Multimodal Communication In Support of Coordinative Functions in Human-Machine Teams

Nadine B. Sarter, Ph.D.

Cognitive Systems Engineering Laboratory and Institute for Ergonomics
The Ohio State University
Columbus, OH
sarter.1@osu.edu

ABSTRACT

The ongoing evolution of modern technology from reactive tool to powerful and independent agent has created problems that are related to breakdowns in human-automation coordination. These breakdowns can be explained, in part, by the fact that machines can initiate actions on their own and without explicit operator consent but do not possess the communication skills that are required to know when and how to share information with operators concerning their intentions, actions, and limitations. One problem in particular is the extensive and increasing use of automation feedback that requires focal visual attention. Such feedback does not support task-sharing and effective attention allocation, especially in the context of unexpected changes and events. One possible solution to the problem is suggested by multiple-resource theory: the distribution of tasks and information across various sensory channels. In this manuscript, we will provide an overview of our recent series of simulator studies on the effectiveness of multimodal feedback for supporting early stages of attention management. Also, some of the many remaining challenges associated with supporting more complex coordinative functions in human-machine teams will be discussed.

INTRODUCTION

The introduction of increasingly complex and autonomous systems to a variety of high-tempo high-risk domains such as aviation has introduced new, and has exacerbated existing, cognitive demands for practitioners. One particularly challenging task in these environments is effective attention management, i.e., the dynamic prioritization and allocation of attentional resources to several parallel threads of activity (Woods, 1995). Even before the introduction of modern automated systems, pilots had to attend to a wide range of tasks and activities, including flying the airplane, navigating, communicating with air traffic control, coordinating with other crewmembers, and managing system failures. With the introduction of highly independent cockpit systems that carry out tasks on their own and can change their status and behavior in the absence of explicit pilot commands, flight crews now also need to supervise and coordinate with these machine agents. Breakdowns in human-automation coordination continue to occur and have been the focus of a considerable body of empirical research (for an overview of problems with human-automation interaction, see Connors in the April 1999 issue of JHPEE; see also Sarter and Woods, 1994, 1995, 1997; Wiener, 1989). These studies have shown that one major problem is pilots' failure to notice changes in the status and/or behavior of their automated cockpit systems. In particular, uncommanded changes that occur due to system coupling, sensor input, or designer instructions are missed. This can lead to a loss of mode awareness, mode errors, and automation surprises (Sarter, Woods, and Billings, 1997).

One major contributor to the observed problems is a lack of communication skills on the part of the automation. Modern systems do not always provide all necessary information to the operator, and/or they fail to present the information in a way that is compatible with human cognitive and perceptual abilities and limitations. For example, many flight deck systems rely primarily and increasingly on focal visual feedback to keep their operators informed instead of utilizing the various sensory channels available to humans (Sarter, 1995, 2000). As a result, they are not very successful at capturing the operator's attention when necessary, as in the case of system-initiated changes and events, and thus fail to support operators in effective attention management. A recent series of simulator studies (Sklar and Sarter, 1999; Nikolic and Sarter, 2000) has shown that multimodal automation feedback (in particular, tactile and peripheral visual cues) is one promising avenue towards addressing this problem.

Findings from these studies and some of the remaining challenges for supporting more complex coordinative functions (such as interruption management) will be discussed in the following sections. First, however, we need to introduce the concepts of attention and interruption management.

ATTENTION AND INTERRUPTION MANAGEMENT

Attention management can be defined as the dynamic prioritization and allocation of attentional resources to several parallel threads of activity (Woods, 1995). Many of the event and activities requiring a pilot's attention throughout a flight can be anticipated as they occur on a regular basis during the same phases of flight. This allows pilots to plan ahead to some extent. They can distribute their workload in an attempt to avoid attentional bottlenecks. However, the high degree of dynamism and uncertainty in event-driven domains such as aviation requires that pilots handle unanticipated externally imposed attentional demands as well. Operators are currently not well supported in performing this task because little is known about effective attention control in the face of multiple signals and tasks that compete for limited attentional resources. As Woods et al. (1994) have pointed out, this cognitive challenge "is the least explored frontier in cognitive science and human-machine cooperation" (Woods et al., 1994).

Alarms represent one of the few cases where the decision about whether or not to shift attention to an incoming signal is made for the operator. Since alarms notify operators of events that are sufficiently critical to warrant an immediate shift in attention under almost all circumstances, they are deliberately designed in a very salient and intrusive manner that makes them difficult or impossible to ignore.

