not guarantee 100% reliability, human performance with the 3D audio system was also collected when they were informed announcement that errors were possible and when the reliability level was less than 100%. This investigation included development of a human performance model, simulation of human performance and workload, as well as a human subject study. Consequently, promising effects of the 3D audio system on multi-aircraft control were found. This novel and unique use of the 3D audio system is discussed, and significant improvements in response time and operator workload are demonstrated through modeling and a human in the loop experiment.

## Acknowledgments

First of all, I would like to thank my family. Thank you to my wife, who came to this distant strange country and had happy but hard times together, for continuous support and dedication. Thank you to my two pretty daughters for deep love and strong ability to adapt to a new world. I am proud of both of you!

I would like to express my deepest gratitude to my research advisor, Dr. Miller, who has supported me throughout my theses with extreme patience. Whenever I visited his office, he welcomed and encouraged me with warm smile. Without him, this research and thesis would not have been completed and written. He was not only my thesis advisor, but also American life advisor. When I had difficulty with American life, he felt anxiety and counselled me as if it were his own difficulty. I would also like to show appreciation to my committee members, Dr. Rusnock and Dr. Elshaw, for their kind and continuous support and feedback with passion.

I want to thank Mrs. Robb for her kindness and support for my family. In addition, to my AFIT friends, Quigg, Zinck, Watson, and Goodman, thank you for all your emotional support and I was always happy when I was with you. Thank you to all my research participants for great help.

Last but not least, I want to thank my parents for everything that helped me get to this point, and parents-in-law for unlimited trust. All the support that they have provided was the greatest gift for me. This thesis is dedicated to them.

Sungbin Kim

## Table of Contents

## List of Figures

## List of Tables

# UNMANNED AERIAL VEHICLE (UAV) OPERATORS' WORKLOAD REDUCTION: THE EFFECT OF 3D AUDIO ON OPERATORS' WORKLOAD AND PERFORMANCE DURING MULTI-AIRCRAFT CONTROL

## I. Introduction

### General Issue

The importance and number of Unmanned Aerial Vehicle (UAV) operations are exponentially increasing not only for military but also for civilian applications. In 2007, the US Department of Defense recorded that UAVs are becoming an increasingly critical aspect of military operations (Calhoun and Draper, 2015:2444). This increase requires more and more UAV operators. However, in practice, the supply of operators cannot keep up with the demand. Therefore, a key obstacle in the growth of the UAV operation is the number of operators required to command and control the vehicles. Still, most UAV systems require two or more operators to operate a vehicle (Calhoun and Draper, 2015:2444). The US Air Force said that it would work to address a shortage of pilots for unmanned aircraft by expanding incentive pay, tapping reserve forces, and working to lure pilots of manned aircraft to move over to drones (Barnes, 2015). While personnel actions such as those listed above should help reduce the shortage of operators in the near term, these actions do not address the significant manpower requirements imposed by the current control system.

Furthermore, the significant manpower requirements can be anticipated by the US drones' global missions. On September 7, 2000, a US Predator flew over Afghanistan for the first time (Bass, 2014). From that first mission, the use of drones overseas has increased exponentially.

Even now, the US military is using UAVs in Iraq and Afghanistan to support ground troops. Thus, many more UAV operators will be required in the future.

Automation, which increases vehicle intelligence and autonomy, could be one of the potential solutions to this problem. However, human beings should not transfer all of their responsibilities to the automated vehicles, because of the automated vehicles' reliability and human safety. That is, machines cannot have 100% reliability and humans may be under threat due to automation failure. Therefore, human judgment is necessary for unpredictable events in which some action must be taken to preserve safety, to avoid expensive failures, or to increase product quality (Shneiderman and Plaisant, 2010:74). Therefore, even with improved automation, human operators must continually supervise the vehicles. It has been proposed that the ultimate goal is to invert the operator/vehicle ratio (Franke and others, 2005:1-11). This means that one operator should control multiple UAVs to continually broaden UAV operations.

Most UAV operations include the three phases as shown in Figure 1. First, one operator at the base handles all ground operations and launches UAVs one-by-one. Then, when a UAV is airborne, the operator makes a hand-off of the UAV to mission operators. The mission operators conduct both a transit mission and the UAV's primary mission. The mission operators shift the UAV up to its mission area and, when the UAV arrives at its mission area, these same operators conduct the drone's real mission. Afterwards, the mission operators transit the UAV back to its base, where it is handed over to ground operations for landing. Currently, one operator controls only one UAV during launching and landing. However, two or more operators may be required for the remainder of the mission. It is notable, that the transit mission (i.e., second phase) requires relatively little operator interaction compared to other phases. So, the mission operators may be utilized inefficiently as they supervise one UAV for the long duration of the transit phase.

Furthermore, they may lose their concentration on their actual mission due to fatigue which is induced during the long transit duration.

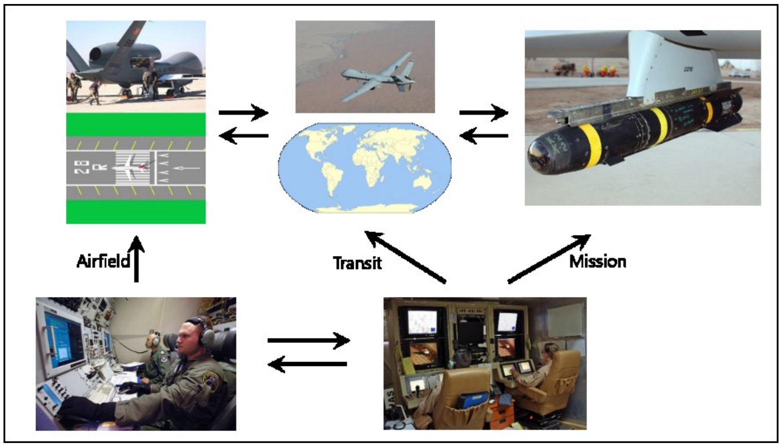


Figure 1. Current UAV Operation Phases

The current research focused on improving the transit phase. If there is an operator dedicated to the transit mission, and the operator controls multiple vehicles as shown in Figure 2, the mission operators can be utilized efficiently and concentrate on their primary mission. In this platform, human resources will be more efficiently assigned and utilized. This platform will also lay the foundation for controlling multiple UAVs during other phases of flight. Ultimately, such a redesign of the mission may reduce the essential number of operators for each UAV.

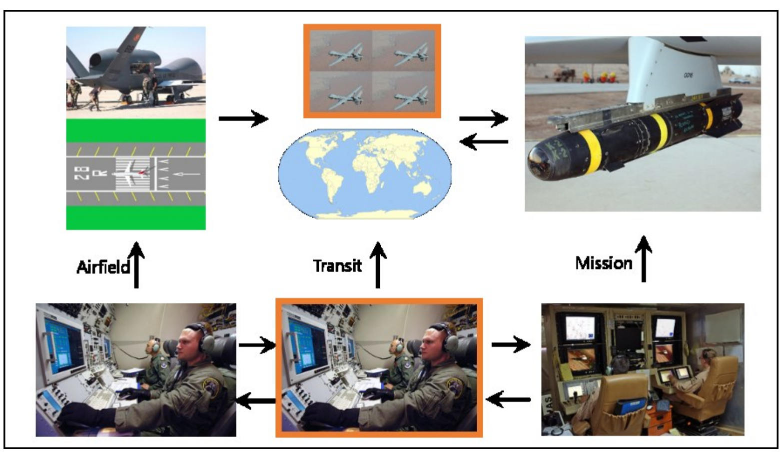


Figure 2. Suggested UAV Operation Phases (Newly Assigned Operator for Transit Mission)

## Problem Statement

For more developed UAVs and their missions, and for future military missions as well, inverting the operator/vehicle ratio is highly desirable. This requires increasing the number of vehicles controlled by an operator. A step towards this goal is to require one operator to control multiple vehicles during less taxing (e.g., transit) phases of flight. If this were undertaken with the current UAV control system, the operators' workload could be increased to unacceptable levels (Colombi and others, 2012:448-460). To reduce their workload, the current UAV-control system must be improved.

During the current transit phase, the UAV operators are exposed to a large amount of information from mission command, ATC (Air Traffic Control), the vehicle itself, and other vehicles. If the operator supervises multiple UAVs, the amount of information that the operator must consider would be proportionally increased, even though the transit missions have relatively less workload compared to the launching/retrieving or primary mission phases. Moreover, the transit duration often requires many hours. This may also cause a negative effect on the operators' performance due to decline in concentration. The operators' increased workload and decreased performance may increase the likelihood of mission failure, which would create not only economic losses but potentially result in fratricide.

## Research Objective

The objective of this research is to reduce the UAV operator's workload during multiple aircraft control under transit missions, by improving the operator's control system. In this research, a three-dimensional (3D) audio system was used to improve the operators' control system by aiding the operator recognition of relevant auditory information. The effect of the 3D audio on the UAV operator's workload and performance was investigated by performing simulations and conducting laboratory experiments.

**Research Focus**

This research focused on the performance of the 3D audio system for UAV operators who control multiple aircraft under transit operations. The 3D audio system provides subjects with separated inputs of critical information (i.e., the operator's information) and distractive information (i.e., other operator's information) to each ear as shown in Figure 3. The system can potentially present information, which the system is unable to differentiate and which is called "ambiguous information" in this research as a future concept, to both ears. This means each of the operator's left and right ear receives different information. The tasks that one operator should conduct in the experiment was simplified as compared to real-world UAV operations, in order to permit reliable measurement of the operator's workload and performance when using the 3D audio system or the current audio system.

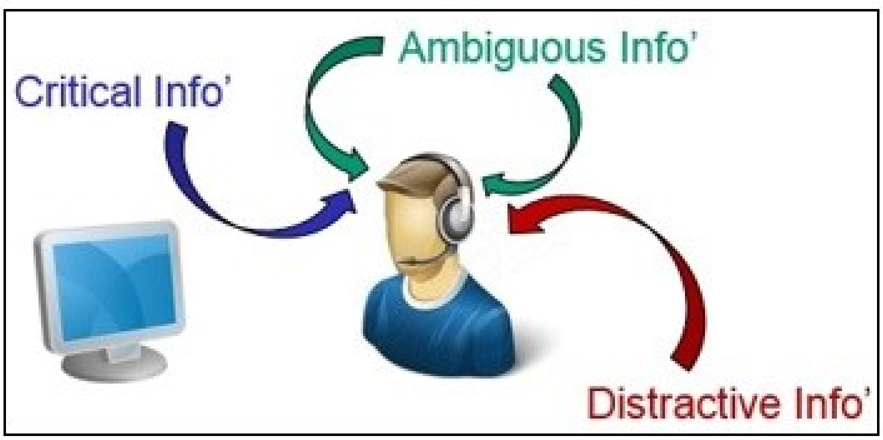


Figure 3. 3D Audio System

In the current audio system, an operator hears all information in both ears, which makes the operator constantly concentrate on all information. In contrast, when an operator uses the proposed 3D audio system, as long as the system can correctly separate relevant from irrelevant information, the operator does not need to concentrate on all information that is provided. The operator can easily distinguish critical or distractive information by determining to which ear the information is provided. Therefore, the operator's task can be simplified to concentrating on only one ear and a limited amount of ambiguous information. Therefore, it is expected that the 3D audio user's workload would be reduced compared to the users of the current audio system.

An alternative solution might be to remove the distractive information entirely from the operator's headset as this manipulation will likely further reduce workload. However, this distractive information could aid the operator in maintaining situation awareness, even when the information is not intended for their use. Although certain information is not directed to the operator, it should be the operator who decides whether the information he or she hears is helpful or not, particularly when the operator has the mental capacity to process this information.

The 3D audio technology has the potential to permit the operator to reduce mental workload by shedding time consuming tasks such as the call sign recognition using his or her notes during times the operator does not have the cognitive resources to process all auditory information, improving the operator's ability to react quickly and to distinguish information intended for them more precisely. It is, therefore hypothesized that the operator's performance, when responding to his or her call signs from among a number of distractor call signs, can be improved by using the 3D audio system compared with using the current audio system.

Under the current audio system, an operator's performance was affected by the number of call signs assigned to the operator (Amaddio and others, 2015:195-200). According to Amaddio and others' research, subjects' response time to their own call signs was increased and their accuracy was decreased when an operator controlled seven UAVs, compared to when the operator controlled five UAVs with different call signs. Likewise, it is expected that the 3D audio user's performance would be affected by the number of call signs, with an increasing number of call signs resulting in lower operator performance.

Additionally, it is likely that the voice recognition system (i.e., parser) cannot guarantee 100% reliability. Although the error rate could be very low, there may be errors while an operator uses the 3D audio system, such as providing critical information to the ear intended to receive distractive information. Since the operators may depend on the 3D audio system and its voice recognition, it is hypothesized that the operator may not easily detect such an error. However, if the operator is warned that these errors are likely, the operator may be able to modify their behavior to detect the errors. In this case, it is possible that the operator's workload will not be reduced significantly as compared to the current audio system.

**Investigative Questions**

This research will achieve the objective when the following questions are answered:

1. How does the 3D audio system affect an operator's workload compared to the current audio system, when the system performs with 100% reliability?
2. How does the 3D audio system affect an operator's performance (i.e., response time and accuracy) compared to the current audio system, when the system performs with 100% reliability?
3. How does increasing the number of call signs affect an operator's workload when the operator performs the task with the 3D audio system, compared to when the operator performs the task with the current audio system, when the 3D audio system performs with 100% reliability?
4. How does increasing the number of call signs affect an operator's performance (i.e., response time and accuracy), when the operator performs a task with the 3D audio system, compared to when the operator performs the task with the current audio system, when the 3D audio system performs with 100% reliability?

5. How does reduction in the reliability of the 3D audio system affect operator's workload?
6. How does reduction in the reliability of the 3D audio system affect operator's performance (i.e., response time and accuracy)?
7. How does the announcing possible errors of the 3D audio system to the subjects affect an operator's workload?
8. How does the announcing possible errors of the 3D audio system to the subjects affect an operator's performance (i.e., response time and accuracy)?

**Methodology**

Before conducting real experiments employing human subjects, a model was constructed and simulations ran in IMPRINT (Improved Performance Research Integration Tool) to explain the anticipated effect of the 3D audio system in an ideal environment. Then, experiment was conducted employing the Air Force Research Laboratory's Multi-Modal Chat (MMC) Monitor Client Program (Finomore and others, 2010). This study employed standard workload assessment methods and measurement of the subjects' response time and accuracy to assess the effect of the 3D audio system.

**Assumptions and Limitations**

It was assumed that there was no error in the 3D audio system for the model and simulations to measure pure effects of the 3D audio system. This model partially answered investigative questions from 1 to 4.

UAV operators could not be employed for the human subjects experiment, because of time and test personnel constraints. Instead, AFIT (Air Force Institute of Technology) student officers were employed as the subjects. For the same reason, the number of available participants was limited, resulting in a relatively small sample size.

This study was not conducted in the real world but was conducted in the synthetic task environment, for the purpose of the measurement under the same environment. The subjects was exposed to only directional instruction from ATC (Air Traffic Control). Other information, which can be provided in the real world like weather, traffic, airport, and mission information, was not provided to the subjects in this experiment. Although this experiment did not reflect real world conditions, a standardized synthetic environment could make it possible to assess the subject's workload, response time, and accuracy under a controlled environment.

**Implications**

Taking advantage of voice recognition technology, a method to improve operator recognition of relevant call signs in the multi-UAV control area, was explored. Most studies related to 3D audio have dealt with spatial information. That is, when a target is on an operator's left side, information related to the target is provided to the operator's left ear. While this information can aid the operator in determining the location of information such as the speaker or the location of the aircraft within the overall space, it may not help the user in distinguishing their call signs from a number of distracting call signs. This study applied the 3D audio in a different way for transit operations to improve the operators' workload and accompanied performance. The results of this study will inform system designers of advanced human-system interfaces. This study will also potentially help future UAV operators to supervise, command, and control multiple vehicles by reducing his or her workload. Ultimately, inverting the operator/vehicle ratio will be achieved, and more unmanned missions will be carried out under advanced technology and its interfaces.

## Preview

The first chapter stated the purpose and objective of this research, an overview of the method, assumptions and limitations, and this study's significance. Chapter 2, Literature Review, contains the theoretical framework for this study. This chapter presents a review of the issues which are relevant to multi-UAV control and the effect of 3D audio systems on UAV operator's performance. Chapter 3, Methodology, describes and justifies the data collection method used for this research. This chapter also outlines how the data will be analyzed. Chapter 4, Results, addresses the results from data analysis. This chapter contains results from the MMC program, including subjects' response times, accuracy scores, their workload, and effects of number of call signs that one operator owns. Finally, Chapter 5, Discussion, Recommendation, and Conclusion, addresses the meaning of the study's findings and contains the overall conclusion and areas for future research.

# II. Literature Review

## Overview

This chapter contains the theoretical framework of this research. First, the importance of operator interface technology will be emphasized by reviewing the use of autonomy in UAV systems. In this section, autonomy concepts for UAV systems and key words will be introduced. Second, necessary issues to be considered during the interface-design phase to achieve multi-UAV control will be discussed. Several paradigms, modes of interaction, automation, and related issues will be described in this part. Next, precedent research, which was conducted at the Air Force Institute of Technology (AFIT), will be reviewed. Then, previous researches addressing the impact of auditory displays and 3D audio for UAV operators were included. This section will explore the performance improvement and workload degradation of 3D auditory cues related to a single operator's supervision of multiple vehicles. Finally, three types of tools (i.e., IMPRINT, NASA-TLX, and SWORD), which were applied to this research for modeling (IMPRINT) and measuring human subjects' workload (NASA-TLX and SWORD), will be briefly introduced.

## Granting Autonomy to UAV Systems: The Importance of Interface Technology

Ultimately, the purpose of this research is to increase the number of vehicles that one operator controls for future UAV missions. To gain the required capability for future UAV missions, granting autonomy to the UAV systems is essential. Although the conceptual future system will possess more intelligent autonomy, the cognitive requirements for the operator responsible for monitoring and commanding these vehicles will not significantly decrease without advances in operator-interface technology (Franke and others, 2005:1-2). To understand this argument, it is first necessary to clearly define and understand the terms: autonomy, authority, and responsibility. These terms, as defined by Patrick (2014:28-29) are:

> *Autonomy* is the capacity of an agent to define its own objectives and to execute them.
>
> *Authority* is the capacity to take responsibility for the final decision, whether this concerns a task carried out in an autonomous manner or orders transmitted to one or several agents.
>
> *Responsibility* is the duty of the agent to answer for his or her actions or decisions in front of a body (agent or group) that possesses oversight authority.

Patrick argues that during the design of a human-machine system, the competence hierarchy, which increases from autonomy to responsibility, must be respected. For example, some assistance tools in automobile driving perform better than any human driver, notably in the avoidance of obstacles, as they are quicker and more efficient. Logic would therefore require that the assistance tools be given the authority that would allow them to make and execute decisions – instead of the human driver in the case of risking an accident. However, for legal reasons, responsibility should remain fully in the hands of the human, as it is not possible to hold the assistance tools accountable for a negative outcome. Similarly, even though a UAV system already possesses or will possess advanced autonomy, its responsibility for missions should remain fully in the hands of the human operator. This responsibility should be considered during the interface-design phase.

**Issues for Successful Interface-Design in Multi-Aircraft Control**

The articles mentioned above emphasize the importance of the interface-design in autonomous system design. As the UAV systems become more autonomous and their use increases, command and control interface concepts become more important for unmanned missions to succeed. For the multi-aircraft control interface concept, several issues should be considered, including operator paradigms, modes of interaction, and control paradigms among others.

***Operator Paradigms***

Many approaches have been formulated toward future operator paradigms, but these efforts can be classified into two main families (Franke and others, 2005:2-3). The first family, referred to as the "Common Operational System," aims to consolidate control functions for multiple types of vehicles under a single-control architecture. Applying this paradigm, multiple vehicles can be controlled using the same control station hardware without significant retooling, as shown in Figure 4. There are several material benefits. The use of a single hardware specification can reduce hardware and training costs and, thus, lead to a more rapid fielding of new systems. The commonality of the control mechanism better supports cross-unit, joint, and/or coalition operations. Note that this family differs from first-generation control stations, in which a unique control station was designed as a part of the UAV acquisition process, producing an operator interface which differs between different models of UAVs.

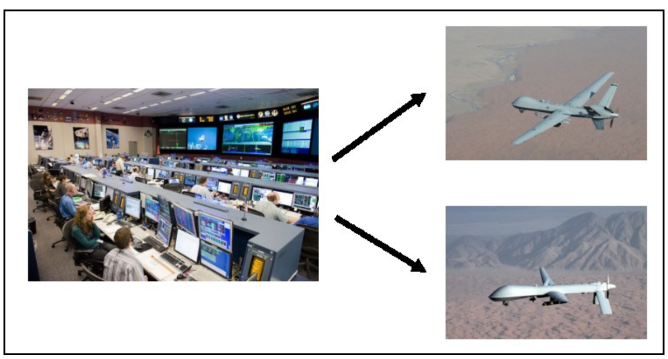

Figure 4. Operator Paradigm: Common Operational System

Another family of approaches is referred to as "Organic Control Systems." UAVs under the organic control will be designed to be controlled by, work in the vicinity of, and interoperate with manned vehicles and infantry. Control interfaces will be portable as shown in Figure 5.

The organic control strategy may reduce the operational timeline for execution of plans, provide better local situation awareness, and reduce hardware cost.

Figure 5. Operator Paradigm: Organic Control System

Notice that these families are not necessarily orthogonal from one another. It is possible to design a system that supports a common operational system for organic control or to design a system in which a single model of UAV can be controlled by a single interface, where this interface is designed for organic control.

### *Modes of Interaction*

To achieve success for inverting the operator/vehicle ratio, both of the operator paradigms require interface equipment that the operator may easily control. The addition of multi-vehicle control requirements overburdens available screen real estate and overtaxes the operator's ability to process visual information. To address these problems, new modes of interaction with UAV systems should be considered.

Multiple Resource Theory argues that task performance in different modalities can result in less cognitive interference, because they use different sets of resources within the cognitive system (Wickens, 2002:159-177). Other research indicated that multimodal feedback can increase situation awareness and reduce workload in certain applications (Philbrick and Colton, 2014:581; Haas, 2007:32-38). Practically, pilots wear head-mounted displays which provide visual and 3D auditory displays at the same time in a cockpit on a combat plane. Yu and

Brewster (2003) experimented using haptic and audio feedback to assist blind people with reading and understanding digitized, scientific charts and graphs. The results indicated that multimodal feedback reduced workload when compared to haptic feedback alone (Philbrick and Colton, 2014:581; Yu and Brewster, 2003:105-124). Haas and Stachowiak explored the use of tactile and 3D audio displays to enhance soldier performance in human-robot interaction tasks while in a moving vehicle, and the results indicated that combined tactile and audio displays had a significantly lower workload than tactile and audio displays used separately (Philbrick and Colton, 2014:581; Haas and Stachowiak, 2007:135-140).

Since the success and performance of Systems of Systems can be significantly impacted by the workload of key operators (Colombi and others, 2012:448-460), this Multiple Resource Theory should be considered during the design phase. In this research, it was assumed that real-time communication is provided through the auditory channel, while other UAV control information is provided to the operator through the visual channel. However, to simplify the experiment, the experimental paradigm did not include the visual control tasks. Instead, the auditory interface for real-time communication allowed the operator's eyes to be free to monitor other specific information of UAV such as attitude, elevation or speed.

### *Control Paradigms and Levels of Automation*

Approaches to vehicle management, or control paradigms, can be divided into three primary categories: direct control, management by consent, and management by exception (Franke and others, 2005:6-7). First, direct control means that the human operator directly commands the vehicle all the time, and the vehicle sends its status to the operator as shown in Figure 6. By conducting direct control, simultaneous control of multiple vehicles is virtually impossible. Because one operator does all of the decision making and information processing, direct control requires the operator to constantly attend to the vehicle. Therefore, it causes high workload for the operator.

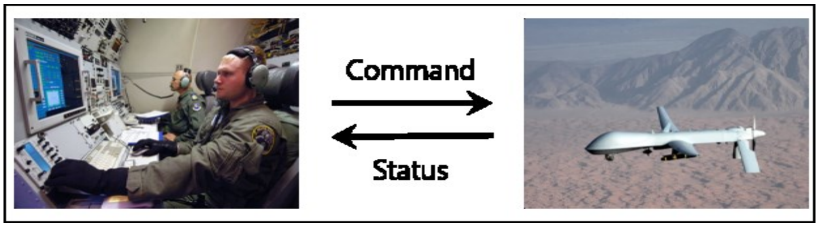


Figure 6. Control Paradigm: Direct Control (after Franke and others, 2005:6)

Under the management by consent control paradigm, vehicles perform planning and information-processing and send such plans to their operator(s) for approval, as shown in Figure 7. They perform no action without obtaining the operator's approval. The operator must react quickly to ensure the vehicle's safety for time-critical actions. This control paradigm produces moderate workload.

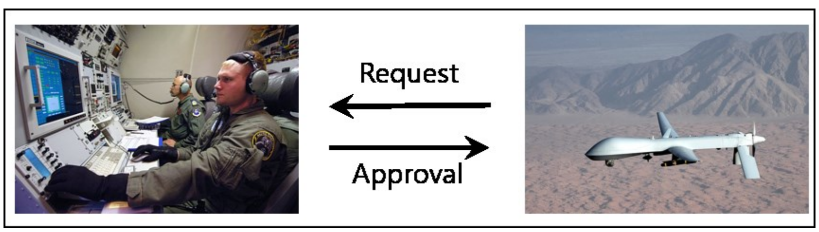


Figure 7. Control Paradigm: Management by Consent (after Franke and others, 2005:6)

The last control paradigm that Franke and others (2005) introduced is management by exception. This means that UAVs not only perform planning and information processing, but they also begin execution. The operator has the ability to override vehicle actions and plans, as shown in Figure 8. This control paradigm requires a high degree of intelligence and autonomy for the vehicle. In addition, it requires the operator to maintain situation awareness. Since the operator does not necessarily need to provide input to the vehicle, this paradigm potentially results in relatively low workload as it likely reduces at least the physical or observable workload.

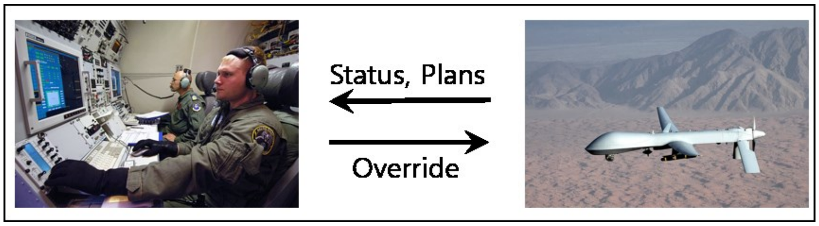


Figure 8. Control Paradigm: Management by Exception (after Franke and others, 2005:6)

However, some decisions cannot be entrusted to the system as mentioned before. For example, autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force (Righetti and others, 2014:8). Franke and others (2005) indicated that systems can employ a mixture of the paradigms for different tasks, while being dynamically configurable to assign which paradigm is used for each type of task or decision at any time during the mission. This is also known as "Adaptive Automation." Adaptive automation has been described as a form of automation that allows dynamic changes in control function allocations between a machine and human operator based on states of the collective human-machine system (Kaber and others, 2001:1). This adaptive automation was defined as a system which varies function allocation during system operation, while minimizing costs (Parasuraman and Wickens, 2008:516-517).

