Relative Effectiveness of Audio Tools for Fighter Pilots in Simulated Operational Flights: A Human Factors Approach

Sylvain Hourlier[#], James Meehan[*], Alain Léger[°], Corinne Roumes

[#]Département de Sciences cognitives – IMASSA – BP 73 – 91223 Brétigny-sur-orge Cedex – France
Tel. 33 1 69 23 78 74 – Fax. 33 1 69 23 70 02

shourlier@imassa.fr

[*] DSTO (Melbourne / Australia)
[°] Thales Avionics (Le Haillan / France)

ABSTRACT

The French-Australian Collaboration on Emerging Technologies (FACET) investigated appropriate means of delivering situation-awareness into the cockpit of fighter aircraft under simulated operational conditions. Increasing use of audio has been suggested as a means to reduce visual workload, to enhance situation awareness, and mitigate the manual and cognitive demands of HOTAS and existing command-and-display concepts. An open design for the pilot interface should incorporate redundant information in the auditory and other sensory modes, while enabling commands to be delivered through voice or manual control interfaces. However, sensory and cognitive resource competition may still limit delivery of the implicit benefits of such a design. The objective of FACET was to investigate this proposition.

Eight military pilots from France and Australia flew four full-mission scenarios in a simulated air combat environment. Auditory signals comprised radio messages, 3-D sounds and alarms, while the pilot's oral responses were verbal responses or direct voice inputs via automatic speech recognition. Some of the signals were drawn from those already in use, but all were selected for their potential to support situation awareness and support visual information, all delivered via a helmet-mounted display.

The four scenarios had specific, embedded combinations of events requiring activation of multiple resources. All events were videotaped. Both planned and unexpected events that were the outcome of human-interface interactions occurred. The pilots' behaviour (e.g., tool selection, prioritization, errors) as well as their efficiency in processing inputs and output control were analysed in order to describe their cognitive resource management. Resource competition imposed by the new technologies was analysed. Interviews were also performed after each run. Following the simulation, Repertory Grids were constructed to elicit each pilot's cognitive representation of the control-and-display concepts.

Pilot performance appeared to be independent of the time course of resource use during an event. Resource management was therefore not based on sequential allocation of resources, but rather on a cognitive prioritization of the task at hand. New audio technologies proved to be easy to use in conjunction with a broader use of the helmet-mounted display. Direct voice input required strict initial training to ensure reliability. Following such training, it was attractive to the pilots, even under high workload. The efficiency of 3-D sound was pilot dependent; one pilot in particular was able to fully use it immediately yielding a significant performance increase with a low cost in cognitive resources. In general, pilots' use of 3-D sound was dependent upon their level of training as well as their trust in, and willingness to experiment with the technology. These results may have implications for the training of pilots in the use of these new technologies.

*Besides the remarkable effectiveness of Franco Australian collaboration and workforce sharing that
needed to be noted, this project highlighted that operational pilots had a relative easy appropriation of
technology they had never used before in realistic simulated air combats.*

*Their strategy of use was workload dependant, yet still efficient at the highest workload levels even for
DVI, revealing the true nature of cognitive resource management in flight. Pilots relied on an in/output
management mastery where "tradeoff guided" resource allocation is the key to success. Interestingly, new
and resource hungry interface systems were not especially discarded at highest levels of workload
because pilots thought their use brought a fair tradeoff. Beyond the investigation of "tradeoff guided"
resource management we should try to establish how to build systems that blends into this high level
resource management strategy.*

1.0   INTRODUCTION

In this paper we describe an evaluation in simulated flight of new control and display technologies for
future combat aircraft. The technologies involved are helmet-mounted display (HMD) delivered visual and
auditory information, and direct-voice commands. This encompasses visor-projected imagery (symbology
and sensor images), advanced auditory displays (3-D sound, auditory icons, synthetic speech) and direct
voice recognition by on-board systems. The essence of the problem is that these technologies are
approaching a level of maturity where plans for their incorporation into new and existing aircraft are
becoming firm, and we do not yet fully understand the likely impact on the pilot. The specific aim of the
study reported here is to understand the behavioral and cognitive outcomes of addressing information to
the different sensory modes and enabling inanimate on-board systems to respond to voice commands of
the pilot. The general objective is the evolution of some principles for design and implementation of such
systems in the next generation combat aircraft. This paper is focusing on the use of a set of audio tools and
the pilot's cognitive resource management that ensued.