In most other cases, the operator needs to decide on his/her own whether and when to attend to an interruption signal. Some of these interruptions are initiated by highly automated systems which need to

- inform the operator of progress on, or difficulties with, performing a task
- request information or approval from the human operator
- negotiate changing goals and intentions (see Ball et al., 1997).

These interactions between human and machine are not always time-critical or of sufficient importance to warrant an immediate shift in the operator's attention away from ongoing tasks and lines of reasoning. Instead, automated systems need to provide the operator with information that allows him/her to make a decision about responding to an interrupt signal in a mentally economical way. Preferably, the decision should be made by preattentive reference, i.e., without the need to orient towards the new stimulus. Woods (1995) outlined the following criteria for supporting preattentive reference which call for attention directing signals to a) be picked up in parallel with ongoing tasks and activities, b) provide information on the significance and meaning of the interruption, and c) allow for evaluation that does not require focal attention.

One type of display that was designed to support preattentive reference is so-called likelihood alarm displays (Sorkin, Kantowitz, and Kantowitz, 1988). With these displays, the likelihood of an event is computed by an automated monitoring system and encoded in an alerting signal. For example, an operator may be informed that a system failure is possible, probable, or certain. Under high workload conditions, this information is more effective than traditional binary alarm signals in helping operators decide whether and when to interrupt ongoing tasks to attend to the problem without imposing undue attentional demands.

Research in our laboratory aims at achieving the same goal. The following sections provide an overview of recent studies that examined the effectiveness of multimodal information presentation for allowing operators to detect and identify uncommanded changes and events in parallel with performing other tasks.

PERIPHERAL VISUAL AND TACTILE FEEDBACK IN SUPPORT OF EFFECTIVE ATTENTION ALLOCATION

As mentioned earlier, current automation feedback relies heavily and increasingly on the focal visual modality. Such feedback forces serial access to the large amount of data that is available in many domains, and it is not well suited for providing attentional guidance through attention capture (Sarter, 1995, 1997). Presenting information via other modalities (e.g., in the form of tactile or peripheral visual cues) appears to be a more promising approach as indicated by a body of empirical evidence from laboratory-based research (e.g., Jonides and Yantis, 1988; Hillstrom and Yantis, 1994; Gilliland and Schlegel, 1994).

Peripheral vision has several properties that make it a likely candidate for capturing and guiding operators' attention. The periphery is highly effective in detecting dynamic discontinuities such as motion or changes in luminance. These features are well suited to produce involuntary shifts of attention (Yantis and Jonides, 1990; Theeuwes, 1991; Folk et al., 1994) and thus help humans detect potentially interesting objects that may warrant subsequent focal attention. It has been suggested that 90% of all visual stimulation is obtained in the periphery (Stokes and Wickens, 1988), helping us perceive the destination of our next saccade and thus serving attentional (re)orientation and control (McConkie, 1983). Another potential benefit of presenting information in peripheral vision is that it can be obtained with little or no conscious effort, thus making peripheral vision a resource-economical channel.

Potential limitations of peripheral visual feedback need to be considered as well. For example, a narrowing of the functional field of view has been shown under conditions of high stress and cognitive task loading - a phenomenon known as "tunnel vision" (Leibowitz & Appelle, 1969). It appears, however, that operators who are highly trained for and experienced in dividing their attention across multiple displays (such as pilots) are affected by this phenomenon to a lesser extent (Williams, 1995).

To date, only few attempts have been made to introduce peripheral visual displays to the aviation domain, and most of those efforts served to support vehicle guidance and control (e.g., Holden, 1964). Examples of such displays are instrument landing aids (e.g., the Para-Visual Director (PVD) (Majendie, 1960)) or peripheral attitude indicators (e.g., the Malcolm horizon (Malcolm, 1984)).

Another form of feedback that is underutilized in many domains is tactile cues. In fact, some of the few sources of tactile and peripheral visual feedback that used to be available on flight decks have been eliminated in recent years in the absence of thorough investigations of the potential impact of these changes. As with peripheral visual feedback, very few efforts have been made to explore the use of tactile feedback in the aviation domain, and these efforts focused primarily on providing navigational guidance (e.g., Gilliland and Schlegel, 1994; Zlotnik, 1988) or presenting warning signals (e.g., the stick shaker indicating a pending stall). More research on the affordances of tactile feedback is needed and feasible thanks to the recent development of small unintrusive tactile interfaces (so-called tactors) that allow designers and researchers fine-tuned control over parameters such as the frequency, duration, and amplitude of the signal.

