2005

The Integration Of Audio Into Multimodal Interfaces: Guidelines And Applications Of Integrating Speech, Earcons, Auditory Icons, and Spatial Audio (SEAS)

David Jones
University of Central Florida

THE INTEGRATION OF AUDIO INTO MULTIMODAL INTERFACES: GUIDELINES AND APPLICATIONS OF INTEGRATING SPEECH, EARCONS, AUDITORY ICONS, AND SPATIAL AUDIO (SEAS)

by

DAVID LEE JONES
B.S. University of Central Florida, 1996

A thesis submitted in partial fulfillment of the requirements
for the degree of Masters of Science
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2005

ABSTRACT

The current research is directed at providing validated guidelines to direct the integration of audio into human-system interfaces. This work first discusses the utility of integrating audio to support multimodal human-information processing. Next, an auditory interactive computing paradigm utilizing Speech, Earcons, Auditory icons, and Spatial audio (SEAS) cues is proposed and guidelines for the integration of SEAS cues into multimodal systems are presented. Finally, the results of two studies are presented that evaluate the utility of using SEAS cues, developed following the proposed guidelines, in relieving perceptual and attention processing bottlenecks when conducting Unmanned Air Vehicle (UAV) control tasks. The results demonstrate that SEAS cues significantly enhance human performance on UAV control tasks, particularly response accuracy and reaction time on a secondary monitoring task. The results suggest that SEAS cues may be effective in overcoming perceptual and attentional bottlenecks, with the advantages being most revealing during high workload conditions. The theories and principles provided in this paper should be of interest to audio system designers and anyone involved in the design of multimodal human-computer systems.

This work is dedicated to my parents, who have wholeheartedly supported me through everything that I have done in life to bring me down this path. Like most children I don't say it enough, but I love you and appreciate everything that you have done to raise me the way you have. I hope you see this as just as much of an accomplishment for you as it is for me.

ACKNOWLEDGMENTS

My appreciation first goes out to my committee, Dr. Kay Stanney, Dr. Linda Malone, and Dr. Dave Graeber. The guidance from each of you has allowed me to complete this hurdle in my academic career and I am grateful to each of you for that.

I am especially grateful to my advisor, Dr. Kay Stanney who has shaped my life in so many ways. As a teacher, she has provided me with invaluable skills that I will use for the rest of my life. As an advisor and mentor she has shaped my academic career and guided me to accomplish what I have today. As a colleague she has pushed me and challenged me to meet a higher standard. As my boss she has helped to direct my career and set professional goals for myself. I am grateful for the effect that she has had on each of these parts of my life but what I am most grateful for is her friendship. To a teacher, a mentor, an advisor, a colleague, and a friend, thank you for everything that you have done for me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

ACRONYM	Definition of Acronym
FOV	Field of View
GUI	Graphical User Interface
HIP	Human Information Processing
IOI	Item of Interest
MRT	Multiple Resource Theory
SEAS	Speech, Earcons, Auditory icons, Spatial audio
TCI	Time Critical Item of Interest
UAV	Unmanned Aerial Vehicle
VHT	Vehicle Health Task
WIMP	Windows, Icons, Menus, and Pointers
WM	Working Memory

CHAPTER ONE: GENERAL INTRODUCTION

As technology advances, computer systems are able to present increasing amounts of information to operators at ever faster rates. Although providing more information to operators does have the potential to allow them to make better decisions, if not presented correctly it can overwhelm operators and have adverse effects. Thus, it is the designer's job to balance the amount of information provided to system users and the presentation method for that data to maximize the amount of information that can effectively be presented.

Visual interfaces, utilizing the common visuo-spatial interaction paradigm of Windows, Icons, Menus, and Pointers (WIMP) to interact with users, generally prove effective when fairly simple tasks are performed, yet they quickly fail when multiple complex tasks are required of users. The primary reason for this failure may be that the visuo-spatial only interaction technique of WIMP interfaces does not take into account the human's ability to time-share human information processing (HIP) resources across multiple modalities (Wickens, 1984), which can lead to users becoming visually overwhelmed. To alleviate such shortcomings, a paradigm shift is required. The current study suggests that the unimodal WIMP design paradigm be extended to a multimodal paradigm that includes Speech, Auditory Icons, Earcons, and Spatial Audio (SEAS) cues.

Support is provided for such a shift in interface design focus by Multiple Resource Theory (MRT; Wickens, 1984). This theory suggests that individuals utilize a multidimensional system of independent resources consisting of distinct stages of processing (encoding, central processing, and responding), which involve various sensory modalities (visual, auditory), working memory (WM) processing codes (spatial, verbal), and response modalities (manual,

vocal). At each stage, resources are thought to be independent (e.g., verbal versus spatial WM resources), it is thus suggested that if tasks are designed to use separate and compatible resources, parallel processing can be preformed and tasks can ultimately be time-shared (Wickens, 1984). For example, individuals find it easier to attend to information when it is presented using multiple modalities (Parkes & Coleman, 1990; Penney, 1989; Rollins & Hendricks, 1980; Seagull et al., 2001; Wickens, Sandry, & Vidulich, 1983). Likewise, individuals perform better when they are required to respond to multiple tasks using separate modalities (Wickens, 1976; Wickens & Liu; 1988).

In light of MRT, it becomes apparent that the integration of audio to offload the visual demands of current interfaces has great potential. It is essential that such bimodal systems be designed around the capabilities and limitations of system users. For example, given audio's transient nature, designers should avoid using audio to present information that is not acted upon quickly, as it has the potential to increase WM demands of users. Instead, audio can be used to present information that requires a fast response, due to its capability to decrease reaction time when compared to visual systems (Bly, 1982; Dix, 1998). To ensure that audio is designed correctly and integrated into systems where it can provide the greatest utility, it is important that human-centered design guidelines be devised and validated. This work is directed at evaluating the utility of audio interaction and compiling such a set of guidelines.

The two studies presented herein provide insight into audio interaction techniques and present SEAS guidelines that can be used to direct the creation and integration of audio into multimodal systems. Specifically, chapter two (Jones, Samman, Stanney, & Graeber, 2005) discusses the results of a pilot study aimed at evaluating the utility of integrating SEAS cues into a primarily unimodal visuo-spatial interface to reduce perceptual and attentional bottlenecks.

Overall, the results of the pilot study showed some utility in the integration of audio, although many of the results were borderline significant. A power analysis performed on the results suggested that the lack of significance in the results may have been due to low observed power of the tests (p between 0.05 and 0.45). The power analysis also suggested that significance could potentially be found if the sample size was increased by a modest amount (14 participants). Based on this analysis the study was extended and the results are presented in chapter three.

In addition to presenting further case study evidence of the utility of integrating audio per the SEAS guidelines, chapter three (Jones, Stanney, & Graeber, submitted) also provides more detail about multimodal HIP and presents a list of theoretically derived SEAS guidelines. Together, these two chapters present scientific support for the integration of audio cues into human-computer systems, guidelines to follow when developing audio interfaces, and the results of two case studies that demonstrate the utility of the guidelines in directing audio system design. The theories and principles provided in this paper should be of interest to audio system designers and anyone involved in the design of multimodal human-computer systems.

CHAPTER TWO: STUDY ONE: SPEECH, EARCONS, AUDITROY SPATIAL SIGNALS (SEAS): AN AUDITROY MULTIMODAL APPROACH

Abstract

The present study examined how visual displays can be augmented with auditory cues to enhance performance. An auditory interactive computing paradigm was proposed utilizing Speech, Earcons, and Spatial signals (SEAS). SEAS cues were suggested to increase human information management capacity by leveraging multiple processing systems. This study focused on the ability of SEAS cues to overcome perceptual and attention processing bottlenecks when conducting Unmanned Air Vehicle (UAV) control tasks. The results demonstrated that SEAS cues enhanced human performance on UAV control tasks, particularly the response accuracy and reaction time on a secondary monitoring task (i.e., vehicle health task). The results suggest that SEAS may be effective in overcoming perceptual and attentional bottlenecks, with the advantages being most revealing during high workload conditions. The results of this study may be of interest to those designing information displays for multitasking environments.

Introduction

Since the 1980's and the instantiation of Graphical User Interfaces (GUI's), a paradigm of using spatial and visual information to influence how users interact with systems has extended across all types of interfaces. The most common interfaces in today's systems fall under a relatively standard set of interaction paradigms, which are collectively referred to as WIMP's

after their basic components of Windows, Icons, Menus, and Pointing Devices. Although this primarily visual interaction style does leverage the human visual system's exceptional ability to aid in the comprehension and understanding of the spatial information presented, restrictions may arise when the visual system is overloaded with information. As technology advances and systems are able to present more information to users at faster rates, this barrage effect is becoming more common.

One of the main goals of designers is to create interfaces that allow individuals to process an optimal amount of essential data while avoiding mental overload. To reach this goal, a paradigm shift from current primarily visual WIMP interactions to the addition of Speech, Earcons, and Auditory Spatial signals (SEAS) may be required. The addition of these auditory cues may serve to improve the information management capacity of the individual by enhancing perception, augmenting sensory processing, and speeding reaction time (Stanney, et al., 2003). In effect, SEAS may help to overcome human information processing (i.e., perceptual, attention, working memory, executive functioning) bottlenecks. The current study serves to evaluate the effectiveness of the SEAS paradigm on overcoming perceptual and attentional bottlenecks.

To reduce cognitive overload, Wickens' (1984) Multiple Resource Theory (MRT) suggests that tasks are more efficiently time-shared when multiple resources are used in terms of sensory/perceptual modalities. It has been found that individuals find it easier to recognize information displayed using multiple modalities (e.g., visual and auditory) than using one modality (Seagull et al., 2001). For instance, Wickens (1980) reviewed several studies and found greater advantages to cross-modal (e.g., visual and auditory) over intra-modal displays (e.g., visual and visual). Furthermore, research has suggested that attention processing is relatively easy when objects are physically distinct from distracters (Proctor & Van Zandt, 1994).

Exploiting an individual's capacity to attend to a wide variety of different sound dimensions in terms of location, pitch, and intensity is suggested to assist in directing attention while enhancing human information processing (Samman, Jones, Stanney, & Graeber, 2004).

Method

An experiment was designed to examine the effectiveness of the SEAS auditory paradigm to offload visual cues in an Unmanned Air Vehicle (UAV) operational setting. Workload was intensified by manipulating the number of UAV groups that were controlled; specifically, the number of vehicles that operators were required to control increased from four to eight to twelve vehicles in each scenario. The challenge of controlling multiple groups in a UAV system (4, 8, and 12 vehicles) was suggested to dramatically increase the mental workload for operators. By adding second and third groups, task conflicts can grow enormously.

Participants

Sixteen university students (3 females and 13 males) were recruited to participate in this study. Participants had a mean age of 19.53 years (SD= 2.17) with a range of 17-25 years. Fourteen participants were right-handed and two were left-handed. The average number of hours playing video games equaled 12.47 hours per week (SD= 9.19) while the average time spent using computers equaled 22.37 hours per week (SD= 19.85).

Apparatus

Tasks were performed on a 3.0 GHz Dell Inspiron 9100 computer with a 128 Mb Radeon 9600 video card. The interface was presented on two 17" NEC Multisync LCD displays at 1280x1024 screen resolution. Audio was presented through a set of Plantronics DSP 500 noise-cancelling headphones, which allowed for spatialized sound presentation. User input was made with a standard 2-button mouse.

Tasks

Main Task

Each participant performed a series of simulated UAV control tasks under various workload conditions. The primary task was to set up sorties on pre-planned items of interest throughout the flight path. For each group of 4 UAVs that were being controlled, twenty pre-planned items of interest were laid out within the environment. To successfully complete a sortie, each UAV had to be paired with an item of interest. Participants were required to search for the type of asset each item of interest needed, denoted by the letters S, M, and H (small, medium, and heavy assets) that appeared in a text box near the items of interest when the mouse was rolled over them. Participants were then required to locate an available UAV carrying the same type of asset and pair UAV to item of interest. The asset type and number of assets carried by each UAV was denoted by a number paired with an S, M, or H that appeared under each UAV (see Figure 1). This search task was performed on a map display that presented all controlled UAVs and items of interest to participants at all times.

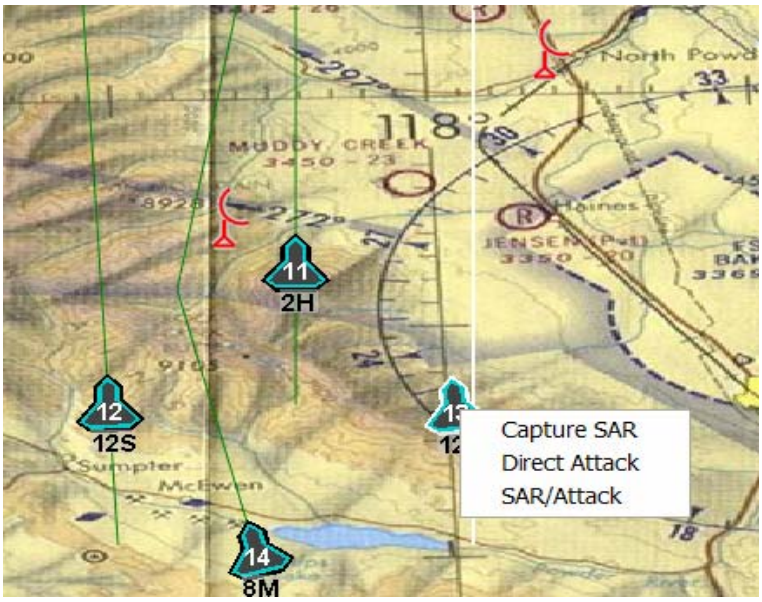


Figure 1: UAV control display

Once a match was found participants were then required to use a context menu selection to take a radar image and to initiate the sortie. Once paired, a line appeared to connect the item of interest with the UAV completing the sortie (Figure 2a). When a UAV was sufficiently close to the item of interest, a radar image was captured. An icon resembling a satellite image replaced the background of the current item of interest icon (Figure 2a). Once the image was available, participants were required to view it by clicking on the icon. An asset allocation window appeared on the right display presenting a detailed radar image of the item of interest. Participants were required to pinpoint the location of the asset drop point on the radar image. Once asset allocation was performed, the item of interest's icon displayed a red triangle depicting the number of assets that were to be dispensed on the item of interest (see Figure 2b). After the precise asset allocation point was selected, the UAV automatically performed the sortie and the participant was not required to monitor it until the asset drop was complete. Following

the release of assets, a visual icon was presented to symbolize the asset drop (see Figure 2c). This icon represented the completion of the sortie and the availability of the UAV to pair with another item of interest. Once asset release was completed, the line connecting UAV to item of interest disappeared, further indicating that the UAV was free to pair with additional items of interest. Upon asset release, the participant was tasked with reallocating the UAV to another item of interest. The subtasks of viewing the radar image and perceiving that a UAV is available for pairing with an item of interest were used to evaluate the usefulness of SEAS cues to increase perception rate and decrease time required to perceive events in the environment.

Figure 2: A) Radar image available Icon B) Number of assets dispensed icon C) Asset release icon

Secondary Tasks

Throughout each mission, two Time Critical Items of Interest (TCIs) appeared suddenly on the map display for each set of 4 UAVs. Participants were required to attend to and

immediately react to these items of interest in the same fashion as the pre-planned items of interest. Time critical icons looked similar to normal item of interest icons but were encircled in red to denote their importance. Participants were required to pair UAVs with TCIs and perform a sortie on them within 10 seconds after they appeared.

In addition, participants were required to detect Vehicle Health Tasks (VHTs) which arose throughout each mission and respond to health questions that were asked after highlighting the VHT issue. This task also required immediate attention. When health problems occurred, the UAV was outlined in red. Once perceived, participants were required to double-click on the UAV that needed attention. This brought up a health text box that displayed a health question that participants were required to respond to. Since both of these tasks required participants to selectively attend to the TCIs and VHTs directly after appearance, they were used to evaluate the effectiveness of SEAS cues to increase the percentage of cues that are attended to and decrease the time required to attend to them.

