Effects of Secondary Task Modality and Processing Code on Automation Trust and Utilization during Simulated Airline Luggage Screening

Rachel Phillips; Poornima Madhavan
Old Dominion University
rphil019@odu.edu; pmadhava@odu.edu

Abstract. The purpose of this research was to examine the impact of environmental distractions on human trust and utilization of automation during the process of visual search. Participants performed a computer-simulated airline luggage screening task with the assistance of a 70% reliable automated decision aid (called DETECTOR) both with and without environmental distractions. The distraction was implemented as a secondary task in either a competing modality (visual) or non-competing modality (auditory). The secondary task processing code either competed with the luggage screening task (spatial code) or with the automation's textual directives (verbal code). We measured participants' system trust, perceived reliability of the system (when a target weapon was present and absent), compliance, reliance, and confidence when agreeing and disagreeing with the system under both distracted and undistracted conditions. Results revealed that system trust was lower in the visual-spatial and auditory-verbal conditions than in the visual-verbal and auditory-spatial conditions. Perceived reliability of the system (when the target was present) was significantly higher when the secondary task was visual rather than auditory. Compliance with the aid increased in all conditions except for the auditory-verbal condition, where it decreased. Similar to the pattern for trust, reliance on the automation was lower in the visual-spatial and auditory-verbal conditions than in the visual-verbal and auditory-spatial conditions. Confidence when agreeing with the system decreased with the addition of any kind of distraction; however, confidence when disagreeing increased with the addition of an auditory secondary task but decreased with the addition of a visual task. A model was developed to represent the research findings and demonstrate the relationship between secondary task modality, processing code, and automation use. Results suggest that the nature of environmental distractions influence interaction with automation via significant effects on trust and system utilization. These findings have implications for both automation design and operator training.

1. INTRODUCTION

Automation, as found in many work environments, is seldom used in isolation. For example, luggage screeners are, by the nature of their task, exposed to the sights and sounds associated with large groups of people. Similarly, fighter pilots are often responsible for monitoring many different systems in various locations. Therefore, it is important to understand the influences of concurrently performed tasks (or secondary tasks) on performance, as well as the impact of specific kinds of tasks on automation use. Achieving this objective would allow for the development of training programs, work environments, and system designs which would maximize human-automation potential.

Multiple Resource Theory (MRT) deals with the theory of multiple task performance and derives importance from the prediction of interference between concurrently performed tasks. The most recent version of the model proposes four categorical and dichotomous dimensions: processing stages, perceptual modalities, visual channels, and processing codes [1], of which we will discuss only the most relevant. MRT postulates that if tasks share a dimensional level, there is greater interference and performance decrement than if the tasks utilize different levels of the same dimension [1], [2], [3].

The perceptual modalities of the MRT are composed of visual and auditory input. It has been found that tasks from different modalities (cross-modal) cause less interference than tasks from the same modality (intra-modal) [4], [5], and that people respond differently to long-term monitoring tasks presented in these two modalities [6]. Processing codes distinguish between analogue/spatial and categorical/symbolic processes. In the model, these modalities are represented as verbal and spatial [1]. The spatial modality is comprised of shape and motion detection [7] as well as sounds, whereas the verbal modality is defined by comprehension of verbal stimuli (either visual or auditory).

Arousal hypothesis basically states that observer alertness is dependent on sensory stimulation [8];

this relationship is generally characterized as an inverted U, with low and high arousal causing performance decrement and moderate arousal leading to the best performance [9]. Over time, if a target occurs rarely among frequent stimuli, accuracy and/or speed of detection generally deteriorates, a phenomenon known as vigilance decrement [10]. In automation interaction, non-vigilance generally manifests as complacency, whereby users assume (incorrectly) satisfactory system state, function, and/or performance, and behave accordingly. Complacency is thought to be influenced by trust, reliability, and confidence [11].

