Applied Ergonomics 70 (2018) 51–58

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Spatialized audio improves call sign recognition during multi-aircraft control

Sungbin Kim, Michael E. Miller*, Christina F. Rusnock, John J. Elshaw

Air Force Institute of Technology, 2950 Hobson Way (AFIT/ENV), Wright Patterson AFB, OH 45433, USA

ARTICLE INFO

Keywords:
Spatial auditory displays
Automation
Call signs

ABSTRACT

We investigated the impact of a spatialized audio display on response time, workload, and accuracy while monitoring auditory information for relevance. The human ability to differentiate sound direction implies that spatial audio may be used to encode information. Therefore, it is hypothesized that spatial audio cues can be applied to aid differentiation of critical versus noncritical verbal auditory information. We used a human performance model and a laboratory study involving 24 participants to examine the effect of applying a notional, automated parser to present audio in a particular ear depending on information relevance. Operator workload and performance were assessed while subjects listened for and responded to relevant audio cues associated with critical information among additional noncritical information. Encoding relevance through spatial location in a spatial audio display system–as opposed to monophonic, binaural presentation–significantly reduced response time and workload, particularly for noncritical information. Future auditory displays employing spatial cues to indicate relevance have the potential to reduce workload and improve operator performance in similar task domains. Furthermore, these displays have the potential to reduce the dependence of workload and performance on the number of audio cues.

1. Introduction

Many domains, including military, aviation, and emergency response, rely upon the transmission of large volumes of auditory communication to permit coordination and interaction among teams of individuals. This communication often involves the transmission of information critical to the performance of a subset of the individuals, as well as, noncritical information, which does not directly affect the performance of these individuals. This noncritical information can provide important contextual information and improve situation awareness, thus elimination of this noncritical information is not desirable. In these domains, individuals often utilize multiple radio channels and communication aids (such as operator call signs) to filter critical information from noncritical information.

This research is motivated by a potential application in Unmanned Aerial Systems (UASs). Current military UASs require two or more people to operate a single Unmanned Aerial Vehicle (UAV) (Calhoun and Draper, 2015). However, a concept referred to as Multi-Aircraft Control (MAC) has been proposed, allowing one UAS operator to control multiple UAVs, (Franke et al., 2005). Operation of military UAVs can be divided into five phases; launch, transit to mission area, mission, transit to base, and recovery. Previous research has shown relatively low workload during the transit phases, making these phases amenable to MAC (Colombi et al., 2012). Aside from manpower reductions, such a concept is appealing because it could alleviate mission operators from the tedious responsibility of long transits, potentially improving operator effectiveness during critical mission segments. However, a number of human factors concerns regarding MAC have yet to be addressed.

The current research recognizes that the UAS operator is exposed to a large amount of auditory information. In the MAC concept, the amount of information one operator must process will increase proportionally to the number of vehicles. Additionally, transit operators would transition aircraft, each with a unique call sign, to and from locations where other operators assume control. Therefore, the call signs of the aircraft the transit operator controls will be ever changing. It has been hypothesized that due to the ever changing call signs controlled by the transit operators, accuracy and response time will be degraded as the number of call signs increases (Amaddio et al., 2015). We concur with this hypothesis.

Significant research has been conducted to understand the utility of spatialized auditory displays in domains where operators must monitor multiple radio channels simultaneously (Haas et al., 1997; MacDonald et al., 2002; McAnally and Martin, 2007; McAnally et al., 2002). This

* Corresponding author.
E-mail addresses: bbinsk@naver.com (S. Kim), michael.miller@afit.edu (M.E. Miller), christina.rusnock@us.af.mil (C.F. Rusnock), john.elshaw@afit.edu (J.J. Elshaw).

https://doi.org/10.1016/j.apergo.2018.02.007
Received 4 June 2017; Received in revised form 30 January 2018; Accepted 12 February 2018
0003-6870/ Published by Elsevier Ltd.

research into auditory displays employing multichannel radios illustrated that spatialized audio aids user comprehension of simultaneously spoken messages when these messages are presented at multiple locations on the azimuthal plane. However, in this system implementation, the operators were expected to listen to all radio channels, even while these channels were simultaneously presenting information, to determine if any of the messages contain their call sign. Changes in speaker spectral composition or intensity also aid comprehension of simultaneous speakers from multichannel radios (Brungart et al., 2002). Unfortunately simple manipulations, such as sound intensity, also reduce the operator's ability to comprehend information from the quieter radio channel.

Spatialized auditory displays have also been used to encode spatial information, such as direction or distance (Cengarle, 2012; Haas and Stachowiak, 2007; Maza et al., 2009; Philbrick and Colton, 2014; Simpson et al., 2005; Trouvin and Schlick, 2004). In these applications, the user's visual system is often cued with sound to reduce the time necessary for visual search.