Mitigations

During UAV task analysis, several perceptual bottlenecks were identified. When participants performed the main task (pair UAV to item of interest, capture radar image, perform sortie on item of interest), visual overload hindered them in perceiving that a radar image was ready for viewing. SEAS auditory cues were used to augment the visual display to prevent this overload. A spatialized auditory icon (camera shot sound) was played to denote the presence and location of the available radar image. The auditory icon was spatialized left, center, or right to guide the user to where they should look within the display. The integration of this spatialized

auditory icon was proposed to offload the visual-spatial load traditionally associated with radar images to an auditory-spatial load. Furthermore, participants were required to monitor the status of UAVs (completion of sortie) to reassign them to new items of interest. Participants were visually saturated with cues and were unable to efficiently perceive when a UAV was free. SEAS auditory cues were augmented to the visual display to alleviate this problem. Spatialized earcons were integrated to denote the location and type of UAV (type denoted by asset carried) that was available. The four UAVs were mapped with distinctly different timbres. The lead vehicle carrying heavy assets was paired with a brass instrument playing a note at two octaves below middle C. The middle UAVs in the diamond shaped formation (see Figure 1), carrying small assets were paired with a vibraphone (left vehicle) and a pan flute (right vehicle). Both played two octaves above middle C. The rear UAV, carrying medium assets was paired with a piano note playing at middle C. Different timbres with various octaves denoted asset size (high octave = small asset, medium octave = medium asset, low octave = heavy asset) and differentiated each UAV in the formation. These earcons were spatialized to originate from the onscreen position of the group that they were in (left, right, center). The integration of spatialized auditory icons and earcons were proposed to transform the task from one of a purely visually scanning search to a tonal cue detection task (cues symbolize that radar image is available and that the UAV is free to pair with additional items of interest).

Attentional bottlenecks were also identified within the secondary tasks. Due to the urgency of TCIs, participants were required to attend to them immediately. Therefore, SEAS augmented display accompanied each TCI with a concise speech message spoken in a natural voice, stating “critical target”. The message was spatialized in accordance with the location of the TCI (left, right, center). In addition, the message was played in different voices depending on

the location of TCI on the display. A male voice was presented if the TCI was included in the set of items of interest closest to the left side of the display. A female voice was played for the center, and a different male voice was used for the right side. These mitigations were expected to transform the purely visual search task to an auditorily guided search task.

Participants were also required to monitor UAVs for health problems while performing the primary task. The SEAS augmented auditory display integrated a spatialized speech cue played to alert participants of health problems. The spatial location of the message coincided with the location of the vehicle (left, right, center). A short concise message stating "Health alert" was played to correspond with the occurrence of health difficulties. The same voice assignments used for the SEAS TCI alert described above were also employed in the VHT task alerts. These mitigations were proposed to transform the task from continuous visual scanning of health alerts to an auditorily directed search of health alerts.

Procedure

Prior to beginning, participants completed an informed consent and demographics questionnaire. Each participant then performed a training session familiarizing them with UAV control tasks. Participants were trained on how to properly pair UAVs to items of interest, how to control UAVs to perform sorties, where to place asset allocation points for each item of interest, and how to recognize and handle TCIs and VHTs. Following training, rules of engagement procedures and strategies were explained to participants. Participants were seated in front of two monitor displays. The monitor situated in front of the participant presented an updating map display to select and pair items of interest and receive information about TCIs and VHTs. The

monitor on the right of the participant was only used to present an asset allocation window that was used to determine the precise asset drop points for each item of interest.

Prior to testing, participants performed a practice trial operating four UAVs at a speed slower than testing speed (700 knots). Participants were then required to perform an additional practice trail at a speed of 800 knots operating four UAVs. Before testing, participants were required to successfully complete 65% of the sorties at a low workload level (four UAVs). This guaranteed that all participants were at the same baseline performance level. Each test session consisted of three test trials evaluating four, eight, and twelve UAVs flying at speed of 800 knots. The number of operated vehicles was used to manipulate participant's workload. Participants were required to perform tasks on two interface conditions (Baseline- visual display, SEAS-augmenting auditory cues to visual display). To reduce order and practice effects, the order of interface presentation was counterbalanced. Prior to performing tasks on the SEAS interface, participants were trained on each sound employed (i.e., camera shot sound, earcons). Accuracy and reaction time were used to assess performance. In addition, following the completion of each UAV interface evaluation (Baseline, SEAS), a workload and situational awareness questionnaire was completed by each participant.

Experimental Design

In order to evaluate the effectiveness of the Baseline and SEAS interfaces at various workloads levels, a 2x3 (interface type x workload) within-subject design was implemented. The two interface types that were compared consisted of the Baseline visual interface and the SEAS augmented auditory interface. The three levels of workload that each interface operated at were

with the control of 4, 8, and 12 UAVs. A one-way repeated measure ANOVA was performed to test for significance on each performance measure. There were both perceptual and selective attention performance measures recorded. Perceptual performance measures included: 1) the effectiveness of SEAS spatialized auditory icon cues to present an update in radar imaging status, which was assessed in terms of the percentage of radar images that were viewed and the time required to view radar images; and 2) the effectiveness of SEAS spatialized earcons to present UAV status updates (when a sortie was completed), which was assessed in terms of the time required to reassign UAVs to new items of interest after becoming available. In order to evaluate the extent that the integrated SEAS cues helped to alleviate potential attentional bottlenecks, performance measures of two attention tasks were compared between the two interfaces and across workload levels. First, to evaluate the effectiveness of SEAS spatialized speech cues to facilitate TCI detection, reaction time and the number of TCIs detected were compared. Second, to evaluate the effectiveness of using SEAS spatialized speech cues to facilitate VHT detection tasks, the number of VHTs detected and time required to detect them were analyzed. In addition, to evaluate overall performance of the UAV/item of interest pairing task, metrics including the percentages of radar images taken, assets used, and items of interest hit were analyzed. Items of interest hit represented the number of items of interest successfully dealt with via UAV sortie.

Results

Perception Evaluation Measures

The percentage of radar images viewed showed no significant main effects and no interaction effects for workload and interface factors between the Baseline and SEAS conditions. Although there was no significant main effect found for the interface type used, the percentage of radar images viewed demonstrated a trend that approached significance ($F(1, 14) = 3.507$, $p = .082$). The percentage of radar images viewed was slightly higher in the SEAS augmented display than the Baseline display, particularly at higher levels of workload (see Table 1). The non-significant results may have been largely due to the low observed power of the test ($p = 0.415$).

The analysis of reaction time to radar images demonstrated a significant main effect of workload level ($F(2, 28) = 14.4$, $p < .05$). A Least Significance Difference (LSD) post-hoc analysis showed that as workload increased, the time required to view radar icons also increased ($p < .05$ for all workload main effect comparisons). No significant main effect based on interface type was found ($F(1, 14) = 0.06$, $p = .81$). As demonstrated in Table 1, the average reaction time to view radar icons was 14.9% lower when using the SEAS interface than the Baseline interface under the highest level of workload. These results support the SEAS principle of using spatialized earcons to reduce perceptual bottlenecks in primarily visual systems.

Table 1: Radar icon detection performance for Baseline and SEAS displays

Workload level	Baseline Performance		SEAS Performance	
	Radar Image Viewed (%)	Radar Image Reaction Time (seconds)	Radar Image Viewed (%)	Radar Image Reaction Time (seconds)
1	97.73	3.57	98.94	4.01
2	98.26	5.67	99.52	5.73
3	91.82	8.47	96.52	7.21
Average	95.94	5.9	98.33	5.65

Results regarding the effectiveness of SEAS spatialized earcons to cue UAV status updates demonstrated a significant main effect for workload ($F(2, 28) = 45.91$, $p < .05$). An LSD post-hoc comparison showed that as workload increased, the time required to reassign an available UAV to items of interest also increased. As demonstrated in Table 2, increases in workload led to increases in vehicle reassignment times ($p < .05$ for all workload comparisons). There was no significant main effect found based on interface type ($F(1, 14) = 0.935$, $p = .35$). Again, a trend was found toward lower vehicle status detection times while using the SEAS display when compared to the Baseline display. The non-significant results may have been largely due to the low observed power of the test ($p = 0.147$). This trend also shows that as workload increased, the performance advantages of using the SEAS interface became more apparent for the vehicle status detection task. These results suggest the potential of SEAS but clearly indicate more research is needed to demonstrate its effectiveness in reducing the perceptual bottlenecks in primarily visual systems.

Table 2: Vehicle status detection for Baseline and SEAS displays

Workload level	Baseline – Vehicle status detection (seconds)	SEAS – Vehicle status detection (seconds)
1	30.65	29.5
2	37.68	35.49
3	54.92	50.59
Average	41.08	38.53

Attention Evaluation Measures

The number of TCIs detected and time required to detect TCIs both demonstrated significant main effects for workload level ($F(2, 28) = 16.16$, $p < .05$, $F(2, 28) = 5.16$, $p < .05$, respectively). LSD post-hoc comparisons showed that as workload increased, the number of TCIs detected significantly decreased ($p < .05$) and the time required to detect TCIs increased ($p < .05$). Trends arose again with regards to the SEAS versus Baseline comparison. The average percentage of TCIs detected was slightly higher and the time required to detect TCIs was slightly lower while using the SEAS display as compared to the Baseline display. In addition, a trend of increased performance differences as workload is increased is also present. The non-significant results may have been largely due to the observed power of the test for TCI reaction time and accuracy ($p = 0.063$; $p = 0.072$ respectively). These results demonstrate the potential of SEAS principles for enhancing attention, but more research is needed.

Table 3: TCI detection in Baseline and SEAS displays

Workload level	Baseline Performance		SEAS Performance	
	TCI detected (%)	TCI detected (seconds)	TCI detected (%)	TCI detected (seconds)
1	100	8.77	100	10.43
2	75	16.94	74.467	15.16
3	67.6	17.33	72.93	14.97
Average	80.87	14.35	82.47	13.52

Analysis of the number of VHTs detected showed significant differences between the performance on the Baseline and SEAS interfaces ($F(1, 14) = 13.01$, $p < .05$), and workload levels ($F(2, 28) = 4.12$, $p < .05$). Furthermore, a significant interaction effect was found between workload level and interface used ($F(2, 28) = 4.476$, $p < .05$). A LSD post-hoc analysis demonstrated that when participants used the SEAS augmented display, more VHTs were handled ($p < .05$) and the performance increases due to the display used increased at higher workload levels. When evaluating the time required to react to VHTs, a significant main effect between interfaces used was found ($F(1, 14) = 18.22$, $p < .05$). As demonstrated in Table 4, the average number of VHT tasks detected was 36.9% higher, on average, and reaction time was 47.1% faster with the SEAS as compared to the Baseline. It is important to note that the performance increases accredited to the use of the SEAS interface are more apparent as workload is increased. A significant difference was also found between subjective perceived mental workload while performing the VHT tasks ($t(14) = 3.51$, $p < .05$) when using the two interfaces. Participants considered the VHT task more demanding in the Baseline condition than in the SEAS condition. These results support the SEAS principle of using concise spatialized speech messages to conveying warning information.

Table 4: VHT detection performance in Baseline and SEAS displays

Workload level	Baseline Performance		SEAS Performance	
	VHT detected (%)	VHT detected (seconds)	VHT detected (%)	VHT detected (seconds)
1	70	6.38	81.67	3.527
2	43.33	8.06	93.33	4.26
3	43.07	9.07	73	4.64
Average	52.13	7.83	82.67	4.14

Other Performance Measures

The percentage of radar images taken showed a significant main effect of workload level ($F(2, 28) = 40.331$, $p < .05$). An LSD post-hoc analysis demonstrated that as workload increased (4 to 12 UAVs), the percentage of radar images taken decreased ($p < .05$ for all comparisons). Although there was not a significant main effect of interface used between Baseline and SEAS ($F(1, 14) = 2.688$, $p = .123$), Table 5 demonstrates a trend, showing that the percentage of radar images taken slightly increased with the use of the SEAS augmented interface and performance differences between the interfaces were slightly more apparent as workload increased. The non-significant results may have been largely due to the low observed power of the test ($p = 0.333$).

Analysis of the percentages of weapons used by participants showed a significant main effect for workload level ($F(2, 28) = 37.673$, $p < .05$). An LSD post-hoc analysis showed a significant decreasing trend in the percentage of weapons used as workload was increased ($p < .005$ for all comparisons). Although there was not a significant main effect for interface used ($F(1, 14) = 1.266$, $p = .279$), a pattern is demonstrated of participants using slightly more weapons in the SEAS augmented interface than in the Baseline.

The percentage of items of interest hit demonstrated a significant main effect of workload level ($F(2, 28) = 23.98$, $p < .05$). An LSD post-hoc analysis of the effect of workload levels on the number of items of interest hit showed that as workload was increased, the percentage of items of interest hit significantly decreased ($p < .05$ for all comparisons). Although the number of items of interest hit did not show a significant main effect based on the interface used, as can be seen in Table 5, on average the number of items of interest hit while using the SEAS augmented display was slightly higher than that found using Baseline displays, thus approaching significance ($F(1, 14) = 2.899$, $p = .111$).

Table 5: Results for Baseline and SEAS TSD performance

Workload level	Baseline Percentage Performance (%)			SEAS Percentage Performance (%)		
	Radar Images Taken	Items of Interest Hit	Assets Used	Radar Images Taken	Items of Interest Hit	Assets Used
1	93.67	88.67	86.4	96.33	91	91.93
2	85.5	72.73	73.6	87.17	80.13	73.53
3	69.87	66.93	61.07	76.31	68.6	62.93
Average	83.01	76.11	73.69	86.6	79.91	76.13

When assessing response accuracy of vehicle health tasks, significant main effects for the interface used ($F(1, 14) = 13.01$, $p < .05$) and workload levels ($F(2, 28) = 4.12$, $p < .05$) were found. A LSD post-hoc comparison showed that when participants used the SEAS augmented display, on average, 37.5% more health tasks were answered correctly ($p < .05$). Average response accuracy for different workload conditions in Baseline and SEAS are presented in Table 6.

Table 6: VHT response accuracy in Baseline and SEAS displays

Workload level	Baseline – VHT response accuracy (%)	SEAS – VHT response accuracy (%)
1	63.33	90.13
2	40	71.67
3	34.27	58.47
Average	45.87	73.42

Subjective Workload and Situational Awareness Measures

Cooper-Harper subjective workload ratings demonstrated that the use of SEAS augmented interface led to a lower perceived mental workload level of the entire task when compared to the use of the Baseline interface ($t(14) = 2.79$, $p < 0.05$). With similar trends, a NASA-TLX subjective workload rating demonstrated that participants considered the SEAS display as less demanding than the Baseline ($t(14) = 1.97$, $p = .065$). As Table 7 demonstrates the NASA-TLX workload factors including mental, temporal, performance, effort, and frustration were perceived as slightly less demanding while using the SEAS display as compared to Baseline display.

Table 7: NASA-TLX subjective perceived mental workload

Interface	Mental	Temporal	Performance	Effort	Frustration	Average
Baseline	16.06	15.88	10.50	16.38	13.56	13.14
SEAS	14.50	14.31	7.50	14.56	12.38	11.74
Difference	**1.56**	**1.57**	**3**	**1.82**	**1.18**	**1.4**

Furthermore, as can be seen in Table 8, Cooper-Harper workload assessment demonstrates that several tasks including radar image detection, vehicle status detection, and VHT detection and response, are perceived as less demanding when using the SEAS interface.

Table 8: Cooper-Harper subjective perceived mental workload levels

Interface	VHT Detection	SAR Detection	Vehicle Status Detection	VHT Response	Overall Task	Average
Baseline	5.31	4.63	4.50	6.13	7.19	5.04
SEAS	3.13	3.44	4.44	5.94	5.40	4.72
Difference	**2.18**	**1.19**	**0.06**	**0.19**	**1.79**	**0.32**

Discussion

The general pattern of results for all of the performance metrics recorded support the concept of integrating SEAS guidelines into visual displays. The objective of SEAS auditory cues in this study is to reduce attentional and visual modality perceptual bottlenecks in an applied system. It was hypothesized that the integration of SEAS cues would reduce perceptual bottlenecks present in the experimental testbed and as a result operators would perform perceptual tasks faster and more consistently. The results of the radar image detection and vehicle status change detection tasks did not support this hypothesis at a significant level. It is expected that this is due to the limited sample size that was evaluated and power analyses suggest that if this sample size was increased, significance would be obtained. The general pattern of the results suggest that as radar images became available to participants, they viewed more of them and viewed them at faster rates when there was a spatial auditory icon used to

direct them to the icons presence and position. Likewise, as UAVs became free to use again, participants utilized them at a faster rate when there was a spatial earcon present to guide them to the location and type of aircraft that was available. Furthermore, as workload was increased, performance gains accredited to the integration of SEAS cues became more apparent, suggesting that the use of such cues may lead to better perceptual gains as workload is increased.