2.0   PURPOSE OF THE STUDY

As a scientific background we used Wickens Multiple Resource Theory [1]. It adds a qualitative aspect to
resources that explains low performance on certain concurrent task because they draw on the same
qualitative pool at the same time. For example, it is most difficult to speak while listening.
Resources are thus defined according to three dimensions; stage defined (early *versus* late processing),
modality defined (auditory *versus* visual encoding) and processing-code defined (spatial *versus* verbal
encoding). For the evaluation of a system, that uses both visual and auditory cues simultaneously with
manual and verbal commands, such qualitative discrimination in resources will reveal incompatibilities in
the system usage. The most expected one being issuing a DVI (direct voice input) command while
listening to audio inputs. In such a frame of reference, the evaluation of the "alternate technology suite"
workload depends on the number of concurrent uses, the respective type of resource they feed on and the
level of ongoing demands due to flight and mission. The idea is that the more an individual feels his
resources are invested, the highest his perceived workload will be. A bench model of resource competition
was designed to represent the most probable resource conflicts to be expected given the HMI at hand
(figure1).
Inspired by Flanagan's [2] critical incidents, critical events were embedded in four operational scenarios
designed by the French Flight test centre (in Istres). Each critical event drew on specific allocation of
resources (auditory, visual, verbal, etc.) predicted by the model to be of a problem for pilots, and were
scheduled along increasing workload phases. Due to the complexity of the missions at hand, some events
combinations were expected to happen during high workload phases but could not be predicted before
hand, so an in-house event videotaping and labeling system (comparable to the "Experimental Video
Annotator" [3] was designed accommodate the appearance of any "unscheduled" event. Each recorded

event is composed of a series of activities (basic interactions with the HMI, either the "standard" ones or the "alternate" ones) in their original order of appearance.
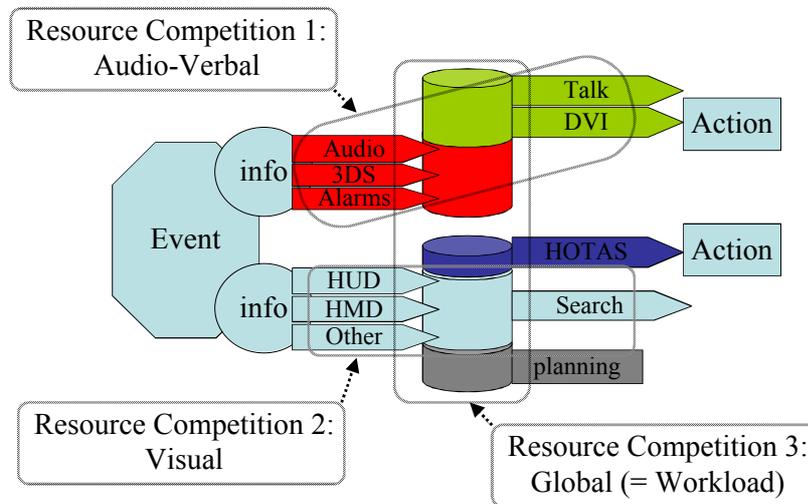
Figure 1: Resource competition model outlining audio-verbal resources competition, visual resources competition and global resource competition (i.e. Workload)

The pilots' behavior (e.g., tool selection, prioritization, errors) as well as their efficiency in processing inputs and output control was analyzed in order to describe their cognitive resource management. Resource competition imposed by the new technologies was analyzed. Interviews were also performed after each run. Following the simulation, Repertory Grids were constructed to elicit each pilot's cognitive representation of the control-and-display concepts.

3.0 METHODS

3.1 Simulation

The Rafale simulation at Istres flight test centre was used for this experiment (full mission sphere).

3.1.1 Alternate technology suite

Thales's alternate HMI suite consisted of a 3DS generating system integrated in the Topsight Rafale helmet (HMD) with a direct voice inputs (DVI) possibility. The standard Rafale system was always available, but pilots were encouraged to use the alternate system whenever possible. Auditory signals comprised radio messages, 3-D sounds and alarms, while the pilot's oral responses were verbal responses or direct voice inputs via automatic speech recognition. Some of the signals were drawn from those already in use, but all were selected for their potential to support situation awareness and support visual information, all delivered via a helmet-mounted display.

