MODALITY INTERACTIONS IN MULTITASKING: SPATIAL AND VERBAL RESOURCE PRIMING VS. CONFLICT

Kathleen M. Stibler, Patrick L. Craven, Joyce Barton, Susan Harkness Regli, and Patrice D. Tremoulet
Lockheed Martin Advanced Technology Laboratories
Cherry Hill, New Jersey

The integration of real-time measures of participants' spatial and verbal working memory usage into operational systems could assist in managing information flow, thereby enhancing human performance of complex tasks. This paper presents an experiment to evaluate participant performance during tasks requiring either spatial or verbal working memory resources. The study was designed to investigate whether an operator performing two tasks tapping the same working memory resource would exhibit an improvement in performance (due to memory resource priming) or a degradation of performance (due to memory resource conflict). Scenarios presented simplified versions of command and control tasks designed to test working memory in operational simulations. Tasks were presented in visual-aural pairs with homogenous or heterogeneous working memory demand. Participants were asked to monitor missile IDs and locations while distinguishing missile status and spatial sounds. Results of the study found that the performance of participants involved in visually presented verbal tasks is degraded when they simultaneously attempt to complete aurally presented verbal tasks. No such degradation was observed for homogenous spatial working memory pairs. Future efforts will incorporate these results into intelligent mitigation strategies that enable operators to maximize their cognitive resources and optimize performance on complex tasks in an operational environment.

INTRODUCTION

The goal of developing advanced technologies to increase productivity, present in both the commercial and military sectors throughout the last century, has generally been approached through two traditional methods: automating tasks that can be optimally performed by machines, and designing intuitive user interfaces to assist operators in efficiently and effectively completing tasks. Within the last decade, a third approach, which focuses upon maximizing human capabilities, has emerged. This approach involves designing and developing technology, such as user interfaces, which can recognize and adapt to differences between users (such as preferred modality for alert notification, e.g. a beep vs. a flashing icon) and to shifts of preferences within individuals (e.g. changing alert notification modality based upon context, such as the criticality of an incoming alert).

One prerequisite for developing this sort of technology is investigating how the modality and context in which a task is presented interact to influence user performance. The study presented in this paper is a first step in this investigation. It is assumed that performance improves by dynamically adjusting a system's interaction with an operator, particularly in domains where an operator is performing multiple tasks (Samman, Stanney, Dalton, Ahmad, & Bowers, 2004). Evidence suggests that when operators are asked to perform multiple tasks simultaneously their performance decreases (Neumann, 2002; Wickens & Hollands, 2000). Consequently, intelligent strategies could adjust the timing of on-screen events to maximize an operator's cognitive resources. The efficacy of various strategies has been examined in other contexts (Daniels, Regli, & Franke, 2002; Hale, Samman, Buff, Stanney, Reeves, & Bowers, 2003), but the research reported here addresses an underlying assumption in this area of research. That assumption regards the underlying cognitive processes that result in diminished performance while conducting multiple tasks: it is unclear whether information using the same working memory capacity results in a priming effect which improves performance or conflict effect which diminishes performance.

Baddeley (1986) proposed a three-component model of working memory which includes two independent storage components designed to process either visual or auditory information. The visual component responsible for processing graphical information is the *visuospatial sketchpad*; the verbal component responsible for processing aurally presented information is the *phonological loop*. The third component, the *central executive*, serves to process information held in either visual or auditory storage. Research examining the reduction of working memory demands has demonstrated that real-time biophysical data could be used to augment the operator's performance during a task requiring high levels of verbal working memory (Daniels et al, 2002). Consequently, developing a deeper understanding of ways to alleviate working memory demands could yield high payoffs in an operational environment.

Previous research suggests that intelligent strategies can be used to improve performance when operators are in a high workload condition (Franke, Daniels, & McFarlane, 2002). A novel research area, however, involves active mitigation strategies that change the way a system interface interacts with a user. Given biophysical input about a user's cognitive state, the system can reduce, increase, or alter the presentation of tasks to maximize the use of the human resources. This research is ongoing and the present study supports the design of effective mitigation strategies. More dramatic improvements in performance through the application of appropriate mitigation strategies can only be achieved through developing a firm base of knowledge of how cognitive capabilities function in high workload conditions.

