Wright State University

CORE Scholar

International Symposium on Aviation Psychology - 2015

International Symposium on Aviation Psychology

2015

The Cognition of Multi-Aircraft Control (MAC): Proactive Interference and Working Memory Capacity

Kelly Amaddio

Michael Miller

John Elshaw

Victor Finomore

THE COGNITION OF MULTI-AIRCRAFT CONTROL (MAC): PROACTIVE INTERFERENCE AND WORKING MEMORY CAPACITY

Kelly Amaddio, Michael Miller, Ph.D. and John Elshaw, Ph.D.
Air Force Institute of Technology, Dayton, Ohio
Victor Finomore, Ph.D.
United States Air Force Academy, Colorado Springs, Colorado

As the number of U.S. Air Force missions requiring UAVs has rapidly increased without commensurate increases in manpower, systems which permit a single operator to supervise and control multiple, highly-automated aircraft are being considered. The operator of such a system may be required to monitor and respond to voice communications for multiple UAVs, each of which can have aircraft specific call signs, which may impose excessive requirements on constrained operator attention, working memory, and cognitive processing. The current research investigates the cognitive load (number of aircraft call signs) an individual can handle and explores the effect of proactive interference (PI) within this application. The results indicate a reduction in performance as the number of call signs are increased from 5 to 7 in the presence of PI. Interestingly performance with 5 call signs without PI is lower than performance with 5 call signs in the presence of PI.

The United States military is currently involved in many conflicts and activities worldwide. As these wars continue and budget pressures forces the decrease of military personnel, technology is relied upon as a force multiplier. Unmanned Aerial Vehicles (UAV) have become increasingly important in recent years as they significantly enhance the gathering of Intelligence, Surveillance and Reconnaissance (ISR) without risking bodily injury to the operators. As a result, the number of UAV sorties has increased exponentially in recent years despite the limited number of pilots available to control them. As a result, new concepts of operation are under consideration wherein a single pilot might control multiple aircraft during certain phases of flight. For example, transit operators may be employed to simultaneously pilot multiple semi-autonomous aircraft between an airbase and the battlespace. If pilots are going to be operating multiple aircraft at once, they will have to monitor and respond to a large throughput of radio communications. Additionally, there is a concern that proactive interference (PI), when previously stored information prevents the learning of new information, may occur when pilots transfer aircraft to other pilots, but still hear the previous aircraft specific radio calls. Several principles related to working memory, interference, and attention are important to the analysis of this issue. The following study is a cognitive laboratory experiment aimed at evaluating cognitive load and the effects of PI.

The ability of an operator to listen to and respond appropriately to radio traffic which contains references to the call signs of the aircraft they are controlling, as well as other entities, is likely to be constrained by their available working memory. Working memory is involved in storing and manipulating information for short-term use in tasks like reasoning and comprehension (Baddeley & Hitch, 1974). A common model of working memory that has been proposed by Baddeley (2000) contains a set of subsystems, including the central executive, which controls attention between the visuospatial sketchpad, episodic buffer, and phonological loop subsystems. The visuospatial sketchpad manipulates visual images while the phonological loop is responsible for storing and replaying words and sounds. The episodic buffer temporarily stores and integrates multimodal information and relays information between the visuospatial sketchpad and phonological loop. The auditory component of this model is important to the current study because participants are asked to listen and respond to a select series of aircraft radio calls.

Although significant research has been conducted on visual working memory, auditory working memory has garnered less attention. Considering this, Kumar et al., 2013 attempted to test auditory working memory over a continuous scale by using sequences of tones in different lengths where participants were asked to adjust a dial to replicate a specific tone that they heard. The findings indicate that increasing the number of tones held in working memory reduced the precision of the memories, much like what is found in visual working memory (Alvarez & Cavanagh, 2004).

Working memory is usually measured by span tasks that require the individual to simultaneously process and remember verbal information, usually words, letters, or numbers. The current study uses a more functional measurement of working memory by requiring the individual to remember a set of words and respond to them when they are spoken in the form of radio calls. They also have to perform this task in the presence of distracting, and sometimes interfering information. This increases their cognitive load, which is considered a measure of the mental effort used to maintain information in working memory (Sweller, 1988), implying that working memory is limited by the amount of information it can hold and process. Miller's (1959) article provides the rule of thumb for information processing capacity: people's ability to process and remember limits them to 7± 2 items. Although the current study only requires participants to recognize call signs (instead of recalling them), the temporal complexity of the task and presence of distracting information causes us to hypothesize that individuals will be able to effectively attend to a similar number of call signs.