It was also hypothesized that the integration of SEAS cues would lead to increased performance on attention tasks. The results based on the vehicle health tasks support this hypothesis. Evaluation of the VHT task demonstrates that operators were able to attend to more health problems at a faster rate when visual cues were augmented with spatialized voice alerts. Redundant multimodal signal effects may have increased parallel processing while perceiving a health problem, attending to the vehicle, and in responding to the problem. Attending to the vehicle health tasks was quicker in the SEAS condition than in the Baseline condition since operators were able to aurally perceive the vehicle that required attention. Although the efficiency of responding to TCIs and VHTs decreased as workload increased for both displays, this falloff was not as substantial with the use of the SEAS display, leading to larger performance differences between the display types in high workload conditions. This suggests that the integration of SEAS cues will be more effective at supporting selective attention tasks as workload levels are increased. Interestingly, a significant difference was also found for the accuracy of vehicle health task. Operators were able to resolve the health problem more accurately in SEAS condition than in Baseline condition. This may be attributed to improved alertness level of operators and the opportunity to process the health alerts and formulate an answer in parallel with other tasks when performing with SEAS interface.

The integration of SEAS cues into the interface also had positive effects for tasks that were not explicitly hypothesized to be affected by the display changes. Although not significantly different, patterns were present showing that when the SEAS interface was used, on average, the number of radar images taken, assets used, and items of interest hit all increased at a level approaching significance. This lack of significance in the results may be due to the small sample size evaluated in this study, causing low power. It is expected that the low power can be increased by increasing the number of participants. During the UAV/item of interest pairing task, operators were required to pair vehicles with items of interest, capture radar images, and perform sorties. In the Baseline condition, operators who were engaged in these tasks were required to divert their gaze continuously, thereby degrading their performance on this primary task. Such diversion did not need to occur when auditory cues denoting when radar images were complete, UAVs were available, TCIs appeared, or VHTs needed attention were integrated. Based on the multimodal resource modal, employing multiple sensory modalities (visual, auditory) increased dual task performance and efficiency. Given the availability of separate visual and auditory perceptual resources, information was better time-shared and that is what led to the performance increases on these tasks while using the SEAS interface.

Subjective workload was also perceived as lower with the use of the SEAS system than with the Baseline system. Operators perceived SEAS as less demanding in terms of mental, temporal, effort, frustration, and overall performance. These findings are comparable to the objective results discussed above.

Conclusions and Future Research

The results of this study suggest that the integration of SEAS cues has the potential to reduce attentional and perceptual bottlenecks when integrated into a primarily visual display. Future research should examine the effects of auditory cues on working memory and executive functioning bottlenecks. In addition, due to increased performance advantages in high workload conditions, the integration of SEAS cues should be studied at increased workload conditions.

Acknowledgement

This work has been sponsored by Dr. David Graeber, Boeing Phantom Works, The Boeing Company, under purchase contract KR7770. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views or endorsement of The Boeing Company.

CHAPTER THREE: STUDY TWO: AUDIO INTERACTION PARADIGMS: GUIDELINES FOR SPEECH, EARCONS, AUDITORY ICONS, AND SPATIAL AUDIO

Abstract

Most human-computer systems designed today utilize visual widgets such as windows, icons, menus, and pointers to interact with system users. This interaction paradigm proves to be effective when single tasks are being performed in low-workload conditions; however, it can quickly lead to visual overload under high workload or multitasking conditions. To alleviate the visual overload problem, the results of this study suggest that designers should consider using the Speech, Earcons, Auditor icons, and Spatial audio (SEAS) paradigm to augment visual-only interfaces. The introduction of this interaction paradigm takes advantage of the human's ability to effectively time-share tasks that utilize multiple sensory and perceptual resources. As a first step to realizing this goal, a number of theoretically derived guidelines for the design of SEAS cues and a case study evaluating their utility are presented. The results suggest that integration of the SEAS paradigm into a formerly visual-only interface can reduce perceptual and attentional overload. The results of this research and the guidelines presented herein should be of interest to developers of UAV control interfaces, in particular, and interactive system designers, in general.

Introduction

The Windows, Icons, Menus, and Pointers (WIMP) interaction paradigm has become the most pervasive interface technique used today. By leveraging the human visual system's ability

to aid in the comprehension and understanding of visuo-spatial information, WIMP interfaces allow users to interact with systems though recognition of graphical widgets instead of requiring users to remember complex interaction commands. Yet, interface systems that restrict interaction to visuo-spatial techniques alone have the potential to quickly overload users' visual perceptual and attentional resources and fail to take into account the human's ability to time-share human information processing (HIP) resources across multiple modalities (Wickens, 1984). In an effort to move toward interactions that leverage multiple human sensory systems, the current work suggests the use of audio interaction and introduces the Speech, Earcons, Auditory icons, and Spatial signals (SEAS) design paradigm.

This article focuses on explaining the utility of, and providing guidelines for the integration of SEAS cues into human-computer systems. The theory behind the utility of the SEAS interaction paradigm is framed in multimodal HIP and is followed by theoretical design guidelines for each aspect of the SEAS paradigm. Finally, the results of a case study in applying a subset of SEAS design guidelines are presented. The theories and principles provided in this article should be of interest to audio system designers and anyone involved in the design of multimodal human-computer systems.

Human Information Processing

Stage theory models of human information processing provide a representation of how humans interact with their environment by processing information in a serial, discontinuous manner (Atkinson & Shriffin, 1968). At a high level, three basic stages are serially performed under stage-theory HIP models. The process generally begins when an individual senses a

stimulus in the environment. Stimulus sensation can occur through multiple sensory processors and requires that the stimulus is of the correct intensity and format to be sensed by the organ that encounters it. Once a stimulus is sensed, it is perceptually encoded and used to make a decision, which is then executed to provide a response back to the environment (Proctor & Van Zandt, 1994). The perception and decision making processes of HIP are supported by a working memory subsystem that exploits long-term memory (Wickens, 1992). During the perception stage of HIP, working memory is utilized to guide bottom-up processing or to pull from long-term memory to guide top-down processing of perceived cues (Woodman, Verca, & Luck, 2003). The decision making stage then requires that working memory pull from long-term memory and actively rehearse information in order to make a decision based on perceived cues.

Due to its central roles in HIP, working memory has been referred to as a functional multiple-component of cognition "that allows humans to comprehend and mentally represent their immediate environment, to retain information about their immediate past experience, to support the acquisition of new knowledge, to solve problems, and to formulate, relate, and act on current goals" (Baddeley & Logie, 1999, p. 29). Many working memory models have been postulated and, in general, they all suggest the existence of different codes or representations (e.g., separate storage buffers) based on sensory modalities (Miyake & Shah, 1999). Thus, from an HIP perspective "multimodal interaction has promise because the WM subsystems are somewhat independent and tend to act cooperatively rather than competitively (i.e., do not entirely compete for the same processing resources)" (Stanney et al., 2004, p. 233). Human-systems interaction designs that take advantage of these multiple HIP resources should yield substantial human performance benefits. Before this can be done, it is important to first develop theories to guide the design of such multimodal systems.

Multiple Research Theory

One of the few theories supporting and guiding the design of systems that support divided attention among multiple modalities is Multiple Resource Theory (MRT), which was originally proposed by Kantowitz and Knight (1976) and extended by Wickens (1980; 1984; 1992). Building from stage-theory HIP models, MRT suggests a multidimensional system of resources consisting of distinct stages of processing (encoding, central processing, and responding), which involve various sensory modalities (visual, auditory), WM processing codes (spatial, verbal), and response modalities (manual, vocal). Each stage's resources are thought to be independent and thus allow parallel processing and time-sharing of tasks with little interference if tasks are designed to use separate and compatible resources (Wickens, 1984). Neuroimaging research (Smith & Jonides, 1998) and neuropsychological studies of brain-damaged patients (Carlesimo, Perri, Turriziani, Tomaiuolo, & Caltagirone, 2001; Mendez, 2001; Pickering, 2001) support this theory, suggesting that separate portions of the brain are activated based on the resources that are being pulled from.

Numerous empirical studies also support the MRT model. In general, these studies suggest that the utilization of the separate resources outlined in MRT leads to individuals being able to process and recall more information as compared to single modality presentation (Baddeley, 1990; Cowan, 2000; Klapp & Netick, 1988; Penney, 1989; Sulzen, 2001). In addition, research also suggests that system users find it easier to attend to information displayed using multiple modalities when compared to unimodal systems (Parkes & Coleman, 1990; Penney, 1989; Rollins & Hendricks, 1980; Seagull et al., 2001; Wickens, Sandry, & Vidulich, 1983). For example, in regards to the sensory stage of the model, Wickens (1980) performed a

review of several studies and found greater utility in the use of cross-modal (e.g. visual and auditory) over intra-modal displays (e.g. visual and visual). In regards to the response stage, Wickens and Liu (1988) have shown that when individuals are required to perform a manual tracking task while simultaneously verbally responding to a tone identification task, they perform better than when a manual response is required on the secondary task. In the same regard, Wickens (1976) has shown that individuals perform better when task responses are distributed across manual and auditory inputs as opposed to requiring two manual responses when simultaneous tasks are performed. Similar results have been found in the design of Unmanned Aerial Vehicle (UAV) interfaces, which are of particular interest to this study. Specifically, Draper, Calhoun, Williamson, Ruff, and Barry (2003) found that by offloading manual response tasks to speech input to take advantage of multiple resources, response time was reduced on average by 40%.

Research has been directed at evaluating the utility of integrating additional modalities and utilizing multiple HIP resources to increase situational awareness (SA) and increase monitoring task detection rates during UAV tasks. For example, Draper and Ruff (2000) have shown that by augmenting a simulated aircraft landing task with haptic cues to indicate turbulence, an increase in controller SA of such events can be obtained. Likewise, Wickens and Dixon (2002) found that by offloading visual UAV control displays to auditory interfaces, increases in detection rates and reductions in response times can be achieved. In an attempt to directly compare the utility of using tactile and auditory alerts in directing attention to visual tasks, Calhoun, Ruff, Draper, and Guilfoos (2005) found that even in high auditory workload conditions, addition of either modality shows potential to direct controllers' attention to activities occurring on a visual display. The performance gains found in each of these UAV interface

studies can generally be attributed to an increase in parallel processing among visual and auditory/haptics sensory and response channels. More rigorous design principles are needed to guide the design of such multimodal interfaces for complex systems.

Each of the studies described in this section provide support for MRT, which strongly advocates the use of multiple modalities to reduce bottlenecks in all stages of human information processing. While MRT provides a solid foundation from which to design bimodal visual-auditory systems, the current study aims to provide guidelines for how to fully exploit the auditory modality in bimodal designs. The following section provides an introduction to the SEAS interaction design paradigm, preceded by an exploration of the general benefits of audio interaction.

SEAS Interaction Design Paradigm

Audio has the potential to enhance HIP and, system interaction in general, because of its omni-directional characteristics, ability to direct attention, and acute temporal resolution. One of the most evident of these advantages is audio's omni-directional characteristics. The human visual field of view (FOV) is restricted to 80º lateral by 60º vertical (Perrott, Sadralobadi, Saberi & Strybel, 1991), with the area of best acuity limited to 2º around the point of fixation (Rayner & Pollatsek, 1989). The auditory system is not restricted by such limitations and can be used to receive cues from 360º around an individual. When coupled with the visual modality, audio can be useful to direct system users' attention to important visual cues in the environment that are both within and outside of their current FOV or area of focus. For example, Perrott, et, al. (1991) demonstrated that using audio to guide visual search led to a reduction in the amount of

time required to detect a target both when a substantial gaze shift was required as well as when the target was in the viewers' initial area of gaze.

Another advantage of audio is its acute temporal resolution (Kramer, 1994). When this is combined with audio's apparent obligatory access to processing (as opposed to written text which must first enter the subvocal rehearsal loop to be recoded; Baddeley, 1986), additional advantages of audio develop. For example, Bly (1982) has shown that reactions to audio are generally faster than reactions to visual cues. More precisely, Dix (1998) suggests that individuals react to visual cues in 200 ms while it only requires 150 ms to react to auditory cues.

In addition to audio being omni-directional and the potential for it to lead to faster reaction times, sound is simply a more natural way to represent some types of information. For example, using a method known as data aurilisation, complex multidimensional data can be presented quickly in a combined form using audio (Gaver, 1997). By mapping data parameters to different parameters of sound, sounds can be created to present patterns and variations within complex data sets. Mansur, Blattner, and Joy (1985) showed the effectiveness of this approach by mapping two-dimensional graphs to audio parameters; although the true utility of data aurilisation becomes evident when more than two dimensions of data must be mapped. For example, Bly (1982) demonstrated that four variables could be mapped to separate sound characteristics to guide the efficient classification of data.

While audio interaction holds great promise, designers must take care when integrating audio into their designs. Many parameters of audio do not allow for the presentation of high-resolution information (Brewster, 1994). For example, only a very limited number of spatial positions of audio can be differentiated by the human listener. Specifically, individuals have a spatial resolution of one degree when audio is presented in front of them but can only

differentiate spatial audio positions separated by 10-15 degrees if presented to the side (Wenzel, 1992). Another issue that must be taken into account when designing audio is that attributes of audio are not orthogonal. For example, changing the pitch of a cue may affect its perceived loudness and vice-versa (Brewster, 1994). Two additional concerns with audio are the potential for it to be annoying and its transient nature (e.g., sound disappears directly after presentation; Jones, 1989). The use of the SEAS audio interaction paradigm promises great utility if these limitations of audio are taken into account during design.

In the following section, audio has been separated into the three categories that make up the SEAS paradigm: speech, earcons/auditory icons, and spatial audio. Each section will explain the utility of these categories of audio and present a number of theoretically derived guidelines for the design and integration of audio into human-systems interaction design. Following the presentation of the guidelines, a case study will be presented that evaluates the utility of applying a subset of these theoretical guidelines.

Speech

Speech is a natural interaction mode. The intuitive nature of speech stems from the fact that humans use speech in their day to day lives to communicate with each other. Thus, integrating speech audio into interfaces pulls from the vast experience that users already have with this interaction technique (c.f. the anthropomorphic approach, Eberts, 1994). Speech shows great potential to be used in warnings, to direct a reaction to a change in system status, and to provide detailed information about a system. Yet, speech has the potential to add to operator

workload and pulls from the limited attentional resources of the human operator (Badeley, 1990); so effective design of speech cues is essential.

One of the first questions that speech system designers must resolve when creating speech output systems is whether to use natural, synthetic, or mixed speech. A number of studies advise against the use of mixed speech systems as both purely natural and synthetic speech systems lead to better performance and a higher level of user trust (Gong & Lai, 2003; Gong, Nass, Simard, & Takhteyev, 2001). Although it is more costly and difficult to implement, research has also shown that the use of human speech is generally liked more, requires less time to become accustomed to (Francis & Nusbaum, 1999), and leads to higher comprehension levels (Tsimhoni, Green, & Lai, 2001) than synthetic speech. Taken together, this research suggests that optimal speech output systems should integrate natural human speech whenever possible and avoid the use of mixed (human/synthetic) speech systems.

Other decisions that have to be made by speech system designers are what volume level, pitch, and speech rate output should be set at. A study by Scherz (2003) examined the effects of varying each of these aspects on the intelligibility of speech output and found that increasing the pitch of messages led to the highest intelligibility levels while increasing the speech rate led to the lowest. These results fall in line with other research that has been performed in the field suggesting that time compressing speech messages places a higher processing load on listeners and leads to negative user opinions of the system (Schwab & DeGroot, 1993). To overcome such issues it is suggested that shorter speech messages be integrated into audio output systems instead of compressing longer messages (Stanton & Edworthy, 1999; Tsimhoni, Green, & Lai, 2001).

It is often the case in complex systems that operators will be required to monitor and listen to multiple speech messages simultaneously. Effective audio system design can facilitate a user in attending to multiple target messages by varying audio characteristics. For example, research supports the technique of spatially separating messages using both 3D audio (Brungart , Ericson, & Simpson, 2002; Drullman & Bronkhorst, 2000) and binaural audio systems (Bolia, Nelson, & Morley, 2001) to increase message intelligibility. In addition, Brungart et al. (2002) have suggested that by presenting multiple messages in voices with different genders associated with them or by increasing the intensity of a target message, intelligibility can be increased in multi-talker displays. It is important to note that if the gender of the voice used is selected to differentiate messages, users will change their interpretations of messages based on gender-based stereotypes and paralinguistic personality cues that are present in the message presented (Nass & Lee, 2000; Nass, Moon, & Green, 1997), and thus such differences should be taken into account when designing speech output systems.