Although automation was developed with the intention of improving human performance, in many instances it has changed the nature of user interaction. Whereas previously, users were responsible for executing a variety of behaviors, now they have been reduced primarily to monitoring positions, interceding only in the event of a problem. Excessive automation has been found to contribute to experiences of sleepiness and fatigue in factory workers [12], as well as decreased arousal and increased frustration in drivers [13]. Decreased arousal is problematic because it is associated with a lack of focus [9], and a decrease in the availability of attentional resources [14], [15].

As stated earlier, vigilance decrement is thought to be a result of lack of sensory stimulation. Sensory stimulation should, therefore, reduce vigilance decrement. Based on this logic, the findings that automation with variable reliability led to greater performance than automation with constant reliability [11] and that adding haptic (tactile) feedback improved performance [16] is not surprising.

Participants completed a luggage screening task with a 70% reliable aid with and without distraction. We hypothesized that tasks which shared two levels (processing code and perceptual modality) with either the primary task or automated aid would cause greater interference than those which shared only one. We further hypothesized that different types of tasks would result in different kinds of interference and this would manifest in different interaction patterns with the automation.

2. METHOD

1. Participants

Eighty-one undergraduate ODU students participated for partial fulfillment of course credit. All participants were 18 years of age or older and had normal or corrected to normal vision and hearing.

2. Materials

Visual Search Task

Participants completed a computer simulation, on two consecutive days, in which they played the role of airline luggage screeners. X-ray images of luggage were presented on a 17 inch color monitor placed approximately 17 inches from the edge of the desk. The computer simulation was developed using Visual Basic for Windows and presented the image, diagnosis of the aid, opportunity for participant input, and feedback. The x-ray images were created using Adobe Photoshop and were comparably cluttered with everyday items (toys, clothes, accessories, etc.). A subset of 20% of the images had one of eight possible knife images digitally superimposed. The participant's task was to indicate which bags contained weapons for one training block (Day 1) of 100 luggage images and two test blocks (Day 2) of 200 luggage images each.

Secondary Tasks

In addition to performing the luggage screening portion of the experiment, for one test block on Day 2, participants were assigned to a secondary task presented in one of two perceptual modalities (visual or auditory) and in one of two processing codes (verbal or spatial). Participants either listened to music (auditory) or read text presented in a text box to the right of the primary task (visual). Participants in the verbal condition were instructed to count the number of times they heard or saw (depending on the modality condition) the word me. Participants in the spatial condition were instructed to count the number of times a specific sound occurred in the auditory condition or, for the visual condition, the number of times a specific symbol occurred. Songs were chosen based on the unlikelihood of familiarity with the artist, the catchiness of the tunes, and the relative clarity of the lyrics. The visual condition presented transcriptions of the auditory-verbal condition (in the visual-spatial condition, symbols were substituted for words). All conditions were appropriately counterbalanced.

Trust Questionnaire

The System Trust Scale (STS) (Jian, Bisantz, & Drury, 2000) was utilized to determine how accurate and dependable participants found the automation. The questionnaire was administered at the end of each block. Participants were asked to respond to twelve statements regarding their feelings for the automation on a scale of 1 (strongly disagree) to 10 (strongly agree).

3. Procedure

Participants were given an informed consent form which explained their rights as participants and further explained that the study lasted for two days. They completed an entrance questionnaire which obtained demographic information such as age, gender, program of study, experience with computers, etc.

Day 1 – The "Training Phase"

Participants completed a luggage screening task in which they attempted to detect 20 hidden knife images in 100 x-ray images of luggage. Prior to beginning, participants were shown an example luggage image and the eight knife images they were searching for. Participants were informed that the experiment was timed and that the computer would be keeping score. On each trial, an x-ray image of passenger luggage appeared for three seconds. After the image disappeared, participants either clicked on "stop bag" if they thought a knife was present or on "pass bag" if they thought the knife was absent. They then rated their confidence in the decision on a scale of 1 (not confident at all) to 5 (extremely confident). After completion of the luggage task, participants were reminded to return the following day.