One of the primary functions of working memory is to navigate the effects of PI (Kane & Engle, 2000) where timely information replaces less recent information to reduce the likelihood of confusion. Therefore, effective working memory will suppress memory of outdated information to prevent it from interfering with the encoding of new information. PI has been shown to affect performance on working memory tasks. May, Hasher, and Kane (1999) found that performance on a working memory span test was improved when measures were taken to prevent PI (e.g., temporally separating trials). Kane and Engle (2000) found that individuals with low working memory spans showed greater susceptibility to PI under low cognitive load conditions, but under high cognitive load conditions, both high and low working memory span individuals showed equal levels of PI. Engle and Oransky (1999) propose that controlled attention is the mechanism by which working memory functions. They describe controlled attention as "an ability to effectively maintain stimulus, goal, or context information in an active, easily accessible state in the face of interference, to effectively inhibit goal-irrelevant stimuli or responses, or both" (Kane, Bleckley, Conway, & Engle, 2001, p.18). Neurological evidence shows that different information (sensory, semantic, etc.) is stored in different areas of the brain (Postle, 2006) suggesting that working memory should be seen as directing attention towards different memory codes stored in long term memory. Although these models of working memory have different implications for the design of interfaces to support MAC, they all support the view that the operator's attention must be divided between the visuospatial tasks necessary to control the aircraft, processing of audio call signs, and the integration of this information.

The current literature has shown that while working memory tests have been applied in numerous laboratory environments, they have not been applied to understand individual differences in real-life applications of working memory. This study will provide a more functional test of working memory by measuring participants' performance (in terms of accuracy and response time (RT)) on a multiaircraft control task in the presence of distracting information. It is predicted that higher cognitive load (created by the addition of more call signs and the presence of PI) will decrease performance.

Method

Participants

Twenty one (5 female and 16 male) volunteers with ages between 22 and 44 (*M* = 27.75, *SD* = 4.96) participated in the study. Participants were required to have a visual acuity of 20/30 or better, determined using a Logarithmic Near Visual Acuity Chart ("New ETDRS" Charts, 2011) and normal color vision, determined using isochromatic plates(Ishihara, 1980). There was no educational requirement, although most participants were graduate engineering students. Participants were recruited through e-mail. A participant number was assigned to each consenting participant's data and no personally identifiable information was retained per Institutional Review Board Protocol.

Apparatus

The experiment was conducted in a 6ft x 6ft cubicle in a quiet laboratory to minimize distractions. The experimental setup consisted of Bose AE2w headphones and a laptop to present the call signs using the Multi-modal Communication (MMC) software (Finomore, Popik, Castle, & Dallman, 2010). Participants were also given a wireless ten-digit number keypad, a clipboard containing a number grid with four rows and three columns, and a clipboard containing the list of call signs. The list of call signs was provided to the participants to remember before the experiment began and attached to the left wall to the cubicle slightly above eye level once the participants indicated their comfort with the call signs. The placement was selected to require the participant to actively turn their head to view the list.

The Multi-modal communication program (MMC) is an Air Force Research Laboratory developed multi-modal, network-centric communication management suite developed to aid Command and Control operators in increasing communication intelligibility and reduce mental workload. This tool combines several features designed to improve the performance of the users, including spatial audio, speech transcription, data capturing and playback, chat messages, and automatic keyword highlighting (for full description of the MMC tool see Finomore et al., 2012). Additionally, this tool has been used extensively as a research tool to evaluate a variety of communication effectiveness questions (Blair, Rahill, Finomore, Satterfield, Shaw, & Funke, 2014; Finomore, et al. 2010; Finomore, Stewart, Singh, Raj, & Dallman, 2012; Finomore, Satterfield, Sitz, Castle, Funke, Shaw, & Funke, 2012; Santana, Langhals, Miller & Finomore, 2013). This experiment utilized monaural sound, a chat window to prompt the participant to enter the appropriate code, and the logging function to record the participants' inputs.

Experimental Procedure

In the design of this experiment, a few assumptions were made regarding the operational components of the UAV control task. Each aircraft was assumed to have a unique call sign and individuals having different voices made radio calls for any of the call signs (one voice was not reserved for each call sign) as is typical in current operational environments. It was also assumed that the workload level was high enough where the participants had to intentionally process the radio calls but not so high that they could not listen to all of the radio calls. Therefore, radio calls were made every five seconds. This differs from the operational environment, which would contain variable levels of workload. Additionally, there were no secondary tasks to accomplish while participants were completing the auditory task, despite the fact that the operators in an operational environment will be responsible for other tasks like navigation, communication, and aircraft monitoring. This simplification of the environment made it possible to assess the ability of the operators to perform this auditory task under near ideal circumstances.