Similar to the control paradigms, Billings (1991) and Kaber (1997) suggested that the level of automation refers to the level of task planning and performance interaction maintained between a human operator and computer in controlling a complex system (Kaber and Endsley, 2004:115). Here, the automation refers to the full or partial replacement of a function previously carried out by the human operator (Wickens and others, 2000:287). The level of automation approach defines the assignment of system control between a human and computer in terms of the degree to which both are involved in system operations. The level of automation approach emphasizes the interaction between a human operator and computer. In 1987, Endsley developed a level of automation hierarchy (Kaber and Endsley, 2004:117):

1. Manual control – with no assistance from the system;
2. Decision support – by the operator with input in the form of recommendations provided by the system;
3. Consensual artificial intelligence – by the system with the consent of the operator required to carry out actions;
4. Monitored artificial intelligence – by the system to be automatically implemented unless vetoed by the operator; and
5. Full automation – with no operator interaction.

Additionally, reliability of automation is usually very important in user-interface design. In cases where the reliability was lower, automation support was found to reduce system performance, as compared to the human use of systems without automation support (Kaber and Endsley, 2004:123-124).

***Other Issues***

There are some other issues that must be addressed to ensure successful multi-UAV operation (Franke and others, 2005:7-10). Interruptions may provide a considerable hazard to both operator workload and effectiveness, because operators may lose their concentration. To prevent the deleterious effects of interruption, an effective interruption management mechanism must be in place. Furthermore, as mentioned above, even during management by consent or management by exception, the operator is still responsible for safety and mission success. Therefore, the operator should be sufficiently trained in understanding and using the features of his or her system. By doing so, the operator can trust his or her system. In addition, predictability is required, where the vehicles behave in a way that the operator can expect.

### *Application to Current Research*

Providing autonomy, or automation, is important to accomplish one operator's control of multiple aircraft. However, as the operator must assume responsibility for multiple aircraft, the command and control interface must permit the operator to control these aircraft effectively. Therefore, the aforementioned issues should be considered and applied when designing the interface. This research considered these issues for experiments. Specifically, a common operational system, where multiple vehicles can be controlled using the same control station hardware, was assumed. An auditory display was used to facilitate communication between the operator and others in the operational environment for each of multiple vehicles. Even more specifically, a 3D audio interface was applied to aid the operator in performing communications relative to multiple aircraft. For the control paradigm, this research assumed management by exception. Detailed assumptions and methodology was described in the next chapter.

## Motivation from Previous Research

Previous research related to UAV operators' workload reduction was conducted at the Air Force Institute of Technology (AFIT). This research investigated the cognitive load (i.e., number of aircraft call signs) that an individual can handle and explored the effect of proactive interference (PI), while conducting communications tasks for multiple aircraft (Amaddio and others, 2015:195-200). Their experiment was conducted using Multi-Modal Chat (MMC) Monitor Client Program (Finomore and others, 2010), which is a Windows software program that monitors and parses messages containing transcriptions of radio communications and text chat messages. The same program was employed for this research.

Amaddio (2015) asked participants to memorize their critical call signs, and to record numbers related to their critical call signs, when they heard these critical call signs among a number of distracters during the experiment. Certain call signs were selected from among the critical call signs during one experimental condition and used as distracters in a subsequent experimental condition – potentially leading to proactive interference (PI), where the participant would recall these distracters as critical call signs in one trial, because they had been critical call signs in the previous trial. The participants were exposed to 4 experimental configurations: 5 call sign without PI, 5 call sign with PI, 7 call sign without PI, and 7 call sign with PI as shown in Table 1. The subjects were divided into two groups. Table 1 presents the trials and the critical and PI call signs for the participant Group 1. The call signs were the same for Participant Group 2, but they experienced the 7 call sign conditions first. The researcher measured the subjects' accuracy scores and response times to explain how the number of call signs and the PI affected the operators' performance.

Table 1. Amaddio's Experiment: Group 1

Call signs experienced by Group 1 during each trial. Call signs which were employed to induce PI during Trials 2 and 4 are shown in Bold-Italics for Trials 1 and 3.

| Participant Group | **Trial 1** (5-NP) | **Trial 2** (5-P) | Break (15 minutes) | **Trial 3** (7-NP) | **Trial 4** (7-P) |
|---|---|---|---|---|---|
| 1 | Laker | Laker | Operations | Charlie | Charlie |
| | Hopper | Hopper | Word Span | Gringo | Gringo |
| | Arrow | Arrow | test followed | Laker | Laker |
| | ***Charlie*** | Tiger | by a break | Raptor | Raptor |
| | ***Gringo*** | Eagle | | Viking | Viking |
| | | | | ***Arrow*** | Thunder |
| | | | | ***Tiger*** | Cobra |

The 5 call sign with PI condition received the highest accuracy score, although it was assumed that the "with PI" condition has higher task load, as shown in the left graph of Figure 9. Amaddio mentioned that this result might be explained by the workload-performance curve similar to the Yerkes-Dodson Law, such as shown in Figure 10 (Teigen, 1994:525-547). In the figure, high and low levels of workload result in low performance, but, medium level of workload results in higher performance (ODonnell, 2011). Although the 7 call sign with PI condition contained the highest task load, it did not produce statistically significantly lower scores than the other 7 call sign condition and the 5 call sign without PI condition.

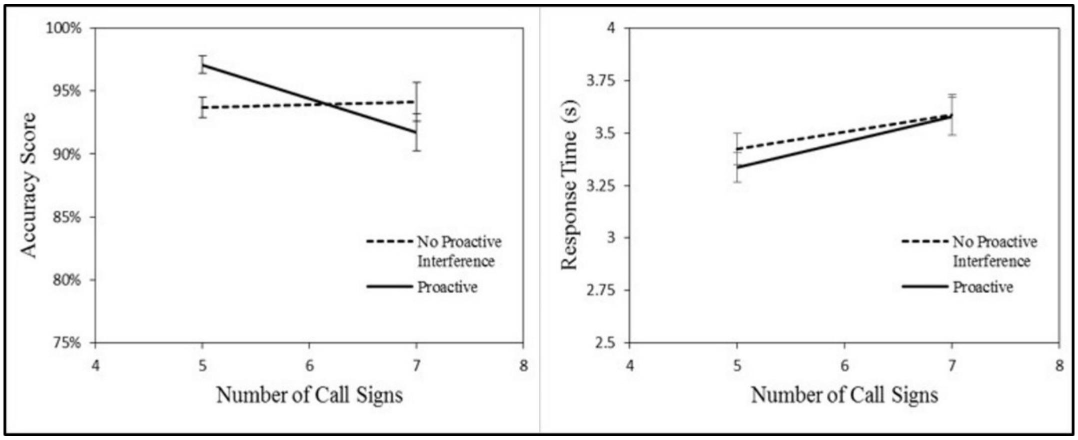


Figure 9. The Results of Amaddio's Experiment (Accuracy Score and Response Time)

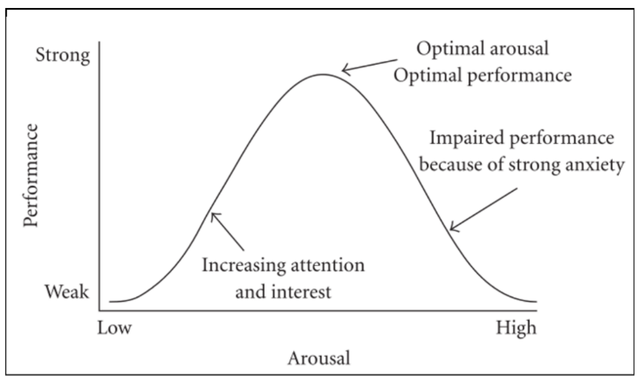


Figure 10. Workload-Performance Curve ("File:HebbianYerkesDodson.svg", 2014)

The highest performing condition (i.e., 5 call sign with PI) had significantly lower response times than the 7 call sign conditions, which were the conditions with the highest task load as shown in Figure 10. Although the accuracy scores of the 5 call sign with PI condition were significantly higher than the 5 call sign without PI condition, their response times were not significantly different.

The results of this precedent research provide conflicting evidence about whether higher task load conditions actually produce lower levels of performance. Participants did not score differently on the highest and the lowest task load condition, suggesting that there may be a non-linear relationship between task load and performance. This research can be helpful to study how many call signs that a single operator can control. However, there may be some gaps with the real-world conditions. UAV operators always receive their critical information with other operators' distractive information during the transit operation from Air Traffic Control (ATC). In addition, they do not need to, and do not have to memorize all of their critical call signs, because the operators' assigned call signs may be changed several times a day. If one operator controls multiple UAVs during the transit operation, the operator may be performing a task similar to the air traffic controller in the control tower. That is, as current air traffic controllers usually refer to their screens and their notes in the control tower, likewise, the operator does not need to memorize all of their critical call signs.

The research discussed within this thesis was motivated by Amaddio's study, thus the general methods resembled those developed within her research. However, this research did not apply any without-PI-conditions. Furthermore, subjects in the current research were not required to memorize their critical call signs.

**Impact of Auditory Displays and 3D (Spatial) Audio**

As mentioned above, this research used auditory displays; specifically, it investigated the impact of a 3D audio interface on multiple UAV radio communications. It is therefore useful to understand the advantages of auditory displays. Neural transmission in the auditory system processing is substantially faster than transmission in the visual system; thus, time-critical warnings are commonly communicated through auditory signals (Simpson and others, 2004:62; Mowbray & Gebhard, 1961:115-149). For this reason, auditory displays can be more applicable to UAV operators' transit missions than the impact of a visual display. Furthermore, Simpson and others described that the auditory system plays a fundamental role in verbal communication, which is in many cases the most direct, efficient, and unambiguous means of information transfer, and that the auditory information can be used even when the sound originates from outside of the operator's visual field of view (Simpson and others, 2004:62).

In 2010, Maza and others researched 3D audio's effect on situation awareness (SA) of UAV operators (Maza and others, 2010:371-391). A simple experiment was conducted. Three screens were installed in front of subjects, and the subjects were provided with a "yes" or "no" signal by several display configurations: touch screen interface only, touch screen with audio, and touch screen with 3D audio. In each trial, only one screen showed a "yes" or "no" signal. When the subject was provided with the "yes" signal, he or she was asked to push the "yes" button on the corresponding screen. Response times and accuracy were measured, and it was also observed that the individuals pointed their head directly on the proper screen after hearing the "yes" message. When the 3D audio was used, according to the location (i.e., left, right, or middle) of the screen which displayed the "yes" signal, the source of audio corresponding to its label was generated on the left, on the right, or in front of the operator respectively through the stereo-headset. As a result, accuracy was almost the same among the three displays. However, as shown in Figure 11, the subjects responded faster when they were exposed to the touch screen with normal audio signal and the touch screen with 3D audio signal, than when exposed to the touch screen only. In the interviews after this test, it was mentioned that the workload was reduced, as the subject was able to be relaxed until the "yes" message was heard. Moreover, subjects performed better, or responded faster, when they were exposed to the touch screen with 3D audio, compared to when they were exposed to the touch screen with normal audio interface. According to Maza and others, workload was reduced due to two different factors. First, there was no need to pay attention while hearing a “no” message.  Second, once the “yes” button appeared, there was no need to search for the button from one screen to another (i.e., focused immediately on the screen which displayed the “yes” message).  This experiment evaluated and explained the potential benefits of the 3D audio with respect to the conventional audio.

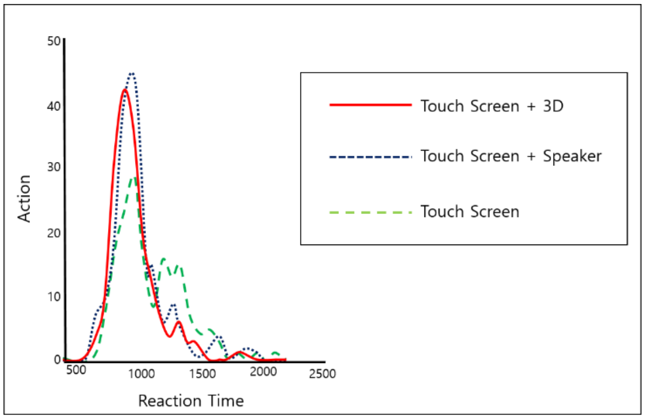


Figure 11. Maza and others’ Experiment: Response Time (after Maza and others, 2010:13)

Guastello described that 3D synthesized audio displays can enhance pilot performance in some types of tasks (Guastello, 2014:95).  He illustrated that Btonkhorst and others (1996) prepared a 3D audio track to accompany a primarily visual task on a flight simulator.  The participating pilots were chasing another aircraft that disappeared at critical points in the flight. The participants were required to locate the target aircraft.  The researchers found that the combination of visual and 3D audio signals produced shorter search times than either visual or 3D audio display alone.  According to their experiment, ratings of workload were not affected by the introduction of 3D audio.

Simpson and others described that spatial auditory display technologies take advantage of the properties of the binaural auditory system by recreating and presenting to an operator the spatial information that would naturally be available in a “real-world” listening environment (Simpson and others, 2004:62).  Therefore, such displays are intuitive and thus impose no additional demands on the information processing capacity of the operator. Therefore, users can gain additional cues based on the location of the sound without devoting additional cognitive resources.

The 3D audio also showed an advantage for detecting infrequent speech signals from a background stream of irrelevant speech (Guastello, 2014:95; McAnally and Martin, 2007:688-695). One website, BeckerUSA.com, also discusses that a user is perceptive to sounds from a predefined direction as a key benefit of the 3D audio ("3-D Audio Technology", 2011). The site described that this capability allows a user to spatially separate simultaneous audio communications, information, and warning tones by focusing his or her attention on the audio source which he or she finds most important. Therefore, it is possible for users to monitor several audio sources in different positions. This effect is generally known as the "Cocktail Party Effect," which is the ability to focus one's listening attention on a single talker among a cacophony of conversations and background noise (Arons, 1992:35).

The U.S. Army Research Laboratory explored the use of advanced technologies such as tactile and spatial (3D) audio displays to enhance soldier performance in human-robot interaction tasks (Haas and Stachowiak, 2007:135). They indicated that spatial audio displays can communicate events, using sound coming from a number of directional sound sources; for example, radio communications from a commander can sound like they originate from the soldier's front, a hazardous agent warning signal may come from the soldier's right, and a signal indicating the position of a remote robot may be heard from the general direction and elevation of that robot. Trouvain and Schlick also demonstrated that with the human ability to separate sound sources, an operator can focus on listening to both left and right channels or exclusively to the left or right channel (Trouvain & Schlick, 2004:2823). Therefore, Haas and Stachowiak explained that spatial audio cues are useful in human-robot interface target search tasks, and that spatial audio displays can increase user situation awareness in target search of unmanned aerial vehicle (UAV) displays. The research also described the use of spatial auditory display cues to enhance 360-degree situation awareness in applications even without a visual display, because they provide positional cues.

Unlike the aforementioned claims, Cengarle mentioned that the 3D audio is “immersive,” in the sense that it brings more involvement to the listener (Cengarle, 2012:137-138). In order to verify this claim, he conducted experiments where subjects watched short movies with 5.1 or 3D audio, while psycho-physiological data such as heart rate, facial electromyography, and electro-dermal activity were recorded. This experiment demonstrated that higher emotional arousal was provoked when the 3D audio was employed. This feature might be seen as both an advantage and a disadvantage of the 3D audio. While appropriate involvement may help the listener concentrate on his or her tasks, excessive immersion may prevent the listener from distributing his or her attention to other critical information within the physical environment.

Furthermore, some articles revealed that the 3D audio has certain limitations. Philbrick and Colton conducted experiments to understand the effects of haptic and 3D audio feedback on operator performance and workload for Quadrotor UAVs in indoor environments (Philbrick and Colton, 2014:580-591). This research suggested that multimodal feedback, specifically 3D audio combined with haptic feedback and a visual interface, can increase situation awareness and reduce workload in a variety of applications. The subjects were asked to guide the UAV in two synthetic indoor environments. They were also asked to complete the course as quickly as possible, with as few collisions as possible. During the experiment, as the time it would take for a UAV to collide with an obstacle decreases, the haptic force increased to warn the operator of an increased chance of collision. The 3D audio was designed to be a tertiary feedback modality, after visual and haptic, with the intent to provide useful warning cues. A discrete audio cue (i.e., a short duration beep) was displayed only in the direction of the UAV velocity and only when the vehicle was within a threshold distance of an obstacle. In addition, the time period between beeps was graded, meaning that the frequency of the beeping increased as the UAV approached an obstacle. The researchers concluded that the 3D audio did not affect the operator's workload. Although the haptic feedback improved the operator's performance, the 3D audio feedback increased the total completion time, without decreasing the number of collisions. Some of their subjects reported that the 3D audio was not as intuitive as the visual or haptic feedback and was frustrating at times. However, many subjects also felt that the audio feedback was helpful. The researchers described that one reason for this conflict was the weakness of his experimental device, which concentrated on haptic feedback. Therefore, Philbrick and Colton emphasized that proper application and improved training could improve the effectiveness of the 3D audio system. Additionally, the cluttered and complex indoor environment may affect the results. Under the cluttered and complex indoor environment, it may be difficult to achieve balance between obstructions to avoid the audio's annoying beep signal.

Vazquez-Alvarez and Brewster stated that listening to concurrent audio increased the effect of cognitive load, and that the use of spatial audio techniques had a negligible impact on reducing this effect (Vazquez-Alvarez and Brewster, 2011:2176). That is, the spatial audio was not helpful for operators' performance and workload when it was used concurrently with other audio sound. This claim also supports the importance of proper application of 3D audio.

Trouvain and Schlick conducted experiments for audio and visual context switch indicators in multi-robot navigation task (Trouvain and Schlick, 2004:2821-2826). In their experiments, three types of interface configuration were compared: "Camera View (CV) only," "CV + Visual Indicator," and "CV + Auditory Indicator." Their results described that "CV + Visual Indicator" had the most benefit for participants' performance, followed by "CV + Auditory Indicator." However, they concluded that the effect of the spatial (3D) audio might be different according to the interface design, because their experimental interface layout featured a very dominant visual indicator, and such a layout may not be possible in all types of interfaces.

According to the researchers mentioned above, the auditory display, especially the 3D audio display, can have positive effects on situation awareness and workload, only when it is used appropriately within a suitable environment. Therefore, more research should be conducted to understand the attributes of 3D auditory displays which are the most useful and effective. Furthermore, most research which applies 3D audio has been focused on encoding spatial information within the sound signal, such as direction or distance information. In this research, however, novel application of the 3D audio was employed; the 3D audio was applied to convey relevance rather than spatial location, relying upon the user's ability to separate signals provided to each ear, to examine the effectiveness in decreasing workload and increasing performance.

**IMPRINT, NASA-TLX, and SWORD**

This research employed IMPRINT (Improved Performance Research Integration Tool) for modeling and simulating the conditions of real experiments to explain anticipated effects of the 3D audio system under an ideal environment. IMPRINT, developed by the Human Research and Engineering Directorate (HRED) of the U.S. Army Research Laboratory (ARL), is a human-system task network modeling tool with specialized analytic capabilities (Allender, 2000:140). The analytical capabilities in IMPRINT include human versus system function allocation, mission effectiveness modeling, maintenance manpower determination, mental workload estimation, prediction of human performance under extreme conditions, and assessment of performance, as a function of varying personal skills and abilities. In this research, mental workload was estimated, and human performance under an ideal environment was predicted by using this software, before conducting the human subjects experiment.

After each condition of the experiment in the current research, each participant rated his or her perceived workload using the NASA-TLX (NASA-Task Load Index). The NASA-TLX is a multi-dimensional scale designed to obtain workload estimates from one or more operators, while they are performing a task or immediately afterwards (Hart, 2006:904). Hart described that the years of research that preceded subscale selection and the weighted averaging approach, resulted in a tool; the tool has proven to be reasonably easy to use and reliably sensitive to experimentally-important manipulations over the past 20 years. By using this tool, participants' workload levels were collected to examine the 3D audio's effect on operators' workload during multi-UAV control.

Each participant also assessed their workload by using SWORD (Subjective Workload Dominance Technique) after completion of all conditions of the experiment. The SWORD is a subjective workload assessment technique, and it uses paired comparison of tasks in order to elicit ratings of workload for individual tasks. The SWORD technique is administered post-trial and requires participants to rate one task's dominance over another in the workload imposed (Stanton and others, 2010:332).

**Summary**

In the near future, UAV systems will have more applications in undesirable or dangerous environments, like military operations such as reconnaissance or long-range and high-altitude missions, as a substitute for manned systems. However, computers cannot always take the place of human decision-making. To keep pace with the rate at which the UAVs are used, its interface should be improved in a common operational system, where a single operator can simultaneously operate multiple vehicles. To improve the interface, auditory display may be considered as one of the possible solutions among several modes of interactions. Specifically, when the 3D audio display is properly applied to a suitable system, it can be expected that not only the operator's performance will be improved, but the operator's workload will also be reduced. Based on this framework, the next chapter will describe the methodology to be employed in this research for modeling and simulation, as well as the human subjects experiment.

# III. Methodology

## Overview

This chapter contains data collection methods used for this research, and outlines how the data will be analyzed. For this research, a model was made and simulations ran in IMPRINT to explain the anticipated effect of the 3D audio system under an ideal environment. From the model simulations, anticipated response times and operators' workload were predicted. Then, a human subjects experiment was conducted to determine accuracy, response time, and workload ratings.

## Model Development and Application

### *Modeling Process*

The development process requires the construction and validation of a model, typically against an existing data set. In the current research, an 'Initial Model' was constructed first. This initial model included development of a basic structure, and the response times from this model was validated against the response times as observed by Amaddio (Amaddio and others, 2015:195-200).

As the conditions of the present experiment did not correspond specifically to the conditions investigated by Amaddio (2015), the initial model was modified to form a baseline model representing the current two-dimensional (2D) sound conditions of the present experiment.

The initial model sought to represent the conditions of Amaddio's research, because the structure of the current research resembled her experiments and because her research included response times which could be used to validate this model. While Amaddio assumed that an operator memorizes his or her critical call signs, this research assumed that an operator does not memorize them. As a result, the tasks to be undertaken by participants in the current study differed from those performed by Amaddio's participants, therefore different response times and workload was applied for the model of this research. By manipulating response times from the initial model, a baseline model was constructed to represent the tasks to be performed by participants in the current experiment. Next, expected workload values were input to the baseline model. Because Amaddio's research did not include workload values, the baseline model cannot be validated by real data from previous experiments. Therefore, the workload values in the baseline model were validated by SME (Subject Matter Expert) data. Finally, for each condition of this research, the baseline model was modified to represent expected participant behavioral changes.

**Overall Scenario**

Similar scenarios were applied when constructing the initial model and the baseline model. However, in the baseline model an operator does not memorize the critical call signs. Instead, an operator checks the critical call sign list to decide whether the call sign is critical or distractive. The different task structures are shown in Figure 12 and Figure 13 for the initial model and the baseline model respectively. A blue box in the Figure 13 indicates an additional task to check the critical call sign list. A detailed explanation of this overall scenario follows.

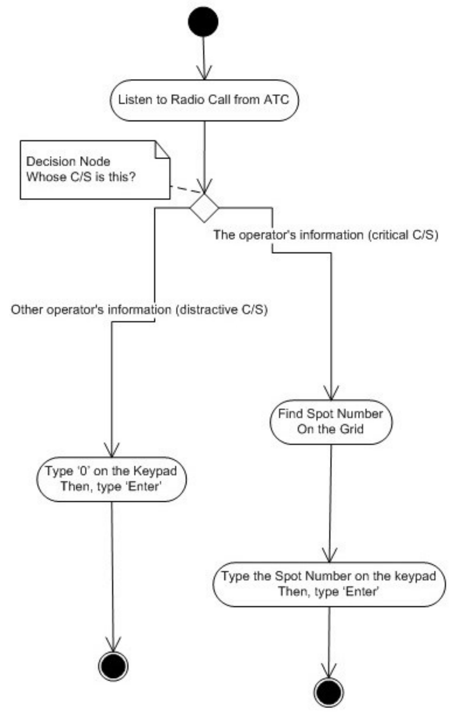


Figure 12. Overall Scenario Employed in the Initial Model

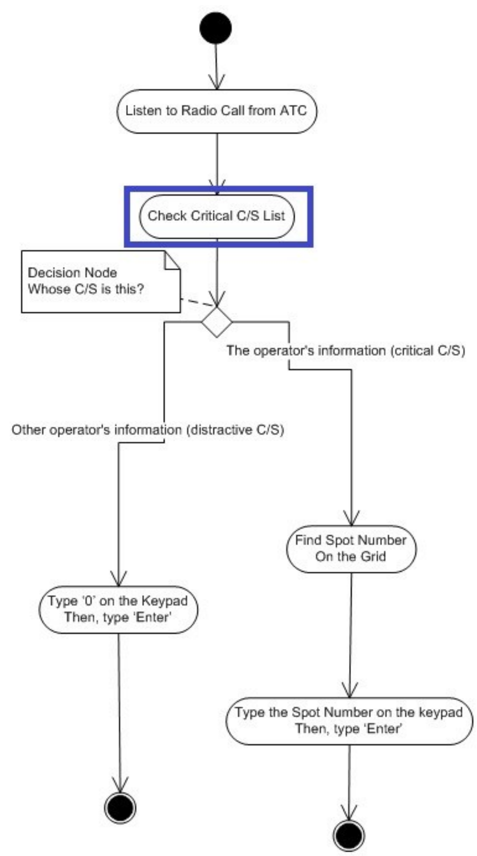


Figure 13. Overall Scenario Employed in the Baseline Model

During a transit operation, one operator controls multiple UAVs. This model starts with the first radio call from ATC. After the operator hears the radio call, he or she decides whether the instruction is intended for him or her through the distinction provided by the call sign; 'Critical' call sign is the operator's call sign and 'Distractive' call sign is another operator's call sign. To categorize the call sign into one of these two categories, the operator refers to the 'Critical Call Sign List' which includes all of the operator's call signs such as shown in Table 2. If the operator hears a distractive call sign, the operator is asked to type '0' on the keypad. In contrast, if the call sign is critical, the operator is asked to type the corresponding 'Spot Number' on the keypad. In this case, to find the two-digit spot number, the operator is required to check a 'Grid', which includes all spot numbers corresponding to the ATC's instructions as shown in Table 3.

Table 2. Critical Call Sign List for Baseline Model

| Critical Call Sign List | |
|---|---|
| 1 | Arrow |
| 2 | Charlie |
| 3 | Eagle |
| 4 | Hopper |
| 5 | Laker |

Table 3. Grid for Spot Number

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Blue | 21 | 81 | 49 | 38 | 95 | 18 | 60 | 98 |
| Red | 72 | 36 | 92 | 07 | 46 | 58 | 30 | 79 |
| White | 90 | 23 | 13 | 86 | 75 | 26 | 71 | 97 |
| Green | 57 | 89 | 52 | 37 | 19 | 83 | 62 | 41 |

For example, if an operator hears an instruction from ATC such as "Ready, Charlie, Go to Green Three, Now," then, the operator would check the critical call sign list to confirm whether the call sign, "Charlie," is among their critical call signs. Table 2 includes "Charlie," so the instruction corresponds to the call sign of a UAV under the operator's control, and the "Charlie" is one of the operator's critical call signs. Next, the operator would check the Grid to find the spot number corresponding to the "Green Three" from ATC's instruction. The operator would identify the row green and the column three, which corresponds to the number "52." The operator would then type "52" on the keypad and press "Enter."

On the other hand, if the instruction was "Ready, Carrier, Go to Blue One, now," then the operator will type "0" on the keypad because the call sign, "Carrier," is not on the critical call sign list. This means that the call sign, "Carrier," is a distractive call sign.

For the purpose of the measurement under the same environment, this model was simplified through the use of some assumptions. Although this model did not completely reflect the real-world environment, this standardized synthetic environment could make it possible to assess the UAV operator's workload and performance under a near-ideal environment. Detailed assumptions for this purpose are described in Appendix A.

## Modeling

### *Initial Model*

As mentioned above, this initial model was based on Amaddio's experiments. Basic structure and response times, collected from these earlier experiments, were included in the initial model. The task network associated with the initial model is shown in Figure 14. In this initial model, one UAV operator controls five UAVs simultaneously, and the operator uses the current audio system with which the operator receives directional instruction from ATC through both ears. Based on Amaddio's protocol, the operators were tasked with memorizing their critical call signs before beginning the experiment, so they did not need to check the critical call sign list.  Therefore, this task was not a required node in this initial model.