The following table presents a number of theoretically derived design guidelines that can be followed to support the development of speech output systems.

Table 9: SEAS speech presentation guidelines

Speech	**Guideline(s)**	**Reference**
SP1	Use speech output and alarms to present detailed information to listeners and when situation can map one-alarm to one-event and fault management is serial in nature.	Stanton and Edworthy (1999)
SP2	Use natural speech interface whenever possible (as opposed to synthetic) as it is more comprehensible.	Tsimhoni et al. (2001)

Speech	Guideline(s)	Reference
SP3	Use a consistent speech output system (voice or synthetic) instead of one that uses a mixture of multiple voice forms.	Nass et al. (2000)
SP4	While performing simple, manual, primarily visual tasks, the type of voice (natural or synthetic) or the complexity of the message should not have an effect on the manual task.	Tsimhoni et al. (2001)
SP5	Use different voices for different interface elements.	ETSI (2001)
SP6	Conform to gender stereotypes when designing speech output since users will apply them to such systems.	Nass et al. (1997)
SP7	When speech is presented dichotically via headphones over single channel, the number of same gender talkers should be kept to a minimum as performance in such systems degrades as each new voice is added across the first three competing voices.	Brungart et al. (2002)
SP8	When providing evaluative information through speech, use a male voice to have a greater influence on users.	Nass et al. (1997)
SP9	Limit human speech to a short message containing a minimum of 5 syllables.	Stanton and Edworthy (1999)
SP9a	Although shorter messages are more accurately comprehended, and are thus preferred, there is not a linear degradation in comprehension of spoken messages.	Tsimhoni et al. (2001)
SP10	Speech output speed should be set at about 160 words per minute and should not exceed 210 words per minute.	ETSI (2001)
SP10a	Do not set speech output at high rates of output (> 210 words per minute) whenever possible.	Scherz (2003)
SP10b	When possible, slow speech messages down below that of normal adult speech output (200 syllables per minute). • Consider making this rate variable as some users may become disinterested in consistently slow messages.	Venkatagiri (1991)

Speech	Guideline(s)	Reference
SP11	Avoid time compressed messages whenever possible as users will perceive them as being too fast and perceived workload may increase. • They provide no advantage when immediate real-time responses are required (as the time saved by compression may lead to longer processing times by users).	Schwab and DeGroot(1993)
SP12	Design multitalker speech displays with a SNR (signal to noise ratio) of +20 dB in the frequency range from 200 Hz to 6100 Hz.	Brungart et al. (2002)
SP13	If binaural speech is used, present the most important (target voice) in one ear and competing voices in the opposite ear as this should have little or no effect on the intelligibility of the target voice.	Drullman and Bronkhorst (2000)
SP13a	If there is only one competing talker, a binaural system can be used, as 3D audio systems provide minimal gains in intelligibility in such cases.	Drullman and Bronkhorst (2000)
SP14	If more than one competing voice is expected, use a 3D audio system, as there are significant increases in intelligibility over binaural and monaural systems in such cases.	Drullman and Bronkhorst (2000)
SP14a	On average, 3D audio systems allow for 1 additional competing talker over binaural systems and 2 additional competing talkers over monaural systems.	Drullman and Bronkhorst (2000)
SP15	Do not require users to make absolute localizations of voices as users perform poorly on such tasks and degrade in performance as additional talkers are added.	Drullman and Bronkhorst (2000)
SP16	Design voice systems to match the expected users of a system in terms of extroversion and introversion, as such systems are typically perceived as being credible and trustworthy.	Nass et al. (2000)
SP16a	If possible, design two systems, one for use by introverts and one for extroverts.	Nass et al. (2000)

Speech	Guideline(s)	Reference
SP16b	The following aspects should be focused on when designing for introverts or extroverts: • Speech Rate: Extroverts speak more rapidly than introverts • Volume: Extroverts speak more loudly than introverts • Pitch: Extroverts speak with higher pitch than introverts • Pitch Range: Extroverts speak with more pitch variation than introverts	Nass and Lee (2000); Pittman (1994); Hall (1980)
SP17	If there is a limited vocabulary size in use, a SNR (signal to noise ratio) of 0dB is acceptable.	Brungart et al. (2002)
SP18	Present speech output systems at high pitch levels since increases in pitch increase intelligibility.	Scherz (2003);
SP19	Avoid the use of preceding earcons if users are expecting auditory cues (aural speech), because they do not have an effect on message comprehension.	Tsimhoni et al. (2001)
SP20	Use speech-based alarms when an immediate response is required.	Stanton and Edworthy (1999)
SP21	Combine speech tasks with non-speech tasks instead of combining them with additional speech tasks because having an additional talker instead of noise (at the same level) is more difficult for intelligibility.	Festen & Plomp (1990); Zatorre (2001)

Earcons and Auditory Icons

The use of non-speech sound has shown utility to present coded information to system users. Currently, such non-speech sounds are grouped into one of two categories, earcons or auditory icons. Each of these types of cues code information in a slightly different manner and thus have associated advantages and disadvantages for presenting various types of data.

Auditory icons are non-speech audio cues that semantically map naturally occurring sounds to objects and events in an interface (Gaver, 1986). The representational mapping allows the cues to be intuitively understood and does not place a demand on the auditory system's cognitive processing resources to understand them. Instead, auditory icons require listeners to draw a connection between the sound that is heard and past experiences. This connection of information to past events is thought to take place in a component of Baddeley's (2000) revised HIP model, specifically placing a memory load on the episodic buffer.

The intuitive nature of auditory icons has led to them being utilized in many software applications today (e.g. the sound of a door opening or closing when someone enters a chat room, America Online, 2004). The use of auditory icons has also shown utility in increasing collaboration in everyday environments. In a study to test the utility of auditory icons, Gaver, Smith, and O'Shea (1991) required teams of users to work together to efficiently ship bottles in a modeled soft drink factory that was augmented with 14 auditory icons to represent machines and events (e.g., bottle dispenser made a clinking bottles sound, smashing bottles sounded to symbolize wasted bottles). Results of this study provided strong support for the use of auditory icons to aid in the secondary task of monitoring background operations without interfering with primary tasks (Gaver, 1991; Gaver et al., 1991).

Earcons use metaphoric or symbolic mappings to relate sounds to objects and events in an interface (Blattner, Sumikawa, & Greenberg, 1989). Although not as intuitive to novice users as auditory icons, the nature of earcons give them the potential to provide structured information to system users. To accomplish this, earcons use parameters such as rhythm, pitch, timbre, register, and other characteristics of sound to differentiate musical messages and to create a hierarchical structure of information that is mapped to audio (e.g., mapping file size to the pitch

of a cue [higher pitch = larger file]) (Brewster, 1994; Brewster, 2003; Brewster, Wright, & Edwards, 1994). These audio parameters can be used to map multiple data dimensions (Blattner et al., 1989; Brewster, 1994; Brewster, 2003; Brewster et al., 1994) or can be combined with visual widgets in graphical user interfaces (e.g., scrollbars, Beaudouin-Lafon & Conversey, 1996; dropdown menus, Brewster, 1994; Maury, Athens, & Chatty, 1999; sonically enhanced buttons, Brewster, 1998; progress bars, Crease & Brewster, 1998).

Due to their symbolic nature, earcons require some training to learn their meanings; however, research has demonstrated that with minimum rehearsal listeners can generally remember earcons, even after performing tasks with additional similar earcons (Brewster, Wright, & Edwards,1993; Brewster, 1994). For example, Brewster, et al. (1993) found that individuals were able to recall over 80% of earcons correctly a week after learning them (without rehearsal). In addition, it has been shown that multiple earcons can be played in parallel with no significant decrement in recall performance (Brewster, 1994) and the recognition of earcons does *not* require musical ability (Brewster, 1994).

The following table presents a number of theoretically derived design guidelines that can be followed to support the integration of earcons into human-computer systems.

Table 10: SEAS earcon presentation guidelines

Earcons	Guideline(s)	Reference
E1	Use earcons to present structured information that can be mapped systematically mapped to various characteristics of audio.	Brewster (1994)
E2	Earcons can be integrated into intermittently used systems since individuals can remember their meanings even after time has passed and after learning similar sounding earcons for other tasks.	Brewster (1994)
E3	Earcons can effectively be integrated into systems used by non-musicians.	Brewster (1994)
E4	Limit the number of tone-based earcon sounds used in a display to seven.	Stanton and Edworthy (1999)
E5	If earcon sounds used are simple, for example just indicating events, then durations can be short (e.g., very simple earcons can be as short as .03 seconds).	Brewster (1994)
E6	When designing serial earcons insert an inter-stimulus interval duration of 0.1 between them so that users can tell where one finishes and the other starts.	Brewster (1994)
E6a	Use spatial location as a method of differentiating earcons, especially when designing serial earcons.	Brewster (1994)
E7	Make earcons demanding (attention grabbing) by using high pitch, wide pitch range, rapid onset and offset times, irregular harmonics and atonal or arrhythmic sounds.	Brewster (1994)
E8	Use timbre to relate high-level organizational earcon concepts, rhythm and tempo to convey relative levels/quantities earcons, and pitch to represent subcomponents of an earcon concept.	Brewster et al. (1993)
E9	Musical instrument timbres should be selected for earcons over simple tones.	Brewster (1994)
E9a	Timbres should be used that are subjectively easy to tell apart.	Brewster (1994)
E9b	Where possible use timbres with multiple harmonics as this helps perception and can avoid masking.	Brewster (1994)

Earcons	Guideline(s)	Reference
E9c	Multi-timbre earcons should be used when long-term memory of earcons is important. • Using multiple timbres per earcon may confer advantages when integrating compound earcons.	Brewster (1994)
E9d	When earcons are expected to be presented in parallel, timbres of earcons that represent similar events should remain the same.	Brewster (1994)
E10	To capture the attention of a listener, consider using changes in the rhythm or pitch of an earcon.	ETSI (2001)
E10a	Make rhythms as different as possible by putting different numbers of notes in each rhythm.	Brewster (1994)
E10b	Do not use notes less than the length of sixteenth notes since small note lengths might not be noticed.	Brewster (1994)
E11	Contain the pitch and register of earcons to: Maximum: 5kHz (four octaves above C3) and Minimum:125Hz - 150Hz (the octave of C4).	Brewster (1994)
E12	Earcons can be combined (more than 2) to create compound parallel earcons with no expected decreases in performance.	Brewster (1994)
E13	Pitch/register changes should only be used alone when relative judgments are to be made among earcons.	Brewster (1994)
E14	When absolute judgment of earcons is required, use a combination of pitch and another parameter to differentiate earcons.	Brewster (1994)
E15	Use large intra-earcon pitch changes in combination with varying the number of notes and rhythm between earcons to enhance user's abilities to differentiate and remember them. • Secondary parameters, such as intensity, stereo position, chords and effects (such as echo or chorus) can be used to help differentiate earcons from each other.	Brewster (1994)

Earcons	Guideline(s)	Reference
E15a	If pitch is used to differentiate earcons, large differences should be used. • More subtle changes in pitch can be utilized if the system is being designed for trained musicians.	Brewster (1994); Scharf & Buus (1986)
E15b	Use wide register ranges to aid in differentiation of earcons.	Brewster (1994)
E15c	If register alone is used for absolute judgments among earcons, then there should be large differences (e.g., 2 or 3 octaves) between each register. • This is not a problem if relative judgments are to be made.	Brewster (1994)
E15d	Earcons should all be kept within a close range according to register when not used as the sole means to differentiate of earcons.	Brewster (1994)
E16	Contain earcon intensity ranges to: Maximum: 20dB above threshold and Minimum: 10dB above threshold.	Patterson (1982)
E16a	The overall sound level should be under the control of the user of a system.	Brewster (1994)
E16b	Earcons should all be kept within a close intensity range so that if the user changes the volume of a system no sound will be lost.	Brewster (1994)
E16c	Intensity should not be used on its own for differentiating earcons.	Patterson (1982)

Spatial Signals

The utility of speech and non-speech audio can be extended by spatializing the cues presented. Research has demonstrated that spatialized audio can serve as a source of localization (Blauert, 1996) to communicate direction, location, movement, and aid in navigation (Mulgund,

Stokes, Turieo, & Devine, 2002). One of the most evident and effective uses of spatialized audio is in the guidance of attention to a target spatial location. Perrott et al. (1991) suggest that audio can be used to guide attention to target locations when targets are both inside and outside of a user's visual FOV. In addition to enhancing situational awareness by guiding users to critical information when visual attention is directed elsewhere (Strybel, Manligas, & Perrott, 1992), spatialized audio can be used to direct visual attention within the area of current visual focus. Spatialized audio is also effective at aiding the differentiation of speech messages in multi-talker displays, as it leads to lower workload levels and increased intelligibility of messages (Bolia, et al., 2001; Brungart et al., 2002; Drullman and Bronkhorst, 2000).

Since distance is generally overestimated when audio is used alone and front-back direction reversals are common (Caelli & Porter, 1980; Kramer, 1994), designers must be cautious when using localized sound for absolute judgment of position and avoid requiring listeners to differentiate between audio positions directly in front of and behind them. In addition, system designers must take note that the audio system has an area of highest acuity, directly in front of individuals, that falls off as sounds are located directly to the right or left of a listener's head (Stevens & Newman, 1936). Areas of low acuity should be avoided when absolute judgment of location is required or supernormal auditory localization can be used to exaggerate normal auditory cues so listeners are better able to localize sounds (Shinn-Cunningham, Durlach, & Held, 1998a; 1998b).

Table 3 presents a number of theoretically derived design guidelines that should be taken into account when designing spatialized audio.

Table 11: SEAS spatial audio presentation guidelines

Spatial Audio	Guideline(s)	Reference
SA1	Auditory cues can be spatialized to indicate direction, location, and movement.	ETSI (2002)
SA2	Spatialized auditory cues can be used to identify approximately 7 directions.	Bushara et al. (1999)
SA3	If using a head related transfer function (HRTF), use a general HRTF as opposed to an individualized HRTF, as there may be little gain from using individualized HRTFs and they take longer to set up.	Drullman & Bronkhorst (2000)
SA4	Position the source of sounds within about 5 deg in elevation and azimuth to aid in accurate identification of localized cues.	Brungart (1998)
SA5	When localization accuracy is important, locate sources within 1 m of the head.	Brungart (1998)
SA6	Locate sources away from the medium plane and within 1 m of the listener to increase localization performance.	Kandel et al. (1995)
SA7	When dynamic localization of sounds is expected, avoid presenting sounds at extreme azimuths (> 40°) and elevations (> 80° off the horizontal plane).	Strybel et al. (1992)
SA8	When using ITD to localize sounds, they should fall in the range of 10 µsec to 50 µsec.	Blauert (1996)
SA9	If using spatialized audio cues to communicate movement, position source in front of the listener (i.e., 0° azimuth; do not exceed ± 40°).	Strybel et al. (1992)
SA10	Add spatialized audio to visual target detection tasks, as it results in decreased search times and lower workload while being just as effective as visual cuing.	Bolia et al. (1999); Flanagan et al. (1998); Nelson et al.(1998)

Spatial Audio	Guideline(s)	Reference
SA11	Use supernormal auditory localization to exaggerate the positions of normal auditory cues.	Shinn-Cunningham, et al. (1998a; 1998b)
SA12	Integrate updating spatial audio cues to represent visual target locations when possible to reduce visual search time and reduce the effect of front-back and up-down reversals.	Flanagan et al. (1998)
SA13	Integrate 2D audio when target detection is on a frontal plane, as there are no advantages in this case of using 3D over 2D audio.	Nelson et al.(1998)
SA14	Use spatialized audio to aid identification of auditory messages in noisy environments.	Mulgund et al. (2002)
SA15	As talkers are added to a system, position them symmetrically around users, especially as the number of talkers increases above 2.	Bolia et al. (2001)
SA16	For tasks that require users to integrate and understand speech input from multiple speakers, separate speakers' spatial positions along the horizontal plane in order to increase identification and comprehension of messages.	Baldis (2001)

The theoretical design guidelines in Tables 1-3 provide designers with a framework to create audio presentation systems around. Following such guidance is expected to lead to interactive systems that distribute information demands across spatial and auditory processing resources such that they reduce the traditionally high visuo-spatial demands placed on users. It is important to note that the guidelines listed in Tables 1-3 are theoretically derived and thus need empirical validation. The case study presented below provides an initial source of validation. This study was carried out to evaluate the utility of the SEAS interaction paradigm described above in reducing perceptual and attentional overloads in a primarily visual interface.