In particular, it would be beneficial to understand whether the types of working memory recruited during tasks makes a difference when an operator is performing two tasks simultaneously. The present study investigated if the type of working memory used by an operator performing two simultaneous tasks impacts performance. It was hypothesized that when an operator is performing two tasks, a conflict would occur during homogenous tasks, resulting in decreased performance compared to heterogeneous tasks.

METHOD

To test this hypothesis, a combination of working memory tasks were used and presented independently and in combination. Baseline performance on graphical working memory tasks tapping either spatial or verbal capacity were recorded. These graphical tasks were then paired with auditory tasks using either heterogeneous or homogenous working memory types.

A PC was used to present stimuli and log participants responses on a variety of working memory tasks. The PC was equipped with an optical mouse, keyboard, and a 3-channel speaker system. Responses were recorded through depression of the spacebar on the keyboard or clicking the mouse.

Participants

Twenty participants were tested using a battery of working memory tasks. The testing system became unstable during one participant's run, resulting in 19 participants with a mean age of 30.6 years. Fourteen participants were male, and one participant was left-handed. The left-handed participant reported no difficulty using the mouse with the non-dominant hand.

Materials

Tests of working memory were developed as simplified versions of command and control tasks which were designed to tax the two types of memory storage, the visuospatial sketchpad and the phonological loop. The tasks were designed to counter-balance the type of information with the mode in which it was displayed. For instance, verbal and spatial information were displayed aurally and graphically. Tasks were designed as memory recognition tasks to yield higher performance than if they were developed as memory recall tasks (Brown, 1976). The tasks were developed as Java software to enable precision presentation and logging.

Graphical Verbal - Missile ID Task This visual test measures a participant's ability to hold a series of graphically presented character groupings in mind and accurately recognize them. The graphical verbal task requires that the participant remember a list of missile IDs displayed on the screen for comparison to IDs presented later in the task. First, a missile ID composed of three alphanumeric characters is presented to the participant for 6 seconds. This list is then removed from the screen, and the participant must remember the missile ID for 6 seconds. At the end of this remember period, four IDs are displayed, each with an associated selection box. The participant then selects the correct ID from this list. The alternate responses consisted of distracter items that shared either two, one, or no character placements in common with the target.

Graphical Spatial - Missile Grid Task This visual test measures a participant's ability to hold a series of graphical groupings in mind and accurately recognize them. The graphical spatial task is a missile grid task, where participants are asked to remember and recognize the location of multiple missiles on a grid. A 5x5 grid is presented to the participant with an array of nine missiles. This array is displayed for 6 seconds, and then the screen is blanked for six seconds while the participant holds the graphical representation in mind. At the end of the period, four grids containing missiles appear for 8 seconds and the participant must select the grid which matches the original. Only one selection grid matches the initial display, and the other three selections contain missiles in different locations. One of the incorrect selections has one missile which is different from the target, another has two which are different, and the third selection had three missiles which were different from the target.

Auditory Verbal - Missile Status Task This test measures one's ability to monitor a series of spoken words and recognize when a word is immediately repeated. This task is a variation of the n-back memory task in which n = 1. Four sets of three rhymed words were used (e.g., tango, mango, dangle) to form a set of 12 possible words. These stimuli were spoken by a voice synthesizer and were randomly presented to the participant using the PC's speakers. The words were screened to ensure that participants could easily discriminate the different words.

Auditory Spatial - Sound Location Task This test measures one's ability to hold a pitch-location in mind and identify when it is immediately repeated. The auditory spatial task requires differentiation of sound generated from three different speaker locations which were arranged in an array designed to maximize discriminatory power. Sounds consisted of four recorded pitches from a MIDI synthesizer. In total there were 12 pitch-location pairs which needed to be discriminated. As in the missile status task, the sound location task is a variation of the n-back task with n = 1. Participants are presented with a particular pitch from one of the six speakers and they press the space bar if what they just heard is the same as the last status/sound that was presented.