Upon arrival, participants were randomly assigned to one of two groups based on their participant number. They were given a quick explanation of the software and task, and then given a three minute practice trial where they were responsible for three call signs. This practice trial was designed to minimize the possibility of a learning effect. Although a hearing test was not administered, participants were encouraged to set the volume of the radio calls to their comfort level during this warm-up period.

Based on their group, participants were asked to attend to either 5 or 7 call signs (out of 13 possible call signs) during each of four 8-minute experimental trials. The trials were counterbalanced to offset a potential learning effect. Participants in Group 1 were assigned five call signs for the first two trials and seven for the second two trials. Participants in Group 2 were assigned seven call signs for the first two trials and five call signs for the second two trials. Each 8-minute trial contained 100 radio calls that were evenly spaced 5 seconds apart. Approximately 50 radio calls were critical and an equal number were distracters. The participants did not know what the ratio was, however. During the second and fourth trials, 20 of the distracters were selected to induce PI as they were among the critical call signs in the previous trial. The order of the radio calls and calls signs was randomized. Table 1 presents the trials and the critical and PI call signs for participant Group 1. The scenarios will be referred to as 5-NP (5 call signs, no PI condition), 5-PI (5 call signs, PI condition), 7-NP (7 call signs, no PI condition), and 7-PI (7 call signs, PI condition).

Table 1.

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

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

The participants were instructed to listen for the commands that contained their call sign. Each radio call began with the word "Ready", which was proceeded by a call sign and a command containing a grid coordinate; for example, "Ready Charlie go to blue one now." The color indicates a column in the grid and the number represents a row in the grid. The grid location would then contain a number. For critical call signs, the participants then found the space on the grid that corresponded with the command, and typed the number from the grid location into the

MMC chat window. For example, when the participant heard “Ready Charlie go to blue one now,” if the participant was responsible for “Charlie” during that trial (Charlie would be on their list of call signs), they would be expected to find the “blue 1” spot on the grid and type the two digit number in that grid location on the keypad. If the participant heard a call sign that was not on their list, they were instructed to type a zero into the chat window. Also, if for some reason they were not sure whether they were responsible for a specific call sign, they were instructed to type a zero. The randomized numbers on the grid were between 10 and 99. Participants were given as much time as they needed to memorize the call sign list before every trial and were instructed to only look at the list of call signs if they forgot them during the trial. The number of times they looked at the call sign list was recorded by the investigator for every trial.

To keep the participants from habituating to certain experimental conditions (call signs and voices), certain measures were taken. First, the list of critical call signs on the clipboard were shuffled for each trial so that they were not in the same order for sequential trials, making it harder to memorize. All trials contained different orders of radio calls, different call signs, and called for different grid locations. Additionally, a new number grid was used for each trial. Finally, a variety of voices made radio calls for every call sign so that the participant could not ignore or attend to a certain call sign based on the speaker. During the experiment, the participant could hear up to 12 different individual’s voices and up to 13 different call signs.

Performance Measures

Data was collected during all trials using the logging function in MMC. After each trial, participants were asked to respond to two 5-point Likert Scale questions: one regarding their workload level (Tattersall & Foord, 1996) and the other regarding the perceived difficulty (1= very easy, 2 = easy, 3 = neutral, 4 = difficult, 5 = very difficult). After the last trial, participants were asked to self report the number of call signs they believed they could reliably monitor.

Numerical responses to the MMC task provided by the participants were evaluated for accuracy and RT. For each trial, the accuracy score was calculated by dividing the number of correct responses by the total number of radio calls and multiplying by 100%. Additionally, a PI accuracy percentage correct score was determined by adding the number of correct responses given for the PI call signs divided by the total number of radio calls expected to induce PI for 5-PI and 7-PI conditions. Finally, the average of the participant’s RTs were calculated for each trial as the average of the amount of time lapse between the time when the radio call was spoken and the time the participant pressed enter after typing their numerical response. This score did not account for RTs for correct and incorrect responses.