Case Study

Method

A case study was conducted to evaluate the utility of applying a subset of SEAS guidelines to the design of a UAV control interface. Operating multiple UAVs requires consideration of many factors, including air traffic separation, target monitoring and identification, weapons deconfliction and release, and battle management integration (Dixon & Wickens, 2003). In such multitasking environments, the allocation of resources to only one task may lead to loss of SA on other potentially important tasks (Riley & Endsley, 2005). For this reason, such control interfaces place a high HIP demand on operators and thus could potentially benefit from offloading of demands from visual displays to auditory displays through the application of SEAS guidelines. To evaluate this supposition, SEAS guidelines were applied to the design of a UAV control interface, specifically, to examine if this would lead to a decrease in attentional and perceptual bottlenecks while individuals operated a UAV flight control interface across various workload levels.

Participants

Thirty students (8 females and 28 males) were recruited from the University of Central Florida to participate in this study. Participants had a mean age of 20.07 years (S.D. = 3.45), with a range of 17-34 years. 26 participants were right-handed and four were left-handed. The average number of hours playing games equaled 7.4 hours (S.D. = 8.62) per week while the

average time spent using a computer outside of game play equaled 17.9 hours (S.D. = 15.51) per week.

Apparatus

UAV control tasks were performed on a 3.0 GHz Dell Inspiron 9100 computer with a 128 Mb Radeon 9600 video card. Two forms of the UAV control interface were developed for this study, one without SEAS cues and one with SEAS cues. The interface was presented on two side-by-side 17" NEC Multisync LCD displays at 1280x1024 screen resolution. Both the Baseline visual control interface and the SEAS enhanced visual-auditory interface presented one window on each display. The display on the left presented a map interface that was used to plan and initiate attacks and the display on the right was used to present radar images of items of interest when they were available, which controllers used to designate points of impact to finalize sorties. Audio was presented through a set of Plantronics DSP 500 noise-canceling headphones which allowed for spatialized sound presentation. All input was made with a standard 2-button mouse.

Three questionnaires were used in this study. Modified versions of the Cooper-Harper questionnaire (Wierwille & Casali, 1983) and NASA-TLX (NASA Human Performance Research Group, 1987) were used to assess workload, while the Situational Awareness Rating Technique questionnaire (SART; Taylor, 1990) was used to assess situational awareness.

Tasks

Participants were required to perform three tasks while interacting with the UAV control system: (1) set up sorties (i.e., operational flights) on preplanned targets; (2) set up sorties on unplanned targets; and (3) detect and resolve vehicle health tasks. For all tasks, UAVs were presented in distinct groups of four arranged in a diamond formation (see Figure 3). The lead UAV carried heavy assets that were used to attack heavy targets. The middle two UAVs carried small assets that were used to attack small targets. The rear UAV carried medium assets that were used to attack medium targets. Workload was increased by increasing the number of UAVs that were controlled by participants from four to eight to twelve UAVs controlled.

The primary task required participants to plan and direct sorties on preplanned items of interest (IOIs) using an updating map display (presented on the left display), which displayed 20 IOIs for each set of four UAVs controlled (see Figure 3). To carry out this task, controllers were required to match a UAV carrying the correct ordnance to the ordnance requirements of the IOIs. Letters located below each IOI indicated the ordnance requirement ([S]mall, [M]edium, or [H]eavy), and the same letters were located below UAVs to designate the ordnance that each carried (see Figure 3).

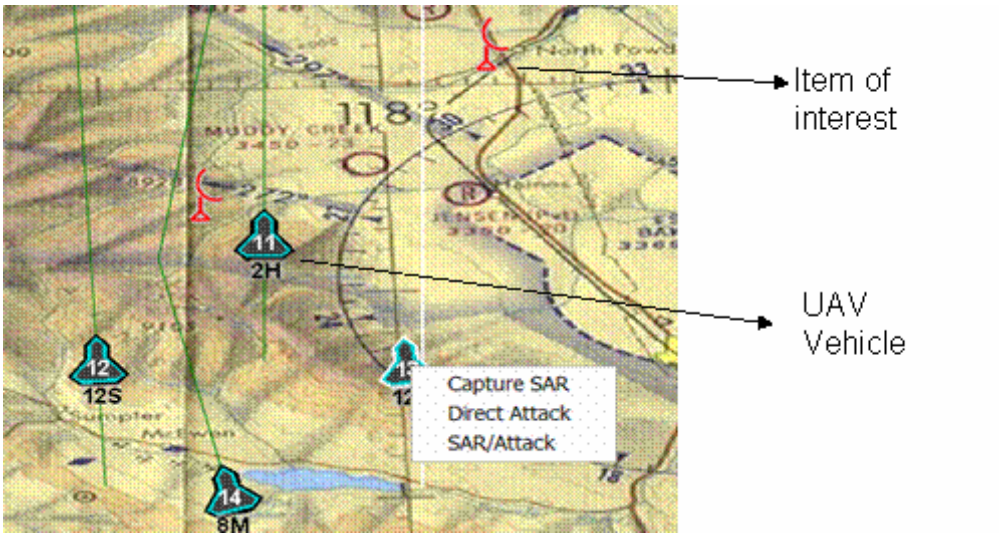


Figure 3: UAV control interface

After IOIs and UAVs were paired, a line appeared connecting them (see Figure 4a) and when close enough, the UAV took a radar image of the IOI. Once the image was available, an icon denoting this was presented under the IOI icon (see Figure 4a), which the controller could click to view the image and finalize the sortie. After clicking the icon, the associated radar image appeared in the display on the right and the controller selected the precise weapon allocation points from this image. Once the allocation points were selected the icon over the IOI changed to depict the number of ordnances allocated to that IOI (see Figure 4b). At this point, the UAV would carry out the sortie without any more guidance from the controller; however, the controller was required to check the status of the UAV to determine when it had completed the sortie to reassign it, an event that was communicated with another change to the icon over the IOI's location (see Figure 4c). This task was performed for every IOI that was presented on the display as the UAVs progressed forward (and were restricted from moving back). The number

of icon changes (task status updates) detected and time required to perceive these changes were recorded.

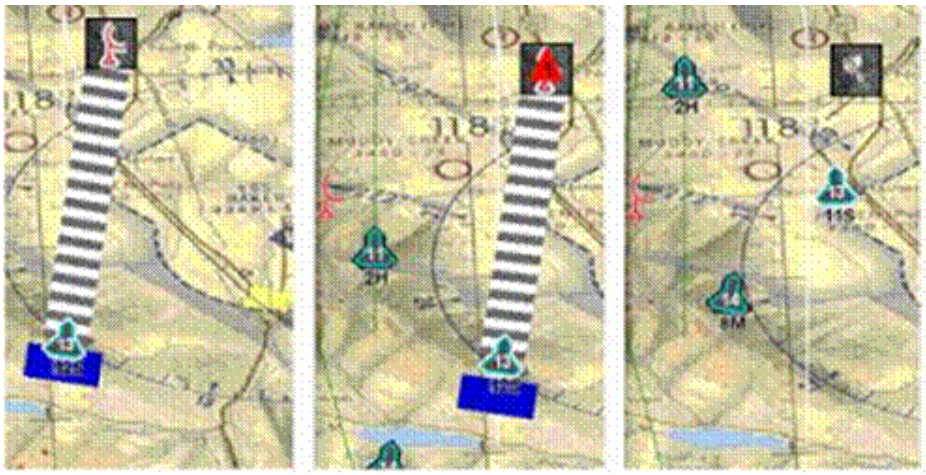


Figure 4: a) Radar image available icon b) Number of assets dispensed icon c) Asset released icon

While performing this primary task, two secondary tasks were also performed. The first required participants to detect and set up sorties on unplanned time critical items of interest (TCIs). Such TCIs appeared throughout each trial and although sorties were set up in the same fashion as any other IOI, these items required participants to quickly identify them and set up a sortie within 10 sec after appearance.

Another secondary task required participants to detect and resolve vehicle health tasks (VHTs) as they occurred throughout scenarios. VHT occurrences were symbolized by presenting a red outline around the UAV that required attention. After detecting this cue, the controllers were required to double-click on the affected UAV and answer questions that were presented in a text box that appeared onscreen.

Since both secondary tasks required participants to selectively attend to the TCIs and VHTs directly after appearance, and thus imposed HIP resource demands during primary task

performance, they were used to evaluate the effectiveness of SEAS cues to increase the percentage of cues that were attended to and decrease the time required to attend to them.

SEAS Cue Design Process

In order to integrate audio cues that had the greatest utility in reducing the perceptual and attentional bottlenecks of the task, a four stage process was followed. First, a task analysis was performed to determine where audio cues should be integrated. Based on the results of the task analysis, portions of the task that placed high demands on perceptual and attentional resources and thus had the potential to lead to perceptual and attentional bottlenecks were determined. Following the SEAS guidelines presented herein (see Tables 9-11), audio cues were then designed to alleviate each of the potential perceptual and attentional bottlenecks. When creating the cues, care was taken to integrate speech, earcons, and auditory icons separately for different tasks in order to make each of them distinct from one another (e.g. speech was used to direct attention while nonspeech cues were used to guide perception). Finally, after all cues were integrated into the display, the augmented interface was evaluated by two pilot participants to determine the usability of the audio cues prior to experimentation. Based on feedback from the pilot, the cues were modified (e.g. volumes changed, earcon timbre selections changed to make them more distinct), and then the new cues were integrated into the display as described below.

SEAS Cues Employed

The task analysis revealed that while performing the primary task of setting up sorties on preplanned targets, visual overload could hinder participants from perceiving that radar images

were ready to be viewed and that UAVs were complete with a sortie and could be applied to another IOI. To reduce these potential perceptual bottlenecks, spatialized auditory icons and earcons were integrated into the UAV control display to aid in the perception of these events. Whenever a radar image was available, a spatialized auditory icon (camera click sound) was played from the position of left, right, or center, depending on the position of the newly available radar image. When UAVs completed sorties, a spatialized earcon was played to denote which UAV was free to be reassigned. The earcons were spatialized left, right, or center to guide the controller to the group of UAVs that the newly available aircraft was in. The structure of the earcon guided the controller to precisely which UAV was available. To do so each group of four UAVs was mapped with distinctly different timbres. The lead vehicle carrying heavy assets was paired with a brass instrument playing a note at two octaves below middle C. The middle UAVs in the diamond shaped formation (see Figure 3), carrying small assets were paired with a vibraphone (left vehicle) and a pan flute (right vehicle) playing two octaves above middle C. The rear UAV, carrying medium assets was paired with a piano note playing at middle C. These cues were expected to reduce perceptual bottlenecks due to visual overload, thereby increasing the number of cues that could be responded to and reducing response time.

In analyzing the secondary tasks, potential attentional bottleneck were identified because both TCIs and VHTs required immediate attention while other tasks were being attended to. To overcome these potential attentional bottlenecks, in the SEAS interface when TCIs or VHTs occurred speech messages were played in a natural voice stating “Critical Target” or “Health Alert” (respectively). The message was spatialized in accordance with the location of the TCI or VHT (left, right, center). In addition, the message was played in different gender voices depending on the location of TCI or VHT. A male voice was presented if the TCI or VHT

occurred on the left side of the display, a female voice was played for the center, and a different male voice was used for the right side. These mitigations were expected to transform the purely visual secondary tasks to auditorily guided search tasks, thereby reducing attentional bottlenecks associated with periodically scanning for the presence of VHTs and TCIs.

Procedure

Prior to beginning, participants completed an informed consent and demographics questionnaire. Each participant then performed a training session familiarizing them with the UAV control tasks, including how to properly pair UAVs to items of interest, how to control UAVs to perform sorties, where to place asset allocation points for each item of interest, and how to recognize and handle TCIs and VHTs. Following training, rules of engagement, procedures, and strategies were explained to participants and they were seated in front of the two monitor displays; the display on the left presented an updating map display to select and pair IOIs and receive information about TCIs and VHTs; the display on the right presented the radar images that were used to determine the precise asset drop points for each item of interest.

Prior to testing, participants performed two practice trials operating four UAVs. During these trials, participants were required to successfully complete 65% of the sorties at a low workload level (four UAVs) before testing. During testing, participants were required to perform tasks on two interface conditions (Baseline- visual display, SEAS- visual display augmented with SEAS cues) under each workload level (four, eight, and twelve UAVs). To reduce order and practice effects, the order of interface presentation was counterbalanced. Prior to performing tasks on the SEAS interface, participants were trained on each sound employed

(i.e., camera shot sound, earcons). Accuracy and reaction time were used to assess performance. In addition, following the completion of each UAV interface evaluation (Baseline, SEAS), workload and situational awareness questionnaires were completed by each participant.

Experimental Design

To evaluate the effectiveness of the Baseline and SEAS interfaces at the three workloads levels observed, a 2x3 (interface type x workload) within-subjects design was implemented. Each participant performed with both the Baseline visual interface and the SEAS augmented interface. Each interface was used to perform the tasks at three levels of workload consisting of the control of four, eight, and twelve vehicles. Several performance measures were recorded and compared using this approach. Perceptual measures included 1) the effectiveness of SEAS spatialized auditory icon cues to present an update in radar imaging status, which was assessed in terms of the percentage of radar images that were viewed and the time required to view radar images; and 2) the effectiveness of SEAS spatialized earcons to present UAV status updates (when a sortie was completed), which was assessed in terms of the time required to reassign UAVs to new items of interest after becoming available. Attentional measures included 1) the effectiveness of SEAS spatialized speech cues to facilitate TCI detection, measured using reaction time and the number of TCIs detected and 2) the effectiveness of using SEAS spatialized speech cues to facilitate VHT detection tasks, measured using the number of VHTs detected and time required to detect them. In addition, the percentage of VHTs correctly answered was also recorded. A repeated measure GLM was performed to test for significant differences across workload and interface types for all performance measures except for the time

to react to VHTs. For the VHT reaction time measure, eight data points were dropped due to missing data (no VHTs detected) and thus this variable was independently evaluated using a separate repeated measures GLM. Least Significant Difference (LSD) post-hoc analyses were then performed on the significant variables. Due to the limited number of responses available for the workload questionnaires (unweighted NASA TLX- 20 point scale; modified Cooper-Harper- 10 point scale), separate Wilcoxon matched pairs signed-rank tests were performed on the answers to each question to compare results reported when using the Baseline to the use of the SEAS interface. The average scores of all questions on each of the tests were then compared using paired-sample T-tests.

Results

Perceptual Evaluation Measures

The percentage of radar images viewed showed significant main effects of interface used ($F(1, 29) = 6.62$, $p = .015$) and workload ($F(2, 58) = 4.866$, $p < .011$). An LSD post-hoc analysis showed that as workload increased, the percentage of radar images viewed decreased ($p < .05$ for all workload main effect comparisons). As can be seen in Table 12, when the SEAS interface was used, on average, 1.9% more SARS were viewed and these differences were more prominent as workload increased.

The analysis of reaction time to radar images demonstrated significant main effects of interface used ($F(1, 29) = 9.87$, $p = .004$) and workload level ($F(2, 29) = 50.09$, $p < .001$). An LSD post-hoc analysis showed that as workload increased, the time required to view radar icons

also increased ($p < .05$ for all workload main effect comparisons). As demonstrated in Table 12, the reaction time to view radar icons was 23.8% lower, on average, when using the SEAS interface than the Baseline interface under the highest level of workload. These results support the SEAS principle of using spatialized auditory icons to reduce perceptual bottlenecks in primarily visual systems.

Table 12: Radar icon detection performance for Baseline and SEAS displays

Workload level	Baseline Performance		SEAS Performance	
	Radar Image Viewed (%)	Radar Image Reaction Time (seconds)	Radar Image Viewed (%)	Radar Image Reaction Time (seconds)
1	98.66 (2.92)	3.9 (2.69)	99.29 (1.85)	2.30 (0.98)
2	97.99 (3.49)	6.33 (2.75)	99.37 (1.31)	5.21 (2.92)
3	93.99 (12.2)	8.54 (3.88)	97.61 (3.69)	6.81 (2.06)
Average	96.88	6.26	98.76	4.77

SD in parentheses

Results regarding the effectiveness of SEAS spatialized earcons to cue UAV status updates demonstrated a significant main effect for workload ($F(2, 29) = 90.936$, $p < .001$). An LSD post-hoc comparison showed that as workload increased, the time required to reassign an available UAV to items of interest also increased. As demonstrated in Table 13, increases in workload led to increases in vehicle reassignment times ($p < .05$ for all workload comparisons). There was no significant main effect found based on interface type ($F(1, 29) = 1.564$, $p = .221$), although a trend is present suggesting lower reassignment times while using the SEAS display when compared to the Baseline display. These results suggest the potential of spatialized earcons but clearly indicate more research is needed to demonstrate their effectiveness in

reducing perceptual bottlenecks in primarily visual systems. In particular, when compared to the effectiveness of auditory icons presented above, it is important to evaluate whether the increased complexity of earcons reduces their capability to alleviate perceptual bottlenecks.