Procedure

Testing was conducted in the Human Testing Laboratory (HTL) of Lockheed Martin Advanced Technology Laboratories (ATL), equipped with separate rooms for testing and observation. The enclosed testing facility reduced extraneous noise interference. The participants were tested individually, and members of the experimental team observed the study from behind a 1-way mirror to ensure system stability and protocol compliance.

Participants were trained on each of the four tasks individually. They were given a written set of instructions, and then were provided the opportunity to train on scenarios which included three presentations of each task individually. Following this, they were given a training scenario that presented the dual-task condition in which a graphical and auditory task was presented simultaneously. Participants were instructed to perform as well as possible on both tasks.

A testing block consisted of six task combinations, and each combination lasted approximately 8.5 minutes. Within each condition, there are 24 iterations of the graphical task with a 1-second pause between the graphical tasks. The auditory tasks occurred such that a word or tone was emitted in regular 4-second cycles throughout the graphical tasks, resulting in approximately 130 presentations with each condition. The first two conditions within each block were the baseline graphical conditions, denoted as V0 or S0 (see Table 1). In these baseline graphical conditions there were no accompanying auditory tasks. The four combination conditions consisted of one graphical with one auditory task. The combinations are denoted as VV, VS, SS, and SV where the first letter indicates the graphical task and the second letter indicates the auditory task. The order of presentation was counterbalanced throughout, with the exception that the baseline conditions always occurred first.

Table 1. Description of Task Conditions Combinations

		Auditory (underlined)	
		Spatial	**Verbal**
Graphical	Spatial	SS	SV
	Verbal	VS	VV

The tasks were presented on an ATL developed testing system that allows for the presentation of the task software and accurate logging task and response capabilities. Sufficient information was logged during each task to track stimulus presentation of both the targets and possible responses, a participant's selection, and whether the selection was correct or not. The log files were collated The following participant performance and system data were logged during the task.

ANALYSIS

In these analyses, raw values for the number of correct response were used for the graphical tasks as there were consistently the same number of items in each condition. Table 2 shows the mean (*M*) and standard deviations (*SD*) of the graphical scores for each condition. Auditory task performance was calculated as a percentage of the number of times that participants corrected identified repeats as the auditory task items were randomly generated. There were between 25-55 auditory stimuli repeats that a participant should have responded to within a condition. Table 2 lists the mean for performance on the auditory task, and Figure 1 visually depicts these means. Note that graphical task scores are number correct out of 24, which auditory task scores are percentage correct. Inter-participant variability was examined to ensure there were no differences in performance by sub-groups. An analysis of the sub-group means revealed no differences according to gender or handedness, therefore scores were grouped as one for subsequent analyses.

Table 2. Means and *SD*s for Performance on Graphical and Auditory Tasks

	Graphical Task		Auditory Task	
	M	***SD***	***M***	***SD***
S0	21.74	1.70		
SS	21.00	2.24	63.9%	8.7%
SV	22.05	1.55	61.7%	11.3%
V0	21.42	2.39		
VS	20.42	2.48	16.2%	7.5%
VV	19.42	2.04	14.6%	7.2%

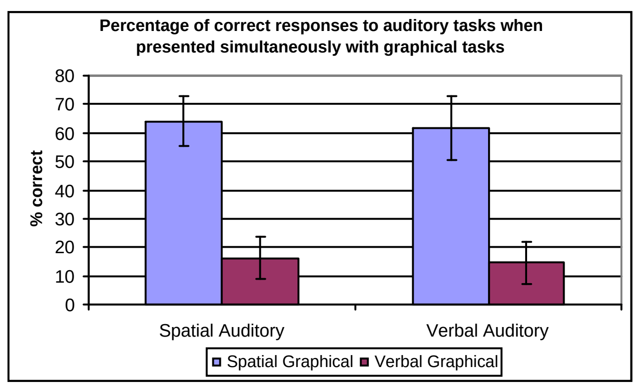

Figure 1. Means and SDs for auditory task performance.