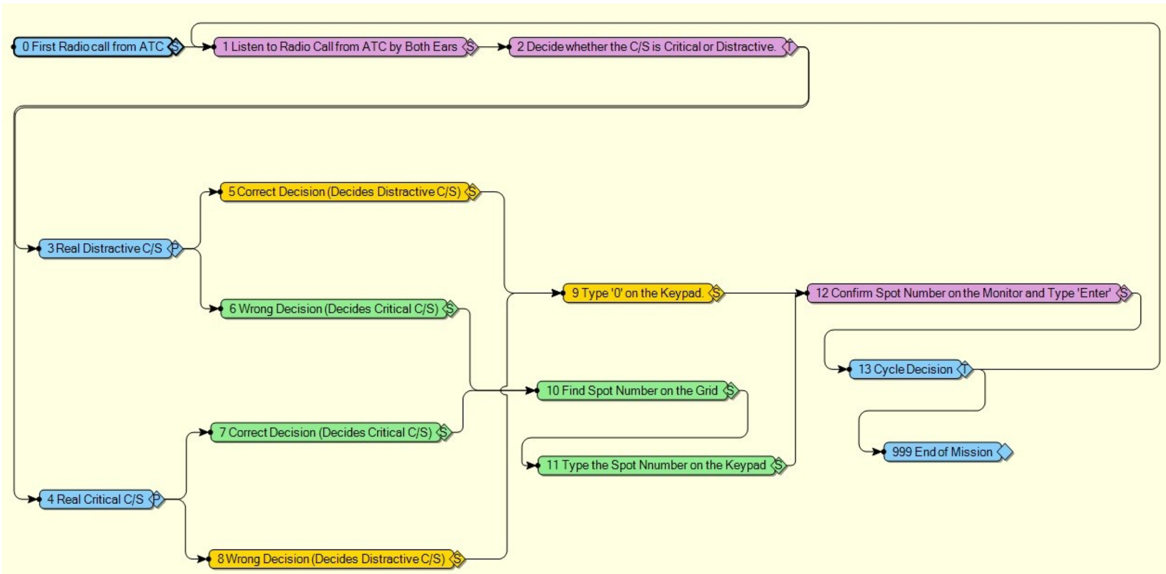

Figure 14. Task Network for Initial Model

Model nodes in the Figure 14 are divided into four types, and each type is depicted with a different color (e.g., blue, plum, gold, or green).  The blue-colored nodes exist only for logic; Node 0 starts this model; Nodes 3 and 4 divide correct or wrong decisions according to probabilities; Node 13 decides how many instructions are provided to one operator; and Node 999 occurs when all instructions provided to the operator were concluded, thus ending the model. At each cycle of this initial model, there are tasks that the operator should always conduct, and these task nodes are shown as plum color nodes.  These plum-colored task nodes include task times and workload, so they affect the operator's performance and workload.  Human tasks that occur periodically, rather than each loop through the model are represented as gold or green nodes.  When the operator decides that the call sign that he or she listened is distractive, the operator conducts the tasks indicated by the gold nodes.  On the other hand, when the operator decides that the call sign is critical, the operator conducts the tasks indicated by the green nodes. These gold and green task nodes occur selectively according to the operator's decision.  Detailed data input modeling and response time validation of this initial model are described in Appendix B and C respectively.

***Baseline Model: Task Network***

The assumption of the initial model was a little different from that of this research. In this research, it was assumed that operators do not memorize their critical call signs, therefore, based on the initial model, an additional task and its task time must be added to the baseline model. Although this may increase overall response time, it may also increase accuracy of the important UAV tasks. In addition, workload was added to this baseline model based on VACP scales as shown in Table 4. The task network for the baseline model is shown in Figure 15. Because the operators do not memorize their critical call signs to increase their accuracy, they need to check the critical call sign list whenever they receive the instruction from ATC. The red box in the Figure 15 reflects this condition; a task, 'Check Critical C/S list', is added to Node 2. Other nodes are not affected by this condition. Detailed data input description of this baseline model and workload validation were explained in Appendix D and E respectively.

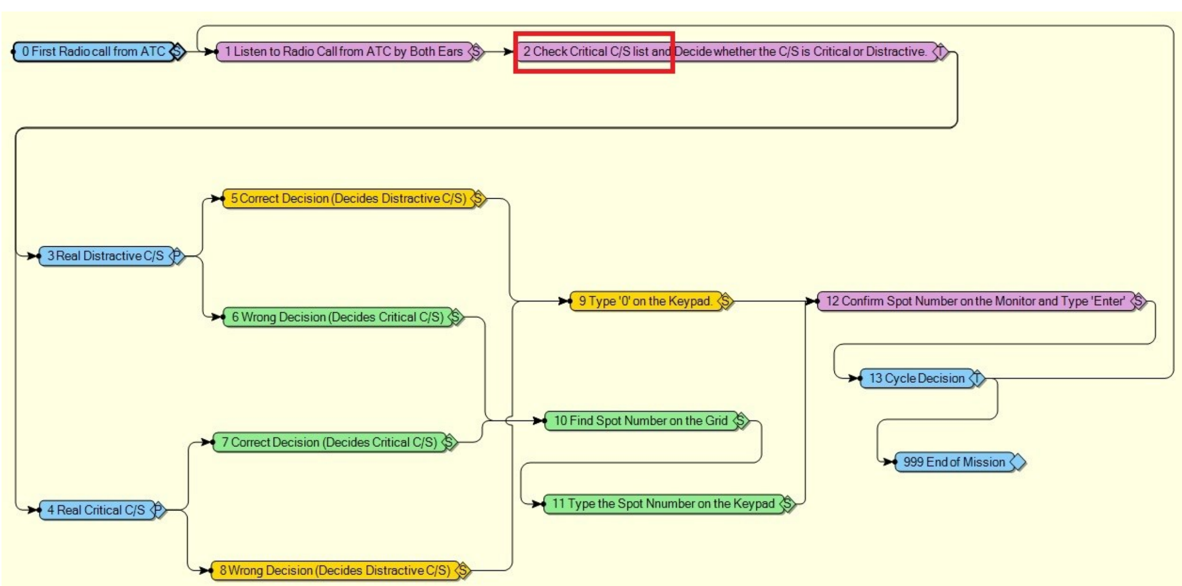


Figure 15. Task Network for Baseline model and Current audio system

Table 4. VACP Scales used in IMPRINT

| Value | Descriptors |
|---|---|
| | **<VISUAL>** |
| 0.0 | No Visual Activity |
| 1.0 | Visually Register/Detect (detect occurrence of image) |
| 3.0 | Visually Inspect/Check (discrete inspection/static condition) |
| 4.0 | Visually Locate/Align (selective orientation) |
| 4.4 | Visually Track/Follow (maintain orientation) |
| 5.0 | Visually Discriminate (detect visual difference) |
| 5.1 | Visually Read (symbol) |
| 6.0 | Visually Scan/Search/Monitor (continuous/serial inspection, multiple conditions) |
| | **<AUDITORY>** |
| 0.0 | No Auditory Activity |
| 1.0 | Detect/Register Sound (detect occurrence of sound) |
| 2.0 | Orient to Sound (general orientation/attention) |
| 3.0 | Interpret Semantic Content (speech, simple, 1-2 words) |
| 4.2 | Orient to Sound (selective orientation/attention) |
| 4.3 | Verify Auditory Feedback (detect occurrence of anticipated sound) |
| 6.0 | Interpret Semantic Content (speech, complex, sentence) |
| 6.6 | Discriminate Sound Characteristics (detect auditory differences) |
| 7.0 | Interpret Sound Patterns (pulse rates, etc.) |
| | **<COGNITIVE>** |
| 0.0 | No Cognitive Activity |
| 1.0 | Automatic (simple association) |
| 1.2 | Alternative Selection |
| 4.6 | Evaluation/Judgment (consider single aspect) |
| 5.0 | Sign/Signal Recognition |
| 5.3 | Encoding/Decoding, Recall |
| 6.8 | Evaluation/Judgment (consider several aspects) |
| 7.0 | Estimation, Calculation, Conversion |
| | **<FINE MOTOR>** |
| 0.0 | No Fine Motor Activity |
| 2.2 | Discrete Actuation (button, toggle, trigger) |
| 2.6 | Continuous Adjustment (flight controls, sensor control) |
| 4.6 | Manipulative (tracking) |
| 5.5 | Discrete Adjustment (rotary, vertical thumbwheel, lever position) |
| 6.5 | Symbolic Production (writing) |
| 7.0 | Serial Discrete Manipulation (keyboard entries) |

### *Alternative Models for Current Research*

For this model, there are two types of stimulus variables (i.e., independent variables): type of audio system and number of call sigs. Each stimulus variable has two levels; the type of audio system includes current audio system and 3D audio system, and the number of call signs includes a 3 call sign condition and a 7 call sign condition, as mentioned above. Therefore, a two-level factorial design with 2 factors was considered for this model, and this is denoted by $2^2$. Thus, four types of alternatives were considered for this model, and to distinguish conditions in the modeling and human subjects experiment, each state was referred as "Alternative" for modeling and "Condition" for the human subjects experiment.

Alternative 1: 3 Call signs with Current audio system

Alternative 2: 7 Call signs with Current audio system

Alternative 3: 3 Call signs with 3D audio system

Alternative 4: 7 Call signs with 3D audio system

By designing the baseline model, the basic configuration of these alternatives was possible. The alternative models to be used in this research further modified the baseline model. The task network for the current audio system (i.e., Baseline, Alternatives 1 and 2) was shown in Figure 15. Similarly, the task network for 3D audio system (i.e., Alternatives 3 and 4) is shown in Figure 16. In the Figure 16, Nodes 14 and 15 were added, and Nodes 1 and 2 were modified. Because of the characteristic of the 3D audio system and the assumptions, distractive information is provided to an operator's left ear, and critical information is provided to an operator's right ear. These are captured as Nodes 14 and 15, and these nodes serve the same role as Node 1 in Figure 15. Therefore, Node 1 in Figure 16 indicates only the start of a new cycle. By using the 3D audio system, if the system has 100% reliability, an operator does not need to check the critical call sign list, thus, the modified Node 2 in Figure 16 reflected this condition. Detailed data input description of these alternative models are explained in Appendix F.

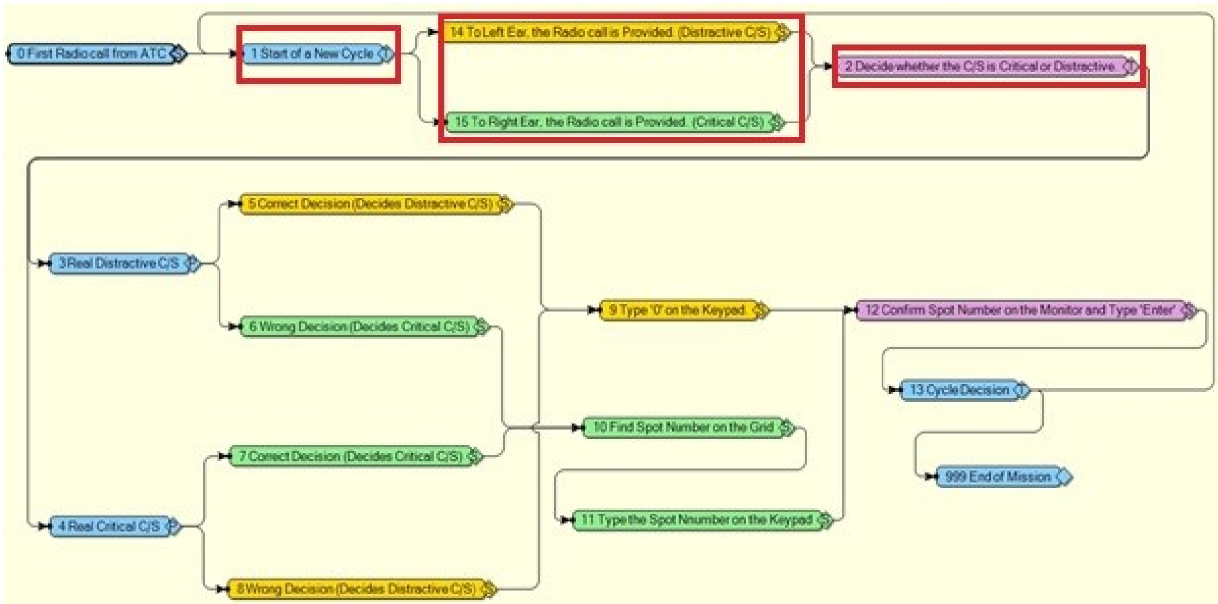

Figure 16. Task Network for 3D audio system

**Model Output Data Analysis**

By analyzing the results of this model, anticipated results of human subjects experiment under an ideal environment can be described. To statistically analyze the resultant VACP values for workload and performance data for response time, two-factor repeated measures analysis of variance (i.e., ANOVA) was applied. This was because the two factors, type of audio system and number of call signs, affected workload and response time results for this model. By using the two-factor repeated measures ANOVA, the effects of the 3D audio system and number of call signs on operators' workload levels and response times in an ideal environment could be predicted.

Specifically, for workload analysis, VACP value for each node was multiplied by task time of the node, then, the resultant values of one operator in one alternative were summed.

After that, the summed value was divided by total time (150 seconds; 30 instructions × 5seconds/instruction). Finally, each value for one alternative of one participant was applied to ANOVA. Although resultant workload values for human subjects experiment cannot be separated according to the type of information (i.e., critical or distractive information), resultant response times for human subjects experiment can be separately analyzed to explain 3D audio's effects on the different type of information. For this reason, all resultant response time data from the modeling results were employed in the statistical analysis, instead of comparing among operators.

## Human Subjects Experiment

The human subjects experiment used the Air Force Laboratory's Multi-Modal Chat (MMC) Monitor Client Program software (Finomore and others, 2010) to measure human subjects' response time and accuracy. In addition, the experiments employed the NASA-TLX (NASA Task Load Index) and SWORD (Subjective Workload Dominance Technique) to assess each participant's subjective workload.

### *Participants*

Twenty four subjects (2 females and 22 males; 3 manned-aircraft pilots and 21 non-pilots) with ages between 22 and 39 (*Mean* = 29.042, *SD* = 4.439) participated in the study. All of the subjects were fluent in English, and had no known hearing deficiency. Due to the characteristic of the 3D audio system, the participants were required to be capable of distinguishing when instructions were provided to the left, right, or both ears, and the ability was evaluated in an early pre-test. Participants were voluntarily recruited through e-mail and notice on a website for company grade officers across Wright-Patterson Air Force Base.

### *Experimental Design*

Four independent variables were manipulated in this human subjects experiment: type of audio system, number of critical call signs, reliability of the 3D audio system, and announcement of possible errors. To measure the effect of the 3D audio system, different types of audio systems and different numbers of critical call signs were provided; the type of audio system included either current or 3D audio system; the number of critical call signs included either 3 or 7. The purpose of this research is not to identify the difference in human response between the 3 and 7 call sign conditions. Instead, the number of call signs was manipulated to determine if the differences between the current audio system condition and the 3D audio system condition were consistent as the number of call signs was increased. This is explained in Figure 17, and this figure describes expected response time results. Under the current audio system, the 7 call sign condition was expected to require a longer response time than the 3 call sign condition as indicated by the blue line in Figure 17. The difference between the response times as a function of the number of call signs is shown as ①. Similarly, under the 3D audio system, the 7 call sign condition was expected to require a longer response time than 3 call sign condition as indicated by gold dotted line. The difference between response time for the 3 and 7 call sign conditions for the 3D audio system is shown as ②. The purpose of this experiment is not to understand the magnitude of ① or ②, but, to compare the magnitude of ① and ②, as it was expected that the difference ② will be less than ①. That is, it was expected that the 3D audio system would permit larger improvements in human response time as the number of call signs increased.

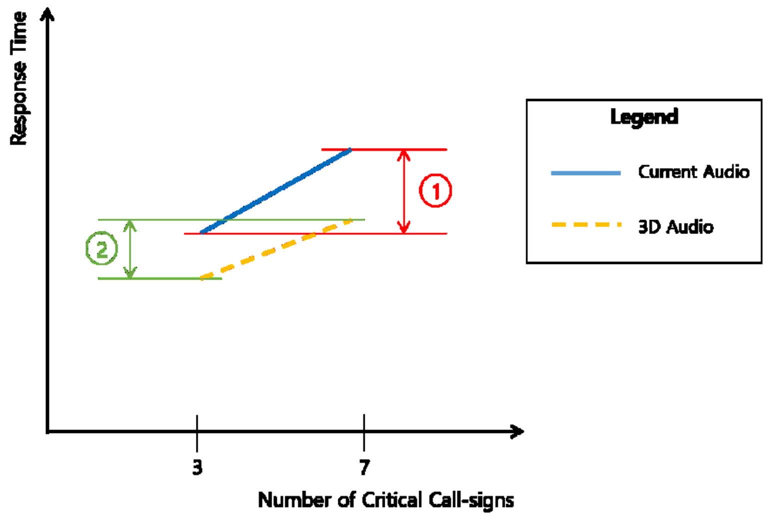


Figure 17. Expected Response Time Results

In addition, the reduction in the reliability of the 3D audio system was considered as a third independent variable. Therefore, different error rates were applied, including no errors and 4 errors per condition (6.7% error rate, with 2 false alarms and 2 misses of the system). Initial pilot experiments included no errors and 6 errors (10% error rate, with 3 false alarms and 3 misses of the system) conditions. However, after the pilot experiments, subjects mentioned that the 10% error rate was too high to trust the system. They also mentioned that they believed that they should have always referred to the critical call sign list after experiencing three or four errors. For this reason, the error rate was reduced. In contrast, if only two errors (3.3% error rate, with 1 false alarm and 1 miss of the system) are applied, the rate was considered to be too low to analyze the results. Additionally, to measure the effect of the fourth independent variable, the announcement of possible errors, in the 3D audio system two different conditions were applied to a subject: no announcement and announcement.

According to the investigative questions introduced in Chapter I, operator's workload, response times, and accuracy were collected from this human subjects experiment, providing the dependent variables. The workload was calculated by subjects' scored subjective assessments from NASA-TLX and SWORD. The response time corresponds to the duration from the start of the ATC's instruction to the operator's completion of his or her tasks for one instruction and was calculated separately for critical and distractive call signs, as well as for all trials within an experimental condition. Accuracy indicates how well an operator conducts his or her tasks. Although the model did not independently produce accuracy because the results would be just from the input probabilities, the human subjects experiment was expected to collect the accuracy results.

The experimental design included a full factorial of the 2 audio systems and the 2 number of call sign conditions such as the modeling. However, as mentioned, each state is referred to as "Condition" for the human subjects experiment to distinguish conditions in the human subjects experiment from the modeling alternatives. To measure the effect of the reduction in the reliability of the 3D audio system and the effect of the announcement of possible errors of the 3D audio system, two conditions were added to the initial 4 model alternatives: Condition 5 and Condition 6, providing the following list of experimental conditions:

Condition 1: 3 Call signs with Current audio system;

Condition 2: 7 Call signs with Current audio system;

Condition 3: 3 Call signs with 3D audio system;

Condition 4: 7 Call signs with 3D audio system;

Condition 5: 7 Call signs with 3D audio system
+ Announcement of possible errors + No errors;

Condition 6: 7 Call signs with 3D audio system
+ Announcement of possible errors + 4 errors.

For the error-related conditions, the 3 call sign conditions were not considered, because the investigative questions did not treat the difference between the numbers of critical call signs under the error-related conditions. Instead, this research focused on the effects of the error announcement and the reliability level. The effects of the announcement of the possible errors were explored by comparing the results between Conditions 4 and 5. The effects of the reduction in the reliability of the 3D audio system were evaluated by comparing the results between Conditions 5 and 6.

***Apparatus***

Experiments were conducted using the Multi-Modal Chat (MMC) Monitor Client Program developed by Air Force Laboratory (Finomore and others, 2010), which is a Windows software program that monitors and parses Extensible Messaging and Presence Protocol (XMPP) messages containing transcriptions of radio communications and text chat messages. The program has several features designed to improve the performance of operators including the 3D audio, chat windows that contain the text version of radio calls, and a logging function that records all data from MMC and outputs them to an Excel spreadsheet. The MMC chat window has the ability to provide a visual indicator; for example, when one ATC instruction is provided to a subject's left ear, a light in the left box is turned on. However, this functions was hidden to the subjects to explore the effect of the auditory indicator only. The only thing that the participants could see on the laptop monitor was the numbers that they were typing by using a keypad. Therefore, they could correct the numbers, when they recognized that they typed wrong numbers before pressing the 'Enter' key. However, after they pressed the 'Enter' key, they were expected to move to the next instruction.

The experiments were conducted in a 6 ft × 6 ft cubicle in a quiet laboratory to minimize distractions. A Bose QC15 noise cancelling headphone and a laptop were used to present the instructions using the Multi-Modal Chat (MMC) Monitor Client Program. A ten-digit number keypad was also given to the participants. The keypad contained a number grid with four rows and three columns, and it was used for participants to type the spot numbers (i.e., location number corresponding to ATC instruction). Figure 18 shows the cubicle laboratory, the headphone, the keypad, and the MMC chat window. Additionally, as mentioned in the overall scenario and assumption, before starting the experiments, the critical call sign list and the grid were provided to the participants, and they were located at a comfortable position for the participants. The participants were instructed not to memorize their critical call signs, but to refer to their call sign list for each experimental condition.


Figure 18. Cubicle Laboratory, Headphone, Keypad, and MMC Window

### *Experimental Procedure*

A within-subject design was applied, thus, each subject was tested in all conditions to minimize individual variations. To minimize learning or fatigue effects, they were randomly assigned to one of four different groups. Group ‘A’ followed the original condition-order: Condition 1-2-3-4-5-6. However, to remove the learning effect of the system error for Group ‘B’, the order of Conditions 5 and 6 was changed, resulting in conditions ordered as 1-2-3-4-6-5. Additionally, before conducting the experiments, participants did not know which condition between Conditions 5 and 6 had real errors, and they did not know the error rate of Condition 6;

the announcements for Conditions 5 and 6 were "This condition may or may not have errors." For Group 'C', to remove the learning effect of the audio systems, a different order was applied, resulting in condition order of 3-4-1-2-5-6. For Group 'D', to remove the learning effect of both the system error and the audio systems, the conditions were ordered as 3-4-1-2-6-5. These orders are arranged as shown in Table 5. Among the independent variables, the number of critical call signs and the announcement of possible errors did not affect the order of the conditions that the participant groups followed. As mentioned before, the number of critical call signs, itself, was not important for this research. In addition, error-related conditions (i.e., Conditions 5 and 6) were intentionally assigned late in the sequence. This assignment was made as the importance of reliability and announcement were secondary to the primary research question and it was believed that other orders would bias the results of the effect of the 3D audio system.

Table 5. Order of Conditions for Each Group

| **Subjects Group** | **Order of Conditions** |
|---|---|
| Group 'A' | 1 – 2 – 3 – 4 – 5 – 6 |
| Group 'B' | 1 – 2 – 3 – 4 – 6 – 5 |
| Group 'C' | 3 – 4 – 1 – 2 – 5 – 6 |
| Group 'D' | 3 – 4 – 1 – 2 – 6 – 5 |

After assignment, participants were provided with the informed consent document and asked if they had any questions after reading the document. The participants were then given a short explanation of the software and their tasks. Before the hearing test, the participants had approximately one minute to experience the 3D audio sound whose sequence was left ear, right ear, and both ears. They were then permitted to adjust the volume of the audio system to their comfort level. Then, the participants received a simple hearing test, and for the hearing test 9 instructions were provided to each participant. The 9 instructions, which had the same format as the real experimental tasks, included 3 left-ear-instructions, 3 right-ear-instructions, and 3 both-ears-instructions, with an order that was randomly assigned. Each participant's ability to hear and respond correctly to the spatial location of the sound was evaluated before continuing with the human subjects experiment. If a participant was unable to perform these tasks correctly, he or she was given the option to adjust the volume of the audio before repeating the trial. If unable to complete the task a second time, the participant was excused from the experiment. After this evaluation, the participant was given two one-minute practice sessions, which included the current audio system and the 3D audio system for one minute each. These practice sessions were designed to minimize the possibility of a learning effect.

After a two-minute break, the participant started the experiments according to the order of conditions of the participant's group. The experiments followed the overall scenario mentioned in this chapter. Each instruction was provided to one participant every 5 seconds. To complete one participant's experiments within one hour, 60 instructions were provided to one participant in every experimental condition; half for critical instructions, and the other half for distractive instructions, but the participants did not know the ratio of the critical to distractive instructions. Because one participant conducted six experimental conditions, they completed 360 experimental trials. And, in every interval between conditions, the participant was requested to conduct NASA-TLX workload assessment and then, he or she received a two-minute break. After completion of all experimental conditions, the participant completed a SWORD workload questionnaire and a brief questionnaire about the usability of the 3D audio system.

To prevent the participants from habituating to certain experimental conditions (i.e., call signs, voices, and grid numbers) and to prevent them from being affected by additional factors, several methods were applied. The kinds of critical call signs and distractive call signs used in each condition were shuffled, and the critical call sign list was provided to a participant immediately before the start of each condition, providing little to no time to memorize the critical call signs. Nineteen different call signs were used for the critical and distractive call signs. Seventeen different voices were recorded for the radio calls as ATC's instructions, and applied throughout the experiment so that the participants could not perform the tasks simply by responding to a given voice. In addition, the same number of occurrences of each voice was assigned to every condition, to minimize the differences of any recorder's speaking speed according to his or her speaking habit. Finally, in every condition, a different grid was used, so that the participants could not memorize the grid numbers. Although the spot numbers in the grid were shuffled in each condition, all grids included the same spot numbers to minimize the effect of typing different combinations of numbers.

**Human Subjects Experiment Output Data Analysis**

As mentioned above, three kinds of data were drawn from the results of human subjects experiment to answer the investigative questions: workload, response time, and accuracy. While the response time and the accuracy were collected from the MMC Monitor Client Program, the workload was collected from additional calculation of NASA-TLX and SWORD values based on the participants' subjective assessment.

After completing all conditions, each participant was required to rank the importance of the 6 NASA-TLX scales (i.e., mental demand, physical demand, temporal demand, performance, effort, and frustration) to determine relative weights. Then, the participant's NASA-TLX ratings were multiplied by the appropriate weight and summed to determine a composite NASA-TLX score for each condition. The SWORD value represents normalized relative workload from each subject. Because the sum of the SWORD values which were assessed by one subject should be 1, the values were re-calculated for each analysis. For example, to analyze the results from Conditions 1 through 4, the sum of all normalized values that one subject assessed for Conditions 1 through 4 should be 1; and to analyze the results from Conditions 4 and 5, the sum of the normalized values for Conditions 4 and 5 should be 1. Therefore, the SWORD values were re-calculated.

The data analysis sought to understand the effect of the type of audio system, number of call signs, reliability of the 3D audio system, and announcement of possible errors on user performance and workload. To statistically analyze the resultant workload and response times, two-factor repeated measures analysis of variance (i.e., ANOVA) was applied. And, accuracy, which was recorded as a binary response for each trial was analyzed using chi-square test or Fisher exact probability test according to a percentage of cells which has an expected frequency of less than 5 (Siegel, 1956: 96-111; 175-179).

First, by comparing Conditions 1 and 3, and by comparing Conditions 2 and 4, the effects of the 3D audio system on UAV operators' workload and performance were explained. Through this analysis, investigative questions 1 and 2 could be addressed. Additionally, the difference between the results for Conditions 1 and 2 was compared with the difference between Conditions 3 and 4 to explain how increasing the number of call signs affects operator's workload and performance under different audio system applications. Investigative questions 3 and 4 could be addressed through this analysis. Further, by comparing Conditions 5 and 6, the effect of the reduction in the reliability of the 3D audio system on operators' workload and performance was explored. Through this comparison, investigative questions 5 and 6 could be explained. Finally, to explain the effects of the announcement of possible errors of the 3D audio system on operators' workload and performance, Conditions 4 and 5 were compared to answer investigative questions 7 and 8.