Table 13: Vehicle status detection for Baseline and SEAS displays

Workload level	Baseline – Vehicle status detection (seconds)	SEAS – Vehicle status detection (seconds)
1	33.6 (11.89)	30.44 (14.54)
2	39.07 (10.29)	38.81 (13.24)
3	60.85 (16.37)	57.33 (14.99)
Average	44.51	42.19

SD in parentheses

Attentional Evaluation Measures

The number of TCIs detected and time required to detect TCIs both demonstrated significant main effects for workload level ($F(2, 58) = 19.65$, $p < .001$, $F(2, 58) = 8.43$, $p = .001$, respectively) and interface type ($F(1, 29) = 5.07$, $p = .032$), $F(1, 29) = 4.71$, $p = .038$, respectively). LSD post-hoc comparisons showed that as workload increased, the number of TCIs detected significantly decreased ($p < .05$) and the time required to detect TCIs increased ($p < .05$). In addition, as can be seen in Table 14, the percentage of TCIs detected was on average, 12.5% higher and the time required to detect TCIs was on average 20.2% lower while using the SEAS display as compared to the Baseline display in the high workload condition. These results support the use of spatialized speech messages for enhancing attention.

Table 14: TCI detection in Baseline and SEAS displays

Workload level	Baseline Performance		SEAS Performance	
	TCI detected (%)	TCI detected (seconds)	TCI detected (%)	TCI detected (seconds)
1	91.67 (18.95)	11.4 (7.65)	98.33 (9.13)	10.12 (5.17)
2	69.17 (32.62)	17.4 (9.22)	77.23 (20.88)	14.63 (9.82)
3	70.9 (27.35)	16.6 (6.83)	79.73 (21.32)	13.25 (4.64)
Average	77.24	15.13	85.09	12.67

SD in parentheses

Analysis of the number of VHTs detected showed significant differences between the performance on the Baseline and SEAS interfaces ($F(1, 29) = 32.29$, $p < .001$), and workload levels ($F(2, 58) = 9.01$, $p < .001$). A significant interaction effect was also found between workload level and interface used ($F(2, 58) = 8.12$, $p = .001$). An LSD post-hoc analysis demonstrated that when workload increased from the low to medium to high levels, fewer VHTs were detected ($p < .05$ for both comparisons). When evaluating the time required to react to VHTs, a significant main effect between interfaces was found ($F(1, 18) = 16.25$, $p < .05$). As demonstrated in Table 15, the use of SEAS cues increased detection rates by 76.13%, on average, and reaction time was 42.11% faster, on average, when compared to the use of the Baseline display. Table 15 also shows that as workload increased to medium and high levels, the performance differences accredited to the use of SEAS cues became more apparent.

A significant difference was also found between subjectively perceived mental workload while performing the VHT tasks ($p < .001$) when using the two interfaces. Participants considered the VHT task, on average, 41.3% less demanding when SEAS cues were integrated into the display. These results support the SEAS principle of using concise spatialized speech messages to conveying warning information to reduce visual attentional bottlenecks.

Table 15: VHT detection performance in baseline and SEAS displays

Workload level	Baseline Performance		SEAS Performance	
	VHT detected (%)	VHT detected (seconds)	VHT detected (%)	VHT detected (seconds)
1	61.67 (42.91)	6.18 (5.72)	74.17 (35.04)	4.11 (3.75)
2	36.67 (39.79)	7.1 (4.64)	83.33 (32.87)	4.05 (2.87)
3	33.53 (31.64)	9.16 (3.57)	71.4 (31.96)	4.77 (2.87)
Average	43.32	7.48	76.3	4.33

SD in parentheses

When assessing response accuracy of VHTs, significant main effects for interface type ($F(1, 29) = 18.28$, $p < .001$) and workload levels ($F(2, 58) = 4.36$, $p < .017$) were found. An LSD post-hoc analysis demonstrated that when participants used the SEAS augmented display, more VHTs were handled correctly ($p < .05$). The percentage of VHTs handled correctly decreased in the highest workload level when compared to the low and medium workload levels ($p < .05$ for both comparisons). As can be seen in Table 16, when participants used the SEAS augmented display, in higher workload conditions, on average 109% more health tasks were answered correctly. In addition, a significant interaction effect ($F(2, 56) = 4.12$, $p < .05$) suggests that as workload increased the utility of SEAS cues to lead to higher VHT response accuracy became more apparent. This is evident when the response accuracy increase of 3.04%, on average, associated with the integration of audio in the low workload condition is compared to the average increase of 109% that was found in the high workload condition.

Table 16: VHT response accuracy in Baseline and SEAS displays

Workload level	Baseline – VHT response accuracy (%)	SEAS – VHT response accuracy (%)
1	55.0 (44.23)	56.67 (40.97)
2	34.17 (36.24)	65.83 (32.49)
3	27.5 (25.54)	57.5 (27.88)
Average	38.89	60

SD in parentheses

Subjective workload and situational awareness

Averaged modified Cooper-Harper subjective workload ratings demonstrated that the use of the SEAS augmented interface led to a lower perceived mental workload when the SEAS interface was used ($t(29) = 2.29$, $p = 0.01$). In particular, the perceived workload level for the entire task was lower when using the SEAS interface ($p = 0.002$). Table 17 shows that individuals found it significantly less demanding to detect VHTs ($p < 0.001$) and perform the overall task ($p = 0.002$) when using the SEAS as compared to the Baseline UAV control interface..

Table 17: Cooper-Harper subjective perceived mental workload levels

Interface	VHT Detection	SAR Detection	Target Pairing	Vehicle Status Detection	Overall Task	Average
Baseline	5.23 (2.25)	4.23 (2.18)	5.17 (2.39)	4.9 (2.35)	7.17 (1.76)	5.00 (0.74)
SEAS	3.07 (1.98)	3.73 (2.18)	5.03 (2.25)	5.27 (2.40)	5.93 (2.06)	4.59 (1.16)
Difference	**2.16***	**0.5**	**0.14**	**-0.37**	**1.24***	**0.41***

*SD in parentheses; * denotes significant difference*

Average unweighted NASA-TLX subjective workload ratings demonstrated that participants found the use of the SEAS interface less demanding than the use of the Baseline interface *(t*(29) = 1.81, p < 0.01). In particular, participants considered the SEAS display as less mentally demanding than the Baseline (*p* = .010). As Table 18 demonstrates, the other NASA-TLX workload factors including temporal, performance, effort, and frustration, though not significant, showed this same pattern of being perceived as slightly less demanding while using the SEAS display as compared to the Baseline display.

Table 18: NASA-TLX subjective workload

Interface	Mental	Temporal	Performance	Effort	Frustration	Average
Baseline	15.5 (3.69)	14.9 (4.57)	10.03 (4.36)	14.93 (3.95)	12.27 (5.26)	12.37 (2.92)
SEAS	13.53 (5.03)	14.2 (4.5)	9.23 (4.45)	13.7 (4.68)	11.2 (5.1)	11.44 (3.4)
Difference	**1.97***	**0.7**	**0.8**	**1.23**	**1.07**	**0.93**

*SD in parentheses; * denotes significant difference*

Comparisons of SART Situational awareness ratings show that there were no significant differences found on any subcomponent of this scale although there was a general pattern of increased SA associated with the integration of SEAS cues.

Discussion

A great deal of emphasis has been placed on research regarding HSI issues related to controlling UAVs due to the unacceptably high percentage (50%) of UAV accident rates attributed to human factors issues (Ferguson, 1999). To aid in reducing UAV accident rates, Draper and Ruff (2000) have developed a research plan to evaluate the utility of using multi-

sensory displays (such as the one described herein) to interface with UAV controllers. They have demonstrated that response times to UAV tasking can be decreased 40% if speech input is used to interact with systems (Draper et al., 2003) and that both tactile and auditory alerts can be used to effectively direct attention to visual UAV warnings (Calhoun et al., 2005). In addition, Draper and Ruff (2000) have shown that the integration of haptics into UAV control interfaces can be used to increase controller SA of turbulence events, though they propose that the use of spatialized audio may lead to the same types of operational benefits. The results of the current study support this proposition by demonstrating that the integration of spatialized audio cues can effectively reduce perceptual and attentional bottlenecks associated with performing UAV control tasks.

Specifically, primary task performance results suggest that spatialized audio icons and spatialized earcons developed in accordance with SEAS guidelines can be integrated into systems to reduce perceptual bottlenecks. The integration of audio icons, and to a lesser extent earcons, led to more effective and faster performance in the perception of visual cues. As radar images became available, participants viewed on average 1.9% more of them and viewed them on average 23.8% faster when spatialized auditory icons guided them to the presence and location of newly available radar images in the highest workload conditions. As workload levels increased, the utility of integrating SEAS cues became more apparent, suggesting that the ability of such cues to lead to perceptual gains is higher as workload is increased.

In terms of attentional bottlenecks, participants attended to, on average, 12.5% more TCIs and attended to them 20.2% faster when guided by spatialized speech messages under high workload conditions. Participants guided by spatialized speech messages also attended to, on average, 76.13% more vehicle health tasks and attended to them, on average, 42.11% faster.

In addition to attending to more secondary tasks at faster rates, participants were on average 109% more accurate at answering vehicle health questions when they were guided to the presence of them by SEAS cues. This may be attributed to an improved alertness level of operators and the opportunity to process the health alerts and formulate an answer in parallel with other tasks when performing with the SEAS interface. Conversely, these performance gains could be a side-effect of decreased response time. Specifically, the SEAS interface allowed participants to attend to VHTs faster, and thus allowed more time to formulate an answer, thus potentially reducing the time pressure of that task.

Taken together, the performance increases in both the perceptual and attentional aspects of the UAV tasks suggest that the use of multiple modalities to guide interaction yields considerable benefits over visual-only interaction. These results fall in line with previous MRT studies that suggest that information displayed using multiple modalities leads to better performance when perceiving and attending to information (Parkes & Coleman, 1990; Penney, 1989; Rollins & Hendricks, 1980; Seagull et al., 2001; Wickens et al., 1983). According to MRT, the benefits of SEAS are likely associated with increased dual task (i.e., visual and auditory) performance efficiency. Specifically, given the availability of separate visual and auditory perceptual and attentional resources, information may have been better time-shared using the SEAS interface.

Although it is apparent that the utilization of multiple resources while using multimodal systems has the potential to lead to performance gains when compared to the use of unimodal systems, a closer look into the audio augmentations made in this study and other UAV studies makes it more evident where these gains may stem from. Wickens, Dixon, and Chang (2003) performed a study much like the one described herein and found the same results; that the

integration of audio (specifically speech output) into primarily visual displays leads to increases in task performance on the task augmented with audio, as well as other tasks concurrently being carried out. Their study required participants to perform three tasks which were very comparable to the tasks performed in the current study. Participants were required to perform a primary task of navigating a UAV to a particular location (comparable to the primary sortie task in this study), while simultaneously scanning for targets of opportunity (comparable to the TCI task), and monitoring a set of gauges for system failures (comparable to the VHT task). Wickens, Dixon, and Chang (2003) augmented the gauge monitoring task with speech output that provided guidance to controllers on what to do whenever a system failure was detected. The primary difference between the study performed by Wickens, Dixon, and Chang (2003) and the current study lie in the number of UAVs controlled and performance gains that were found. The former study demonstrated performance increases due to the integration of non-spatial speech audio when the number of UAVs controlled varied between one and two. The current study demonstrated performance increases due to the integration of spatialized speech as well as spatialized earcons and spatialized auditory icons when the number of UAVs controlled was increased from four to 12. Taken together, the results of these studies suggest that if systems are designed to support multimodal HIP, the capabilities and performance levels of UAV controllers could be increased dramatically. This result is of utmost importance, given the military's reduced manning and minimum-crew multitasking objectives.

Table 19 compares the results that were found in this study to three other UAV studies that were discussed. The results presented in this table are based on the highest workload conditions in each study. The presented results are based on the integration of audio as an output source to guide controllers for all studies presented except for Draper et al.'s (2003) study which

compared the utility of using speech input to manual input systems. The general conclusion that can be drawn from these studies is that the integration of both audio output and speech input systems have great potential when used to augment visual interface systems.

Table 19: UAV study results comparison

Study	Min/Max UAVs Controlled	Response Time Decrease Due to Audio*	Detection Rate Increase Due to Audio*	Workload Decrease Attributed to Audio
Current study	4/12	42%	76%	42%
Wickens & Dixon	1/2	60%	33%	N/A
Calhoun et. al (2005)	1	17%	N/A	62%
Draper et. al (2003)	1	40%*	N/A	Significant but unspecified

More detailed analysis of the augmentations made in the present study and the study performed by Wickens and Dixon (2002) makes it apparent that the true utility of audio may lie in its ability to eliminate secondary visual monitoring tasks. By augmenting visual displays with auditory alarms, the task of periodically visually scanning areas of a display for secondary task updates while monitoring information and searching for changes is totally removed. Instead, this task is replaced by the task of monitoring an auditory channel for discrete, readily detectable changes. This may be why subjective workload was perceived as lower with the use of the SEAS as compared to the Baseline system. Each periodic scan of the environment that was required in the Baseline interface required visuo-spatial resources that were already in use by the primary UAV-IOI pairing task. By transferring the resource requirements of the monitoring task

to the audio channel, thus splitting the resource requirements, this associated workload may have been reduced.

Given that individuals perform better on both primary and secondary tasks when multiple response modalities are used during dual-task performance (Wickens, 1976; Wickens and Liu, 1988), it could be assumed that even greater advantages would be seen in the current study and in the study performed by Wickens, Dixon, and Chang (2003) if speech input was used as the response modality for the secondary task, thus allowing for total parallel performance of the tasks from perception to response. Research by Draper et al. (2003) supports this supposition, suggesting that the integration of speech input systems has the potential to lead to substantial time savings when compared to using manual input for secondary UAV control tasks.

Each of the studies discussed herein and the results of the current study support the integration of audio to create multi-sensory interfaces. In particular, by following guidelines such as the ones presented in this article, audio can be designed and integrated into primarily visual interfaces to alleviate critical perceptual and attentional bottlenecks and allow for multitasking by utilizing multiple HIP resources. In turn decreases in workload can be achieved while allowing operators to perform additional operations at higher performance levels. It is imperative that system designers expand the WIMP interaction paradigm to include SEAS cues to support this end. This is especially true for military system designers to meet the military's reduced manning and minimum-crew multitasking objectives.

Conclusions and Future Research

The results of this case study support the use of the SEAS guidelines to reduce attentional and perceptual bottleneck caused by visual overload. In general, this case study supports the idea of integrating additional modalities into system designs to take advantage of multimodal HIP capabilities.

In regards to UAV control system interface design, this study demonstrated that audio designed using SEAS guidelines can be used to alleviate the demands of secondary tasks that are typically performed using a visual interface. These results show strong support for the research plan proposed by Draper and Ruff (2000) that is focused on the integration of multi-sensory displays for UAV workstations. To aid such a research plan, guidelines such as the ones presented in this study for audio should be developed for all modalities.

Future research should also examine the effects of auditory cues designed using the SEAS guidelines on working memory and executive functioning bottlenecks. In addition, given that the use of auditory icons led to performance increases when compared to the no audio conditions, while the integration of earcons did not, future research should focus on determining what causes such differences in the utility between the two types of audio cues and what types of tasks (e.g. decision making, working memory) earcons are better suited to guide. Due to increased performance advantages in high workload conditions, the integration of SEAS cues should be studied at increased workload conditions.

CHAPTER FOUR: GENERAL CONCLUSIONS AND FUTURE DIRECTION

MRT (Wickens, 1984) suggests that individuals utilize a multidimensional system of independent resources consisting of distinct stages of processing (encoding, central processing, and responding), which involve various sensory modalities (visual, auditory), working memory (WM) processing codes (spatial, verbal), and response modalities (manual, vocal). It is further suggested that if tasks are designed to utilize separate resources that they can be successfully performed in parallel (Wickens, 1984). In light of this theory, the two studies presented herein focused on evaluating the utility of adding audio in alleviating perceptual and attentional bottlenecks associated with the use of unimodal visuo-spatial interfaces.