Effects of order of the graphical tasks were investigated using a 1-way ANOVA, and results suggested there was no significant difference in the graphical task means. A follow-up correlation analysis between the graphical tasks and the task order suggested that the placement of the VV condition might have been mildly sensitive to learning effects. The scores on VV were significantly positively related to the order in which VV appeared ($r = .49$, $p < .05$), negatively related to the order in which SS appeared ($r = -.63$, $p < .01$), and positively related to the order in which V0 appeared ($r = .71$, $p < .001$). However, any relation with order appears to have no impact on later analyses and their interpretation as any learning effects within the VV condition would serve only to reduce the difference between the VV and V0 condition means.

In order to test whether there was a genuine difference between the baseline and their respective dual-task conditions, the individual baseline scores were subtracted from their corresponding dual-task scores to generate mean difference scores (see Table 3). A negative value indicates that a particular graphical performance mean in a dual task was greater than the baseline condition mean. However, the SV condition is positive, suggesting that it might be easier than the baseline condition. These difference scores were then tested to see if they significantly differed from zero (see Table 3 for related *t*-tests). The results reveal that both the SV and SS were not significantly different from zero, and this indicates that they were of similar difficulty to the S0 baseline condition. However, The VS condition was marginally significant and the VV was significant. This indicates that the graphical verbal dual-task conditions were more difficult than the graphical verbal single-task baseline condition. However, no such difference was observed between the graphical spatial dual-task conditions and its graphical spatial single-task counterpart.

Table 3. *T*-test of Graphical Scores Means in Dual-Task Condition with Respective Baseline Score Means

	Δ*M*	***SD***	***t***	***p***
SS – S0	-.74	2.33	-1.38	*ns*
SV – S0	.32	1.63	.84	*ns*
VS – V0	-1.00	2.16	-2.02	=.059
VV – V0	-2.00	2.43	-3.60	<.01

Homogeneity of Task Difficulty

Prior to analyzing the principal research question, it was important to establish that the two graphical and two auditory tasks were of comparable difficulty. The difference between the two single-task baseline conditions was calculated (DM = .30), and a t-test of this difference was not significant (t = .584, ns). This result indicates that the two graphical tasks were of similar difficulty.

Independent scores for the auditory tasks were not included as part of the testing battery because, during piloting it was discovered that the tasks were found to be simple and easily accomplished when performed in isolation. Auditory task difficulty was independently validated on a small group of pilot participants, and these results revealed that participants were consistently exceeding 99% accuracy across hundreds of trials. Additionally, auditory task performance was estimated using the current data by generating composites of the auditory tasks (spatial composite = SS + VS, and verbal composite = SV + VV). Analysis of the difference between the auditory spatial task (M = 39.2%) and auditory verbal task (M = 37.7%) revealed no significant differences (t = .57, *ns*). Together, the above results confirm similar task difficulty between the two graphical and two auditory tasks.

Analysis of Hypothesis

The hypothesis that homogenous dual task conditions (e.g., VV, or SS) would induce conflict compared to their respective heterogeneous pairs was investigated. First, single-task graphical baseline scores were subtracted from the dual-task graphical scores. These differences were analyzed using paired-sample *t*-tests in which the SS and VS, and then the SV and VV conditions were compared. In the SS – VS comparison the mean difference was .26 (t = .49, *ns*). In the SV – VV comparison the mean difference was 2.32 ($t = 3.84$, $p < .001$). These results suggest when performing a verbal auditory task simultaneously with a graphical task, performance on the graphical task suffers when the graphical task is verbal but not when the graphical task is spatial. This suggests a verbal conflict exists. In contrast, when performing a spatial auditory task simultaneously with a graphical task, there is little impact on performance on the graphical task regardless of whether the visual task is verbal or spatial.

Auditory Task Analysis

Analysis of the auditory tasks revealed that there were no differences between auditory task means when one was performing the same type of graphical task. This indicated that conflict conditions did not result in diminished auditory task performance. However, it was observed that task performance dropped considerably on both auditory tasks when one was simultaneously performing a verbal graphical task as compared to a spatial graphical task. Pair-wise comparisons (see Table 4) revealed that these differences were each significant. As previously noted, this difference was not due to a difference in task difficulty, as there were no significant differences either between the two auditory or between the two graphical tasks.