Day 2 – The "Test Phase"

Exactly 24 hours after the initial portion of the experiment, participants returned to their seats from the previous day. Participants were informed that they were, once again, playing the role of luggage screeners and that the task and targets were the same. It was explained that we were interested in their ability to multi-task and therefore, either in the first or second half of the experiment they would complete a secondary task and the luggage screening task simultaneously. Participants completed two test blocks of 200 images each. The secondary tasks were counterbalanced so half of the participants performed the secondary task in the first test block and the other half completed the secondary task in the second.

In addition, on this day, the participants were assisted by a text-based automated decision aid. The aid provided a diagnosis of knife presence or absence in the form of a text message at the top of the screen at the end of each trial prior to participant input. Unbeknownst to participants the aid was designed to be only 70% accurate.

After the completion of the first trial block of 200 images, participants completed the STS and Secondary Task Questionnaire as appropriate. After a short break, participants resumed the screening task. After completing the second set of 200 images, participants again completed the scales as appropriate. They were thanked and debriefed before leaving.

3. RESULTS

Day one was included in the experiment primarily to facilitate similar baseline levels of performance for the second day. Although training is undoubtedly an interesting and important area of research, it is beyond the scope of this paper.

System Use Measures

Participants were measured for both compliance (probability of agreeing with the aid when it said target present) and reliance (probability of agreeing with the aid when it said target absent). A 2 (distraction: distracted vs. undistracted) x 2 (modality: visual vs. auditory) x 2 (code: verbal vs. visual) mixed ANOVA for compliance revealed a significant three way interaction, $F(1, 77) = 3.84$, $p = .054$ indicating that compliance levels were influenced by distraction differently depending on the distractor modality and processing code. Interestingly, those in the auditory-verbal distraction condition demonstrated a decrease in compliance when distracted (M = .703, SE = .037) versus undistracted (M = .740, SE = .038) which was contrary to the increase demonstrated by the auditory-spatial (distracted: M = .717, SE = .037; undistracted: M = .646, SE = .038), visual-verbal (distracted: M = .648, SE = .037; undistracted: M = .578, SE = .038), and visual-spatial (distracted: M = .682, SE = .036; undistracted: M = .659, SE = .037) conditions.

A 2 (distraction: distracted vs. undistracted) x 2 (modality: visual vs. auditory) x 2 (code: verbal vs. spatial) mixed ANOVA for reliance yielded a marginally significant three-way interaction between distraction, modality, and processing code, $F(1, 77) = 3.59$, $p = .062$. Participants in the auditory-verbal and visual-spatial condition decreased reliance when distracted (auditory-verbal: M = .668, SE = .014; visual-spatial: M = .686, SE = .013) versus undistracted (auditory-verbal: M = .702, SE = .016; visual-spatial: M = .689, SE = .015), but those in the auditory-spatial and visual-verbal conditions increased reliance when distracted (auditory-spatial: M = .702, SE = .014; visual-verbal: M = .678, SE = .014) versus undistracted (auditory-spatial: M = .696, SE = .016; visual-verbal: M = .659, SE = .016).

Confidence

We divided confidence into two variables, confidence when agreeing with the aid and confidence when disagreeing with the aid. A 2 (distraction: distracted vs. undistracted) x 2 (modality: visual vs. auditory) x 2 (code: verbal vs. spatial) mixed ANOVA for confidence when agreeing indicated a significant main effect for distraction, $F(1, 77) = 4.89$, $p = .030$. Participant confidence when agreeing with the aid actually decreased when distracted ($M = 2.35$, $SE = .09$) as compared to undistracted ($M = 2.45$, $SE = .09$).

A 2 (distraction: distracted vs. undistracted) x 2 (modality: auditory vs. visual) x 2 (code: verbal vs. spatial) mixed ANOVA for confidence when disagreeing revealed a significant interaction between distraction and the modality of the secondary task, $F(1, 77) = 5.13$, $p = .026$. This indicated that participant confidence when disagreeing varied as a function of distraction and modality. For the auditory condition, participant confidence levels remained relatively stable when distracted ($M = 2.26$, $SE = .13$) versus undistracted ($M = 2.20$, $SE = .12$). However, in the visual condition, participant confidence when disagreeing decreased with the addition of the distraction ($M = 2.04$, $SE = .13$) as compared to undistracted confidence levels ($M = 2.20$, $SE = .12$).