Results

A two-factor repeated measures ANOVA revealed that there was a significant main effect of the number of call signs as well as the interaction between the number of call signs and the presence of PI on accuracy scores on the MMC task ($F(1, 19) = 7.631$, $p = 0.012$), as shown in the left panel of Figure 1. The interaction was further analyzed by applying a single factor repeated measures ANOVA. This analysis revealed that the accuracy scores were significantly different across trials ($F(2.28, 43.31) = 4.307$, $p = 0.016$, partial eta squared = 0.19). Post hoc tests using the Bonferroni correction determined that scores in the 5-PI condition (M = 97.11%, SD = 3.75%) were statistically higher than scores in the 5-NP condition (M = 93.70%, SD = 3.16%) and 7-PI conditions (M = 91.73%, SD = 6.48%). The scores for 5-NP, 7-NP (M = 94.14%, SD = 6.44%), and 7-PI were not significantly different from one another. Therefore, we can conclude that the highest scoring condition occurred when the participants were tasked with 5 call signs in the PI condition. A paired samples t-test indicated that PI accuracy scores were not significantly different between 5-PI (M = 95.29%, SD = 12.63%) and 7-PI (M = 90.25%, SD = 15.27%). Additionally, an independent samples t-test showed that accuracy scores were not significantly different based on the order the participants experienced those conditions, indicating that there was not a significant learning effect.

A two-factor repeated measures ANOVA revealed that the number of call signs had a significant effect on RT ($F(1, 17) = 11.786$, $p = 0.003$, partial eta = .409), but there was no significant effect of PI (although it approached significance at $p = .073$) or the interaction on RTs, as shown in the right panel of Figure 1. A repeated measures single factor ANOVA with a Greenhouse-Geisser correction revealed that the RTs across trials were significantly different, ($F(1.7, 28.5) = 8.520$, $p = 0.002$, partial eta = 0.334). Post hoc tests using the Bonferroni correction determined that RTs in 5-PI (M = 3.338 SD = .342) were statistically significantly lower than RTs in 7-NP (M = 3.587, SD = .405) and 7-PI (M = 3.579, SD = .430). The RT for 5-NP (M = 3.425, SD = .316) was not significantly different from the others.

Additionally, an independent samples t-test indicated that RTs were significantly different based on the order participants experienced the 5 versus 7 call sign condition ($t(76) = 3.034$, $p = .003$) where those experiencing the 5-CS conditions first had a significantly higher RT ($M = 3.601$, $SD = .376$) than those who experienced the 7-CS conditions first ($M = 3.352$, $SD = .349$).

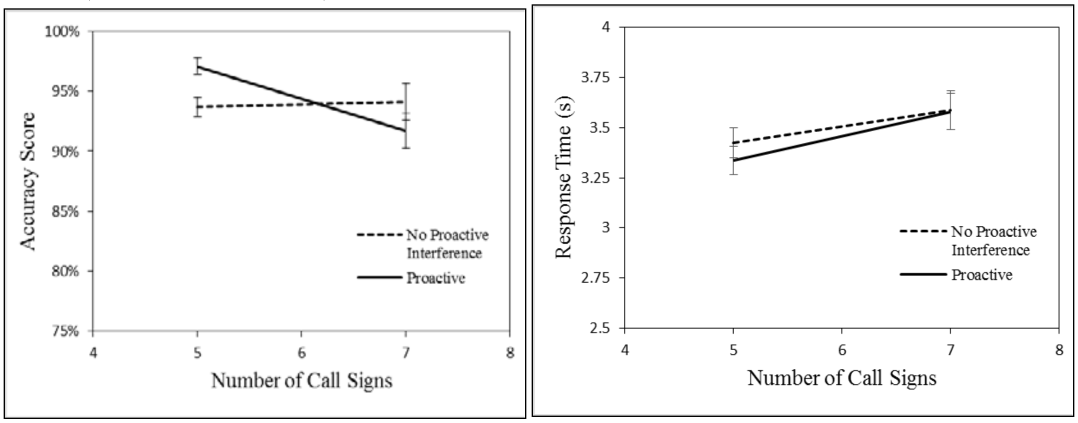

Figure 1. Interaction of number of call signs on accuracy scores for both PI conditions,(left panel) and the interaction of number of call signs on response times for both PI conditions (right panel).

A repeated measures ANOVA determined that there was no significant difference between workload or difficulty measures across all trials. Additionally, when asked "based on your experience today, how many call signs do you think you could monitor comfortably before you would begin missing time critical information?" after all experimental trials, participants responded with a mean of 5.86 ($SD = 1.35$). Responses ranged from 3 to 8 call signs.

Discussion

Overall, the results show that the participants' accuracy and response time was degraded as the number of call signs increased from 5 to 7, as expected. However, the results with respect to proactive interference differed from expected as accuracy and response time were not consistently degraded in the presence of proactive interference. Specifically, with respect to the accuracy scores, the 5 call sign PI condition was the highest scoring even though it was not the lowest taskload condition. A few possible explanations could be offered.