# IV. Results

## Overview

This chapter details the results of the simulation modeling and the human subjects experiment. As mentioned before, to distinguish conditions in the modeling and human subjects experiment, each state was referred as "Alternative" for the modeling and "Condition" for human subjects experiment. First, the modeling results will be discussed, including results for "Alternative 1" through "Alternative 4." Then, the human subjects experiment results will be described, including results for "Condition 1" through "Condition 6." Alternatives 1 through 4 are directly comparable to Conditions 1 through 4. However, Conditions 5 and 6 are error-related conditions for which performance was not predicted through model results. In each section, overall results will be first shown, then statistical comparison will be conducted to answer the investigative questions mentioned in Chapter I. Discussions and conclusions will be provided in the subsequent chapter.

In addition, two types of charts will be used in this chapter to visualize the results: boxplots as shown in the left panel of Figure 19, and line graphs as shown in the right panel of Figure 19. As shown, the box plots will represent the mean (circle), median (center line), upper and lower quartile (box limits), as well as minimum and maximum value as indicated by the extent of the error bars. For these box plots, outliers were defined as any value greater than 3/2 times of upper quartile and less than 3/2 times of lower quartile, but these were omitted for visual simplification. In the line graph, as depicted in the right panel of Figure 19, the line connects the means of conditions, and the error bars indicate plus and minus one standard error from the mean.

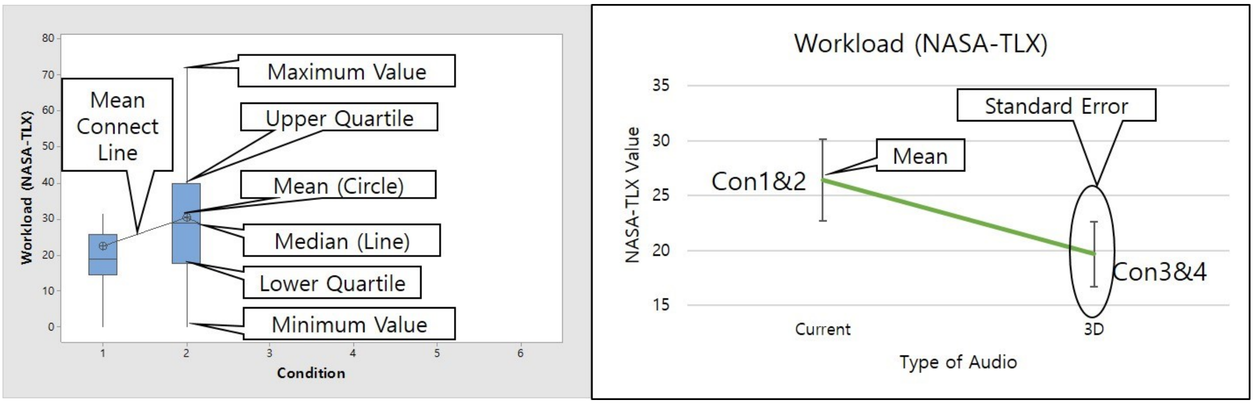


Figure 19. Example Charts used in Chapter IV and Definition

## Modeling Results

Expected results of human subjects experiment under 100% reliability were collected from the model results. As mentioned before, in this modeling exercise, four alternatives which are exactly the same as the first four conditions in the human subjects experiment were modeled to produce estimates of workload and response time for each alternative.

### *Predicted Workload*

As mentioned, to reflect workload in the model, VACP values were input for each node of the task network shown in Figure 15 for Alternatives 1 and 2, and the task network shown in Figure 16 for Alternatives 3 and 4. Table 6 shows means and standard deviations of resultant VACP values from modeling. These data are plotted using box plots as shown in Figure 20.

Table 6. Means and Standard Deviations of VACP values for Model

| Alternative | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Mean | 4.889 | 5.311 | 3.548 | 3.548 |
| Standard Deviation | 0.074 | 0.110 | 0.061 | 0.061 |

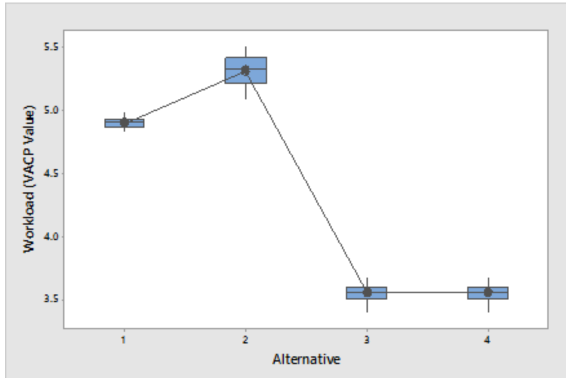

Figure 20. Boxplots of Predicted Workload

To statistically analyze VACP values, results from Alternatives 1 through 4 were subjected to a two-factor repeated measures ANOVA. The repeated measures ANOVA with type of audio system and number of call signs that one operator handled indicated significant main effects of type of audio ($F(1,24)=11876.81$, $p=0.000$) and number of call signs ($F(1,24)=220.49$, $p=0.000$). As shown in the left panel of Figure 21, VACP values were lower for the 3D audio system than the current audio system. Although VACP value increased as a function of the number of call signs, this finding does not inform the utility of the 3D audio system in current research, because in this analysis one circumstance has both types of audio system. The ANOVA also showed an interaction between type of audio and number of call signs ($F(1,24)=220.49$, $p=0.000$) as shown in the right panel of Figure 21. Post hoc Pairwise Tukey Comparisons indicated that the VACP value for Alternative 2 was significantly higher than that for Alternative 1 ($p=0.000$), while VACP values for Alternatives 3 and 4 were not significantly different ($p=1.000$). Furthermore, the Tukey Pairwise Comparison also showed that the VACP value for Alternative 1 was significantly higher than for Alternative 3 ($p=0.000$), and the VACP value for Alternative 2 was significantly higher than for Alternative 4 ($p=0.000$).

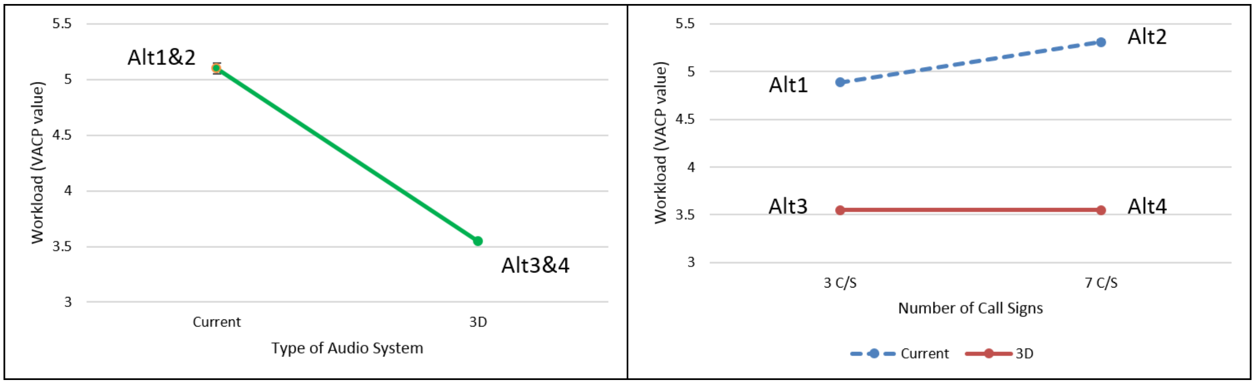


Figure 21. Model Workload Comparison

***Predicted Response Time***

Response times for critical call signs and distractive call signs were analyzed separately. To do so, all data were included in the statistical analysis, instead of comparing just among subjects. Table 7 shows mean response times and standard deviations for overall call signs, critical call signs, and distractive call signs. Figure 22 shows boxplots of response times for critical and distractive information as predicted by this model.

Table 7. Mean Response Times (seconds) and Standard Deviation for Overall, Critical and Distractive Information for Model

| Information | Alternative | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Overall | Mean | 3.493 | 3.636 | 2.577 | 2.577 |
| | Standard Deviation | 0.571 | 0.621 | 1.231 | 1.231 |
| Critical | Mean | 4.008 | 4.096 | 3.786 | 3.786 |
| | Standard Deviation | 0.238 | 0.428 | 0.284 | 0.284 |
| Distractive | Mean | 2.978 | 3.175 | 1.368 | 1.368 |
| | Standard Deviation | 0.257 | 0.403 | 0.157 | 0.157 |

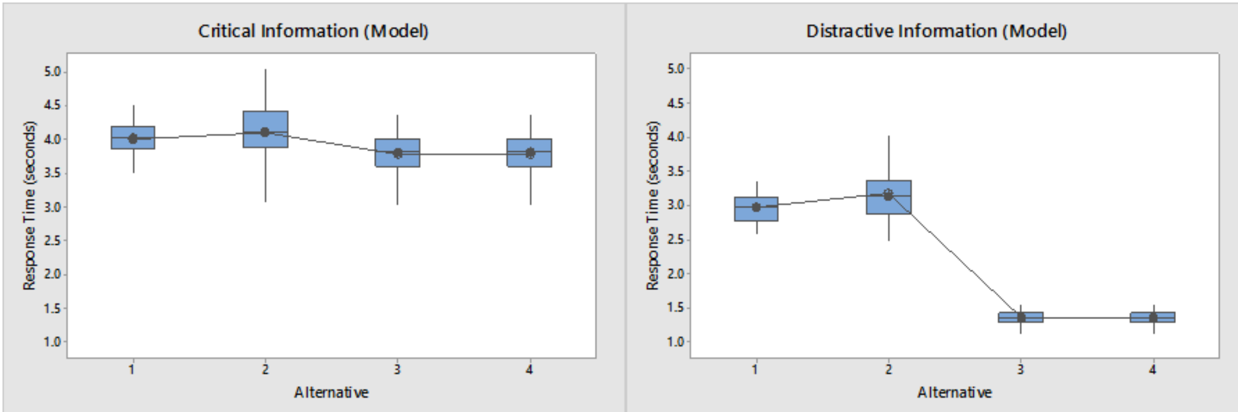


Figure 22. Boxplots of Predicted Response Times for Critical and Distractive Information

First, response times for critical call signs were analyzed. A repeated measures ANOVA with type of audio system and number of critical call signs as a within-subjects factors showed significant main effects of type of audio (F(1,24)=266.61, p=0.000) and number of call signs (F(1,24)=7.39, p=0.007). The means and standard errors as a function of the type of audio system are shown in the left panel of Figure 23. The finding that increasing the number of call signs significantly increased response time was expected but does not have significant implications for the current research, because in this analysis one circumstance has both types of audio system. Importantly, however, the ANOVA also indicated an interaction between type of audio and number of call signs (F(1,24)=7.39, p=0.007), as shown in the right panel of Figure 23. Post hoc Pairwise Tukey Comparisons showed that response time for Alternative 2 was significantly longer than response time for Alternative 1 (p=0.001), while response times for Alternatives 3 and 4 were not significantly different (p=1.000). The Tukey Pairwise Comparison also showed that the response time for Alternative 1 was significantly longer than the response time for Alternative 3 (p=0.000), and that response time for Alternative 2 was significantly longer than response time for Alternative 4 (p=0.000).

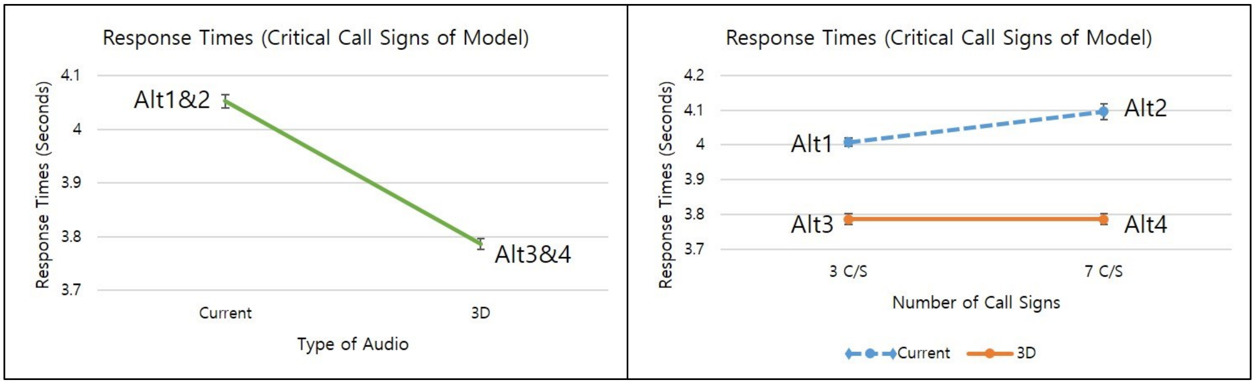


Figure 23. Model Response Times Comparison (Alternatives 1 through 4) for Critical C/S

Next, response times for distractive call signs were analyzed. A repeated measures ANOVA with type of audio system and number of critical call signs that one operator owns as a within-subjects factor showed significant main effects of type of audio ($F(1,24)=15839.45$, $p=0.000$) as shown in the left panel of Figure 24, and number of call signs ($F(1,24)=52.43$, $p=0.000$). Further, interaction between type of audio and number of call signs ($F(1,24)=52.43$, $p=0.000$) was also significant as shown in the right panel of Figure 24. Post hoc Pairwise Tukey Comparisons showed that response time for Alternative 2 was significantly longer than response time for Alternative 1 ($p=0.000$), while response times for Alternatives 3 and 4 were not statistically different ($p=1.000$). The Tukey Pairwise Comparison also showed that response time for Alternative 1 was significantly longer than response time for Alternative 3 ($p=0.000$), and that response time for Alternative 2 was significantly longer than response time for Alternative 4 ($p=0.000$).

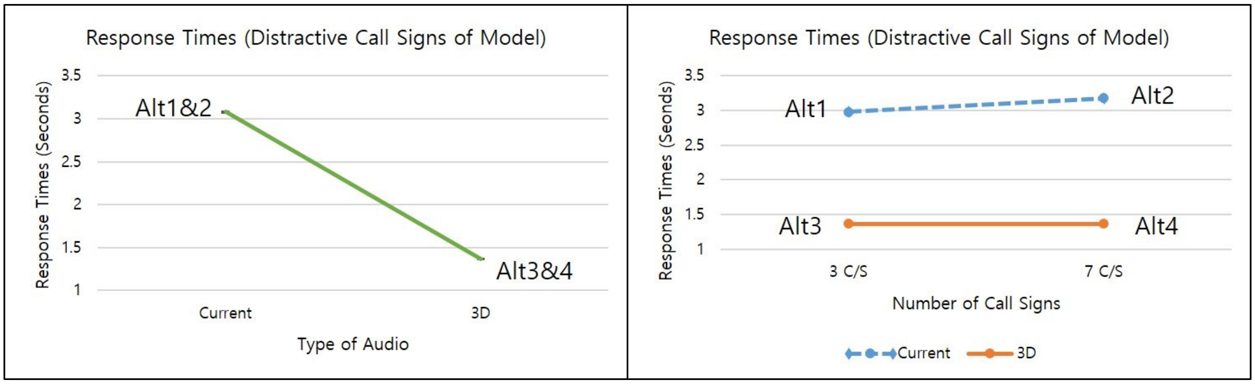


Figure 24. Model Response Times Comparison (Alternatives 1 through 4) for Distractive C/S

### *Necessity of Error-Related Alternatives in Modeling*

In addition to the four modeled alternatives, two other error-related conditions were applied in the human subjects experiment. To anticipate expected results of the error-related conditions from modeling, many cases can be considered. Some participants may entirely rely on the 3D audio system despite being informed that it may present information with errors. Others may completely disregard the 3D audio system and use the same procedure as they apply with the current audio system. Yet others may apply a hybrid approach. Therefore, without any data associated with error-related conditions any model would be constructed based on presumption and is unlikely to be of value. Therefore, the error-related alternatives were not modeled. Instead, the results of error-related conditions from the human subjects experiment were relied upon to understand this effect. However, the expected values were anticipated to lie within the envelope defined by the lines for the interaction in Figure 23 and Figure 24. That is, in the human subjects experiment, the participants would rely on the automation, producing results similar to the 3D audio condition or disregard the automation, producing results similar to the current audio condition.

## Human Subjects Experiment Results

### *Workload*

Workload was measured from each subject in two ways: NASA-TLX and SWORD. Table 8 shows means and standard deviations for both NASA-TLX and SWORD. These data are plotted using box plots as shown in Figure 25.

Table 8. Means and Standard Deviations for NASA-TLX and SWORD for Each Experimental Condition.

| Condition | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| NASA-TLX | Mean | 22.389 | 30.403 | 19.361 | 19.986 | 30.750 | 31.431 |
| | Standard Deviation | 16.748 | 19.630 | 14.905 | 14.722 | 18.942 | 19.297 |
| SWORD | Mean | 0.109 | 0.220 | 0.061 | 0.085 | 0.232 | 0.293 |
| | Standard Deviation | 0.061 | 0.088 | 0.027 | 0.051 | 0.067 | 0.085 |

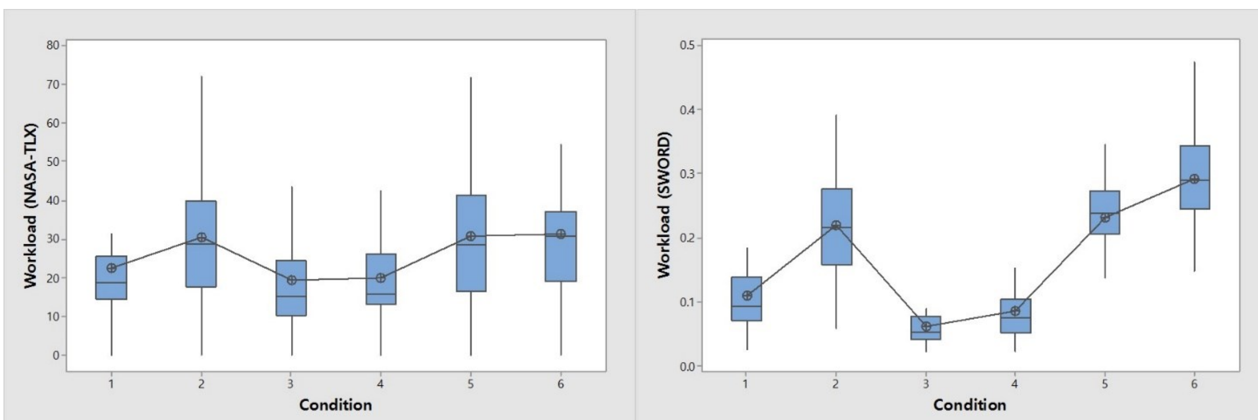


Figure 25. Boxplots for NASA-TLX and SWORD Values

To statistically compare NASA-TLX values in conditions which are not related with errors, results from Conditions 1 through 4 were subjected to a two-factor repeated measures ANOVA. The repeated measures ANOVA with type of audio system and number of critical call signs indicated significant main effects of type of audio ($F(1,23)=27.66$, $p=0.000$) and number of call signs ($F(1,23)=11.42$, $p=0.001$). As shown in the left panel of Figure 26, NASA-TLX values were lower for the 3D audio system than the current audio system. Although NASA-TLX value increased as a function of the number of call signs, this finding does not inform the utility of the 3D audio system. The ANOVA also showed an interaction between type of audio and number of call signs (F(1,23)=8.35, p=0.005) as shown in the right panel of Figure 26. Post hoc Pairwise Tukey Comparisons indicated that the NASA-TLX value for Condition 2 was significantly higher than that for Condition 1 (p=0.000), while NASA-TLX values for Conditions 3 and 4 were not significantly different (p=0.986). Furthermore, the Tukey Pairwise Comparison also showed that the NASA-TLX value for Condition 2 was significantly higher than for Condition 4 (p=0.000), while the NASA-TLX values for Conditions 1 and 3 were not significantly different (p=0.345). Additionally, NASA-TLX values for Conditions 1 and 4 were not significantly different (p=0.548).

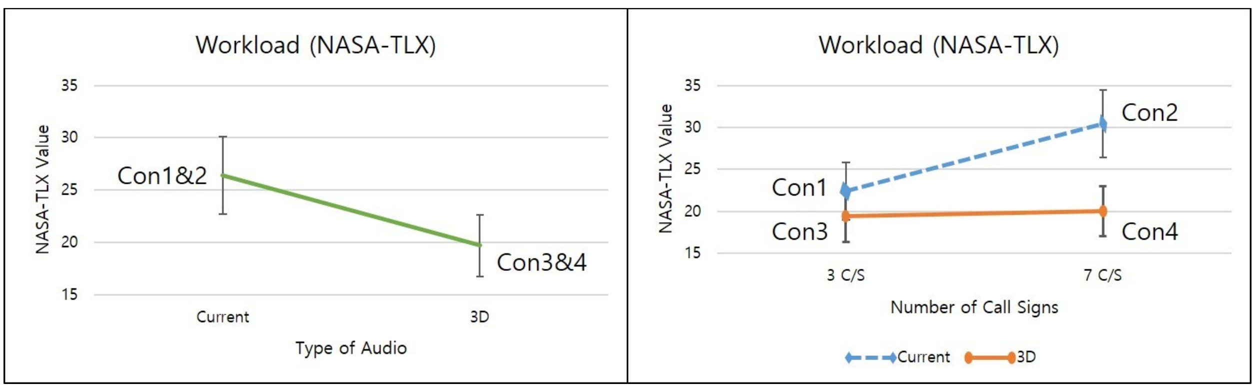


Figure 26. NASA-TLX Values Comparison (Conditions 1 through 4)

Similarly to the NASA-TLX analysis, a repeated measures ANOVA was conducted for SWORD results. It indicated significant main effects for type of audio (F(1,23)=70.49, p=0.000) and number of call signs (F(1,23)=45.24, p=0.000). As shown in the left panel of Figure 27, the SWORD value was lower for the 3D than for the current audio system. There was also an interaction between type of audio and number of call signs (F(1,23)=17.28, p=0.000) as shown in the right panel of Figure 27. Post hoc Pairwise Tukey Comparisons showed that the mean SWORD value for Condition 2 was significantly higher than the mean SWORD value for

Condition 1 (p=0.000), while the mean SWORD values for Conditions 3 and 4 were not significantly different (p=0.274). The Tukey Pairwise Comparison also showed that the mean SWORD value for Condition 1 was significantly higher than that for Condition 3 (p=0.019), and that the mean SWORD value for Condition 2 was significantly higher than that for Condition 4 (p=0.000). Additionally, mean SWORD values for Conditions 1 and 4 were not significantly different (p=0.641).

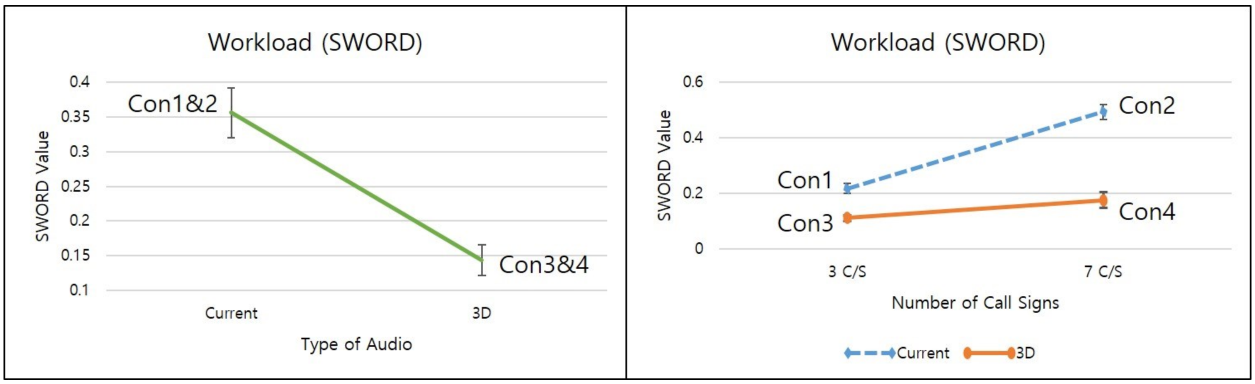


Figure 27. SWORD Values Comparison (Conditions 1 through 4)

To investigate the effect of announcement of possible errors on operators' workload, mean NASA-TLX and SWORD values were compared for Conditions 4 and 5. Paired t-tests indicated that mean NASA-TLX (t(23)=-5.06, p=0.000) and SWORD (t(23)=-6.69, p=0.000) values for Condition 5 were significantly higher than those values for Condition 4, as shown in Figure 28. More specifically, additional paired t-tests between Conditions 2 and 5 were conducted to investigate the extent of increased workload level for Condition 5; the NASA-TLX value for Condition 5 was not significantly different from the NASA-TLX value for Condition 2 (t(23)=-0.24, p=0.816); but the SWORD value for Condition 5 was significantly higher than the SWORD value for Condition 2 (t(23)=-2.48, p=0.021).

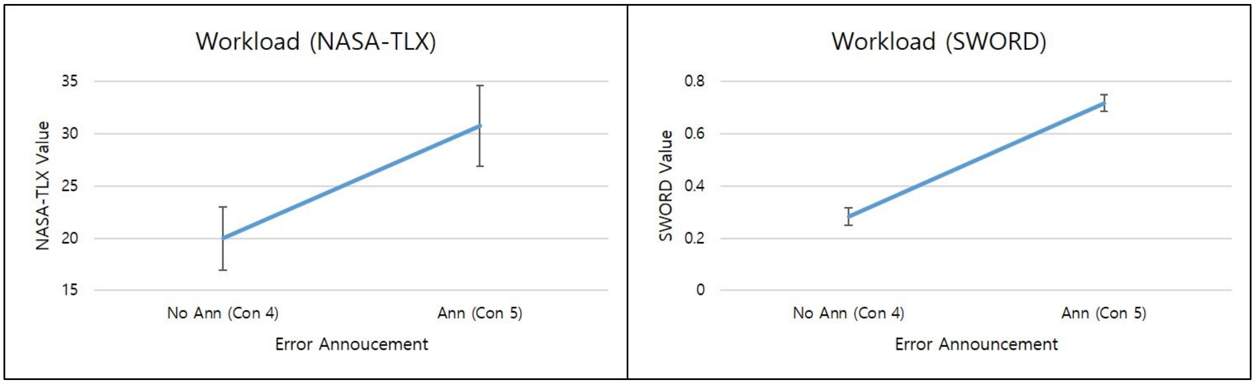


Figure 28. NASA-TLX and SWORD Values Comparison (Conditions 4 and 5)

To explain the effect of reduction in reliability on operators' workload, the results from Conditions 5 and 6 were compared. For the NASA-TLX results, paired t-tests indicated that mean NASA-TLX values were not significantly different between Conditions 5 and 6 ($t(23)=-0.62$, $p=0.538$), as shown in the left panel of Figure 29, and additionally, mean NASA-TLX values were not significantly different between Conditions 2 and 6 ($t(23)=-0.77$, $p=0.450$). However, for the SWORD results, paired t-tests indicated that mean SWORD values were significantly different ($t(23)=-2.97$, $p=0.007$), and the mean SWORD value for Condition 6 was significantly higher than the mean SWORD value for Condition 5 as shown in the right panel of Figure 29, and additionally, the mean SWORD value for Condition 6 was significantly higher than the mean SWORD value for Condition 2 ($t(23)=-3.77$, $p=0.001$).