Trust Measures

A 2 (distraction: distracted vs. undistracted) x 2 (modality: auditory vs. visual) x 2 (code: verbal vs. spatial) mixed ANOVA revealed a significant three-way interaction between distraction, distractor modality, and processing code, $F(1, 72) = 6.32$, $p = .014$. For the visual condition, participant trust in the aid increased with the addition of the verbal stimuli (distracted: $M = 49.9$, $SE = 4.44$; undistracted: $M = 47.0$, $SE = 4.58$), but decreased with the addition of spatial stimuli (distracted: $M = 49.3$, $SE = 4.22$; undistracted: $M = 54.0$, $SE = 4.36$), a pattern directly opposite that demonstrated by those in the auditory condition [spatial: (distracted: $M = 50.6$, $SE = 4.44$; undistracted: $M = 46.2$, $SE = 4.58$); verbal: (distracted: $M = 59.8$, $SE = 4.69$; undistracted: $M = 63.5$, $SE = 4.84$)].

Participants were asked to estimate the reliability of the aid when the target was present. A 2 (distraction: distracted vs. undistracted) x 2 (modality: auditory vs. visual) x 2 (code: verbal vs. spatial) mixed ANOVA revealed a significant interaction between distraction and distractor modality for estimated system reliability, $F(1, 73) = 4.26$, $p = .043$. When distracted, participants in the visual condition estimated slightly higher accuracy ($M = 56.75$, $SE = 3.75$) than when undistracted ($M = 55.20$, $SE = 3.61$) and had higher estimates than distracted participants in the auditory condition ($M = 50.17$, $SE = 4.00$) but slightly lower estimates than the auditory participants when undistracted ($M = 56.94$, $SE = 3.86$).

Participants were also asked to estimate the reliability of the system when the target was absent. A 2 (distraction: distracted vs. undistracted) x 2 (modality: visual vs. auditory) x 2 (code: verbal vs. spatial) mixed ANOVA for estimated reliability when the target was absent revealed no significant main effects or interactions.

4. DISCUSSION

Performance is impacted by arousal [9]. Over time, vigilance tasks become boring and repetitive, leading to vigilance decrement [10] which manifests with automation as complacency [11]. Complacency can result in potentially dangerous human-automation interaction and is thought to be influenced by trust, system reliability, and user confidence [11]. Our findings suggest that this may be mediated by resource demands resulting from job requirements (simultaneous multiple task performance) or environmental factors (background music, noise, etc.). This would suggest the interaction model shown in Figure 1.

As hypothesized, the perceptual modality and processing code of the secondary task influenced automation use. Based on the MRT [1] we expected that the visual distractor tasks would be more detrimental to performance than the auditory. As indicated by the decrease in confidence when disagreeing with the aid and increase in the perceived reliability of the system with the addition of a visual distractor (see Figure 1), this seems to be the case. Especially when one considers that participants in the auditory conditions actually demonstrated a decrease in perceived system reliability, and an increase in confidence when disagreeing with the aid.

The processing codes of the secondary task further influenced performance depending on the presentation modality. The spatial task was expected to interfere with primary task performance. Contrary to our predictions, participant trust decreased with the addition of the visual-spatial distractor but increased with the addition of the auditory-spatial distractor. It is possible that the visual-spatial task was challenging enough to prevent complacency, increasing participant awareness of automation errors and thus leading to decreased trust. On the other hand, the auditory-spatial condition interfered with primary task performance but not with complacency, leading to increased system trust.