First, the workload-performance curve (similar to the Yerkes-Dodson Law) shows that high and low levels of workload result in low performance, but medium levels of workload result in higher performance (Teigen, 1994) creating an inverted-U shaped relationship. One potential explanation is that the workload was so low that the participants' performance did not reach its optimal level. This, however, was not supported by the reported workload and difficulty scores which did not significantly differ across the experimental conditions.

As it is necessary for the participants to be exposed to a set of call signs before these same call signs can induce proactive interference, another possible explanation stems from the need to present the PI conditions after the NP conditions. The results indicated that RT was influenced by whether the participants experience the 5 or the 7 call sign condition first, potentially indicating that the participants who experienced the 7 call sign condition first underwent a higher rate of learning than the participants who experienced the 5 call sign condition first. It is possible that negative effects of proactive interference were offset by learning effects within the current experiment.

Sampling error could have also contributed to the unexpected outcomes. For most variables, there was data from only 21 participants (due to missing data). Because of this small sample size, irregular data points could have been magnified in the results. Although the trials were kept to a short length, fatigue could have been a factor in this study, as some participants reported feelings of boredom. Additionally, there were a limited number of call signs used in this experiment, with only 13 call signs available for use in the trials. As a result, on trials where participants were supposed to remember 7 call signs, some reported that instead of listening for the call signs on the list, they listened for the ones not on the list since they believed (correctly) that there were fewer of those. Ideally, a new set of call signs would be used on each trial to prevent habituation.

Conclusion

The results of this study provide conflicting evidence about whether higher taskload conditions actually produce lower levels of performance. This study indicated that increasing the number of call signs from 5 to 7 reduced the participants' accuracy and increased their response time. However, the results do not support the hypothesis that performance will be reduced by proactive interference, a result which has multiple potential explanations including learning, workload, and sample bias effects. Further research is recommended which include additional task load levels (more call signs/PI conditions), more participants, less overlap in call signs between conditions, and potentially enhanced training. Data from this research could give insight into a relationship that exists among these variables.

Acknowledgements: The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Air Force, Department of Defense, nor the U.S. Government. The authors would like to thank Colonel Anthony Tvaryanas for motivating the current study.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15(2),* 106-111.

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in Cognitive Sciences, 4(11),* 417-423.

Baddeley, A.D. and Hitch, G. (1974) *Working memory*. In The Psychology of Learning and Motivation (Bower, G.A., ed.), pp. 48-79, Academic Press.

Blair, E. A., Rahill, K. M., Finomore, V., Satterfield, K., Shaw, T., & Funke, G. (2014, September). Best of Both Worlds Evaluation of Multi-Modal Communication Management Suite. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 58, No. 1, pp. 410-414). SAGE Publications.

Engle, R. W., & Oransky, N. (1999). Multi-store versus dynamic models of temporary storage in memory. *The Nature of Cognition,* 515-555.

Finomore, V.S., Popik, D.K., Castle, C.E., & Dallman, R.C. (2010). Effects of network-centric multi-modal communication on a communication-monitoring task. *Proceedings of the Human Factors and Ergonomics Society, 54,* 2125- 2129.

Finomore, V., Stewart, J., Singh, R., Raj, B., & Dallman, R. (2012). Demonstration of advanced multi-modal, network-centric communication management suite. *Proceedings of the Interspeech, 13.*

Finomore, V., Satterfield, K., Sitz, A., Castle, C., Funke, G., Shaw, T., & Funke, M. (2012). Effects of the multi-modal communication tool on communication and change detection for command and control operators. *Proceedings of the Human Factors and Ergonomics Society, 56, 1464-1465*.

Ishihara, S. (1980). *Ishihara's design charts for colour-blindness of unlettered persons*. Kanehara & Company.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130(2),* 169.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(2),* 336.

Kumar, S., Joseph, S., Pearson, B., Teki, S., Fox, Z. V., Griffiths, T. D., & Husain, M. (2013). Resource allocation and prioritization in auditory working memory. *Cognitive Neuroscience, 4(1),* 12-20.

Logarithmic Near Visual Acuity Chart 2000 "New ETDRS" Charts (2011). Precision Vision.

May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory & Cognition, 27(5),* 759-767.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63(2),* 81.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience, 139(1),* 23-38.

Santana, L., Langhals, B., Miller, M., & Finomore, V. (2013). Does supplementary computer generated cueing enhance controller efficiency in a congested communication environment? *Proceedings of the International Symposium on Aviation Psychology, 17,* 226-231.

Sweller, J (1988). "Cognitive load during problem solving: Effects on learning". *Cognitive Science 12(2)*: 257–285.

Tattersall, A. J., & Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics, 39*(5), 740-748.

Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. Theory & Psychology, 4(4), 525-547.