3.1.2 Scenarios

Four full-mission scenarios were designed in a simulated air combat environment.

3.2 Population

Seven military pilots from France and Australia (4 French, 3 Australians) participated in the study, though they had various background, all were flying combat aircraft at the time of the experiment.

3.3 Event analysis

Each event (expected and unexpected) was named after its activity composition; for example if an in coming communication (A) appeared when the pilot was issuing a DVI call (DVI) that happened on a preexistent cognitive load (G), the resulting event was labeled "G-DVI-A".

Events were analyzed with regards to their composition (number and order of activities), each event and activity being assessed in terms of pilot realization (good or bad performance for the event and its activities, as well as for the impact of an activity over a preexisting event).

3.4   Controlled debriefings

The "on the spot" debriefings enabled understanding of event performance as well as insights on acceptance and usability, both with regards to the systems (HMD, DVI, 3DS) used.

3.5   Collected Data

The data collected consist of:
- 28 debriefed runs,
- 57 types of events by combination of 6 types of activities: A (alarms, com-in), 3DS, DVI, HMD, Ve (com-out), G (cognitive workload), Vi (visual search).
- 271 recorded events (135 predictable and136 unpredictable) over 5 levels of workload

4.0   SELECTED RESULTS AND DISCUSSION

4.1   Resource management strategy

One could think that the arrangement of activities over time that create an event can influence the performance of the included activities or on the performance of the subsequent event. To address this, the performance of every activity was tested with regards to its position in an event. Only 3DS, DVI and HMD had sufficient occurrences to be statistically analyzed with an ANOVA. In fact, the evaluation of event performance with regards to the position of activities shows no significant impact whether the activity position (last or first, or in the middle): DVI (F^3_{79}=0.26; p=0.853; NS), 3DS (F^3_{186}=0.59; p=0.619; NS) and HMD (F^2_{128}=0.38; p=0.686; NS). One could have thought that interface management information had to be "cumulative". Prioritization strategy dominates thus sequentiallity of activities. The prioritization is governed by task and objective at hand and activities are evaluated in terms of "resource investment demand" versus "performance gain" and integrated to or discarded from the ongoing event, whatever their timing position. (i.e. an incoming communication can lead to the abandon of another ongoing activity if the pilot thinks the tradeoff is worth it). Pilot resource management seems to be based on a cognitive prioritization of the task at hand.

4.2   Effect of 3DS and DVI use on performance

To test the effect of DVI or 3DS use on the performance of an ongoing event, the relative performances of events associated with their respective use (failed or successful) were compared. It appears that when a DVI or 3DS command is issued, its performance does not influence the performance of the ongoing event.
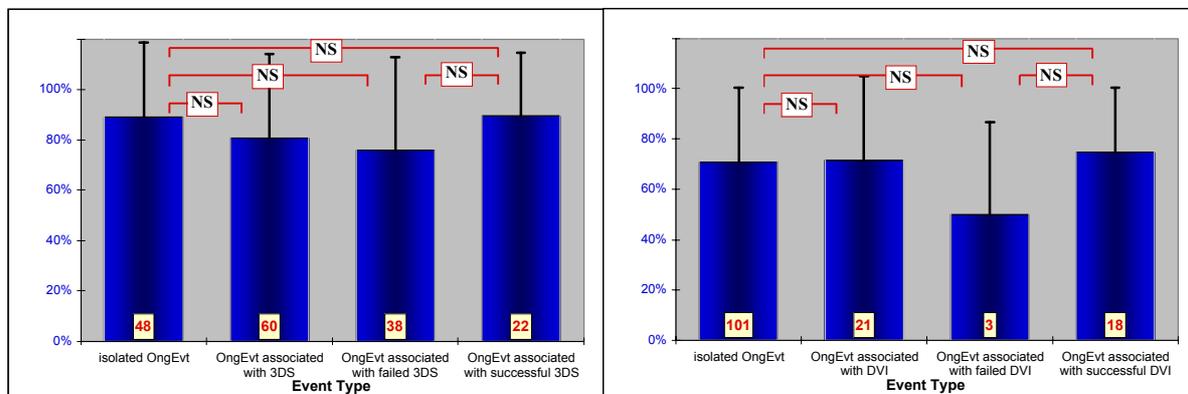

Figure 2: effect of a DVI (left) or 3DS (right) activity on an ongoing event (performance is in blue, number of events in red, black bars show standard deviation), ANOVA tests are shown in red brackets.