Table 4. Pair-Wise Mean Difference Comparisons of Auditory Task Performance

Pairs	**Δ*M***	***SD***	***T***	***p***
SV – VV	47.16%	10.61%	19.37	<.001
SS – VV	49.39%	11.74%	18.34	<.001
SV – VS	45.48%	13.56%	14.62	<.001
SS – VS	47.70%	11.20%	18.58	<.001

DISCUSSION

These results support the idea that a conflict of working memory resources arises in dual-task conditions when the tasks are homogenous, but only when the tasks are verbal. The current data do not support the idea that conflict arises during dual task conditions when the information is spatial. Additionally, results indicate a graphical verbal bias such that performance on both auditory tasks dropped considerably when one was performing a graphical verbal task. Participants in this study could only successfully recognize simple auditory stimuli approximately 15% of the time when simultaneously performing the graphical verbal task. This is a reduction from approximately a 99% success rate when the auditory tasks were performed alone. Although research supports the idea that multimedia interfaces can increase cognitive throughput (Samman, et al., 2004), the current results clearly indicate that graphical verbal information demands a large quantity of cognitive resources.

Although the current data do not support the notion of a spatial working memory conflict, it is possible that task design limitation prevented a proper investigation of this question. The Sound Location task, designed to tap spatial working memory, consisted of four pitches of sounds emanating from only three locations. Technical difficulties during development of this task prevented utilization of a 3D sound system that would have enabled the placement of a single sound to a much greater number of perceived sound source locations. The current task confounds memorizing the pitch with the location, which might mask a greater difference between the graphical spatial task conditions. Additional research will need to further investigate the existence of a spatial working memory conflict. Future research might also investigate the extent to which there is a difference in the verbal working memory conflict when dual tasks are presented in the same modality, graphically or aurally.

The results of the auditory task analysis were unanticipated, but support current understanding of multimedia interface design. These data support the idea that essential information should not be conveyed to individuals aurally when they are simultaneously processing graphical verbal information (reading). This finding will be incorporated into the development of intelligent mitigation strategies. For example, ensuring that secondary task activities are scheduled to support maximum operator performance by timing the presentation of verbal and spatial secondary tasks to prevent the operator from becoming overloaded verbally or spatially, respectively. Another possible mitigation strategy is modality-based task switching. This strategy involves developing alternate display strategies to invoke sensory modalities with spare capacity and/or which are best suited for the information to be communicated will be employed. Employing these intelligent mitigation strategies may yield payoffs through optimization of human capabilities resulting in greater overall operational efficiency.

REFERENCES

Baddeley, A. (1986). *Working memory*. New York, NY: Clarendon Press/Oxford University Press.

Brown, J. (1976). An analysis of recognition and recall and of problems in their comparison. In J. Brown (Ed.), *Recall and recognition* (pp. 1-35). New York: Wiley.

Daniels, J., Regli, S., & Franke, J. (2002). Support for Intelligent Interruption and Augmented Context Recovery. *Proceedings of 2002 IEEE Seventh Conference on Human Factors and Power Plant*s, Scottsdale, AZ, September 15-19, 2002.

Franke, J., Daniels, J, & McFarlane, D. (2002). Recovering Context After Interruption. *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Societ*y, CogSci-2002, Fairfax, VA, August 8-10, 2002.

Hale, K., Samman, S., Buff, W., Stanney, K.M., Reeves, L., & Bowers, C. (2003). Multimodal, multitask interaction design: Challenging long-standing unimodal design assumptions. *Proceeding of the 47th Annual Human Factors and Ergonomics Society Meeting*, Denver, CO, October 13-17, 2003 (pp. 586-590).

Neumann, D. L. (2002, June). Effect of varying levels of mental workload on startle eyeblink modulation. *Ergonomics 45*(8), 583-602.

Samman, S., Stanney, K., Dalton, J., Ahmad, A. & Bowers, C. (2004) *Multimodal interaction: Multi-capacity processing beyond 7 +/- 2*. Unpublished manuscript.

Wickens, C. D. & Hollands, J. G. (2000). *Engineering psychology and human performance.* Upper Saddle River, NJ: Prentice-Hall Inc.