The target systems for the current research will employ multichannel radios and could employ spatialized audio to aid the separation of information on separate channels as employed by Haas et al. (1997) and others. Alternately, the UAV's spatial location with respect to the operator or operator's area of control could be encoded with location in a spatialized auditory display, similar to the systems evaluated by Cengarle (2012) and others. However, neither of these manipulations are likely to aid the operator in remembering, recognizing, or responding to multiple, critical UAV call signs. The current research sought to utilize a spatialized auditory display system in a different way. Specifically, the system was assumed to employ a voice recognition system capable of automatically recognizing call signs on each of the multiple radio channels and to determine whether each call sign was critical, i.e., corresponded to a UAV under control of the operator. As shown in the left side of Fig. 1, the voice recognition system then parsed critical from noncritical radio calls and encoded relevance through spatial location.

The objective of this research was to investigate whether an operator would benefit from a system which automatically classifies call signs based on relevance, and then encodes relevance using spatial location. Specifically, the system presented call signs the operator is responsible for monitoring (critical information) in a different spatial location than call signs of aircraft controlled by other operators (noncritical information). We hypothesize that operators will rely upon the spatial location cue to simplify the task structure necessary to distinguish critical from noncritical information. This simplification is expected to decrease task times and operator workload compared to the traditional system, as shown in Fig. 1. Specifically, we hypothesize that the operator will rely upon the spatial information rather than their memory or memory aids to determine the relevance of auditory information. This benefit is expected to be particularly impactful during high workload conditions. Therefore, we expected this influence to be greater as the number of call signs increased and it might also depend upon information criticality.

2. Method

Our method consisted of three phases. First, an experiment was designed. Second, a model was constructed to simulate human performance during the experiment. This model assumed certain human behavioral changes in response to the number of critical call signs and the auditory display system, permitting response time and workload to be estimated for each experimental condition. The results of the model were used to establish clear hypotheses for changes in human performance and workload during the human-in-the-loop experiment. Finally, the experiment was conducted to test the hypothesized changes in response time and workload.

2.1. Human-in-the-loop experiment

2.1.1. Experimental design

The human-in-the-loop experiment employed a two-by-two-by-two, within-subjects, experimental design including the independent variables of number of critical call signs (3 or 7), auditory display system (traditional or spatialized audio) and call sign relevance (critical or noncritical). The traditional audio system employed monaural, diotic sound. The order of the number of call signs and display system type was counterbalanced. However, the participants always completed both call sign conditions for one audio display system type before experiencing the other. Call sign relevance was randomized with equal numbers of critical and noncritical call signs within each block, where each block was formed from a combination of display system type and number of call signs.

When using the spatialized audio display, critical call signs were presented to the observer's right ear as most individuals show a left hemisphere (Jung et al., 2003) and right-ear advantage for speech (Studdert-Kennedy and Shankweiler, 1970). Therefore, most listeners are more likely to remember information from their right ear better than their left (Kimura, 1967).

2.1.2. Participants

Twenty four individuals (2 females, 22 males) volunteered from among the military and civilian workforce on Wright-Patterson Air Force Base. The ages of the participants ranged from 22 to 39 with a mean of 29 and a standard deviation of 4.4 years. The educational background of these participants is representative of U.S. Air Force UAS operators. All subjects were fluent in English, had obtained or were near completion of a bachelor's degree, and self-reported having no known hearing deficiency.

2.1.3. Experimental procedure

At the start of the experiment, each participant was exposed to 1 min of spatialized audio, during which time they were told which ear the signal was present in. Next they completed 9 trials indicating whether the voice was present in their left, right, or both ears and were required to successfully indicate 8 of 9 correct responses. The participants were then instructed on and given two, 1-min practice sessions for

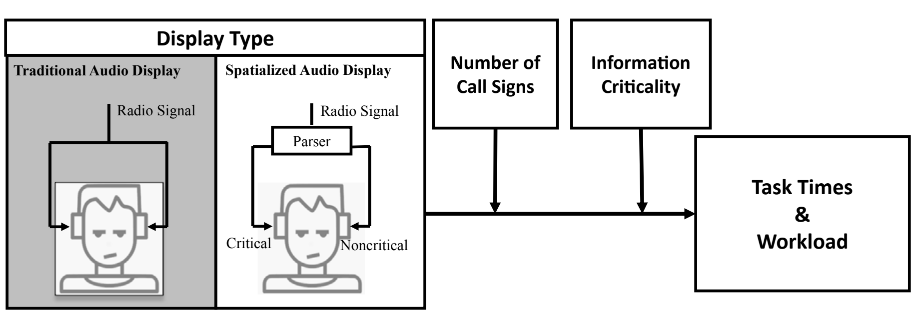

Fig. 1. Graphical depiction of the experimental conditions and their hypothesized relationship with task time and workload.