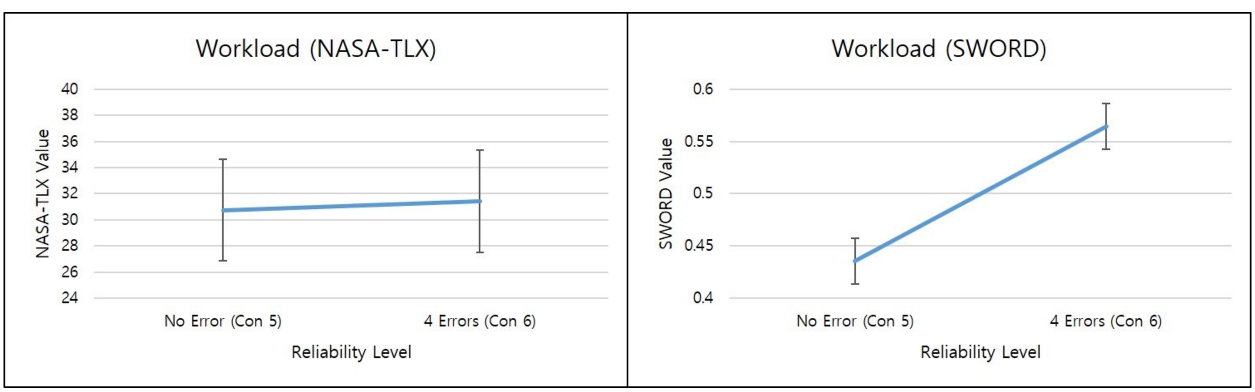


Figure 29. NASA-TLX and SWORD Values Comparison (Conditions 5 and 6)

### ***Response Times***

In this section, response times are analyzed across all subject responses to provide an overall value. Critical and Distractive call sign conditions are then separated and the response times are analyzed separately for each condition category. Table 9 shows means and standard deviations for response times for overall call signs, critical call signs, and distractive call signs. And, Figure 30 shows boxplots of response times for critical and distractive information.

Table 9. Mean Response Times (seconds) and their Standard Deviations for Overall, Critical and Distractive Information

| Information | Condition | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Overall | Mean | 3.156 | 3.451 | 2.877 | 2.954 | 3.404 | 3.368 |
| | Standard Deviation | 0.287 | 0.375 | 0.284 | 0.267 | 0.410 | 0.330 |
| Critical | Mean | 3.958 | 4.051 | 3.921 | 4.037 | 4.358 | 4.299 |
| | Standard Deviation | 0.322 | 0.358 | 0.337 | 0.312 | 0.478 | 0.346 |
| Distractive | Mean | 2.354 | 2.851 | 1.833 | 1.871 | 2.450 | 2.436 |
| | Standard Deviation | 0.312 | 0.441 | 0.274 | 0.292 | 0.383 | 0.356 |

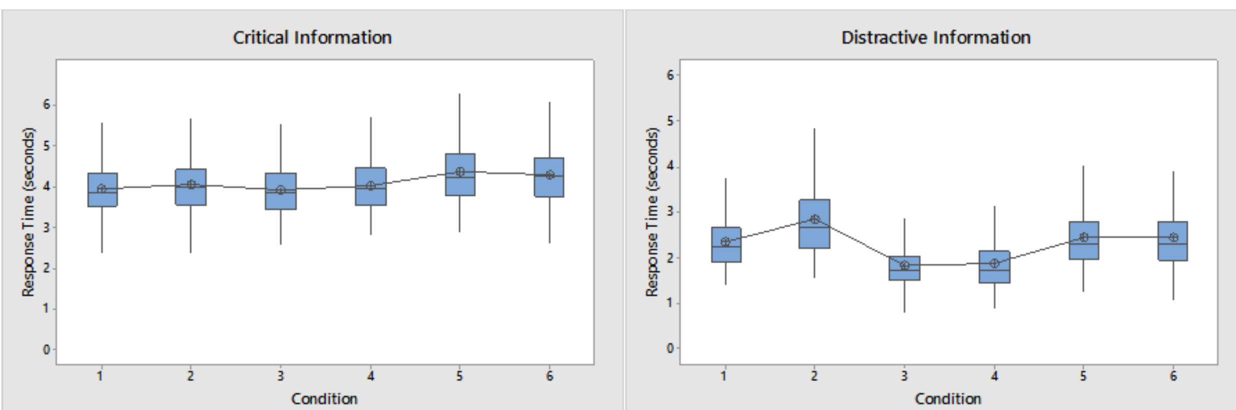


Figure 30. Boxplots of Response Times (Critical Information and Distractive Information)

First, response times for critical call signs were analyzed. To statistically compare response times in conditions which are not related with automation errors, results from Conditions 1 through 4 were subjected to a two-factor repeated measures ANOVA. The repeated measures ANOVA with type of audio system and number of critical call signs that one operator owns as a within-subjects factor showed a significant main effect of number of call signs (F(1,23)=21.00, p=0.000). However, as mentioned above, this does not inform the utility of the 3D audio system. No significant differences were found between conditions for type of audio (F(1,23)=1.28, p=0.258) as shown in the left panel of Figure 31. And, there was not an interaction between type of audio and number of call signs (F(1,23)=0.24, p=0.626) as shown in the right panel of Figure 31. Post hoc Pairwise Tukey Comparisons showed that response time for Condition 2 was significantly longer than response time for Condition 1 (p=0.020), and response time for Condition 4 was significantly longer than response time for Condition 3 (p=0.002). Additionally, the Tukey Pairwise Comparison also showed that response times for Conditions 1 and 3 were not significantly different (p=0.662), and response times for Conditions 2 and 4 were not significantly different either (p=0.969).

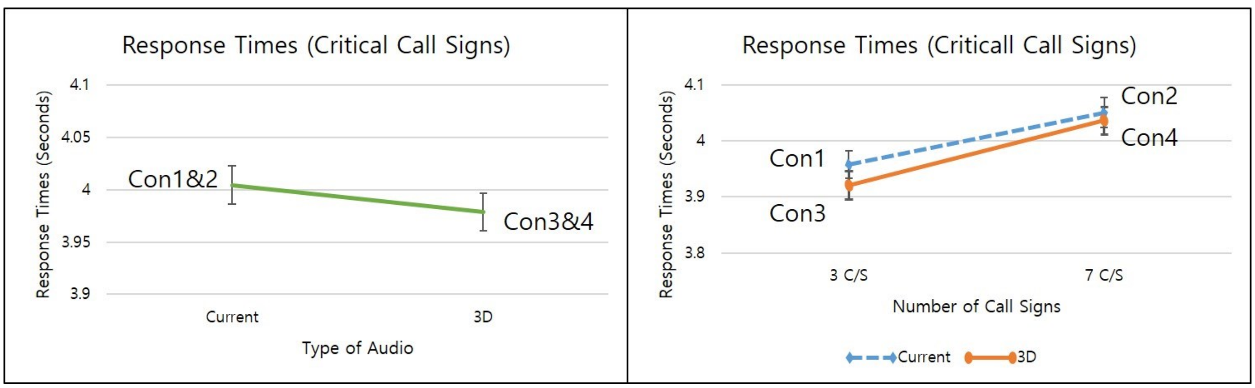


Figure 31. Response Times Comparison (Conditions 1 through 4) for Critical Call Signs

To draw the effect of announcement of possible errors on an operator's response times for critical information, the results from Conditions 4 and 5 were compared. A paired t-test indicated that response times were significantly different between Conditions 4 and 5 (t(719)=-8.90, p=0.000), and Condition 5 took significantly longer than Condition 4 as shown in the left panel of Figure 32. And, additional t-test was conducted to investigate the extent of the increased response time for Condition 5; the response time for Condition 5 was significantly longer than the response time for Condition 2 (t(719)=-9.02, p=0.000).

In addition, to explain the effect of the reduction in reliability on an operator's response times for critical information, the results from Conditions 5 and 6 were compared. A paired t-test indicated that response times were not significantly different between Condition 5 and Condition 6 (t(719)=-1.65, p=0.099), as shown in the right panel of Figure 32. Also, the response time for Condition 6 was significantly longer than the response time for Condition 2 (t(719)=-7.53, p=0.000).

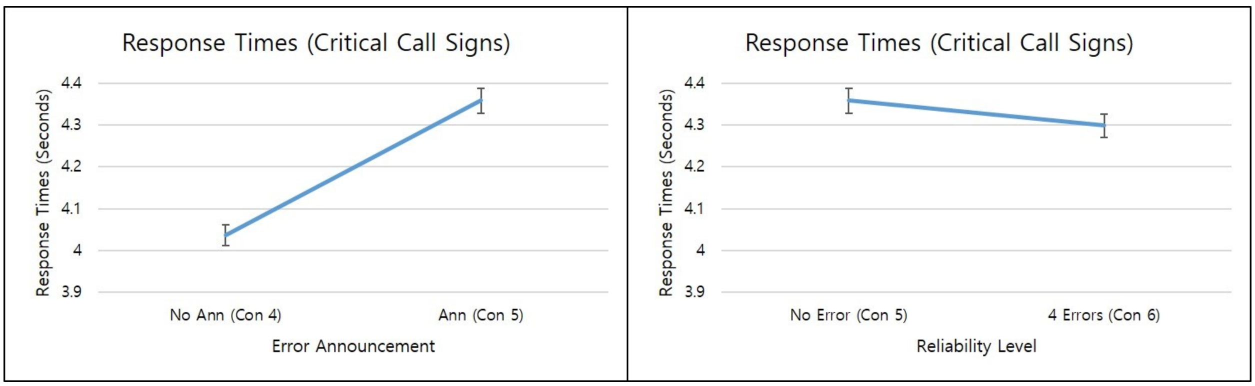


Figure 32. Response Times Comparison (Conditions 4&5, and Conditions 5&6) for Critical C/S

Next, response times for distractive call signs were analyzed. To statistically compare response times in conditions which are not related with errors, results from Conditions 1 through 4 were subjected to a two-factor repeated measures ANOVA. The repeated measures ANOVA with type of audio system and number of critical call signs showed significant main effects of type of audio (F(1,23)=1028.48, p=0.000) as shown in the left panel of Figure 33, and number of call signs (F(1,23)=130.55, p=0.000). There was also an interaction between type of audio and number of call signs (F(1,23)=96.19, p=0.000) as shown in the right panel of Figure 33. Post hoc Pairwise Tukey Comparisons showed that response time for Condition 2 was significantly longer than response time for Condition 1 (p=0.000), while response times for Conditions 3 and

4 were not significantly different (p=0.662). The Tukey Pairwise Comparison also showed that response time for Condition 1 was significantly longer than response time for Condition 3 (p=0.000), and that response time for Condition 2 was significantly longer than response time for Condition 4 (p=0.000).

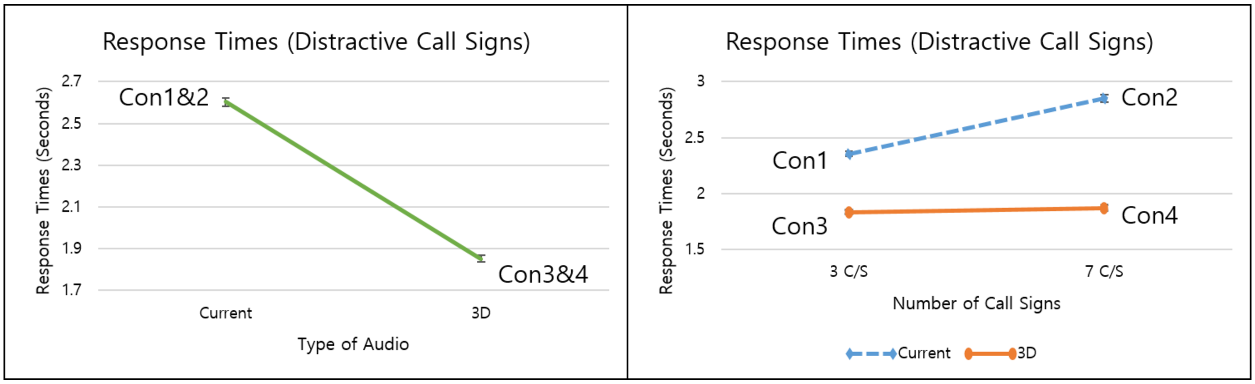


Figure 33. Response Times Comparison (Condition 1 through 4) for Distractive Call Signs

To describe the effect of announcement of possible errors on an operator's response times for distractive information, the results from Conditions 4 and 5 were compared. A paired t-test indicated that response time was significantly different between Conditions 4 and 5 (t(719)=-18.17, p=0.000), and response time for Condition 5 was significantly longer than response time for Condition 4 as shown in the left panel of Figure 34. However, the response time for Condition 5 was still significantly shorter than the response time for Condition 2 (t(719)=11.31, p=0.000), but it was significantly longer than the response time for Condition 1 (t(719)=-3.01, p=0.003); that is, the response time for Condition 5 lied between response times for Conditions 1 and 2.

Additionally, to explain the effect of reduction in reliability on an operator's response times for distractive information, the results from Conditions 5 and 6 were compared. A paired t-test indicated that response times were not significantly different between Conditions 5 and 6 (t(719)=0.48, p=0.628), as shown in the right panel of Figure 34. Once again, the response time for Condition 6 was significantly shorter than the response time for Condition 2 (t(719)=11.43, p=0.000), but it was significantly longer than the response time for Condition 1 (t(719)=-2.53, p=0.011); that is, the response time for Condition 6 also lied between response times for Conditions 1 and 2.

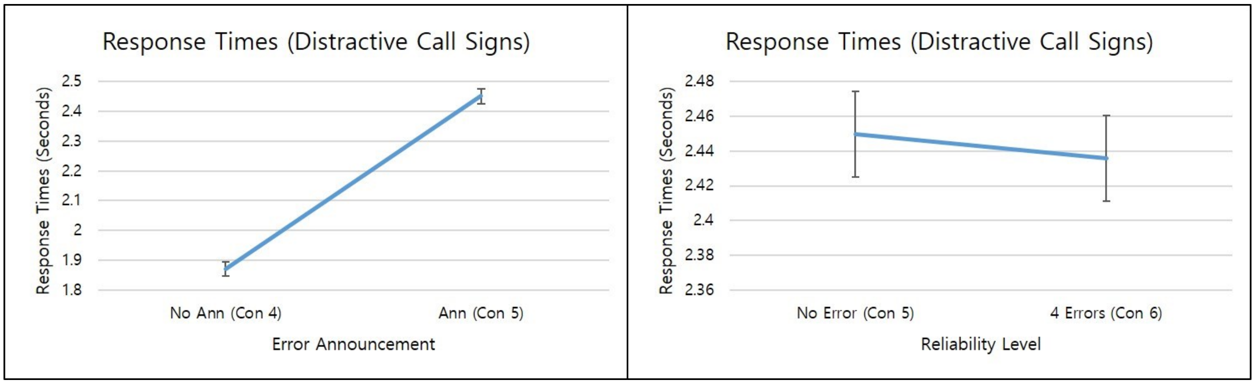


Figure 34. Response Times Comparison (Condition 4&5 and Condition 5&6) for Distractive C/S

Additionally, only for Condition 6, response time results were analyzed according to the "Signal Detection Theory for the voice recognition system": "Hit" represents the voice recognition system's critical output from real critical call sign; "Miss" represents the system's distractive output from real critical call sign; "Correct Rejection" represents the system's distractive output from real distractive call sign; and "False Alarm" represents the system's critical output from real distractive call sign. And, as mentioned in the previous chapter, two "Misses" and two "False Alarms" were applied to this Condition 6. Table 10 shows means and standard deviations for response times according to the signal detection theory for the voice recognition system, and Figure 35 shows their boxplots.

Table 10. Means and Standard Deviations for Response Times according to Signal Detection Theory for Voice Recognition System

| Signal Detection Theory for Voice Recognition System | Hit | Miss | Correct Rejection | False Alarm |
|---|---|---|---|---|
| Mean (seconds) | 4.271 | 4.694 | 2.406 | 2.857 |
| Standard Deviation (seconds) | 0.742 | 0.743 | 0.645 | 0.767 |

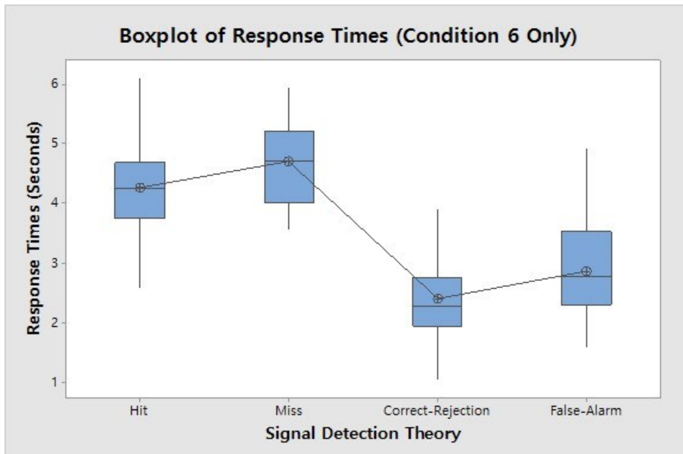


Figure 35. Boxplot of Response Times according to Signal Detection Theory (Condition 6)

A repeated measures ANOVA with type of signal as a within-subjects factor showed a significant main effect of type of signal (F(3,23)=1076.40, p=0.000) as shown in Figure 36. Post hoc Pairwise Tukey Comparisons showed that response time for Misses was significantly longer than response time for Hits (p=0.000), and response times for False Alarms were significantly longer than response times for Correct Rejections (p=0.000).

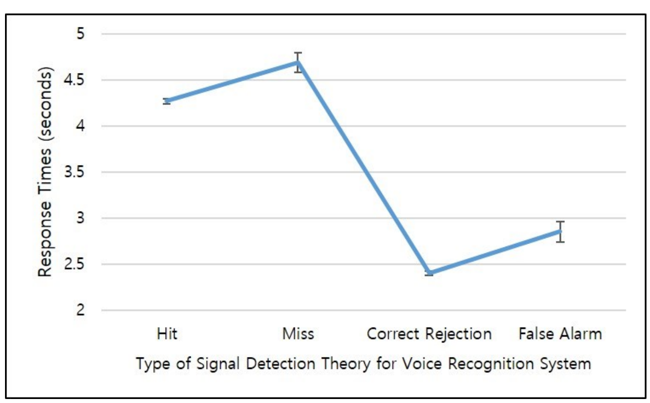


Figure 36. Response Times Comparison (Condition 6) according to Signal Detection Theory

***Accuracy***

In this section, accuracies for critical and distractive call signs are analyzed separately to explain the 3D audio's different effects on the different types of information. Table 11 shows means and standard deviations for accuracy for overall call signs, critical call signs, and distractive call signs. Figure 37 shows graphs of accuracy for each condition according to the distinction of information (i.e., critical and distractive information).

Table 11. Accuracy (%) for Overall, Critical and Distractive Information

| Information | Con 1 | Con 2 | Con 3 | Con 4 | Con 5 | Con 6 |
|---|---|---|---|---|---|---|
| Overall | 98.06 | 98.89 | 99.51 | 98.96 | 98.82 | 99.17 |
| Critical | 96.25 | 98.89 | 99.31 | 98.75 | 97.92 | 98.47 |
| Distractive | 99.86 | 98.89 | 99.72 | 99.17 | 99.72 | 99.86 |

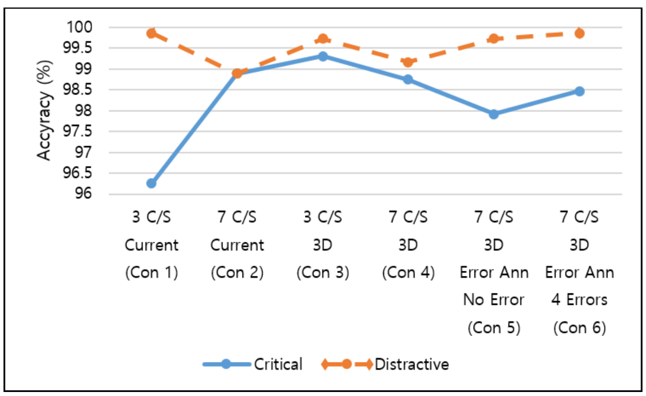


Figure 37. Accuracy for Critical and Distractive Information

First, accuracies for critical call signs were analyzed. To statistically compare the results for accuracy between conditions, chi-square tests were employed. The chi-square tests showed that the accuracy for Condition 3 was significantly higher than the accuracy for Condition 1 ($p=0.000$), while no significant difference was found between the results for Conditions 2 and 4 ($p=0.807$). In addition, chi-square tests revealed that the accuracy for Condition 2 was significantly higher than that for Condition 1 ($p=0.001$), while no significant difference was found between the results for Conditions 3 and 4 (p=0.283). Furthermore, a chi-square test showed that there was no significant difference between accuracies for Conditions 4 and 5 (p=0.217). A chi-square test also revealed that there was no significant difference between accuracies for Conditions 5 and 6 (p=0.429).

Next, accuracies for distractive call signs were analyzed. A chi-square test showed that there was no significant difference between accuracies for Conditions 2 and 4 (p=0.591). For the chi-square tests, fewer than 20% of the cells should have an expected frequency of less than 5. If more than 20% of the cells have an expected frequency of less than 5, Fisher exact probability tests should be employed (Siegel, 1956: 96-111; 175-179). Therefore, other statistical analyses for distractive information were conducted by this Fisher exact probability test. The Fisher exact probability test revealed that no significant difference can be found between accuracies for Conditions 1 and 3 (p=0.250). In addition, Fisher exact probability tests showed that the accuracy for Condition 2 was significantly lower than the accuracy for Condition 1 (p=0.017), while no significant difference was found between the results for Conditions 3 and 4 (p=0.109). Additionally, a Fisher exact probability test showed that there was no significant difference between accuracies for Conditions 4 and 5 (p=0.109). A Fisher exact probability test also revealed that there was no significant difference between accuracies for Conditions 5 and 6 (p=0.375).

Finally, similarly to the response time analysis, the results of accuracy were analyzed according to the Signal Detection Theory categories for the voice recognition system for Condition 6. Table 12 shows the accuracies according to the signal detection theory for the voice recognition system. A chi-square test showed that the accuracy for Correct Rejections was significantly higher than that for Hits (p=0.001). Except for this combination, all the other combinations had no significant difference by Fisher exact probability tests.

Table 12. Accuracies according to Signal Detection Theory for Voice Recognition System

| Signal Detection Theory for Voice Recognition System | Hit | Miss | Correct Rejection | False Alarm |
|---|---|---|---|---|
| Accuracy (%) | 98.36 | 100.00 | 100.00 | 97.92 |

***Results from Survey***

After completing all conditions, every subject was provided with 6 survey questions about the usability of the 3D audio system.

First question was “Do you think the 3D audio is helpful for a UAV operator to reduce his or her workload? And why?” For the question, 22 subjects out of 24 subjects (91.7%) answered “Yes.” However, 8 subjects among the 22 subjects who answered “Yes” (36.4%) qualified their response with the statement “If no errors are present.” Most answerers mentioned that the 3D audio would help catch only relevant information, and operators can easily ignore distractive information. However, there were two subjects who answered “No.” They mentioned that they were more focused on which ear was hearing instructions, so that the subjects were rushed in acting on the information, or even missed the information.

Second question was “Do you think the 3D audio is helpful for a UAV operator to reduce his or her response times? And why?” and all subjects (100.0%) answered “Yes.” However, similarly to the 1$^{st}$ question, 6 of them (25.0%) also mentioned “If no errors are present.” They stated that they can easily and quickly disregard irrelevant information when they used the 3D audio system.

Third question was “Do you think the 3D audio is helpful for a UAV operator to increase his or her accuracy? And why?” For this question, 22 subjects (91.7%) answered “Yes.” And, similarly to the previous questions, 10 of the 22 subjects who answered “Yes” (45.5%) mentioned “If no errors are present.” Participants mentioned that the 3D audio gave less chance of misreading or mishearing, and sometimes it provided them with a double check. However, one participant among the 2 subjects who answered “No” mentioned that the accuracy would depend on the reliability of the system.

Fourth question was “If you were a UAV operator and the 3D audio system does not have any error, would you want to use the 3D audio system? And why?” And, all subjects (100.0%) answered “Yes.” They mentioned the reasons as the 3D audio decreased workload, stress, and response times, and as the 3D audio made their job easier. Some of them mentioned that the 3D audio would be helpful for long term jobs, and that it would provide a high degree of confidence in the performance of the operator.

Fifth question was “If you were a UAV operator and the 3D audio system may have errors, would you still want to use the 3D audio system? And why?” For this question, only 7 subjects (29.2%) answered “Yes.” Among the participants who answered “Yes,” three subjects (42.9%) stated that it would depend on the error rate, and only if the errors are very rare, they would use the 3D audio system. But, the remaining four of them indicated that the 3D audio could still give them some general information whether a call sign is critical or distractive, and it could be a good initial indicator of whether the information has importance or noise. Seventeen of the subjects (70.8%) did not want to use the 3D audio system, if it may have errors. They explained that if the 3D audio may have errors, the 3D audio system would induce feelings of tiredness, confusion, or distraction, because they would either follow the same procedure as they had with the current audio system or even double-check everything to confirm. So, they indicated that the 3D audio had no benefit or it made the task more difficult than just using current audio, when errors could be present.

Finally, sixth question was “If you have any other comments about the 3D audio system and/or this experiments, please feel free to write them.” For this question, some subjects mentioned that if the 3D audio has benefits from the results of this research, it should be applied to other communication platforms with much higher workload levels.

# V. Discussion, Recommendation, and Conclusion

## Discussion

### *Workload*

This research employed VACP values for modeling, and NASA-TLX and SWORD values for human subjects experiment to assess workload. As predicted based upon the VACP results produced by the model, the 3D audio decreased operators' workload as compared to the current audio conditions when no errors were present, as indicated by the statistically lower mean NASA-TLX and SWORD values that were observed for the 3D audio system as compared to the current audio system. Further, as predicted from the model's VACP values, the operators' workload, as measured using both NASA-TLX and SWORD, did not change as a function of the number of call signs in the 3D audio system condition, while the operators' workload increased as a function of an increasing the number of call signs when using the current audio system. Additionally, based upon the results of the experimentally-obtained NASA-TLX and SWORD values, it would appear that when the operator must respond to 7 call signs, workload for the 3D audio condition can be reduced to a value as low as that produced for the 3 call sign condition when using the current audio system.

Although the modeled VACP results, measured NASA-TLX, and measured SWORD results were in general agreement, their results differed when comparing the 3 call sign current audio condition to the 3 call sign 3D audio condition, with VACP and SWORD indicating that the workload was lower for the 3D audio condition and the NASA-TLX indicating that no difference was present. This difference may be because the NASA-TLX was conducted directly after every completion of each condition, so subjects focused on the condition that they had just experienced and did not compare the relative workload across condition. This argument can be supported by Gluckman who indicated that NASA-TLX does not provide information concerning the relative change in workload under varying conditions, while alternative measures of workload such as SWORD do (Gluckman and others, 1993:8). Therefore, it is possible that NASA-TLX did not provide the ability to reliably differentiate the difference in workload between these two conditions.

The NASA-TLX and SWORD values for Conditions 4 and 5 indicated that the announcement of possible errors increased operators' workload. Specifically, from the NASA-TLX results, the workload for the 3D audio with announcement of possible error condition was increased to a value as high as that produced for the 7 call sign condition using the current audio system, but from the SWORD results, it was increased to a value higher than the 7 call sign condition using the current audio system.

In addition, NASA-TLX did not indicate a significant difference in workload between Condition 5, where participants were told that errors might be present but errors were not, and Condition 6, where the participants were told that errors might be present and errors existed; from the NASA-TLX results, the workload for Condition 6 was as high as the 7 call sign condition using the current audio system. However, SWORD indicated that the workload for Condition 6 was higher than condition 5. Further, the SWORD results indicated the workload for Condition 6 was higher than that produced for the 7 call sign condition using the current audio system. Again, for the same reason mentioned above, the SWORD may be considered more reliable.

***Response Time***

According to the results from modeling, response times were expected to be significantly shorter for the 3D audio system than the current audio system, regardless of the distinction of instructions (i.e., critical or distractive) and the number of call signs. The results from modeling also showed that regardless of the distinction of instructions, the response time is not expected to increase as a function of the number of call signs for the 3D audio system, while the response time is expected to increase as the number of call signs increases for the current audio system.