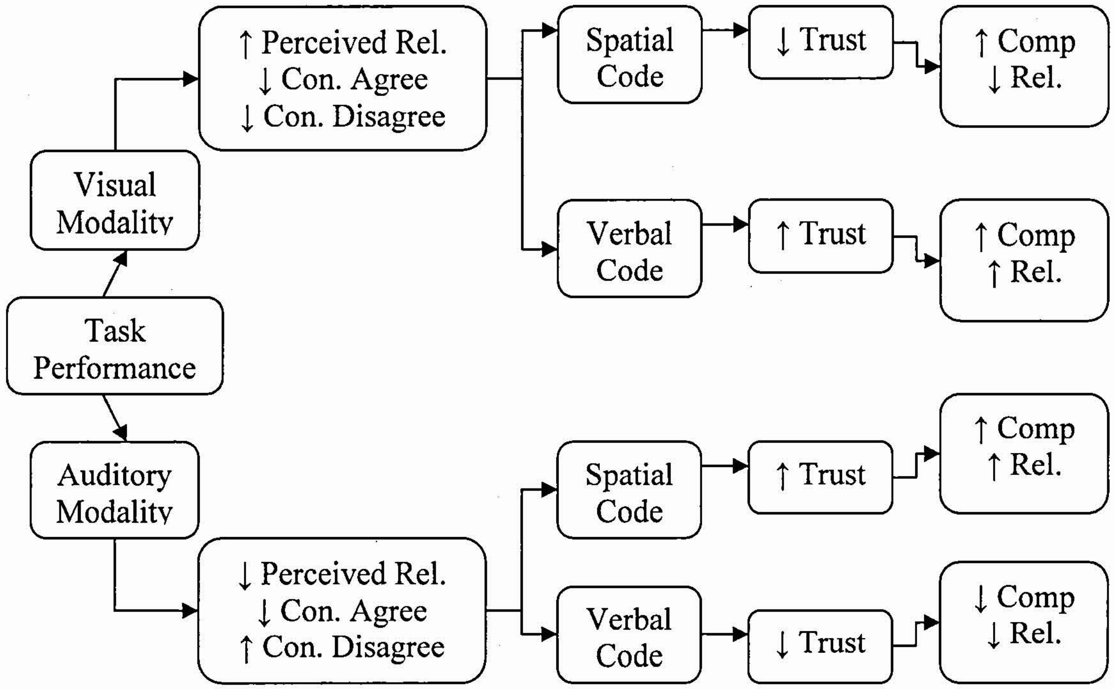

Figure 1. Suggested model for the influence of processing codes and perceptual modality on automation use.

The verbal condition was expected to interfere with aid comprehension. In the visual-verbal condition, participant trust increased, while in the auditory-verbal condition, participant trust decreased (see Figure 1). It is again possible that the visual-verbal condition interfered with task performance but not complacency, leading to increased trust, but the auditory-verbal condition interfered with both task performance and complacency, thereby decreasing trust.

It should be noted that the aid made two kinds of errors, misses and false alarms; however, because we had a 20% weapon base-rate, it actually presented a greater number of false alarms than misses. Additionally, the aid said "target absent" much more than it said "target present" giving participants greater opportunity to demonstrate reliance. It is possible that the difference in compliance (probability of agreeing with the aid when it said target present) and reliance (probability of agreeing with the aid when it said target absent) patterns between the auditory-verbal and visual-spatial conditions may have been due to the nature of the task. However, since reliance decreased (as illustrated in Figure 1) it seems more likely that the results were a product of interactions between perceived reliability, confidence, and trust.

In conditions for which trust increased with the addition of the secondary task (visual-verbal and auditory-spatial), compliance and reliance both increased. For the visual-spatial condition, compliance increased while reliance decreased suggesting that participants noticed the misses more than the false alarms. The combination of decreased confidence and trust may have also played a role. Because participants were less confident in their own abilities they may have been unwilling to disagree with the aid when it said "target present" leading to increased compliance. However, since they distrusted the aid (despite finding it more reliable), they were more likely to disagree with the "target absent" diagnosis resulting in decreased reliance. In the auditory-verbal condition, participants decreased in both compliance and reliance indicating that they were more leery of the system overall, and more confident in their own abilities.

As shown clearly in the model, participant interaction with the automation differed not only as a function of secondary task modality (visual or auditory) but also as a function of the processing code (verbal or spatial) utilized by the secondary task. Regardless of the reason for the resulting differences in perceived reliability, confidence when disagreeing, trust, and compliance and reliance patterns, these differences demonstrate the inaccuracy of assumptions regarding the stability of human-automation interaction. In order to be used correctly, different automation systems may require a specific pattern of use. Once validated, the model would be useful for designing workspaces and responsibility combinations to elicit the desired interaction pattern, thus improving human-automation performance.