Two recent simulator studies in our laboratory have confirmed that peripheral visual and tactile feedback can effectively support attention capture in data-rich environments involving multiple tasks and complex systems. Subjects in both studies were instructor pilots who were asked to fly a desktop simulator of a modern glass cockpit aircraft. In addition to their flight-related tasks, which differed in terms of difficulty (from easiest (none – autopilot active) to most difficult (automation management)), pilots had to detect uncommanded changes in the status of an automated system (so-called mode transitions) as well as other events (such as traffic) that occurred either in isolation or concurrently with a mode transition. The first study compared the effectiveness of current foveal visual feedback (FMA) with two types of peripheral visual cues (1) a more salient version of the current FMA and 2) an ambient strip extending across the bottom of both monitors) for indicating the occurrence of uncommanded and unexpected transitions (Nikolic and Sarter, 2000). The second study examined the effectiveness of using vibrotactile cues, presented via tactors to the subjects' wrist, for the same purpose (Sklar and Sarter, 1999). Participants were presented either with tactile cues only or with tactile cues in combination with current visual feedback (FMA).

Both peripheral visual and, even more so, tactile cues resulted in significantly improved detection performance for the experimenter-induced mode transitions (see Figure 1) without affecting the performance of concurrent tasks such as flight path tracking or the detection of other important events (e.g., traffic).

Insert Figure 1 about here

Figure 1. Detection performance for experimenter-induced mode transitions for a baseline group (FMA - receiving currently available feedback) and two groups each receiving either peripheral visual (Enhanced FMA and Ambient Strip) or vibrotactile (Tactile + FMA and Tactile Only) feedback

The advantage of tactile feedback over visual cues was particularly pronounced in conditions of high concurrent load (see Automation mgmt block) where tactile feedback yielded near perfect detection rates and significantly reduced reaction times to transitions. Pilots receiving peripheral visual cues still missed a considerable number of events. A recent follow-up study has shown that this may be explained, in part, by the fact that the attention capture power of peripheral visual cues is affected considerably by the background or context in which these cues appear (Orr, Nikolic, and Sarter, 2001). Not surprisingly, if onsets are embedded in a dynamic background involving objects of similar color as the onset itself (which is typically the case for indications on modern flight decks), a significant decrease in detection performance can be observed.

The above studies examined not only whether or not pilots could detect changes in the status of the automated system (see the above-mentioned first criterion for supporting preattentive reference) but also whether different peripheral visual and tactile cues enabled them to distinguish between two different kinds of transitions (a step towards addressing the second criterion which calls for providing information on the significance and meaning of an interruption signal). The peripheral visual cues varied in terms of hue whereas the same tactile feedback was presented in different locations (inner and outer wrist) to indicate two different transitions. Out of the total of 576 transitions presented to pilots in the two peripheral visual feedback conditions, only one was misidentified. And only seven of the 168 transitions presented to pilots in the tactile condition were misinterpreted. This is an encouraging finding as it indicates that some limited information about the nature of an interrupt signal can be picked up without affecting performance on other primary tasks.

SUPPORTING ATTENTION AND INTERRUPTION MANAGEMENT: SOME REMAINING ISSUES AND QUESTIONS

The above research has shown that multimodal information representation is one promising avenue towards improving the communication and coordination skills of highly autonomous systems. Tactile and peripheral visual cues were successful in informing pilots reliably and rather unintrusively about the occurrence and, to a very limited extent, the nature of potentially attention-demanding events (in this case, mode transitions). Follow-on studies are currently under way to examine in more detail the affordances and possible limitations of the two sensory channels and to develop a robust integrated implementation of multimodal feedback.

The above research focuses primarily on the early stages of attention management, such as attention capture. Surprisingly little is still known about how to manage later stages of attention and interruption management effectively despite the fact that interruptions are widely considered one of the major threats to safety in a variety of domains (e.g., Dornheim, 2000). Some of many issues that remain to be addressed include

- The type of information that is needed for making a decision about responding to an interruption signal.

- For example, operators could be informed about the source, urgency, and nature of an interruption. They may need to know how much time they have until they need to respond and how much time it will likely take to handle the interruption (see Latorella, 1999). We need to identify what information is most useful under what circumstances.