The results of the pilot study presented in chapter two provides only borderline support for the integration of spatialized earcons and auditory icons to direct perception and spatialized speech to direct attention. Further analysis suggested that the lack of significance in the data may have been due to low power, which suggested a need to extend the study to determine the true utility of integrating audio cues to relieve perceptual and attentional demands. As was expected, when the pilot study was extended in study two (presented in chapter 3), the borderline results became significant, suggesting that the integration of audio designed using the SEAS guidelines presented herein has the potential to decrease both perceptual and attentional bottlenecks associated with the use of unimodal visuo-spatial interferences. The performance increases found are likely due to the use of separate modalities that take advantage of human's abilities to time-share tasks when separate MRT resources are used.

Overall this research effort provides four overarching guidelines for the integration of audio into primarily visual interfaces. First, although spatialized auditory icons can be used to

aid in the perception of visual cues, earcons may not be effective at doing so. Second, spatialized speech messages are effective at guiding visual attention. Third, SEAS audio cues are effective at guiding interactions with visuo-spatial displays, particularly for secondary tasks. Finally, the utility of integrating spatial audio into primarily visual interfaces increases with workload.

Although this effort does provide evidence that audio can be used to reduce perceptual and attentional bottlenecks and provides a source of validation for a subset of the audio design guidelines presented herein, it does not validate all of them or provide insight on the utility of audio or other modalities to reduce working memory or executive function bottlenecks. For this reason, there are a number of areas of future research that are essential in achieving a multimodal design science.

First, there is a need to validate the remainder of the SEAS audio design guidelines presented in this work. Such an effort would assure that each of the guidelines provides utility to audio designers and helps to support the second area of essential future research, the evaluation of audio cues to reduce working and executive function bottlenecks. The current work focuses on attentional and perceptual overloads and ignores the other stages of HIP. To determine where the integration of audio is most useful, research must be focused on the integration of audio to support working memory and executive function bottlenecks.

Given that this study and others (Lemmens, Bussemakers & de Haan, 2001) have shown that there is greater utility to using auditory icons over earcons to guide performance, future research needs to be directed at determining precisely what causes the differences in utility between the two types of audio cues and what types of tasks (e.g. decision making, working memory, high complexity) earcons are better suited to guide. In addition, before earcons and

auditory icons can be effectively integrated into systems that are used by diverse cultures, it is important to determine what are the cultural limitations each type of cue. For example, given that people learn the natural mappings of auditory icons from past interactions with the world (Gaver, 1986), it is likely that if cues are heard by people who have not experienced a similar audio cue in the past, the meaning will be lost and the audio cue will lose its utility and may ultimately add to the complexity of a task. Likewise, given that most earcons are created using Western tonal scales (Blattner, Sumikawa, & Greenberg, 1989), the utility of using such cues could be diminished or lost if they are heard by someone unfamiliar with Western music. In order to assure that audio cues are designed to be useful to all users, cultural differences such as these need to be taken into account and further studied.

To establish a comprehensive multimodal design science, research efforts should also be extended to additional modalities. Although high level guidelines do exist for the integration of multiple modalities (Stanney et al., 2004), detailed design guidelines need to be created for haptics, olfaction, and gustatory information. Such guidelines then need to be validated and combined with current visual and auditory guidelines to build a set of optimum multimodal interface design guidelines.

Finally, the utility of each modality in reducing HIP bottlenecks needs to be evaluated across various workload conditions to determine if benefits change based on operator load. Given that the current studies suggested that the integration of audio was more useful at alleviating perceptual and attentional bottlenecks under high workload conditions, it is likely that the utility of other modalities to affect these and other stages of HIP may also vary with the conditions that the interface is operated under (e.g., operator workload, stress).

APPENDIX A: CONSENT FORM

INFORMED CONSENT TO PARTICIPATE

Study Introduction

You are being asked to voluntarily participate in a research study titled, "**SEAS Audio Interface Evaluation"**. In this study, you will participate in an Unmanned Combat Air Vehicle (UCAV) interface evaluation. The task will require you to direct a set of 4, 8, and 12 UCAVs through a series of bombing missions. During each mission you will be required to pair each UCAV with a number of targets, take an image of the target, evaluate the image, and attack the target. Throughout each scenario these tasks will be performed multiple times on each UCAV that you are operating. You will be required to perform these tasks on two different system interfaces. The study will consist of approximately two hours. You will be asked to complete an informed consent form, a demographic questionnaire, and a training session to help in familiarizing you with the scenarios. The experimental session will require you to perform six missions using two different interfaces, followed by several questionnaires. You must be 18 years of age or older to participate.

Risks and Benefits

This experiment poses no risks or discomforts to you as a participant other than those associated with the following: working on any desktop computer application with a mouse/keyboard, audio headset, or; playing an interactive video game. If you do experience any discomfort, you may stop the research at any time.

As a research participant you will not benefit directly from this research, besides learning more about how research is conducted.

Compensation

Class extra credit or a monetary compensation of $20 will be given to each participant after completion of the experiment. Participants from the psychology department, signed up through experimetrack will be compensated with extra credit through that system and all other participants will be compensated monetarily.

If you believe you have been injured during participation in this research project, you may file a claim with UCF Environmental Health & Safety, Risk and Insurance Office, P.O. Box 163500, Orlando, FL 32816-3500 (407) 823-6300. The University of Central Florida is an agency of the State of Florida for purposes of sovereign immunity and the university's and the state's liability for personal injury or property damage is extremely limited under Florida law. Accordingly, the university's and the state's ability to compensate you for any personal injury or property damage suffered during this research project is very limited.

Information regarding your rights as a research volunteer may be obtained from:
Barbara Ward
Institutional Review Board (IRB)
University of Central Florida (UCF)
12443 Research Parkway, Suite 302
Orlando, FL 32826-3252
Telephone: (407) 823-2901

(Continued on next page)

Confidentiality of Personal Data:
All data you contribute to this study will be held in strict confidentiality by the researchers and your individual data will not be revealed to anyone other than the researchers and their immediate assistants.

To insure confidentiality, the following steps will be taken: (a) only researchers will have access to the data; (b) data will be stored in locked facilities; (c) all electronically stored data will be held on secure unnetworked computers in locked facilities (d) the actual forms will not contain names or other personal information. Instead, the forms will be matched to each participant by a number assigned by and only known to the experimenters; and (e) only group means scores and standard deviations, but not individual scores, will be published or reported.

YOUR PARTICIPATION IN THIS RESEARCH IS COMPLETELY VOLUNTARY. YOU MAY WITHDRAW FROM PARTICIPATION AT ANY TIME WITHOUT PENALTY - THIS INCLUDES REMOVAL/DELETION OF ANY DATA YOU MAY HAVE CONTRIBUTED. SHOULD YOU DECIDE NOT TO COMPLETE THE STUDY, YOU WILL RECEIVE FULL REMUNERATION.

You will be given a copy of the informed consent form to take with you.

_____ _____

Experimenter Date Participant Date

Please direct any questions about this study to:

David Jones (c/o Kay Stanney)
Research Assistant
University of Central Florida
Industrial Engineering
4000 Central Florida Blvd.
Orlando, FL 32816
david@mail.ucf.edu
(407) 823-4689

Kay Stanney
Professor (Supervisor)
University of Central Florida
Industrial Engineering
4000 Central Florida Blvd.
Orlando, FL 32816
stanney@mail.ucf.edu
(407) 823-5582

APPENDIX B: DEMOGRAPHICS QUESTIONAIRE

Demographics Questionnaire

Participant #_____

Please take a few moments to complete the following.

1. **Gender:**
____ Female
____ Male

2. **Age:** _____

3. **Education:**
Major:_____
___ Freshman
___ Sophomore
___ Junior
___ Senior
___ Graduate

4. **Vision:**
___ Normal/Corrected Vision
___ Vision problems (please describe) _____

5. **Hearing:**
___ Normal/Corrected Hearing
___ Hearing Problems (please describe) _____

6. **Handedness:**
___ Right-handed
___ Left-handed
___ Ambidextrous

7. **Computer Experience:**
___ Low (used 1 to 2 software applications)
___ Medium (used 3 to 10 software applications)
___ High (programming skills)

8. **Music Experience:**
___ None (never played a musical instrument)
___ Somewhat (took some lessons)
___ Experienced (can play an instrument)

9. Hobbies:
____ Art
____ Music
____ Sports
____ Reading
____ Other (please describe) _____

10. Do you play video games?
____ No
____ I have tried it, but I do not play regularly
____ Yes, I play regularly

If you answered "yes, I play regularly" go to question 11. If not, go to question 12.

11. How much do you play?

Days per week (please mark appropriate response)
____ 1-2 days
____ 3-4 days
____ 5 or more days

Hours each day of play (please mark appropriate response)
____ Less than 2 hours
____ 2-4 hours
____ More than 4 hours

Please estimate hours per week ____

How long have you been playing regularly? (please answer in months and/or years)
_____

12. Did you play video games as a child?
____ No
____ Yes

What was the amount of play?
____ Occasionally
____ Regularly

At what age did you play? _____

13. How many hours a week do you spend on the computer, aside from playing video games?
_____

14. List the six games tat you currently play the most.

_____

_____

_____

_____

_____

_____

APPENDIX C: MODIFIED COOPER-HARPER QUESTIONAIRE

Measure of Perceived Mental Workload

Modified Cooper-Harper (1969) Scale

Instructions: To Answer the following questionnaire refer to the definition of Mental Demand given.

1. Place an "**A1**" on the scale below where you feel the overall Mental Demands of detecting a Time Critical Target exists

2. Place an "**A2**" on the scale below where you feel the overall Mental Demands of **detecting** a Vehicle Health Task exists.

3. Place an "**A3**" on the scale below where you feel the overall Mental Demands of the Target pairing task exist.

4. Place an "**A4**" on the scale below where you feel the overall Mental Demands of detecting a SAR image exists.

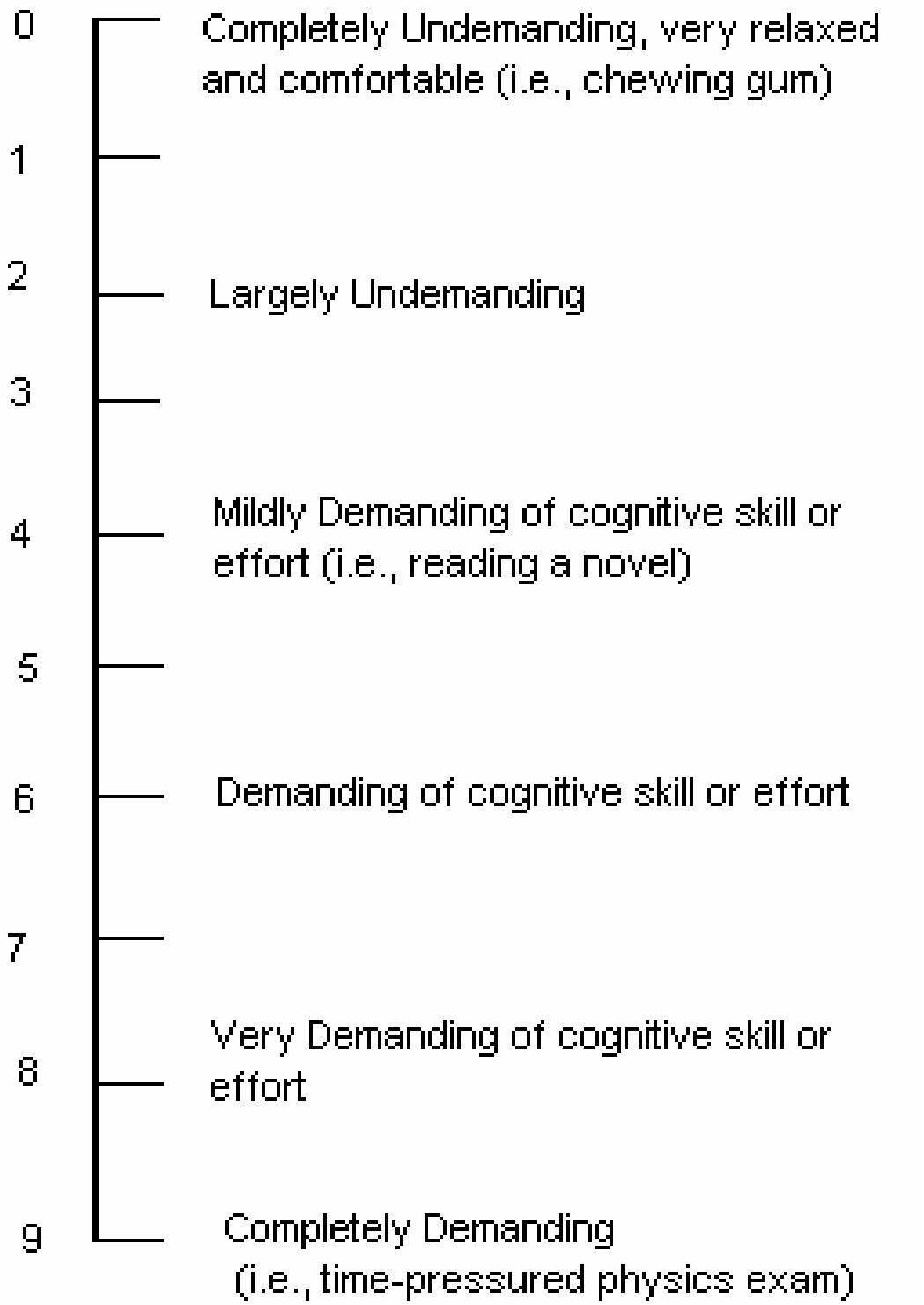
5. Place an "**A5**" on the scale below where you feel the overall Mental Demands of detecting the weapons release icon exists.

6. Place an "**A6**" on the scale below where you feel the overall Mental Demands of recognizing the status of the vehicle (free, busy) exists

7. Place an "**A7**" on the scale below where you feel the overall Mental Demands of the Target selection task (selecting the target off of the SAR image) exists.

8. Place an "**A8**" on the scale below where you feel the overall Mental Demands of **responding** to the Vehicle Health Task question exist.

9. Place an "**A9**" on the scale below where you feel the overall Mental Demands of the whole task exist.


0
Completely Undemanding, very relaxed and comfortable (i.e., chewing gum)
1
2
Largely Undemanding
3
4
Mildly Demanding of cognitive skill or effort (i.e., reading a novel)
5
6
Demanding of cognitive skill or effort
7
8
Very Demanding of cognitive skill or effort
9
Completely Demanding (i.e., time-pressured physics exam)

APPENDIX D: IRB APPROVAL LETTER

University of Central Florida

Office of Research & Commercialization

September 27, 2005

David L. Jones
475 Eagle Circle
Casselberry, FL 32707

Dear Mr. Jones:

With reference to your protocol #05-2880 entitled, **"SEAS Audio Interface Evaluation,"** I am enclosing for your records the approved, expedited document of the UCFIRB Form you had submitted to our office. **This study was approved by the Chairman on 9/23/05. The expiration date for this study will be 9/22/06.** Should there be a need to extend this study, a Continuing Review form must be submitted to the IRB Office for review by the Chairman or full IRB at least one month prior to the expiration date. This is the responsibility of the investigator. **Please notify the IRB when you have completed this study.**

Please be advised that this approval is given for one year. Should there be any addendums or administrative changes to the already approved protocol, they must also be submitted to the Board through use of the Addendum/Modification Request form. Changes should not be initiated until written IRB approval is received. Adverse events should be reported to the IRB as they occur.

Should you have any questions, please do not hesitate to call me at 407-823-2901.

Please accept our best wishes for the success of your endeavors.

Cordially,

Barbara Ward

Barbara Ward, CIM
UCF IRB Coordinator
(FWA00000351, IRB00001138)

Copy: IRB file
Kay Stanney, Ph.D.

BW:jm

LIST OF REFERENCES

America Online Instant Messenger (2004). [Computer Software]. Dulles, VA.

Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. In K Spence & J Spence (Eds.). *The psychology of learning and motivation: Advances in research and theory* (Vol. 2). New York: Academic Press.

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, A. D. (1990). *Human memory: Theory and practice*. Boston: Allyn & Bacon.

Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science*, 4, 417-423.

Baddeley A., & Logie, R.H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp.28-61). New York: Cambridge University Press.