These model results were predictive of the human subjects experiment results for the distractive information. However, for the critical information, the model results differed from the human subjects experiment results. This difference could be due to smaller variance, which were present in the model results than the results from the current experiment. When the model was constructed for this research, response times and their minimum values were based on and validated by means and standard deviations from earlier research (Amaddio and others, 2015:195-200). In Amaddio's experiment, each participant's mean response time was calculated across all responses for each experimental condition from each participant. These mean values were then subjected to analysis. In contrast, this research analyzed all data from all participants as repeated measures during the analysis. Therefore, much of the variability in Amaddio's data was removed in calculating the mean response time, thus the variability in response time within the current analysis is significantly larger than reported by Amaddio. The larger variance in the present study then reduced the power of the current statistical analysis resulting in the finding that, no significant differences were found between response times for the two types of audio system for critical information. Additionally, for distractive information, the response time was not affected by the increased number of call signs under the 3D audio system, while the response time did increase as the number of call signs was increased under the current audio system. However, for critical information, response times increased as a function of increasing the number of call signs both under the current audio system and the 3D audio system. Furthermore, for distractive information, the 3D audio significantly reduced response time under the same number of call sign conditions, regardless of the number of call signs. On the other hand, for critical information, the 3D audio did not reduce response time significantly.

With these results, it can be concluded that the participants applied the 3D audio system to filter out distractive information. This interpretation is consistent with the results of survey from participants, as 11 subjects out of 24 (45.8%) stated for the 2nd survey question (the usability of the 3D audio for reducing response time) that they could easily and quickly disregard irrelevant information when using the 3D audio system. However, they likely confirmed the presence of critical call signs on the critical call sign list rather than entirely relying upon the 3D audio cue to answer the critical instructions, because theoretically there would be no difference in response time for both critical and distractive instructions between the 3 and 7 call sign conditions under the 3D audio system, as mentioned in the description for creating Alternative 4 model in Appendix F.

Based upon this finding, during the design of a system to parse an incoming audio stream to present the information to either of the operators' ears, if the parser is not completely reliable, it might be desirable to bias the parser towards providing distractive call signs in the ear intended to receive critical call signs. Such a bias should then be more likely to present distractive information to the ear the user expects critical call signs and not to present critical information to the ear the user expects to receive distractive call signs. Under these conditions, although the system makes an error that distractive information is provided to the operator's one ear intended to receive critical information, the operator will likely detect the error as they likely confirm the presence of the call sign that he or she heard on the critical call sign list rather than entirely relying upon the 3D audio cue. In contrast, if the system makes an error that critical information is provided to an operator's ear intended to receive distractive information, the operator may not be able to detect the error, because he or she disregards the information presented to the ear intended to receive distractive information.

In addition, regardless of the distinction of information, the announcement of possible errors made operators' response times longer, and the response time was not affected by the reliability level. Specifically, for critical information, the response times for error-related conditions (i.e., Conditions 5 and 6) were longer than the response times for current audio conditions. However, for distractive information, the response times for error-related conditions were still shorter than the response time for the 7 call sign condition using the current audio system, but they were longer than the response time for the 3 call sign condition using the current audio system; that is, they lied in between response times for the 3 and 7 call sign conditions using the current audio system. Moreover, it took more time for the participants to respond to the information which had automation induced errors, compared to the information which did not have any errors, regardless of the real distinction of the information.

#### *Accuracy*

According to the results from the human subjects experiment, while the 3D audio increased accuracy for critical information compared to current audio under the 3 call sign conditions, the 3D audio did not affect accuracy for critical information under the 7 call sign condition. There were slightly different results for distractive information; the 3D audio did not affect accuracy for distractive information, regardless of the number of call signs. It should be pointed out, however, that the 3D audio's effect on operators' accuracy was minor as accuracies were very high as shown in Table 11 and Figure 37, for all conditions with the lowest accuracy value exceeding 96%.

Although the number of call signs was increased, the accuracy for critical information did not significantly change under the 3D audio system, and the same result could be shown for distractive information under the 3D audio system. However, for critical information under the current audio system, the accuracy for the 7 call sign condition was significantly higher than the accuracy for the 3 call sign condition. This might be attributed to learning as the 7 call sign condition was always provided to the subjects after the 3 call sign condition. In contrast, for distractive information under the current audio system, the accuracy was significantly lower for the 7 call sign condition than the 3 call sign condition.

Not only the announcement of possible errors but also the reliability level did not affect operators' accuracy, regardless of the real distinction of information. Furthermore, the results from the Signal Detection analysis for the voice recognition system showed that for instructions without errors, accuracy for distractive information was significantly higher than accuracy for critical information, but the other combinations did not have any significant difference.

**Recommendation**

The human subjects experiment indicated that the 3D audio cues provided by the proposed system can reduce UAV operators' workload and response times when having to listen for and respond to multiple call signs among a large number of distractors. One especially interesting discovery was that the operators' workload and performance generally were not influenced by the number of call signs while using the 3D audio system. That is, the cues provided by the 3D audio system permits the operator to respond to the perceptual cues rather than to perform the time consuming task of comparing the call sign to a list of critical call signs. This modification of the work process permits the operators' workload and performance to be constant, regardless of the number of UAVs the operator controls. Although it would be necessary to demonstrate this result in a more realistic environment, the results are encouraging in that it would indicate a technology to aid operator performance and workload to be leveraged during re-design of future multi-aircraft control systems. Rather than increasing the number of UAVs that one operator controls by merely adding an operator for transit mission as described in Figure 2, letting a UAV operator be in charge of an assigned airspace with the 3D audio system by dividing territory in the air such as current air traffic controllers might be possible, making the most of the characteristics of the 3D audio system. For example, if a UAV passes a boundary for an operator, the UAV would be handed over to the operator who is in charge of the territory as shown in Figure 38. Then, the operator would control the UAV until it moves out from his or her territory. Considering the 3D audio's characteristic (i.e., constant workload and performance regardless of the number of UAVs), this re-design of the UAV transit mission could be sensible, if the number of UAV missions is explosively increased in the future.

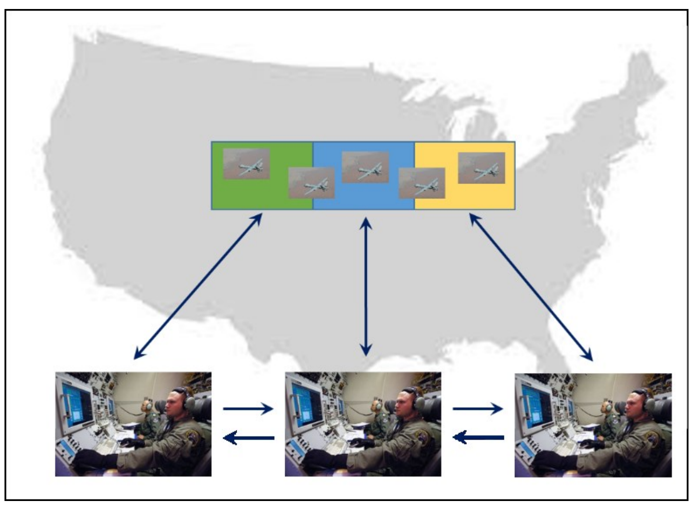

Figure 38. Re-Design of UAV Transit Mission

Although the current research sought to investigate the use of 3D audio in UAV operations, the system described in this research might have application in other domains. For example, the same system might be useful to manned-aircraft pilots. Currently, most military manned-aircraft pilots hear two or more radio frequencies simultaneously during their missions, such as control tower frequency with UHF-1 (Ultra High Frequency), mission frequency with UHF-2, squadron frequency with VHF-1 (Very High Frequency), and emergency frequency (i.e., guard) with VHF-2. Pilots often adjust the volume of these different radio frequencies to make it easier to distinguish those frequencies, but they may miss their critical information because of overlapped radio communications. However, if this 3D audio system is used, they may more easily and quickly disregard irrelevant information, regardless of the number of frequencies that they are using simultaneously, thus, allowing them to concentrate on their critical information. In a slightly different way, if a military manned-aircraft pilot can control a direction of each frequency toward his or her ear(s) such as UHF-1 to the pilot's right ear, UHF-2 to the pilot's both ears, and VHF frequencies to the pilot's left ear, the pilot may also easily and quickly distinguish those frequencies. Similarly, because civil manned-aircraft pilots also use two or more frequencies while they are flying such as control tower frequency with VHF-1, company frequency with VHF-2, and emergency frequency with VHF-3, this technology can also be employed to increase the pilots' performance with reduced workload within these environments.

However, to continually develop this 3D audio system, assuring the reliability of the 3D audio (i.e., voice recognition technology) is absolutely necessary. This is not only because many participants for this research mentioned for the survey that the 3D audio can be helpful only when the 3D audio does not have errors, but also because all participants wanted to use the 3D audio system if no errors are assured. Furthermore, although the system would have very low error rate, a function should be provided to permit the operator to select an audio system from among the current and the 3D. This function would meet the demands of some operators who do not want to use the 3D audio system in the presence of possible error conditions.

To develop these findings further, more research should be conducted, because this research based on the basic step of the 3D audio system for multi-aircraft control. First, this 3D audio system should be applied to the real console for current UAV missions, as one of participants mentioned. Also, when the 3D audio system is used with visual reference, the effects of visual reference should be investigated, because real UAV consoles usually use both visual and auditory information. Additionally, 3D audio's effects should be investigated for multi-tasking environments such as manual piloting of a UAV while hearing the 3D audio sound. UAV operators can be exposed to the multi-task environments and their willingness to rely upon the automation in the 3D audio system may differ, producing different effects when an operator is concentrating on another task. Furthermore, from the concept of this 3D audio system as shown in Figure 3, ambiguous information (i.e., both ears) should be considered together, and then the 3D audio might have more power even though it may not be able to assure a non-error state. In addition, more conditions should be tested such as larger numbers of call signs and more error rates to expose this 3D audio system to diverse environments. It would be also helpful to know which strategy each participant uses for a particular environment. Further, other media to increase UAV operators' performance with reduced workload during multi-aircraft control such as tactile signal could be considered. For the media, text and radio volume might be also considered. However, as cited in the literature review, auditory display is better for time-critical information than visual display (Simpson and others, 2004:62; Mowbray & Gebhard, 1961:115-149). And, as mentioned above for manned-aircraft pilots, when an operator hears multiple frequencies simultaneously, the volume may not that helpful to distinguish information due to overlapped communications; specifically critical information in low volume may be disregarded due to distractive information in high volume. Furthermore, low volume may be another factor to increase operators' workload, because operators may concentrate on the low volume to hear clearly to maintain situation awareness. For these reasons, it is believed the 3D audio system will provide benefits over systems which manipulate the text and radio volume to differentiate critical and distractive information.

**Conclusion**

The 3D audio technology is maturing and the implemented solutions are growing fast; at the same time, the potential is promising but still largely hidden and unexplored; and, under these premises, 3D audio is still a fertile field for research in the near future (Cengarle, 2012:138). In addition, it is also a promising field to increase human performance with reduced workload.

Based on this research, a different approach to the application of the 3D audio system in multi-aircraft control was explored, and the promising effects of the 3D audio system on multi-aircraft control were evaluated. Specifically, with the 3D audio system, UAV operators' performance could be increased with reduced workload during multi-aircraft control under transit operations. Consequently, our goal of inverting the operator/vehicle ratio could be achieved during the transit phases, and this wishful achievement could inspire other UAV mission phases' multi-aircraft control. Ultimately, more unmanned missions could be carried out under advanced technologies and interfaces. As many well-known and eminent scientists did, small changes such as the one explored in this thesis can make our future much better and more efficient.

## Appendix A. Research Assumptions

The model and experiments used in this research described a synthetic task environment, not a real-world situation. The real-world states were simplified and standardized for this research as described in the following description.

Before starting, it was assumed that one operator is seated in a fully equipped UAV-control station, which is able to control multiple UAVs. In the current research, it was also assumed that an operator controls either 3 or 7 UAVs, which are already assigned to the operator. Note that these conditions differ slightly from those applied by Amaddio (2015), who employed 5 or 7 UAVs in her research. The larger difference in the number of UAVs to be employed in the current research was anticipated to create a larger effect.

It was further assumed that the operator is provided with his or her critical call signs, and the operator recorded them on a written critical call sign list. Operators are asked not to rely on their memory but rely on the list to improve accuracy. Therefore, the operators always refer to the critical call sign list while using the current audio system. In addition, during the human subjects experiment, the critical call signs changed in every experimental condition, so participants did not have enough time to memorize their critical call signs. The operator also has a grid, and the spot numbers on the grid changed in every experimental condition during the human subjects experiment. It was assumed that the operator has already placed the critical call sign list and the grid at certain position where he or she can easily read them. Thus, the participants for human subjects experiment could read them with minimized movements of their body.

The model did not consider specific differences between critical call signs and spot numbers. Instead, the model only considered the distinction of instructions (i.e., critical or distractive) when determining tasks to be performed or the workload and time required to perform each task. Finally, this model did not consider any operator learning effects, assuming the operator's performance is constant throughout the experiment.

Assumptions were also made regarding the UAVs. Specifically, it was assumed that the assigned UAVs are moving separately under transit operations. During the transit operation, the operator receives only directional instructions from ATC every five seconds, and other information such as weather, traffic, base condition, or mission information are not provided to the operator. ATC provides one instruction for one UAV at one time. Although all UAVs are conducting automatic navigation, when an instruction is provided to the operator for one UAV, the instruction requires immediate action by the operator, and there are no execution delays. The operator is asked to type number(s) corresponding to the distinction of the call sign and the spot number in the grid, as soon as possible. There is no error in the UAV's movement, so if one operator types certain spot number, the UAV goes there without any exception.

During the model and experiments, the operators, or subjects, received only auditory information; visual or tactile information, other than reading the call signs and spot numbers was not considered. The format of the instructions was "Ready, *Charlie*, Go to *Blue One*, Now," and the italic words were flexible according to its call sign and position instruction, but consistent in all other respects.

In this research, it was assumed that typing errors do not occur. Instead, the typing errors were considered as 'wrong decisions' (i.e., bad performance). This is reasonable because the typing errors are expected to increase with increasing time pressure (i.e., high workload), and the time pressure affects the operator's bad performance.

For modeling, it was assumed that the 3D audio system does not have any system errors. That is, critical call signs were provided to the correct ear only; there was no chance for the critical call signs to be provided to the opposite ear for the model. In addition, for both model and human subjects experiment, it was assumed that there was not any ambiguous call signs which are provided to operator's both ears as mentioned in the system concept in the first chapter. By excluding these possibilities, the pure effects of the 3D audio on the operator's performance and workload could be obtained under the near-ideal conditions.

## Appendix B. Initial Model Data Input Description (Basic Structure and Response Time)

Detailed descriptions about the initial model's basic structure and data input (i.e., response times input) will follow the sequence of the nodes from Task Network shown in Figure 14.

**0. First Radio Call from ATC:** This is the starting point of this initial model. This initial model provides one operator with 30 instructions. That is, additional instructions are not provided to the operator, after one operator's completion of the 30 instructions. These 30 instructions are assigned in this node as 'Critical' or 'Distractive', and they are provided to one operator according to the sequence as shown in Table 13. This table shows that there are 15 critical instructions, and 15 distractive instructions, so the ratio of the critical instructions to the distractive instructions was 1:1. While each call sign was named specifically such as 'Charlie' or 'Eagle' in the human subjects experiment, the naming was ignored in this model, as it is only the decision of instruction distinction that affects operator workload and performance in the model, not the individual call signs. Task time and workload were not allocated to this node, because this node describes only the starting point of this model.

Table 13. Sequence of the Instruction Distinction

| Sequence | Distinction | Sequence | Distinction | Sequence | Distinction |
|---|---|---|---|---|---|
| 1 | Critical | 11 | Distractive | 21 | Distractive |
| 2 | Distractive | 12 | Distractive | 22 | Critical |
| 3 | Critical | 13 | Critical | 23 | Distractive |
| 4 | Distractive | 14 | Critical | 24 | Distractive |
| 5 | Critical | 15 | Distractive | 25 | Critical |
| 6 | Critical | 16 | Distractive | 26 | Critical |
| 7 | Critical | 17 | Critical | 27 | Distractive |
| 8 | Distractive | 18 | Distractive | 28 | Distractive |
| 9 | Distractive | 19 | Critical | 29 | Distractive |
| 10 | Critical | 20 | Critical | 30 | Critical |

**1. Listen to Radio Call from ATC by Both Ears:** This node shows that the operator listens to the directional information from ATC. One of the important variables, response time, starts to be measured from the beginning of this node. The measurement of this response time ends at the end of Node 12, where this response time corresponds to one operator's time to complete all tasks for one instruction (i.e., one cycle of the Task Network). Task time for this node was calculated by IMPRINT's 'MicroModel' tool because the response times from Amaddio's experiment did not include each task time. From the assumption, the format of this radio call was "Ready, Charlie, Go to Blue One, Now." Although seven words were used for this format, the operator may carefully listen to the first six words because the operator does not need to listen to the "now" word. 'MicroModel' tool calculated this speaking time for six words as 2.07 seconds. However, the task times should have variability, that is, some instructions take less than 2.07 seconds, and other instructions take more than 2.07 seconds. The 2.07-second can be considered as a mean time. For this, the IMPRINT provides a distribution of task time. Based upon the results from Amaddio's experiments, the response times were distributed as approximately normal as shown in Figure 39 and Figure 40. However, if normal distribution is used, theoretically, infinite positive or negative time may be applied to the model, which is not practical. Therefore, it was important to include a time limit for the response time variable including both a maximum and minimum.

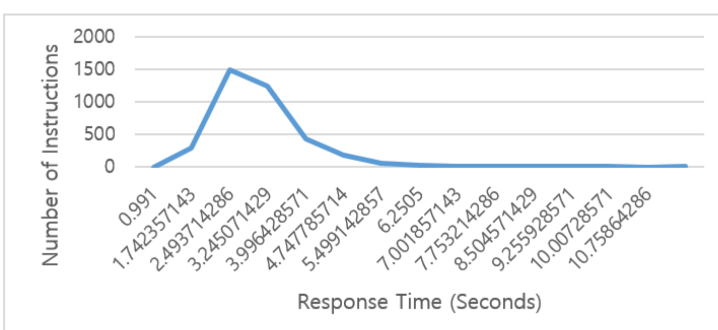


Figure 39. Distribution of Response Times for Typing '0' from Amaddio's Experiments

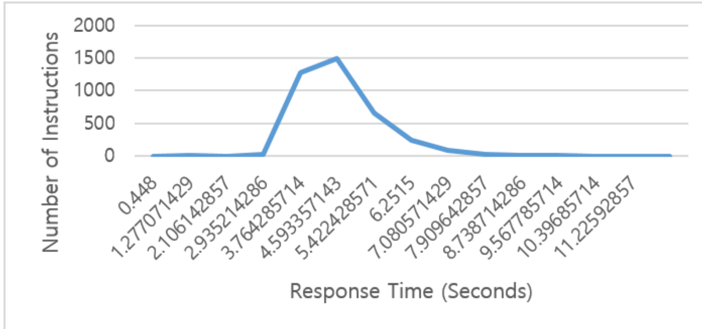


Figure 40. Distribution of Response Times for Typing 'Spot Number' from Amaddio's Experiments

Both triangular distribution and rectangular distribution provide a minimum and maximum. Further, each of these distributions results in an approximately normal distribution in IMPRINT. However, if a triangular distribution is used, the resulting variance of the distribution is too smaller than the variance observed from the prior experiment. A more representative variance is provided when the input distribution is rectangular. Therefore, a rectangular distribution was used for this model's response time variable input. This made the distribution of the response time variable similar to the normal distribution, however, there were still limitations such as maximum and minimum.

For the rectangular distribution in IMPRINT, mean and minimum values for each task time were required. To calculate these values, the 'Empirical Rule' was used (Milton and Arnold, 2003:118-120). According to the rule, 95% of values are within $\mu \pm 2\sigma$ where the population is approximately normal, therefore, the 95% interval can be collected from Amaddio's research. From the results of her 5 call sign with PI condition, $\mu$ (i.e., mean) of response time variable was 3.338, and $\sigma$ (i.e., standard deviation) was 0.342. By applying the rule, the 95% interval was between 2.654 and 4.022. The minimum value, 2.654, was almost 20% less than the mean; the maximum value, 4.022, was almost 20% more than the mean. Therefore, 95% of values were within $\pm$20% of the mean. This $\pm$20% was applied to each task time's mean in the initial model for calculating the minimum and maximum values. In the Node 1, 1.66sec was 20% less than the mean time (i.e., 2.07sec), so the 1.66sec was applied as a minimum value of this node's task time in this initial model.

**2. Decide whether the C/S is Critical or Distractive:** It was assumed that the operator already memorized his or her critical call signs in this initial model as mentioned before. Therefore, after ATC's instruction, the operator can decide whether the C/S is critical or distractive without referring to the critical call sign list. According to the IMPRINT's 'MicroModel' tool, task time was calculated; 0.1sec (perceptual process) + 0.07sec (decision process) = 0.17sec. This means that the operator takes 0.17 second on average to perceive the instruction and decide the distinction of a call sign. For rectangular distribution, 0.14 second was used as a minimum value (i.e., 20% less than the mean). This node includes logic, that is, this node distributes call signs to the next nodes (i.e., Nodes 3 and 4) according to the pre-assigned sequence in Node 0 (i.e., Table 13). If a call sign of certain sequence assigned at Node 0 is distractive, the next path should be Node 3 (i.e., Real Distractive C/S); and if the call sign is critical, the next path should be Node 4 (i.e., Real Critical C/S).

**3. Real Distractive C/S:** This is not a real action or task. This node exists only for logic; it does not include workload demands. However, it includes probabilistic decision. Although an operator listens to a distractive call sign, he or she may decide it as a critical call sign by mistake. That is, the operator may make wrong decisions, and this node reflects the situation. These Nodes 3 and 4 apply probabilities for Nodes 5 through 8. Node 5, Correct Decision (Decides Distractive C/S), describes that the operator made correct decision when the operator listened to a distractive call sign. Node 6, Wrong Decision (Decides Critical C/S), describes that the operator made wrong decision when the operator listened to a distractive call sign. Node 7, Correct Decision (Decides Critical C/S), describes that the operator made correct decision when the operator listened to a critical call sign. Finally, Node 8, Wrong Decision (Decides Distractive C/S), describes that the operator made wrong decision when the operator listened to a critical call sign. And the probabilities for the nodes could be derived from Amaddio's results. Her results indicated that mean probability of correct answers under the 5 call sign with PI condition was 97.11%. However, this was not enough information for this model, because her results included all correct answers for critical and distractive call signs. The whole data were analyzed again, then it was found that there was almost 1% difference between Node 6 (5.07%) and Node 8 (6.05%). Based on these results, probabilities for this initial model were calculated as below:

1. The mean probability for correct answers was 97.11% as mentioned above. From this, mean fault was 2.89%: 100% - 97.11% = 2.89%. The 97.11% was applied to the mean between Nodes 5 and 7. And, the 2.89% was applied to the mean between Nodes 6 and 8.

2. The difference between Nodes 6 and 8 was 1% as mentioned above. To match the value, Node 6 should have less probability by 0.5% than the mean probability, and Node 8 should have more probability by 0.5% than the mean probability. By doing this, the difference between Nodes 6 and 8, and their mean probability could be maintained.
   - probability of Node 6 = mean–0.5 = 2.89–0.5 = 2.39%
   - probability of Node 8 = mean+0.5 = 2.89+0.5 = 3.39%

3. From the probabilities for Nodes 6 and 8, probabilities for Nodes 5 and 7 can be drawn:
   - probability of Node 5 = 100 – 'Node 6' = 97.61%
   - probability of Node 7 = 100 – 'Node 8' = 96.61%

Consequently, if a call sign that the operator heard was distractive, the probability of the operator's correct decision is 97.61%, and the probability of the operator's wrong decision is 2.39%. According to the assumption, mistyping was considered as wrong decision.

**4. Real Critical C/S:** Such as Node 3, this is not a real action, and this does not include workload demand. This exists only for logic. As described above, if a call sign that the operator heard was critical, the probability of the operator's correct decision is 96.61%, and the probability of the operator's wrong decision is 3.39%. According to the assumption, mistyping was considered as wrong decision.

**5. Correct Decision (Decides Distractive C/S) to 8. Wrong Decision (Decides Distractive C/S):** By using IMPRINT's 'Snapshot' tool, accuracy for each operator can be collected. However, in this model, the accuracy was directly affected by input data, or input probabilities. For this reason, the accuracy was ignored for the results of this modeling, so these task nodes (i.e., Node 5 through Node 8) provide just conceptual tasks. However, the accuracy was an important variable in the human subjects experiment.

**9. Type '0' on the Keypad:** This node represents that the operator types '0' button on the keypad. This is the former step of pressing 'Enter' key. After confirming the spot number on the monitor in Node 12, then the operator would press 'Enter' key. This Node 9 occurs immediately after the operator decides that the call sign he or she heard is distractive, regardless of the real distinction of the call sign. From the 'MicroModel' tool of IMPRINT, task time for this node was calculated: expected duration for typing 1 letter was 0.21 second. Because operators might generally put their fingers on the keypad and type numbers without seeing each number in the keypad, duration for eye movement and eye fixation was not considered. For rectangular distribution, 0.17 second which is 20% less than 0.21 second, was used as a minimum task time.

**10. Find Spot Number on the Grid:** This node shows the situation that the operator finds two-digit spot number on the grid corresponding to the ATC's directional instruction, when the operator decided that the call sign he or she heard was critical, regardless of the real distinction of the call sign. Its task time was calculated from the 'MicroModel' tool as: Eye movement (0.1sec) + Eye fixation (0.3sec) + Simple Reaction (Class match) (0.45sec) = 0.85 second. Prior to this task, the operator checked critical call sign list. To read the grid for this node, eye movement and eye fixation should be considered. After eye fixation, the operator would find out spot number on the grid, and this situation can be considered as class match. For rectangular distribution, 0.68 second was used as a minimum task time.

**11. Type the Spot Number on the Keypad:** This node describes that the operator types the spot number which was found at the Node 10, on the keypad. This task also occurs only when the operator decided that the call sign he or she listened was critical, regardless of the call sign's real distinction. Two-digit number is typed because one spot number consists of two digits according to the assumption. From the 'MicroModel', expected duration for typing the 2-digit number was 0.42 second. For rectangular distribution, 0.34 second was used as a minimum task time.

**12. Confirm Spot Number on the Monitor and Type 'Enter':** After typing the number(s), the operator checks the monitor to confirm whether his or her typing is correct or not. When the operator's typing is correct, the operator will type 'Enter' key. Current node describes this situation. While subjects may find mistyped number on the monitor and correct the number for this task in the human subjects experiment, the mistyping was not considered in this model, because the mistyping rate is not known.

From the 'MicroModel' tool, its task time was calculated as: Decision process (0.07sec) + Typing rate (1 letter) (0.21sec) = 0.28 second. The eye movement and eye fixation are required for this task, however, their task times can be ignored for this node because the operator conducts simultaneous handling. To be specific, if its former node was Node 9, that is, if the operator typed '0', the operator does not need to move his or her gaze from monitor. This is not only because the operator does not need to move gaze to the critical call sign list or grid, but also because he or she can type '0' without seeing the keypad. In addition, if its former node was Node 11, that is, if the operator typed spot number on the keypad, the operator could move gaze while typing the number, because he or she could type the number without seeing the keypad. While it took 0.42 second for typing the 2 digit number from the Node 11, the eye movement and fixation takes only 0.4 second from the 'MicroModel' tool. Therefore, the 0.42 second is enough time for the operator to conduct simultaneous handling (i.e., eye movement and fixation during typing). Then, deciding whether the operator's typing is correct takes 0.07 second, and typing the 'Enter' key (i.e., only one letter) takes 0.21 second.