- The use and integration of different modalities for supporting interruption management.
    - Latorella (1999) conducted one of the few studies that have looked at the effects of modalities on interruption management. She found that interruptions to auditory tasks were acknowledged significantly slower than interruptions to visual tasks due to the transient nature of auditory information. Interruption initiation times to crossmodal conditions were significantly slower than to same-modality conditions. And the performance of interrupting tasks was begun significantly slower when the interruption was presented visually and when the interruption occurred to an auditory task. In this case, the long-term availability of the visual reference allows for delays in attending to an interruption. More research is needed to include other sensory channels and examine their bandwidth and interference. Also, most research on multimodal communication to date has considered the concurrent use of two modalities only. Many of these studies included visual and auditory

feedback but failed to include other senses (for exceptions, see e.g., Spence et al., 1998 on crossmodal cueing between touch, audition, and vision). Another important question in this context relates to modality expectations. Expecting a cue to appear in a certain modality increases the detection rate and reduces the response time to that stimulus (see Post and Chapman, 1991). Conversely, such expectations will hurt detection performance for cues that appear in a modality other than the expected. Modality expectations may be created by the higher frequency of cues in one modality compared to others or as the result of the actual or perceived importance of cues in that modality. This emphasizes the need to examine the entire task environment to avoid interference and imbalance.

- Effective protocols for interruptions

It will also be important to identify and compare the benefits and disadvantages of different protocols for handling interruptions. For example, McFarlane (1999) has examined four possible approaches: immediate, negotiated, mediated, and scheduled interruptions. Overall, he has shown that the preferred method depends on the nature of tasks and associated performance measures. In his study, negotiated interruptions worked best for accuracy on continuous tasks but carry the risk that people may not handle interruptions in a timely manner. Immediate interruptions, on the other hand, lead to prompt and effective handling of the interruption at the expense of the interrupted task where more errors are made. The entire range of possible protocols for handling interruptions needs to be explored in order to develop context-sensitive recommendations for their use and combination.

CONCLUSION

Breakdowns in human-automation coordination can be explained, to a large extent, by the limited communication and coordination skills of modern technologies. A recent line of research indicates that one promising avenue towards improving the situation is multimodal information representation. In particular, the use of peripheral visual and tactile cues - two currently underutilized means of communication - was shown to enable the automation to inform operators reliably and without significant attentional costs about the occurrence and, to a limited extent, the nature of system-initiated changes and events. More research on the affordances and limitations of these channels and their interference with other modalities is needed to be able to develop theory-based principles guidance for the use of tactile and peripheral visual feedback (see Wickens, Sandry, and Vidulich, 1983).

In addition to developing guiding principles for the use of each individual modality, we need to consider their functional and temporal coordination with other modalities. As Oviatt (1998) points out, the design of multimodal systems needs to be based on knowledge about how people use and switch between modalities for different tasks and purposes. Modality shifting, or the "contrastive functional use of modes", has been shown to serve as a means to designate consequential shifts in the content or functionality of expression (see Oviatt, 1998; VanGent, 1995).

In conclusion, the goal of much of today's research is to address existing difficulties with human-automation interaction in a variety of domains. Relatively few efforts are under way to anticipate and invest in a better understanding of future challenges that are likely to result from the continued evolution of technology from passive tool to independent agent. In particular, future highly autonomous systems can be expected to increase the need for supporting effective attention and interruption management. In this manuscript, we have shown that multimodal communication is a promising avenue towards achieving this goal but that a considerable number of issues remain to be resolved before we can hope to create truly cooperative human-machine teams.

ACKNOWLEDGMENT

The preparation of this manuscript was supported, in part, by a grant from the National Science Foundation (grant-# NSFII597-33100CAR; Technical Monitor: Dr. Ephraim Glinert).

REFERENCES

Ball, G., Ling, D., Kurlander, D., Miller, J., Pugh, D., Skelly, T., Stankosky, A., Thiel, D., Van Dantzich, M. and Wax, T. (1997). Lifelike Computer Characters: The Persona Project at Microsoft. In J.M. Bradshaw (Ed.), Software Agents (pp. 191-222). AAAI Press: Menlo Park, CA.

Connors, M.M. (1999). Teaming Humans and Automated Systems in Safely Engineered Environments. Journal of Human Performance in Extreme Environments, 4(1), 84-90.

Dornheim, M.A. (2000). Crew Distractions Emerge As New Safety Focus. Aviation Week and Space Technology, July 17, pp. 58-60.

Folk, C. L., Remington, R. W., and Wright, J. H. (1994). The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and color. Journal of Experimental Psychology: Human Perception and Performance, 20, 317-329.

Gilliland, K. and Schlegel, R. E. (1994). Tactile stimulation of the human head for information display. Human Factors, 36(4), 700-717.

Hillstrom, A.P. and Yantis, S. (1994). Visual Motion and Attentional Capture. Perception and Psychophysics, 55(4), 399-411.

Holden, K.J. (1964). Instrument Displays for Blind Flying. Journal of the Royal Aeronautical Society, 68, 833.

Jonides, J. and Yantis, S. (1988). Uniqueness of Abrupt Visual Onset in Capturing Attention. Perception and Psychophysics, 43, 346-354.