The relative high performances (80 to 90%) of the ongoing events presented on figure 2 (left) show that the complementary activities in an event including 3DS are highly successful whatever the success of the 3DS. A 3DS does not alter the prioritization of ongoing events. 3DS is generally ignored and the performance of the ongoing event is not altered because pilots (but one) did not delegate resources to address the 3DS information.

As for DVI use, when a command is issued, its performance does not influence the performance of the ongoing event (figure 2 right). This could be interpreted as a transfer of resources from one standard previous activity (HOTAS for example) to DVI. The ongoing performance variation is not identifiable but gain comes through ease of use as pilots reported.

4.3 Effect of training with DVI and 3DS over runs

The DVI learning curve happens during all 3 first scenarios and the effect is significant (F^3_{79}=6.34; p=0.0007). The effect is though not significant for 3DS. The only activity of the HMI suite that has a strong potential for early learning is the DVI (figure 3 left). As for the 3DS it seems that the progress was not consistent enough to be noticeable. There has been no significant progress of the 3DS activity over the four scenarios (F^3_{186}=1.57; p=0.19; NS): It is possible that the learning curve was uneven throughout pilots (figure 3 right). Though, the less than 45% overall success with 3DS is noteworthy.
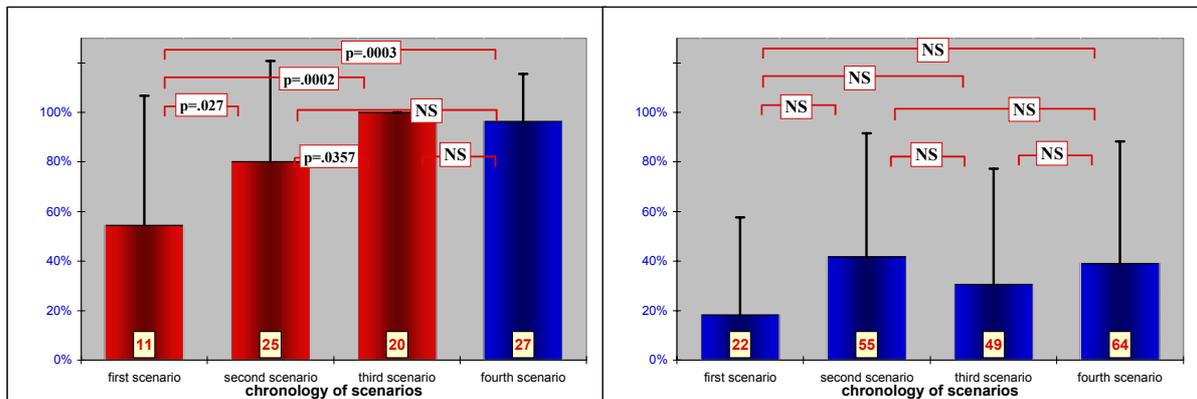
Figure 3: effect of training through successive scenarios on performance of DVI (left) and 3DS (right). The red columns identify the one statistically different than the others (Fisher test shown in red brackets).

4.4 Performance with DVI and 3DS with regards to workload

It appeared that there is no significant effect of workload on event performance (when they are all aggregated) (F^4_{266}=1.27; p=0.28; NS).
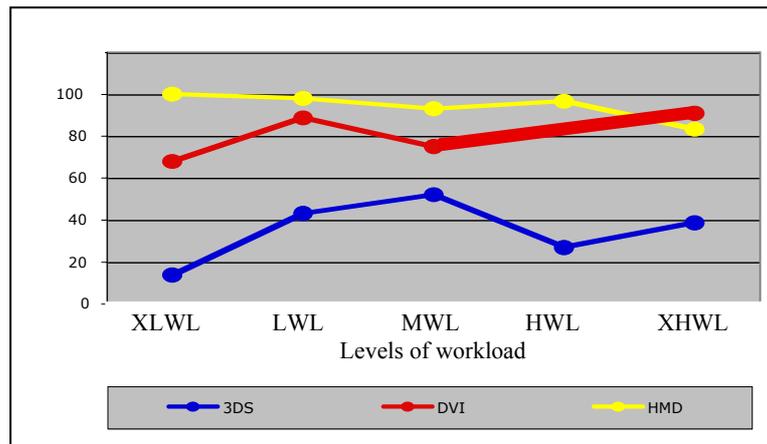
Figure 4: evolution of HMI suite activity performance with regards to workload levels.