**13. Cycle Decision:** The color of this node is blue, which means that this node does not require real action and task time, and that this is a logical node. This node decides the number of ATC's instructions, and the operator's response time for handling one instruction. For the modeling of this research, each instruction was provided to an operator every 5 seconds including the time for instruction, and one operator handled 30 instructions. This node captures the "30 instructions" and "5 seconds." This node calculates how many instructions one operator handled. So, if an operator handled less than 30 instructions, the operator should return to Node 1, but if an operator handled 30 instructions, the operator can go to Node 999 (i.e., final node). The response time which started to be measured from Node 1, is finally collected in this node. And the remaining time until 5 seconds is the operator's recess. For example, if time for an instruction took 2 seconds and time for an operator's handling of the instruction took 1.3 second, the operator spent 3.3 seconds for the instruction's handling: 2sec + 1.3sec = 3.3sec. Therefore, the operator's response time is 3.3 seconds, and recess time is 1.7 second: 5sec - 3.3sec = 1.7sec. After the 1.7 seconds, the operator receives the next instruction.

**999. End of Mission:** After one operator conducts 30 instructions (i.e., 30 cycles of the Task Network of this model), this model is completed. This means the operator completed this model. However, more operators are required to be observed to increase the credibility of the results of this model. Therefore, it is assumed that 25 operators are observed for the model of this research. Therefore, total 750 cycles were run for this model: 30 cycle/operator × 25 operators = 750 cycles.

## Appendix C. Initial Model Validation (Response Times)

The response time variable of this initial model was validated by Amaddio's results which had collected from 21 participants. The procedure and assumptions of this initial model are exactly same as the 5 call sign with PI condition of Amaddio's experiment. However, while this model provides 30 instructions to one operator, her experiment provided almost 100 instructions to one operator.

Her results provided mean and standard deviation, and the data were derived from a sample population similarly to this model. To find out statistically significant difference between her experiments and this initial model, 'Comparing Means' method (i.e., T-test) was applied (Milton and Arnold, 2003:338-349). Table 14 shows the mean response times, standard deviations, and sample sizes for this initial model and for her experiments.

Table 14. Initial Model Response Times Result and Amaddio's Result

| | Initial Model | Amaddio's Research |
|---|---|---|
| Sample Size | 25 | 21 |
| Mean Response Time | 3.256 | 3.338 |
| Standard Deviation | 0.059 | 0.342 |

There are two methods to compare means: Comparing means with equal variances (i.e., pooled test) and Comparing means with unequal variances. To decide what method should be applied, F-test should be conducted first. And the F-test was conducted as follows;

1. For the F-test, hypotheses were made: $H_0$: $\sigma_1^2 = \sigma_2^2$ and $H_1$: $\sigma_1^2 \neq \sigma_2^2$, where the $\sigma$ denotes population variance.
2. To compare variances, the ratio $S_A^2/S_B^2$ should be formed as a test statistic where $S_A^2$ is the larger of the two sample variances. In this case, $S_A$ is the sample standard deviation for Amaddio's experiments, 0.342. And, $S_B$ is the sample standard deviation for this initial model, 0.059. The observed value of the test statistic is $S_A^2/S_B^2$ = 34.14973.

3. The p value should be calculated. The number of degrees of freedom associated with the test statistic are $n_A$-1 = 21-1 = 20, and $n_B$-1 = 25-1 = 24. From the F distribution, P($F_{20,24}$>2.207) = 0.05. The probability of seeing a value larger than 34.14973, test statistic, is even smaller than this. Therefore, the p value is smaller than 0.05. However, because this test is two-tailed, this value is doubled.

So, the null hypothesis of equal variances was rejected. And, it can be concluded that the two variances are different. Next, with the unequal variances, means should be compared. The procedure was as follows;

1. To compare the means, hypotheses should be made: $H_0$: $\mu_1 = \mu_2$ and $H_1$: $\mu_1 \neq \mu_2$, where the $\mu$ denotes population mean.

2. To know the number of degrees of freedom, γ, Smith-Satterthwaite Degrees of Freedom was used. The value for γ is not necessarily an integer. If it is not, it is rounded down to the nearest integer. As shown below, this value is rounded down to 20.

$$\gamma \cong \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}} = \frac{\left(0.059^2/25 + 0.342^2/21\right)^2}{\frac{(0.059^2/25)^2}{25 - 1} + \frac{(0.342^2/21)^2}{21 - 1}} \cong 20.98542$$

3. The test statistic for this unequal variance is observed as below:

$$\text{Unequal Variance Test Statistic} = \frac{(\bar{x}_1 - \bar{x}_2) - (u_1 - u_2)_0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} = \frac{(3.256 - 3.338) - 0}{\sqrt{0.059^2/25 + 0.342^2/21}} \cong -1.08193$$

4. Based on the $T_{20}$ distribution, $t_{0.75}$=0.687, and $t_{0.9}$=1.325. Test statistic, 1.08193, is between them. And, because this is two-tailed test, the p value can be calculated as follow:

$$0.75 < 1 - \frac{p\ value}{2} < 0.9 \;\leftrightarrow\; -0.25 < -\frac{p\ value}{2} < -0.1 \;\leftrightarrow\; 0.5 > p\ value > 0.2$$

5. Therefore the p value lies between 0.2 and 0.5. Since this p value is big enough, the null hypothesis cannot be rejected. That is, it is plausible that the means are same; $\mu_1 = \mu_2$.

By applying this method, the response time variable in this initial model was validated.

## Appendix D. Baseline Model Data Input Description (Response Time Modification and Workload Input)

The baseline model's general structure and task times follow the initial model. However, the task time for Node 2 should be changed because it was modified. For the situation that an operator checks the critical call sign list and decides whether the call sign is critical or distractive, Choice Reaction Time (5 alternatives) was calculated as 0.39 second from the 'MicroModel' tool, and this was added to the Node 2. Because this baseline model also assumes that one operator controls 5 UAVs, the number of alternatives used for the 'MicroModel' was also 5. Because the 0.39-second was added, the decision process was excluded from the initial model. Therefore, the final task time of the Node 2 in this baseline model was calculated as: perceptual process (0.1sec) + Choice Reaction Time (5 alternatives) (0.39sec) = 0.49 second. For the rectangular distribution, 0.39 second was used as a minimum task time.

Workload was initially assessed according to the VACP values shown in Table 4, and then peer review was conducted by 4 AFIT students involved in the modeling class. To employ the peer review for this model, the ratio of the initial VACP values to the peers' VACP values was applied as 4:6. That is, the initial workload assessment received a weight of 40%, and each peer's workload assessment received a weight of 15%. This was because the initial workload assessment included the most knowledge about this model.

Workload data input is described according to the sequence of the task network as shown in Figure 15. Some nodes which is not mentioned below does not include workload data because they express only logic, not operator's real task.

**1. Listen to Radio Call from ATC by Both Ears:** Workload for this node was input as 6.50 according to the VACP values. Initial workload was assessed as 5.30 VACP values: Visual(0) + Auditory(4.3) + Cognitive(1.0) + Fine Motor(0) = 5.30. Gross Motor, Speech, and

Tactile values are not considered in this entire model because they are not related to the conditions for this research. The peers assessed this node's workload as 7.30 VACP values on average: Visual(0) + Auditory(4.65) + Cognitive(2.65) + Fine Motor(0) = 7.30. Cognitive value was relatively higher than the initial assessment. By applying the ratio of 4:6 as mentioned above, this node's revised VACP value was calculated: 5.30 × 40% + 7.30 × 60% = 6.50.

**2. Check Critical C/S List and Decide whether the C/S is Critical or Distractive:** Workload was input as 8.74 VACP value. Initial workload assessment was 10.10: Visual(5.1) + Auditory(0) + Cognitive(5.0) + Fine Motor(0) = 10.10. Peers' mean workload assessment was 7.825: Visual(3.025) + Auditory(0) + Cognitive(4.8) + Fine Motor(0) = 7.825. Peers' visual value was relatively lower than the initial assessment. The revised VACP value for this node was calculated: 10.10 × 40% + 7.825 × 60% = 8.735 ≅ 8.74.

**9. Type '0' on the Keypad:** Workload was input as 6.77 VACP value. Initial workload was assessed as 8.30: Visual(5.1) + Auditory(0) + Cognitive(1.0) + Fine Motor(2.2) = 8.30. Peers assessed this node's workload as 5.75 on average: Visual(2.5) + Auditory(0) + Cognitive(1.05) + Fine Motor(2.2) = 5.75. Peers' visual workload assessment was relatively lower than the initial assessment. Revised VACP value for this node was calculated: 8.30 × 40% + 5.75 × 60% = 6.77.

**10. Find Spot Number on the Grid:** Workload was input as 7.42 VACP values. Initial workload was assessed as 9.70: Visual(5.1) + Auditory(0) + Cognitive(4.6) + Fine Motor(0) = 9.70. Mean VACP value of peers' workload assessments was 5.90: Visual(3.75) + Auditory(0) + Cognitive(2.15) + Fine Motor(2.2) = 5.90. Peers' assessment of visual and cognitive workload was lower than initial value, but their fine motor assessment was higher than the initial value. Applying the 4:6 ratio, the 7.42 VACP value was obtained: 9.70 × 40% + 5.90 × 60% = 7.42.

**11. Type the Spot Number on the Keypad:** Workload was input as 7.16 VACP values. Initial workload was assessed as 8.30: Visual(5.1) + Auditory(0) + Cognitive(1.0) + Fine Motor(2.2) = 8.30. The mean of the peers' assessment was 6.40: Visual(3.0) + Auditory(0) + Cognitive(1.1) + Fine Motor(2.3) = 6.40. Peers' visual workload assessment was relatively lower than the initial value. Revised VACP value was calculated as: 8.30 × 40% + 6.40 × 60% = 7.16.

**12. Confirm Spot Number on the Monitor and Type 'Enter':** Workload was input as 9.59 VACP values. Initial workload was assessed as 11.90: Visual(5.1) + Auditory(0) + Cognitive(4.6) + Fine Motor(2.2) = 11.90. Peers' assessment was 8.05 on average: Visual(3.0) + Auditory(0) + Cognitive(2.85) + Fine Motor(2.2) = 8.05. Applying the 4:6 ratio, 9.59 value was obtained: 11.90 × 40% + 8.05 × 60% = 9.59.

## Appendix E.  Baseline Model Validation (Workload)

There was no previous data about workload, as mentioned above.  Therefore, workload in this model was validated by SME (Subject Matter Expert) data, and its results are shown in Figure 41.

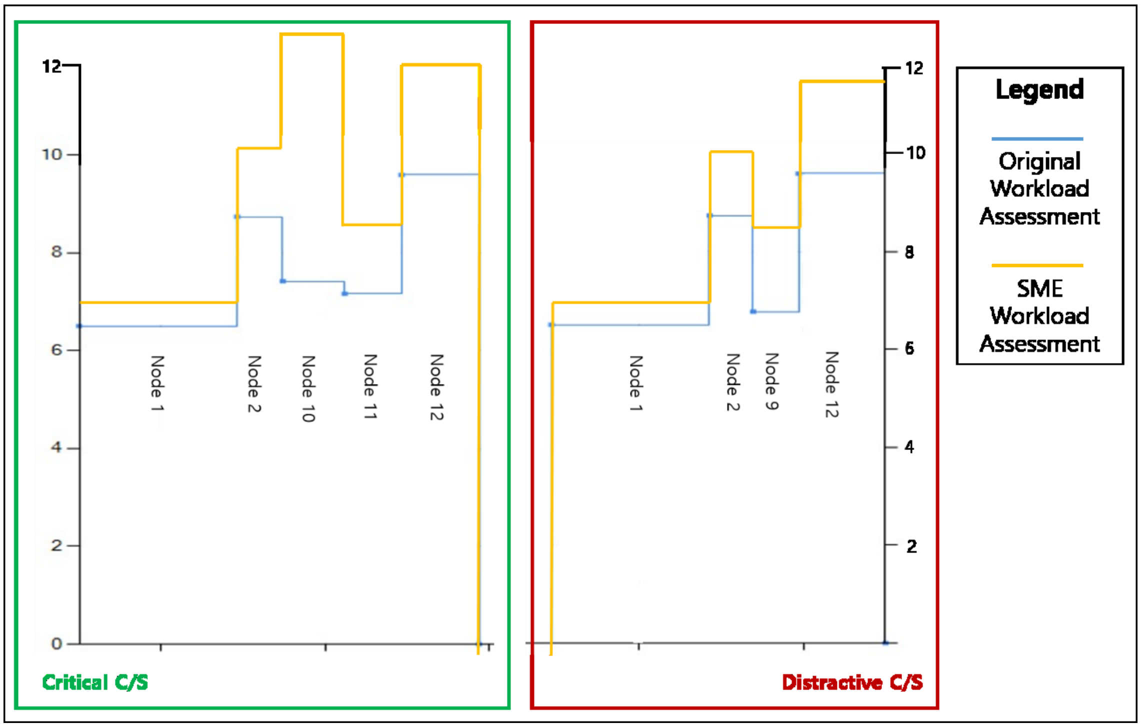


Figure 41. Workload Validation

In the Figure 41, left graph shows workloads for an operator's handling of a critical instruction, so it indicates workloads for Nodes 1-2-10-11-12.  Right graph shows workloads for an operator's handling of a distractive instruction, so it indicates workloads for Nodes 1-2-9-12. Blue lines in the graphs mean original workload assessment which was already calculated in Appendix D and applied to this model.  Yellow lines in the graphs mean the SME data.

In the left graph, that is, the operator's handling of a critical call sign, overall relative workload assessment was similar except for Node 10.  While original workload for Node 10 is lower than Nodes 2 and 12, the SME data assessed it as the highest workload.  For the operator's handling of a distractive call sign in the right red rectangle, all relative workload assessments were same. The ranking of these assessed workloads is shown in Table 15.

Table 15. Ranking of Assessed workloads

| **Critical Call Sign** | |
|---|---|
| Assessment | Workload Ranking |
| Baseline Model | Node 12 > Node 2 > $\boxed{\text{Node 10}}$ > Node 11 > Node 1 |
| SME Data | $\boxed{\text{Node 10}}$ > Node 12 > Node 2 > Node 11 > Node 1 |
| **Distractive Call Sign** | |
| Assessment | Workload Ranking |
| Baseline Model | Node 12 > Node 2 > Node 9 > Node 1 |
| SME Data | Node 12 > Node 2 > Node 9 > Node 1 |

From these Figure 41 and Table 15, it was found that the workload for Node 10 should be corrected. The reason why the assessments were different was that the SME data assumed that an operator finds the spot number on the grid by pointing the numbers with his or her finger. Thus, the SME data assigned fine motor value to this Node 10. On the other hand, the initial assessment did not consider this fine motor value, so '0' fine motor value was assigned for Node 10. To resolve this problem, 2.4 fine motor value was added to the Node 10: 7.42 + 2.4 = 9.82. This revised workload for this model is shown in Table 16. After this revision, workload for this model is shown as Figure 42, and it has same relative workload assessment with the SME data. That is, this baseline model's workload variable was validated.

Table 16. Revised Workload Assessment

| Node | Applied Workload | Revised Workload |
|---|---|---|
| 1 | 6.5 | 6.5 |
| 2 | 8.735 | 8.735 |
| 9 | 6.77 | 6.77 |
| 10 | 7.42 | **9.82** |
| 11 | 7.16 | 7.16 |
| 12 | 9.59 | 9.59 |

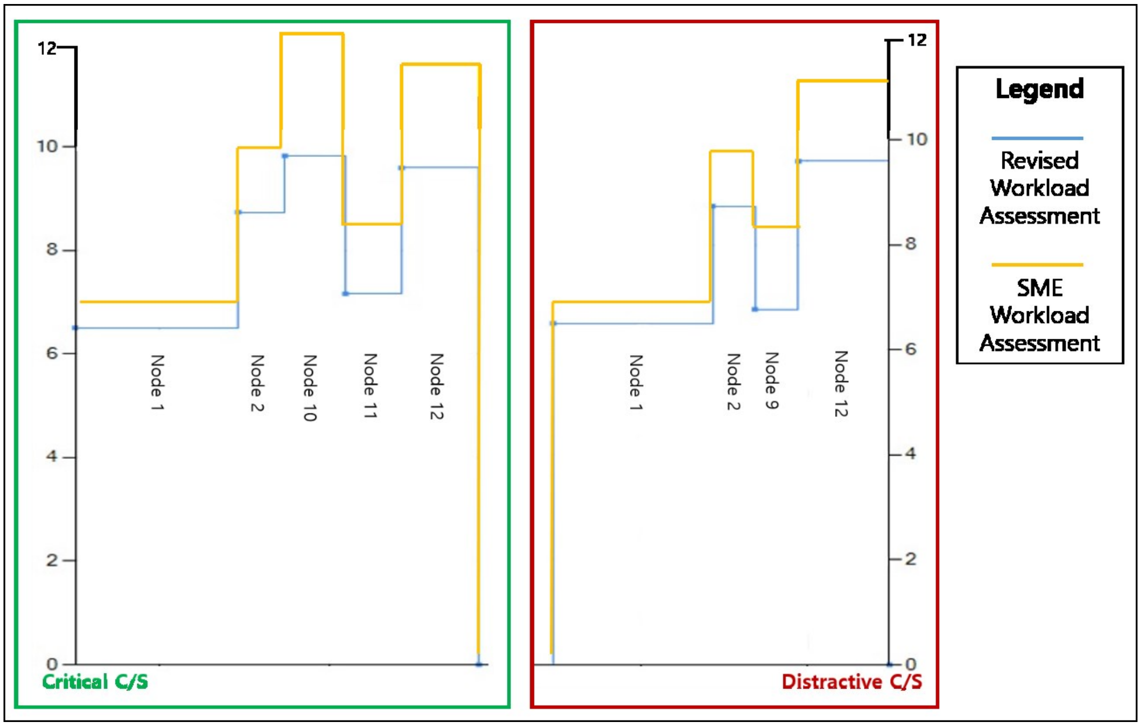


Figure 42. Revised Workload Validation

Although this model's workload was validated, it was required to reach a consensus with the SME data. As a result, Node 1's VACP value was reduced to the initial workload: from 6.5 to 5.3. Node 1 does not require high cognitive value because this task requires only listening and this is a prior step to decide whether the call sign is critical or distractive. Node 2's VACP value was increased, because cognitive value was underestimated. Node 9's VACP value was increased from 6.77 to 7.2, because visual value was underestimated. Node 10's VACP value was increased, because cognitive value was underestimated. Node 11's VACP value was increased because of underestimation of visual value. And, Node 12's VACP value was increased because cognitive value was underestimated. These values are shown in Table 17, and the values were validated once again for confirmation as shown in Figure 43.

Table 17. Agreed Workload Assessment

| Node | | Auditory | Cognitive | Fine Motor | Visual | Total |
|---|---|---|---|---|---|---|
| 1 | From | 4.51 | 1.99 | 0 | 0 | 6.5 |
| | **To** | **4.3** | **1.0** | **0** | **0** | **5.3** |
| 2 | From | 0 | 4.88 | 0 | 3.855 | 8.735 |
| | **To** | **0** | **5.3** | **0** | **4.0** | **9.3** |
| 9 | From | 0 | 1.03 | 2.2 | 3.54 | 6.77 |
| | **To** | **0** | **1.0** | **2.2** | **4.0** | **7.2** |
| 10 | From | 0 | 3.13 | 2.4 | 4.29 | 9.82 |
| | **To** | **0** | **4.6** | **2.4** | **4.0** | **11.0** |
| 11 | From | 0 | 1.06 | 2.26 | 3.84 | 7.16 |
| | **To** | **0** | **1.0** | **2.2** | **4.4** | **7.6** |
| 12 | From | 0 | 3.55 | 2.2 | 3.84 | 9.59 |
| | **To** | **0** | **4.6** | **2.2** | **4.0** | **10.8** |

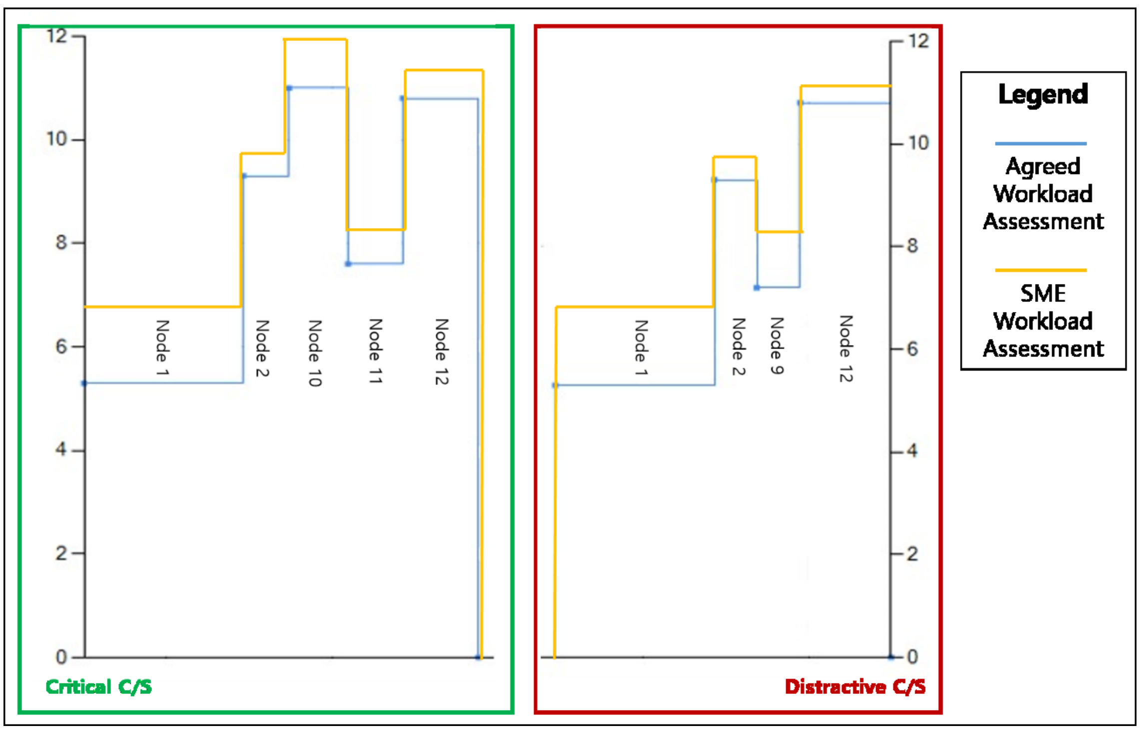


Figure 43. Agreed Workload Validation

## Appendix F. Alternative Models Data Input Description

***Alternative 2 (7 Call signs with Current audio system)***

Because Amaddio's research had the 7 call sign conditions, it was easier to make Alternative 2 first, rather than Alternative 1 which has 3 call sign condition. Therefore, detailed data input explanation of Alternative 2 is treated first, then, that of Alternative 1 would be described.

While in the baseline model one operator handled 5 critical call signs (i.e., one operator controls 5 UAVs), in this Alternative 2 one operator handles 7 critical call signs (i.e., one operator controls 7 UAVs). The ratio of instruction distinction that one operator is provided is not changed and it is maintained as 1:1; number of critical instructions that one operator is provided in this model is 15, and number of distractive instructions is 15.

Among the nodes in the Task Network shown in Figure 15, some details of Nodes 2, 3, and 4 were changed. Other nodes did not have any changes. Because Node 0 shows this model's starting point and it includes the sequence and ratio of instruction distinction, the Node 0 did not have any change.

Node 1 shows instruction from ATC, and the Node 1 is the same as in the baseline model because the format of the instruction was not changed. To calculate minimum task times for rectangular distribution in IMPRINT, 'Empirical Rule' was used again such as in the baseline model. From the results of the 7 call sign with PI condition in Amaddio's research, $\mu$ (i.e., mean) of response time variable was 3.579, and $\sigma$ (i.e., standard deviation) was 0.430. By applying the rule, the 95% interval was between 2.719 and 4.439. The minimum value, 2.719, was almost 24% less than the mean; the maximum value, 4.439, was almost 24% more than the mean. Therefore, 95% of values were within $\pm$24% of the mean under this condition. This $\pm$24% was applied to each task time's mean in this alternative for calculating the minimum times. In this Node 1, 1.57-second is 24% less than the mean (i.e., 2.07 seconds), so it was applied as the minimum value of the task time for this node.

If there are more kinds of critical call signs in the list, it may take more time for the operator to check the critical call sign list and to decide whether the call sign is critical or distractive. So, task time for Node 2 may be increased as compared to the baseline model. This could be explained by the results of Amaddio's research. According to the results of her research, while mean duration for the 5 call sign with PI condition was 3.338 seconds, mean duration for 7 call sign with PI condition was 3.579 seconds. Among the nodes in the Task Network shown in Figure 15, this node was the only one whose task time was affected by the number of critical call signs. Choice Reaction Time (7 alternatives) was calculated as 0.45 second from the 'MicroModel' tool, and this was reflected in the Node 2. Because this Alternative 2 assumes that one operator controls 7 UAVs, the number of alternatives used in the 'MicroModel' was 7. Therefore, the task time of this Node 2 in the Alternative 2 was calculated as: perceptual process (0.1sec) + choice reaction time (7 alternatives) (0.45sec) = 0.55 second. For the rectangular distribution, 0.42 second was used as a minimum task time.

For the same reason, an operator's workload was increased in Node 2. Because 2 critical call signs were added to the list, the operator's Visual and Cognitive workload values were increased, while Auditory and Fine motor values were not affected by the number of critical call signs. According to Visual VACP table shown in Table 4, if one operator handles only one UAV, the operator's visual VACP value can be considered as 3.0. However, the baseline model assumed that the operator controls 5 UAVs, and this was assessed as 4.0 visual workload value. Therefore, it can be considered that one call sign has 0.25 visual VACP value: $(4.0 - 3.0) \div$

$(5 - 1) = 0.25$. Therefore, for this alternative, 4.5 visual VACP value was applied: $3.0 + [0.25 \times (7 - 1)] = 4.5$. Similar approach was used for calculating cognitive value. From Cognitive VACP table shown in Table 4, if one operator handles only one UAV, the operator's cognitive VACP value can be considered as 4.6. The baseline model was assessed to have a cognitive VACP value of 5.3. Therefore, it can be considered that one call sign has 0.175 cognitive VACP value: $(5.3 - 4.6) \div (5 - 1) = 0.175$. Therefore, a cognitive VACP value of 5.65 was applied to Alternative 2: $4.6 + [0.175 \times (7 - 1) = 5.65$.

Nodes 3 and 4 have probabilities, and the probabilities may be affected by the number of critical call signs that one operator handles. If the number is increased, the operator's accuracy may be decreased. This could also be explained by the Amaddio's research. While in the 5 call sign with PI condition the operator's accuracy was 97.11%, it was decreased to 91.73% in the 7 call sign with PI condition. So, mean fault was 8.27%: $100\% - 91.73\% = 8.27\%$. The 91.73% was applied to the mean between Nodes 5 and 7, and the 8.27% was applied to the mean between Nodes 6 and 8. As mentioned in the baseline model, the difference between Nodes 6 and 8 was almost 1%. To maintain the 1%, the Node 6 has less probability by 0.5% than mean probability, and the Node 8 has more probability by 0.5% than the mean. Thus, the probability of Node 6 was applied as 7.77%: $8.27 - 0.5 = 7.77\%$. And, the probability of Node 8 was applied as 8.77%: $8.27 + 0.5 = 8.77\%$. From these, probabilities for Nodes 5 and 7 could be drawn. The probability of Node 5 was applied as 92.23%: $100 - 7.77 = 92.23\%$. And, the probability of Node 7 was applied as 91.23%: $100 - 8.77 = 91.23\%$. As a result, if a call sign that an operator heard was distractive, the probability of the operator's correct decision is 92.23% and the probability of the operator's wrong decision is 7.77%. In contrast, if a call sign that an operator heard was critical, the probability of the operator's correct decision is 91.23% and the probability of the operator's wrong decision is 8.77%. According to the assumption, mistyping was considered as a wrong decision.