Latorella, K.A. (1999). Investigating Interruptions: Implications for Flightdeck Performance. NASA Technical Memorandum TM-1999-209707. Langley Research Center: Hampton, VA.

Majendie, A.M.A. (1960). Para-Visual Director. Journal of the Institute of Navigation (London) 13, 447-454.

Malcolm, R. (1984). Pilot disorientation and the use of a peripheral vision display. Aviation, Space, and Environmental Medicine, 55 (3), pp. 231-238.

McConkie, G. (1983). Eye Movements and Perception During Reading. In K. Rayner (ed.), Eye movements in reading: Perceptual and language processes. New York: Academic Press.

McFarlane, D.C. (1999). Coordination the Interruption of People in Human-Computer Interaction. In M.A. Sasse and C. Johnson (eds.), Interact 99 (Vol. 1; pp. 295-303). Amsterdam, The Netherlands: IOS Press.

Nikolic, M.I. and Sarter, N.B. (2000). Peripheral Visual Feedback: A Powerful Means of Supporting Attention Allocation and Human-Automation Coordination In Highly Dynamic Data-Rich Environments. Human Factors. In press.

Orr, J., Nikolic, M., and Sarter, N.B. (2001). Supporting Human-Automation Communication Through Display Design: Do Visual Onsets Always Capture Attention? In Proceedings of the 11th International Symposium for Aviation Psychology. Columbus, OH. March 2001.

Post, L.J. and Chapman, C.E. (1991). The effects of cross-modal manipulations of attention on the detection of vibrotactile stimuli in humans. Somatosensory and Motor Research, 8, 149-157.

Sarter, N. B. (1995). "Knowing when to look where": Attention allocation on advanced automated flight decks. In R. S. Jensen (Ed.), Proceedings of the 8th International Symposium on Aviation Psychology (pp. 239-242). San Diego, CA: Academic.

Sarter, N.B. (2000). The Need for Multisensory Feedback in Support of Effective Attention Allocation in Highly Dynamic Event-Driven Environments: The Case of Cockpit Automation. *International Journal of Aviation Psychology, 10(3), 231-245.*

Sarter, N. B. and Woods, D. D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the Flight Management System (FMS). International Journal of Aviation Psychology, 4(1), 1-28.

Sarter, N. B. and Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. Human Factors, 37, 5-19.

Sarter, N. B. and Woods, D. D. (1997). Team play with a powerful and independent agent: Operational experiences and automation surprises on the Airbus A-320. Human Factors, 39, 553-569.

Sarter, N. B., Woods, D. D., and Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), Handbook of human factors and ergonomics (2nd ed., pp. 1926-1943). New York: Wiley.

Sklar, A. E., and Sarter, N. B. (1999). Good Vibrations: Tactile feedback in support of attention allocation and human-automation coordination in event-driven domains. Human Factors, 41(4), 543-552.

Sorkin, R.D., Kantowitz, B. H., and Kantowitz, S.C.(1988). Likelihood Alarm Displays. Human Factors, 30(4), 445-459.

Spence, C., Nicholls, M.E.R., Gillespie, N., and Driver, J. (1998). Cross-modal links in exogenous covert spatial orienting between touch, audition, and vision. Perception and Psychophysics, 60(4), 544-557.

Stokes, A.F. and Wickens, C.D. (1988). Aviation Displays. In E.L. Wiener and D.C. Nagel (eds.), Human Factors in Aviation (pp. 387-432). New York: Academic Press.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effects of visual onsets and offsets. Perception and Psychophysics, 49, 83-90.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman and R. Davies (Eds.), *Varieties of attention (pp. 63-101).* Orlando, FL: Academic Press.

Wickens, C.D., Sandry, D.L., and Vidulich, M. (1983). Compatibility and Resource Competition between Modalities of Input, Central Processing, and Output. Human Factors, 25(2), 227-248.

Wiener, E.L. (1989). Human factors of advanced technology ("glass cockpit") transport aircraft. (NASA Contractor Report No. 177528). Moffett Field, CA: NASA-Ames Research Center.

Williams, L.J. (1995). Peripheral target recognition and visual field narrowing in aviators and nonaviators. The International Journal of Aviation Psychology, 5(2), 215-232.

Woods (1995). The Alarm Problem and Directed Attention in Dynamic Fault Management. Ergonomics, 38(11), 2371-2393.

Yantis, S. and Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. Journal of Experimental Psychology: Human Perception and Performance, 16, 121-134.

Zlotnik, M. A. (1988). Applying electro-tactile display technology to fighter aircraft—flying with feeling again. Proceedings of the IEEE 1988 National Aerospace and Electronics Conference, pp.191-197.