Baldis, J. J. (2001). Effects of Spatial Audio on Memory, Comprehension, and Preference during Desktop Conferences. *Proceedings of the SIGCHI conference on Human factors in computing systems.*

Beaudouin-Lafon, M., & Conversy, S. (1996). Auditory illusions for audio feedback. *Paper presented at the ACM CHI'96* Conference Companion, Vancouver, Canada.

Blattner, M., Sumikawa, D., & Greenberg, R. (1989). Earcons and icons: Their structure and common design principles. *Human Computer Interaction*, 4(1), 11-44.

Blauert, J. (1996). *Spatial hearing*. Cambridge, MA: The MIT Press.

Bly, S. (1982). *Sound and computer information presentation* (Unpublished PhD Thesis UCRL53282): Lawrence Livermore National Laboratory.

Bolia, R.S., D'Angelo, W.R., & McKinley, R.L. (1999). Aurally aided visual search in three-dimensional space. *Human Factors, 41*(4), 664-669.

Bolia, R. S., Nelson, W. T., and Morley, R. M. (2001). Asymmetric performance in the cocktail party effect: Implications for the design of spatial audio displays. *Human Factors, Vol 43, No. 2*.

Brewster, S.A. (1994) *Providing a structured method for integrating non-speech audio into human-computer interfaces*. PhD Thesis, University of York, UK.

Brewster, S. A. (1998). Using Non-Speech Sounds to Provide Navigation Cues. *ACM Transactions on Computer-Human Interaction,* 5(3), 224-259.

Brewster, S. (2003). Nonspeech auditory output. *In J.A. Jacko and A. Sears (Eds.), The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (pp. 220-239). Mahwah, NJ: Lawrence Erlbaum.

Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1993). An evaluation of earcons for use in auditory human-computer interfaces. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel, & T. White (Ed.), *INTERCHI'93,* Amsterdam: ACM Press, Addison-Wesley, pp. 222-227.

Brewster, S. A., Wright, P. C., & Edwards, A. D. N. (1994). A detailed investigation into the effectiveness of earcons. *Paper presented at the Proceedings of ICAD'92,* Santa Fe Institute, Santa Fe.

Brungart, D. (1998). "Near-Field Auditory Localization." PhD thesis, Massachusetts Institute of Technology.

Brungart, D. S., Ericson, M. A., and Simpson. B. D. (2002) Design Considerations For Improving The Effectiveness Of Multitalker Speech Displays, *in Proceedings of ICAD* 2002, Kyoto, Japan, 2002, pp. 424-430.

Bushara, K. O., Weeks, R. A., Ishii, K., Catalan, M. J., Tian, B., Rauschecker, J. P., Hallett, M. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nature Neuroscience*, *2*(8), 759–766.

Caelli, T., & Porter, D. (1980). On difficulties in localizing ambulance sirens. *Human Factors,* 22, 719-724.

Calhoun, G., Ruff, H., Draper, M., & Guilfoos, B. (2005). Tactile and aural alerts in high auditory load uav control environments. *Proceedings of the human factors and ergonomics society,* p. 145-149.

Carlesimo, G. A., Perri, R., Turriziani, P., Tomaiuolo, F., & Caltagirone, C. (2001). Remembering what but not where: Independence of spatial and visual working memory in the human brain. *Cortex*, *36*, 519-534.

Cooper, G.H. and Harper, R.P. (1969). *The use of pilot ratings in the evaluation of aircraft handling characteristics*. NASA, Report No. TN-D-5153, Washington, DC.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences, 24*, 87-185.

Crease, M. C., & Brewster, S. A. (1998). *Making progress with sounds - The design and evaluation of an audio progress bar*. Paper presented at the Proceedings of ICAD'98, Glasgow, UK.

Dix, A., Finlay, J., Gregory A., & Beale, R. (1998) *Human Computer Interaction*. 2nd. Ed. London: Prentice Hall.

Dixon, S.R., & Wickens, C.D. (2003). Control of multiple-UAVs: A workload analysis. *Paper presented at the 12th International Symposium on Aviation Psychology*, Dayton, OH, April 14-17.

Draper, M.H. & Ruff, H.A. (2000). Multi-sensory displays and visualization techniques supporting the control of unmanned air vehicles. *IEEE International Conference on Robotics and Automation*, San Francisco, California, 2000.

Draper, M. H., Ruff, H. A., Repperger, D. W., & Lu, L. G. (2000). Multi-sensory interface concepts supporting turbulence detection by UAV controllers. *Proceedings of the Human Performance, Situational Awareness and Automation Conference*, p. 107-112.

Draper, M.H., Calhoun, G.L., Williamson, D., Ruff, H.A., & Barry, T. (2003). Manual versus speech input for unmanned aerial vehicle control station operations. *Proceedings of the Human Factors and Ergonomics Society, p.* 109- 113.

Drullman, R., Bronkhorst, A. W. (2000). Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation. *The Journal of the Acoustical Society of America*, Volume 107, Number 4 (April 2000), pp. 2224-2235.

Eberts, R. (1994). *User interface design*. Englewood Cliffs, NJ: Prentice Hall.

European Telecommunications Standards Institute (2002). Human factors (HF); Guidelines on the multimodality of icons, symbols, and pictograms. (Report No. ETSI EG 202 048 v 1.1.1 (2002-08). Sophia Antipolis, France: ETSI.

Ferguson, M.G. (1999). Stochastic modeling of Naval unmanned aerial vehicle mishaps: assessment of potential intervention strategies. Master's thesis, Naval Postgraduate School, Monterey, CA.

Festen, J.M., Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing. *Journal of the Acoustic Society of America*, 88, 1725-1736.

Flanagan P., McAnally K.I., Martin R.L., Meehan J.W. (1998). Oldfield SR. Aurally and visually guided visual search in a virtual environment. *Human Factors 1998; 40*:461-468.

Francis, A.L., & Nusbaum, H.C. (1999). The effect of lexical complexity on intelligibility. *International Journal of Speech Technology, 3,* 15-25.

Gaver, W. (1986). Auditory icons: Using sound in computer interfaces. *Human Computer Interaction*, 2(2), 167-177.

Gaver, W. W. (1991). Sound support for collaboration. *Proceedings of ECSCW 91***,** 293-308. Dordrecht, The Netherlands: Kluwer Academic Publishers. Reprinted in Baecker, R. M. (Ed.) (1993), *Readings in groupware and computer-supported cooperative work: Assisting human-human collaboration***,** 355-362. San Francisco, CA: Morgan Kaufmann Publishers.

Gaver, W. W. (1997). Auditory interfaces. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of human-computer interaction* (pp. 1003–1041). Amsterdam, Netherlands: North-Holland.

Gaver, W.W., Smith, R.B., & O'Shea, T. (1991). Effective use of sounds in complex systems: The ARKola simulation. In *Proceedings of the CHI, ACM Conference on Human Factors in Software* (pp. 85–90). New York: ACM Press.

Gong, L. Lai, J. (2003).To mix or Not Mix Synthetis Speech and Human Speech? Contrasting Impact on Judge-Rated Task Performance versus Self-Rated Performance and Attitudinal Responses. *International Journal of Speech Technology 6(2), 123-131.*

Gong, L., Nass, C., Simard, C., Takhteyev, Y. (2001). When non-human is better than semi-human: Consistency in speech interfaces. Pp. 1558-1562 in M. J. Smith, G. Salvendy, D. Harris, & R. Koubek (Eds.), *Usability evaluation and interface design: Cognitive engineering, intelligent agents, and virtual reality*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hall, J. A. (1980). *Nonverbal sex differences.* Johns Hopkins University Press, Baltimore, MD.

Jones, D. (1989). The Sonic Interface. In M. Smith & G. Salvendy (Eds.), *Work with computers: Organizational, Management, Stress and health aspects*, Amsterdam: Elsevier Science publishers.

Jones, D., Samman, S., Stanney, K., Graeber, D. (2005). Speech, Earcons, Auditory Spatial signals (SEAS): An auditory multimodal approach. *Proceedings of the 11th International Conference on Human-Computer Interaction.*

Jones, D., Stanney, K., Greaber, D. (submitted). Audio interaction paradigms: guidelines for speech, earcons, auditory icons, and spatial audio. *International Journal for Human-Computer Interaction.*

Kandel, E.R., Schwartz, J.H., & Jessell, T.M. (1995). *Essentials of neural science and behavior*. Norwalk, CT: Appleton & Lange.

Kantowitz, B. H., & Knight, J. L. (1976). Testing tapping time-sharing: I. Auditory secondary task. *Acta Psychologica*, 40, 343- 362.

Klapp, S. T., & Netick, A. (1988). Multiple resources for processing and storage in short-term working memory. *Human Factors, 30,* 617-632.

Kramer, G. (1994). An introduction to auditory display. In G. Kramer (Ed.), *Auditory Display* (pp. 1-77). Reading, MA: Addison-Wesley.

Mansur, D. L., Blattner, M., & Joy, K. (1985). Sound-Graphs: A numerical data analysis method for the blind. *Journal of Medical Systems,* 9, 163-174.

Maury, S., Athenes, S., & Chatty, S. (1999). Rhythmic menus: toward interaction based on rhythm. Paper presented at the Extended Abstracts of ACM CHI'99, Pittsburgh, PA.

Mendez, M. F. (2001). Visuospatial deficits with preserved reading ability in a patient with posterior cortical atrophy. *Cortex, 36,* 535-543.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

Mulgund, S., Stokes, J., Turieo, M. and Devine, M. (2002). *Human/Machine Interface Modalities for Soldier Systems Technologies (Report No. 71950-00).* Cambridge, MA: TIAZ LLC. (DTIC No. ADA414918)

NASA Human Performance Research Group (1987). *Task Load Index (NASA-TLX) v1.0 computerized version.* NASA Ames Research Centre.

Nass, C., Lee, K. M (2000). Does computer-generated speech manifest personality? An experimental test of similarity-attraction, *Proceedings of the CHI 2000 conference on human factors in computing systems*, p.329-336, April 01-06, 2000, The Hague, The Netherlands.

Nass, C., Moon, Y., & Green, N. (1997). Are computers gender-neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology,* 27:864–876.

Nass, C., Simard, C., & Takhteyev, Y. (2000). Should Recorded and Synthesized Speech be Mixed. Retrieved August 8, 2004, from http://www.stanford.edu/~nass/comm369/pdf/MixingTTSandRecordedSpeech.pdf

Nelson, W. T., Hettinger, L. J., Cunningham, J. A., Brickman, B. J., Haas, M. W., & McKinley, R. L. (1998). Effects of localized auditory information on visual target detection performance using a helmet-mounted display. Human Factors, 40(3), 452-460.

Parkes, A. M., & Coleman, N. (1990). Route guidance systems: A comparison of methods of presenting directional information to the driver (pp.480-485). In E. J. Lovesey (Ed.), *Contemporary Ergonomics 1990*. London: Taylor & Francis.

Patterson, R.D. (1982). *Guidelines for auditory warning systems on civil aircraft* (CAA Paper No. 82017). Civil Aviation Authority, London.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17,* 398-422.

Perrott, D., Sadralobadi, T., Saberi, K. & Strybel, T. (1991). Aurally aided visual search in the central visual field: Effects of visual load and visual enhancement of the target. *Human Factors,* 33(4), pp. 389-400.

Pickering, S. J. (2001). Cognitive approaches to the fractionation of visuo-spatial working memory. *Cortex, 37,* 457-473.

Pittam, J. (1994).*Voice in social interaction: An interdisciplinary approach.* Sage, Thousand Oaks, CA.

Proctor, R. W., & Van Zandt, T. (Eds). (1994). *Human factors in simple and complex systems*. Needham heights, MA: Allyn and Bacon.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* New York: Prentice-Hall.

Riley, J., & Endsley, M. (2005). Situation awareness in HRI with collaborating remotely piloted vehicles. *Proceedings of the human factors and ergonomics society*, p. 407-411.

Rollins, R. A., & Hendricks, R. (1980). Processing of words presented simultaneously to eye and ear. *Journal of Experimental Psychology: Human Perception & Performance, 6*, 99-109.

Samman, S. N., Jones, D. L., Stanney, K. M., & Graenlber, D. (2004). *Speech, Earcons, Auditory Spatial signlas (SEAS): An auditory interactive computing paradigm*. Technical Report submitted to Boeing, December, 2004.

Scharf, B., & Buus, S. (1986). Audition I. In K. Boff, L. Kaufman, & J. Thomas (Eds), *Handbook of Perception and Human Performance Vol.* I, New York: John Wiley & Sons.

Scherz, J. A. (2003). Intelligibility of Synthesized Speech When Output Parameters Are Varied**.** *Poster session at the annual meeting of the American Speech-Language-Hearing* Association, Chicago, IL.

Schwab, E., DeGroot, J. (1993)**.** Listener response to time-compressed speech. *ERACT '93 and CHI '93 conference companion on Human factors in computing systems*, Amsterdam, The Netherlands.

Seagull, F. J., Wickens, C. D., & Loeb, R., G. (2001). When is less more? Attention and workload in auditory, visual and redundant patient-monitoring conditions. *Proceedings of the 45th Annual Meeting of the Human Factors & Ergonomics Society.* Santa Monica, CA: Human Factors & Ergonomics Society.

Shinn-Cunningham, B.G., Durlach, N. I., & Held, R. (1998a). Adapting to supernormal auditory localization cues I: Bias and resolution. *Journal of the Acoustical Society of America, 103*(6), 3656–3666.

Shinn-Cunningham, B.G., Durlach, N. I., & Held, R. (1998b). Adapting to supernormal auditory localization cues II: Changes in mean response. *Journal of the Acoustical Society of America*, *103*(6), 3667–3676.

Smith, E.E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Science, 95*, 12061-12068.

Stanney, K., Samman, S., Reeves, L., Hale, K., Buff, W., Bowers, C., Goldiez, B., Nicholson, D., & Lackey, S. (2004). A paradigm shift in interactive computing: Deriving multimodal design principles from behavioral and neurological foundations. *International Journal of Human-Computer Interaction*, 17(2), 229-257.

Stanton, N., Edworthy, J. (1999). *Human Factors in Auditory Warnings*. Ashgate Publishing, Aldershot.

Stevens, S.S. and Newman, E.B. (1936). The localization of actual sources of sound, *American Journal of Psychology*, 48, 297-306.

Strybel, T. Z., Manligas, C. L., & Perrott, D. R. (1992). Minimum audible movement angle as a function of the azimuth and elevation of the source. *Human Factors, 34*(3), 267–275.

Sulzen, J. (2001). *Modality based working memory*. School of Education, Stanford University. Retrieved, February 5, 2003 from http://ldt.stanford.edu/~jsulzen/james-sulzen-portfolio/classes/PSY205/modality-project/paper/modality-expt-paper.PDF.

Taylor, R. M. (1990). Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. *Paper presented at the Situational Awareness in Aerospace Operations*, Copenhagen, Denmark.

Tsimhoni, O., Green, P., and Lai, J. (2001). Listening to Natural and Synthesized Speech while Driving: Effects on User Performance. *International Journal of Speech Technology, 4*, 155–169.

Venkatagiri, H. S. (1991). Effects of rate and pitch variations on the intelligibility of synthesized speech. Augmentative and Alternative Communication, 7, 284–289.

Wenzel, E. M. (1992) Localization in Virtual Acoustic Displays. *Presence*, 1, 80-107.

Wierwille, W., & Casali, J. (1983). A validated rating scale for global mental workload measurement applications. *Proceedings of the Human Factors Society 2 th Annual Meetin*g, 129-133.

Wickens, C. D. (1976). The effects of divided attention on information processing in tracking. *Journal of Experimental Psychology: Human Perception & Performance*, 2, 1-13.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (ed.), *Attention and performance VIII* (pp. 239-257). Hillsdate, NJ: Erlbaum.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasurraman & R. Davies (eds.), *Varieties of attention* (pp. 63-101). New York: Academic Press.

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

Wickens, C. D., & Dixon, S. (2002). Pilot control of multiple UAVs: Predicting performance through operator and task interference models. *Proceedings of the Army Science Conference 23rd Annual Meeting*.

Wickens, C. D., & Liu, Y. (1988). Codes and modalities in multiple resources: A success and a qualification. *Human Factors,* 30, 599- 616.

Wickens, C. D., Sandry, D., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, output, and central processing. *Human Factors, 25,* 227-248.

Woodman, G., Vecera, S., & Luck, S. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin and Review, 10 (1),* 80-87.

Zatorre, R. J. (2001). Neural specializations for tonal processing. In R. J. Zatorre & I. Peretz (Eds.), *The biological foundations of music* (pp.193-210). New York: The New York Academy of Sciences.