In Node 9, an operator types '0' on the keypad. This action is not related with the number of critical call signs, so task time and workload for this node are not affected by it. In Node 10, an operator finds spot numbers on the grid. Task time and workload for this node may depend on how complex the grid is, but they do not depend on the number of critical call signs. Node 11 is similar to Node 9, the operator's action is just typing 'spot numbers' on the keypad. Although task time and workload for this Node 11 may be affected by the digits of the spot number, they are not affected by the number of critical call signs. In Node 12, an operator confirms the spot number on the monitor, and then, types 'Enter' key. This action may be also affected by the digits of the spot number, however, the number of critical call signs does not affect the task time and workload for this node. Node 13 exists to calculate cycle number and response time, so it is not affected by this alternative condition.

***Alternative 1 (3 Call signs with Current audio system)***

While in the Alternative 2 one operator handled 7 critical call signs, one operator handles 3 critical call signs in this Alternative 1. This smaller number was selected to provide a larger difference in human performance as compared to the Amaddio's research which included 5 call sign conditions.

Similarly to Alternative 2, some details of Nodes 2, 3, and 4 were changed. Additionally, to calculate minimum task times for a rectangular distribution in IMPRINT, the 'Empirical Rule' was also applied in this alternative. However, because there are no previous data related to this 3 call sign condition, linear assumption was applied for the 'Empirical Rule' of this alternative. To calculate the minimum task times, two standard deviations (i.e., $2\sigma$) were calculated to be $\pm 24\%$

of the mean for the 7 call sign with current audio condition (i.e., Alternative 2), and $\pm$20% for the 5 call sign with current audio condition (i.e., Baseline model). Therefore, linearly, it could be assumed that the two standard deviations for the 3 call sign with current audio condition (i.e., Alternative 1) would be $\pm$16% of the mean. Therefore, this $\pm$16% was applied to each task time's mean in this Alternative 1 for calculating minimum times.

If there are fewer critical call signs in the list, it may take less time for the operator to check the critical call sign list and to decide whether the call sign is critical or distractive. So, task time for Node 2 may be decreased. Similar method as used in Alternative 2 was applied to calculate Node 2's duration. Choice Reaction Time (3 Alternatives) was calculated as 0.3 second from the 'MicroModel' tool, therefore, the task time of this Node 2 was calculated as: perceptual process (0.1sec) + Choice Reaction Time (3 alternatives) (0.3sec) = 0.4 second. For the rectangular distribution, 0.34 second was used as a minimum task time.

For the same reason, an operator's VACP value for Node 2 would be decreased in Alternative 1. Based on the method used to calculate VACP value for Node 2 in Alternative 2, visual and cognitive VACP values were calculated as 3.5 and 4.95 respectively, and these were calculated as: $3.0 + [0.25 \times (3 - 1)] = 3.5$, and $4.6 + [0.175 \times (3 - 1)] = 4.95$.

Probabilities applied to Nodes 3 and 4 may also be affected by the number of critical call signs. Because the number is decreased, the operator's accuracy may be increased. However, the accuracy point cannot be 100%, because people make mistakes. For these reasons, it was assumed that for this Alternative 1, 1% accuracy may be increased from the results of 5 call sign with PI condition in the Amaddio's experiments: 97.11% + 1% = 98.11%. Such as in the baseline model, the probability for Node 8 may be bigger than the probability for Node 6, because the operator's probability to mistype the spot numbers would be higher than the probability to mistype '0' key. Therefore, 1% difference between them was applied to this alternative in the same way. As results, probabilities in these nodes are: Node 5(98.61%), Node 6(1.39%), Node 7(97.61%), and Node 8(2.39%). As mentioned before, however, these probabilities were calculated only to consider anticipated results, and the results of these accuracies from this model were not considered as output data, because the results were just affected by the input probabilities.

***Alternative 3 (3 Call signs with 3D audio system)***

For Alternatives 3 and 4, it was assumed that there was no ambiguous call sign which is potentially provided to both of the operator's ears as mentioned from the concept, and the 3D audio's reliability was 100%. That is, every piece of information was provided to only one of the operator's ears, and this system did not have any error. To make an ideal environment and to draw pure effects of the 3D audio system, these assumptions were made.

Among the nodes in the baseline model, details of Nodes 1, 2, 3, and 4 were changed, and Nodes 14 and 15 were added such as shown in Figure 16. Node 0 did not change because the ratio of instruction distinction that one operator is provided was not changed by the audio system.

In the baseline, Node 1 included task time and workload. However, in this Alternative 3, Nodes 14 and 15 are conducting the Node 1's role. So, in this Alternative 3, Node 1 just shows the start of a new cycle. It was renamed as 'Start of a New Cycle', and color of this node was changed from plum to blue, indicating it only exists for logic.

Nodes 14 and 15 were added to the Alternative 3 to distinguish required task times and workload. When an operator starts to hear a distractive instruction through his or her left ear, the operator does not need to hear the remaining instruction. That is, if an operator hears the first word from ATC's instruction through his or her left ear, the operator immediately perceives from which ear this information is provided. Then, the operator would confirm it by hearing the second word. Therefore, by listening only first two words, the operator can believe from which ear the information was provided. From the 'MicroModel' tool of IMPRINT, the duration for speaking two words was 0.69 second, so this 0.69-second was applied to the mean task time for Node 14. As a minimum task time for rectangular distribution, 0.55-second, 20% less than the mean task time, was applied. Because there is no data related to the response time for the 3D audio system, the two standard deviations (i.e., 2σ) were equally used as in the baseline model. However, if the information is provided to the operator's right ear (i.e., critical call sign), the operator should listen to the remaining instruction because it includes important position information. So, Node 15 had the same condition as the Node 1 in the baseline model. For 6 words, 2.07-second was applied to the Node 15's mean task time, and 1.66-second was applied to the Node 15's minimum task time. The color of the Node 14 is gold, and that of the Node 15 is green. When an operator decides the call sign that he or she listened was distractive, this model follows gold task nodes. On the other hand, when the operator decides that the call sign was critical, this model follows green task nodes. Based on VACP values shown in Table 4, 3.0 auditory VACP value and 1.0 cognitive value were applied to the Node 14, and 4.3 auditory VACP value and 1.0 cognitive value were applied to the Node 15.

Because this system does not have any error according to the assumption, the operator does not need to check the critical call sign list. This assumption made Node 2 modified; 'Check Critical C/S List' was deleted from the baseline model. When the operator perceives from which ear the information is provided, the operator can immediately decide whether the call sign is critical or distractive. Therefore, task time and workload for this Node 2 were reduced. The choice reaction time among 5 alternatives (i.e., 0.39 second) was excluded from the baseline model, therefore, only 0.17-second was applied as mean task time: perceptual process (0.1 sec) + decision process (0.07 sec) = 0.17 second. As minimum task time, 0.14-second which is 20% less than the mean task time was applied. For workload, visual VACP value is not required because an operator does not need to see the critical call sign list. And, cognitive value was decreased to 4.6 because the operator is required to decide only from which ear the information was provided. Because fine motor and auditory VACP values are not required, the 4.6 value was applied to the VACP value for this Node 2.

Nodes 3 and 4 have probabilities, and the probabilities may be affected by the type of audio systems. It was expected that when an operator uses the 3D audio system, distinguishing the distinction of the call signs would be easier than when the operator uses the current audio system. This was because the operator does not need to check the critical call sign list. However, the accuracy cannot be 100%, because people make mistakes. For these reasons, it was assumed that for the 3D audio system, 1.5% accuracy may be increased from the results of 5 call sign with PI condition in the Amaddio's experiments: 97.11% + 1.5% = 98.61%. Such as in the baseline model, the probability for Node 8 may be bigger than the probability for Node 6, because the operator's probability to mistype the spot numbers would be higher than the probability to mistype '0' key. Then, 1% difference between them was applied to this alternative in the same way. As results, probabilities in these nodes are: Node 5(99.11%), Node 6(0.89%), Node 7(98.11%), and Node 8(1.89%).

Nodes 5 through 8 exist only to draw the number of operators' faults for each situation, so they did not have any change. In the Node 9, the operator types '0' on the keypad, and because this action is not related with the type of audio systems, its task time and workload were not affected. In the Node 10, the operator finds out spot numbers on the grid. The task time and workload for this Node 10 may depend on how complex the grid is, but they would not depend on the audio types. This was the reason why the Node 10 was not affected by the audio type. Node 11 is similar to Node 9. The operator's action is just typing 'spot numbers' on the keypad. Although the task time and workload for this Node 11 may be affected by the digits of the spot number, they are not affected by the audio type. In the Node 12, the operator confirms the spot number on the monitor and presses 'Enter' key. This action may be affected by the digits of the spot number, however, the type of audio systems does not affect the time and workload for this node. Node 13 exists to calculate the number of instructions from ATC and to collect the operator's response time. The node is not affected by this alternative condition.

***Alternative 4 (7 Call signs with 3D audio system)***

Before creating this model, it was expected that as the number of critical call signs that one operator handles is increased, task times and workload may be increased even though the operator uses the 3D audio system, such as in the current audio system. However, if the ratio of the number of critical instructions to the number of distractive instructions provided to an operator is not changed as compared the current audio system, the task times and workload would not be affected by the number of critical call signs when the 3D audio system is used, because of the assumption that this 3D audio system does not have any error. That is, because the operator completely believes this 3D audio system and does not need to check the critical call sign list, he or she would react only according to the perception from which ear the information is provided. Even though the operator handles 100 UAVs, the only thing that the operator needs to do is to react to his or her right ear. Similarly, it was expected initially that accuracy may be decreased, as the number of critical call signs is increased under 3D audio condition. However, the accuracy would not be affected by the number of call signs either, because the operator does not need to check the critical call sign list under the 3D audio condition with 100% reliability. Therefore, in all nodes of this Alternative 4, the operator's workload and performance (i.e., task time and accuracy) were not affected by the number of critical call signs that one operator handles. That is, Alternatives 3 and 4 are same. However, the previous discussion assumes that the ratio of critical to distractive instructions is not excessively large. If the ratio of critical to distractive instructions were increased significantly, such that most of the instructions were critical, the operator would need to respond to most incoming instructions and the change in behavior would not reduce workload as the participant would need to respond to a large proportion of the instructions.

## Appendix G. Questionnaire

**<u>Before the Experiment:</u>**

1) **Do you have any hearing deficiency?**

   Yes_____ No _____

2) **Are you fluent in English?**

   Yes_____ No _____

3) **Are you a pilot?**

   Yes_____ No _____

4) **Please indicate your age:** _______years

5) **Please indicate your gender:** Male _____ Female_____

**<u>Middle of the Experiment: NASA-TLX</u>**

## *NASA Task Load Index*

*Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.*

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

**<u>After Completion of Experiment:</u>**

**NASA-TLX Mental Workload Rankings:**

For each of the pairs listed below, please circle the scale title that represents the more important contributor to workload in the experiments.

| | | |
|---|---|---|
| Mental Demand | or | Physical Demand |
| Mental Demand | or | Temporal Demand |
| Mental Demand | or | Performance |
| Mental Demand | or | Effort |
| Mental Demand | or | Frustration |
| Physical Demand | or | Temporal Demand |
| Physical Demand | or | Performance |
| Physical Demand | or | Effort |
| Physical Demand | or | Frustration |
| Temporal Demand | or | Performance |
| Temporal Demand | or | Frustration |
| Temporal Demand | or | Effort |
| Performance | or | Frustration |
| Performance | or | Effort |
| Frustration | or | Effort |

**SWORD (Subjective Workload Dominance Technique)**

Today, you were exposed to 6 conditions. Based on your today's trials, please check subjective relative workload of the conditions.
For example, if you feel that the two conditions imposed a similar level of workload, you can mark the 'EQUAL' point on the rating sheet, and if you feel that 'C2' imposed a slightly higher level of workload than 'C1' did, you can move toward 'C2' on the sheet and mark the 'Weak' point on the rating sheet.

- Condition 1 **(C1)**: 3 Call Signs + Current Audio
- Condition 2 **(C2)**: 7 Call Signs + Current Audio
- Condition 3 **(C3)**: 3 Call Signs + 3D Audio
- Condition 4 **(C4)**: 7 Call Signs + 3D Audio
- Condition 5 **(C5)**: 7 Call Signs + 3D Audio + Announcement of Possible Errors + No Real Error
- Condition 6 **(C6)**: 7 Call Signs + 3D Audio + Announcement of Possible Errors + 4 Real Errors

| **Tasks** | **Absolute** | | **Very Strong** | | **Strong** | | **Weak** (Slight) | | **EQUAL** | **Weak** (Slight) | | **Strong** | | **Very Strong** | | **Absolute** | | **Tasks** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ← | | | | | | | | | → | | | | | | | | |
| | **9** | **8** | **7** | **6** | **5** | **4** | **3** | **2** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | |
| **C1** (3C/S, Current) | | | | | | | | | | | | | | | | | | **C2** (7C/S, Current) |
| **C1** (3C/S, Current) | | | | | | | | | | | | | | | | | | **C3** (3C/S, 3D, No Error) |
| **C1** (3C/S, Current) | | | | | | | | | | | | | | | | | | **C4** (7C/S, 3D, No Error) |
| **C1** (3C/S, Current) | | | | | | | | | | | | | | | | | | **C5** (7C/S, 3D, Error-Announcement, No Real Error) |
| **C1** (3C/S, Current) | | | | | | | | | | | | | | | | | | **C6** (7C/S, 3D, Error-Announcement, 4 Real Errors) |
| **C2** (7C/S, Current) | | | | | | | | | | | | | | | | | | **C3** (3C/S, 3D, No Error) |

| Tasks | Absolute | | Very Strong | | Strong | | Weak (Slight) | | EQUAL | Weak (Slight) | | Strong | | Very Strong | | Absolute | | Tasks |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ← | | | | | | | | | → | | | | | | | | |
| | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| **C2** (7C/S, Current) | | | | | | | | | | | | | | | | | | **C4** (7C/S, 3D, No Error) |
| **C2** (7C/S, Current) | | | | | | | | | | | | | | | | | | **C5** (7C/S, 3D, Error-Announcement, No Real Error) |
| **C2** (7C/S, Current) | | | | | | | | | | | | | | | | | | **C6** (7C/S, 3D, Error-Announcement, 4 Real Errors) |
| **C3** (3C/S, 3D, No Error) | | | | | | | | | | | | | | | | | | **C4** (7C/S, 3D, No Error) |
| **C3** (3C/S, 3D, No Error) | | | | | | | | | | | | | | | | | | **C5** (7C/S, 3D, Error-Announcement, No Real Error) |
| **C3** (3C/S, 3D, No Error) | | | | | | | | | | | | | | | | | | **C6** (7C/S, 3D, Error-Announcement, 4 Real Errors) |
| **C4** (7C/S, 3D, No Error) | | | | | | | | | | | | | | | | | | **C5** (7C/S, 3D, Error-Announcement, No Real Error) |
| **C4** (7C/S, 3D, No Error) | | | | | | | | | | | | | | | | | | **C6** (7C/S, 3D, Error-Announcement, 4 Real Errors) |
| **C5** (7C/S, 3D, Error-Announcement, No Real Error) | | | | | | | | | | | | | | | | | | **C6** (7C/S, 3D, Error-Announcement, 4 Real Errors) |

**Survey Questions about Usability**

1) **Do you think the 3D audio is helpful for a UAV operator to reduce his or her workload? And, why?**

   Yes_____ No _____

   Reason:

2) **Do you think the 3D audio is helpful for a UAV operator to reduce his or her response times? And, why?**

   Yes_____ No _____

   Reason:

3) **Do you think the 3D audio is helpful for a UAV operator to increase his or her accuracy? And, why?**

   Yes_____ No _____

   Reason:

4) **If you were a UAV operator and the 3D audio system does NOT have any error, would you want to use the 3D audio system? And, why?**

   Yes_____ No _____

   Reason:

5) **If you were a UAV operator and the 3D audio system MAY HAVE errors, would you still want to use the 3D audio system? And, why?**

   Yes_____ No _____

   Reason:

6) **If you have any other comments about the 3D audio system and/or this experiments, please feel free to write them.**

**----- Thank You Very Much -----**

## Appendix H. Vita

### Sungbin Kim, Major, Republic of Korea Air Force

#### CONTACT INFORMATION


| | |
|---|---|
| Address: | Department of Systems Engineering and Management<br>Air Force Institute of Technology<br>2950 Hobson Way<br>WPAFB, OH 45433-7765 |
| Phone: | 1-937-374-9797 |
| Email Address: | sung.kim@afit.edu; bbinsk@naver.com |


#### MILITARY INFORMATION

| | |
|---|---|
| Rank: | Major |
| Date of Current Rank: | 1 December 2014 |
| Date of Commission: | 1 March 2004 |
| Military Specialty: | (Manned-Aircraft) Pilot |

#### FORMAL EDUCATION

M.S. Systems Engineering, Human Factors, 2016
Air Force Institute of Technology, Wright-Patterson AFB, OH

B.S. Industrial Engineering, 2004
Republic of Korea Air Force Academy, Cheongju-si, Chungbuk, Republic of Korea

## Bibliography

Allender, L. "Modeling Human Performance: Impacting System Design, Performance, and Cost," *Proceedings of the Military, Government and Aerospace Simulation Symposium, 2000 advanced Simulation Technologies Conference*. 139-144. Washington, D.C., 2000.

Amaddio, K. M., Miller, M. M., Elshaw, J., and Finomore, V. "The Cognition of Multi-Aircraft Control (MAC): Proactive Interference and Working Memory Capacity," *Proceedings of the 18th International Symposium on Aviation Psychology (ISAP 2015)*. 195-200. Dayton, 2015.

Arons, B. "A Review of The Cocktail Party Effect," *Journal of the American Voice I/O Society*, 12: 35-50 (July 1992).

Barnes, J. E. "U.S. Air Force Moves to Retain a Crucial Resource: Drone Pilots," *The Wall Street Journal*, 15 January 2015.

Bass, W. "How to U.S. Stumbled Into the Drone Era," *The Wall Street Journal*, 24 July 2014.

Calhoun, G. L. and Draper, M. H. "Display and Control Concepts for Multi-UAV Applications," In *Handbook of Unmanned Aerial Vehicles*. New York: Springer Science+Business Media Dordrecht, 2015.

Cengarle, G. *3D Audio Technologies: Applications to Sound Capture, Post-Production and Listener Perception.* Ph.D. disseration. Universitat Pompeu Fabra, 2012.

Colombi, J. M., Miller, M. E., Schneider, M., McGrogan, J., Long, D. S., and Plaga, J. "Predictive Mental Workload Modeling for Semiautonomous System Design: Implications for Systems of Systems," *Journal of Systems Engineering*, 15:448-460 (January 2012).

Finomore, V., Popik, D. K., Castle, C. E., and Dallman, R. C. "Effects of Network-Centric Multi-Modal Communication on a Communication-Monitoring Task.," *Proceedings of the Human Factors and Ergonomics Society*. 2125-2129. San Francisco, 2010.

Franke, J. L., Zaychick, V., Spura, T. M., and Alves, E. E. "Inverting the Operator/Vehicle Ratio: Approaches to Next Generation UAV Command and Control," *Proceedings of Association for Unmanned Vehicle Systems International*. 1-11. 2005.

Gluckman, J. P., Carmody, M. A., Morrison, J. G., and Hitchcock, E. M. "Effects of Allocation and Partitioning Strategies of Adaptive Automation on Task Performance and Perceived Workload in Aviation Relevant Tasks," *The Adaptive Function Allocation for Intelligent*

*Cockpits (AFAIC) Program: Interim Research Guidelines for the Application of Adative Automation.* 7-12. Naval Air Warfare Center, 1993.

Guastello, S. J. *Human Factors Engineering and Ergonomics.* New York: CRC Press, 2014.

Haas, E. C. "Emerging multimodal technology: Role in enhancing combat and civilian system safety," *Professional safety*, 52: 32-38 (December 2007).

Haas, E. C. and Stachowiak, C. "Multimodal Displays to Enhace Human Robot Interaction On-the-Move," *Proceeding of the 2007 Workshop on Performance Metrics for Intelligent Systems*. 135-140. Gaithersburg, 2007.

Hart, S. G. "Nasa-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting (50)*. 904-908. Santa Monica: Human Factors & Ergonomics Society, 2006.

Kaber, D. B. and Endsley, M. R. "The Effects of Level of Automation and Adaptive Automation on Human Performance, Situation Awareness and Workload in a Dynamic Control Task," *Theoretical Issues in Ergonomics Science*, 5(2): 113-153 (April 2004).

Kaber, D. B., Riley, J. M., Tan, K. and Endsley, M. R. "On the Design of Adaptive Automation for Complex Systems," *International Journal of Cognitive Ergonomics*, 5: 37-57 (December 2001).

Maza, I., Caballero, F., Molina, R., Pena, N., and Ollero, A. "Multimodal interface technologies for UAV ground control stations," *Journal of Intelligent and Robotic Systems Manuscript,* 57: 371-391 (August 2010).

McAnally, K. I. and Martin, R. L. "Spatial Audio Displays Improve the Detection of Target Messages in a Continuous Monitoring Task," *The Journal of the Human Factors and Ergonomics Society*, 49: 688-695 (August 2007).

Milton, J. S. and Arnold, J. C. *Introduction to Probability and Statistics.* New York: McGraw-Hill, 2003.

Mowbray, G. H. and Gebhard, J. W. "Man's Senses as Informational Channels," In W. H. Sinaiko, *Human Factors in the Design and Use of Control Systems*. New York: Dover Publications, 1961.

ODonnell, L. "Yerkes Dodson Law." *Aviation Knowledge*. Wikidot.com. 18 August 2011: http://aviationknowledge.wikidot.com/aviation:yerkes-dodson-law, Retrieved on 2 March 2015.

Parasuraman, R. and Wickens, C. D. “Humans: Still Vital after All These Years of Automation,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3) 511-520 (June 2008).

Patrick, M. *Designing Human-Machine Cooperation Systems*. Hoboken: Wiley, 2014.

Philbrick, R. M. and Colton, M. B. “Effects of Haptic and 3D Audio Feedback on Operator Performance and Workload for Quadrotor UAVs in Indoor Environments,” *Journal of Robotics and Mechatronics*, 26(5): 580-591 (December 2014).

Righetti, L., Sharkey, N., Arkin, R., Ansell, D., Sassoli, M., Heyns, C., Asaro, P., and Lee, P. “Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects,” *Proceedings of International Committee of the Red Cross*. Geneva, Switzerland, 2014.

Shneiderman, B. and Plaisant, C. *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* Boston: Addison Wesley. 2010.

Siegel, S. *Nonparametric Statistics for the Behavioral Science*. New York: McGraw-Hill Book Company, 1956.

Simpson, B. D., Bolia, R. S., and Draper, M. H. “Spatial Audio Display Concepts Supporting Situation Awareness for Operators of Unmanned Aerial Vehicles,” *Human Performance, Situation Awareness, and Automation*, 61-65. (March 2004).

Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C. and Jenkins, D. P. *Human Factors Methods*. Famham: Ashgate, 2010.

Teigen, K. H. “Yerkes-Dodson: A Law for all Seasons,” *Theory & Psychology*, 525-547. 1994.

Trouvain, B. and Schlick, C. “A Study of Audio and Visual Context Switch Indicators in a Multirobot Navigation Task,” *Proceedings of the 2004 IEEE International Conference on Systems (3)*. 2821 - 2826. The Netherlands, 2004.

Vazquez-Alvarez, Y. and Brewster, S. A. “Eyes-Free Multitasking: The Effect of Cognitive Load on Mobile Spatial Audio Interfaces,” *Proceedings of the Conference on Human Factors in Computing Systems (CHI) 2011*. 2173-2176. New York, 2011.

Wickens, C. D. “Multiple resources and performance prediction,” *Theoretical Issues in Ergonomics Science*, 3: 159-177 (2002).

Wickens, C. D., Parasuraman, R., and Sheridan, T. B. “A Model for Types and Levels of Human Interaction with Automation,” *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions,* 30: 286-297 (May 2000).

Yu, W. and Brewster, S. "Evaluation of multimodal graphs for blind people," *Universal Access in the Information Society,* 2: 105-124 (May 2003).

"3-D Audio Technology." *BeckerUSA.com*. Becker Avionics, Inc. 10 November 2011: http://www.beckerusa.com/3-d-audio-technology-advantages/, Retrieved on 2 March 2015.

"File:HebbianYerkesDodson.svg" *Commons.Wikimedia.org.* Wikimedia Commons, 29 November 2014: https://commons.wikimedia.org/wiki/File:HebbianYerkesDodson.svg, Retrieved on 7 March 2016.

| **REPORT DOCUMENTATION PAGE** | *Form Approved OMB No. 074-0188* |
|---|---|



| **1. REPORT DATE** *(DD-MM-YYYY)* | **2. REPORT TYPE** | **3. DATES COVERED** *(From – To)* |
|---|---|---|
| 24-03-2016 | Master's Thesis | Aug 2014 - Mar 2016 |

| | |
|---|---|
| **4. TITLE AND SUBTITLE**<br>Unmanned Aerial Vehicle (UAV) Operators' Workload Reduction:<br>The Effect of 3D Audio on Operators' Workload and Performance during Multi-Aircraft Control | **5a. CONTRACT NUMBER**<br>**5b. GRANT NUMBER**<br>**5c. PROGRAM ELEMENT NUMBER** |
| **6. AUTHOR**<br>Sungbin Kim, Major, Republic of Korea Air Force | **5d. PROJECT NUMBER**<br>**5e. TASK NUMBER**<br>**5f. WORK UNIT NUMBER** |
| **7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)**<br>Air Force Institute of Technology<br>Graduate School of Engineering and Management (AFIT/EN)<br>2950 Hobson Way<br>WPAFB OH 45433-7765 | **8. PERFORMING ORGANIZATION REPORT NUMBER**<br>AFIT-ENV-MS-16-M-163 |
| **9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**<br>Col Anthony Tvaryanas, MD<br>711th Human Performance Wing/HP SG<br>2510 Fifth St Building 840 room W413.32<br>Wright-Patterson AFB, OH 45433<br>(937) 798-3253 anthony.tvaryanas@us.af.mil | **10. SPONSOR/MONITOR'S ACRONYM(S)**<br>711th HPW/HP SG<br>**11. SPONSOR/MONITOR'S REPORT NUMBER(S)** |

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
DISTRIBUTION STATEMENT A: APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

**13. SUPPLEMENTARY NOTES**


**14. ABSTRACT**
The importance and number of Unmanned Aerial Vehicle (UAV) operations are rapidly growing in both military and civilian applications. This growth has produced significant manpower issues, producing a desire that multiple aircraft are controlled by a single operator as opposed to the current model where one aircraft may require multiple operators. A potential issue is the need for an operator to monitor radio traffic for the call signs of multi-aircraft. An investigation of the use of 3D sound was undertaken to investigate whether an automatic parser, which preselected the spatial location of relevant versus irrelevant call signs, could aid UAV operators in increasing performance with reduced workload. Furthermore, because the 3D audio system may not guarantee 100% reliability, human performance with the 3D audio system was also collected when they were informed announcement that errors were possible and when the reliability level was less than 100%. This investigation included development of a human performance model, simulation of human performance and workload, and a human subject study. Consequently, promising effects of the 3D audio system on multi-aircraft control were found. This novel and unique use of 3D sound is discussed, and significant improvements in response time and workload are demonstrated.

**15. SUBJECT TERMS**
Multi-UAV control, 3D audio, Workload, Performance, Automation, Interface design

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| | | | | | Michael E. Miller, PhD, AFIT/ENV |
| **REPORT** | **ABSTRACT** | **c. THIS PAGE** | | | **19b. TELEPHONE NUMBER** *(Include area code)* |
| U | U | U | UU | 140 | (937) 255-3636, x 4651; Michael.Miller@afit.edu |