REFERENCES

1. Wickens, C. D. (2002) "Multiple resources and performance prediction," *Theoretical Issues in Ergonomics Science*, vol. 3, no. 2, Apr-Jun, pp. 159-177.
2. Leigh, J. H. (1992) "Modality congruence, multiple resource theory, and intermedia broadcast comparisons: An elaboration," *Journal of Advertising*, vol. 21, no. 2, Jun, pp. 55-62.
3. Smith, R. E. & Buchholz, L. M. (1991) "Multiple resource theory and consumer processing of broadcast advertisements: An involvement perspective," *Journal of Advertising*, vol. 20, no. 3, Sep, pp. 1-7.
4. Sklar, A.E. & Sarter, N.B. (1999) "Good vibrations: Tactile feedback in support of attention and human automation coordination in event driven domains," *Human Factors*, vol. 41, no. 4, Dec, pp. 543-552.
5. St. John, M.; Kobus, D.A.; Morrison, J.G.; & Schmorrow, D. (2004). "Overview of the DARPA augmented cognition technical integration experiment," *International Journal of Human-Computer Interaction*, vol. 17, no. 2, pp. 131-149.
6. Szalma, J. L.; Warm, J. S.; Matthews, G.; Dember, W. N.; Weiler, E. M.; Meier, A.; & Eggemeier, F. T. (2004) "Effects of sensory modality and task duration in performance, workload, and stress in sustained attention," *Human Factors*, vol. 46, no. 2, Sum, pp. 219-233.
7. Weinstein, L. F. & Wickens, C. D. (1992) "Use of nontraditional flight displays for the reduction of central visual overload in the cockpit," *International Journal of Aviation Psychology*, vol. 2, no. 2, pp. 121-142.
8. Davenport, W. G. (1972) "Vigilance and arousal: Effects of different types of background stimulation," *The Journal of Psychology*, vol. 82, no. 2, Nov, pp. 339-346.
9. Bodner, E.; Gilboa, A.; & Amir, D. (2007) "The unexpected side effects of dissonance," *Psychology of Music*, vol. 35, no. 2, Apr, pp. 286-305.
10. Caggiano, D. M. & Parasuraman, R. (2004) "The role of memory representation in the vigilance decrement," *Psychometric Bulletin & Review*, vol. 11, no. 5, Oct, pp. 932-937.
11. Singh, I. L.; Molloy, R.; & Parasuraman, R. (1993) "Individual differences in monitoring failures of automation," *The Journal of General Psychology*, vol. 120, no. 3, Jul, pp. 357-373.
12. Persson, R.; Garde, A. H.; Hansen, A. M.; Ørbæk, P.; & Ohlsson, K. (2003) "The influence of production systems on self-reported arousal, sleepiness, physical exertion, and fatigue—consequences of increasing mechanization," *Stress and Health*, vol. 19, no. 3, Aug, pp. 163-171.
13. Stanton, N. A. & Young, M. S. (2005) "Driver behavior with adaptive cruise control," *Ergonomics*, vol. 48, no. 10, Aug, pp. 1294-1313.
14. Chase, B.; Irwan-Chase, H.; & Sonico, J. (2004) "A methodology to control for individual differences in dual-task performance," *Theoretical Issues in Ergonomic Science*, vol. 5, no. 5 , Sep-Oct, pp. 445-451.
15. Young, M. S. & Stanton, N. A. (2002) "Malleable attentional resources theory: A new explanation for the effects of mental under load on performance," *Human Factors*, vol. 44, no. 3, Fall, pp. 365-375.
16. Griffiths, P. G. & Gillespie, R. B. (2005) "Sharing control between humans and automation using haptic interface: Primary and secondary task performance benefits," *Human Factors*, vol. 47, no. 3, Fall, pp. 574-590.