When focusing on activity performances, though, there is a decrease in performance with the increase of workload ($F^1_{657}=5.04$; $p=0.025$), but only when Xtra Low, & Low levels are grouped as well as High & Xtra High levels. When analyzing HMI suite related activities with regards to workload (figure 4), it appears that 3DS and DVI performance vary differently. 3DS (in blue) rises from a very low performance to an approximate 50% at medium workload level and then recedes, while DVI (in red) seems to have an improvement of its performance with workload. Note the absence of DVI commands at HW Level.

Because performance aspects do not seem all that descriptive, pilots use over workload levels needs to be looked into. To examine this, activity counts were made on DVI use by pilots with regards to workload.

4.5   DVI and 3DS use with regards to workload

It seems that pilot use of DVI varies according to increased workload (figure 5). As a surprise, at higher levels, all pilots but one, use DVI. The reconnaissance level was good but not 100% and usually pilots, at the highest workload levels, are not inclined to use non-fully reliable systems.
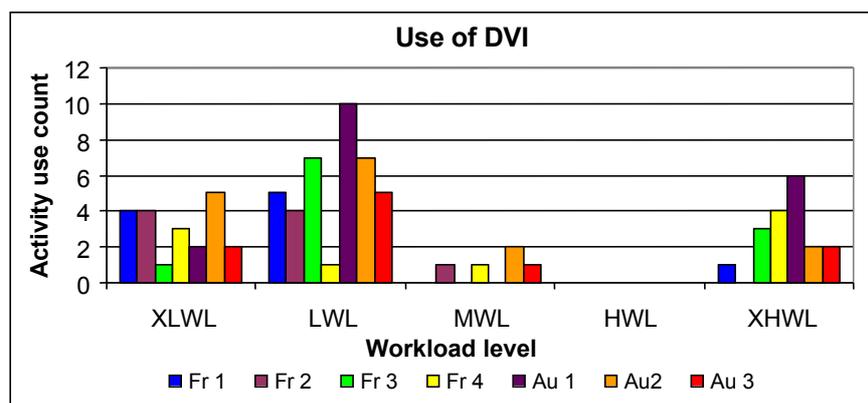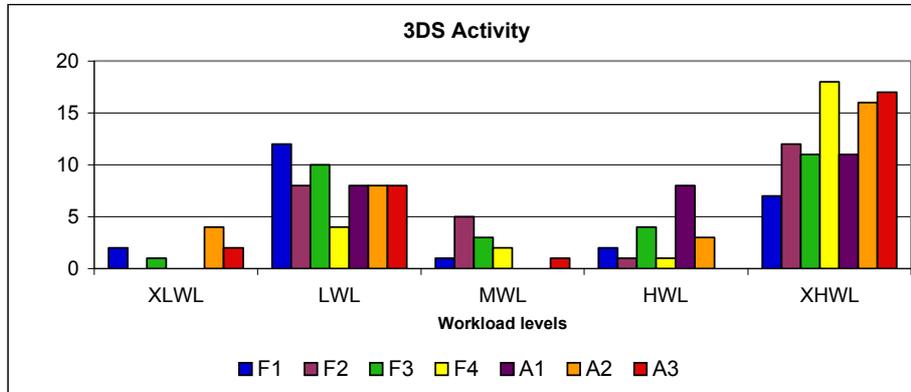

Figure 5: use of DVI with regards to pilots and workload level (F = French pilots; A = Australian pilots).

The surprising lack of use of DVI at the high workload (HWL) level can be explained three ways:
- First, pilots cannot, for some reason, delegate enough resources at that specific level for DVI use. It is a possibility, but then why would they be able to do it at the highest level?
- Second, a collection bias for events including DVI. The evaluator missing events with DVI at the high workload level. But then, this would not have happened for *all* of the pilots and over *all* runs.

- Third, and it is the explanation that is favored, the so called high level work load phase is always likely to be very short because workload increase is very abrupt and can jump directly from Xtra Low, Low & Medium to Xtra High. For instance, pilots consider enemy engagement as Xtra High workload, the build up phase before that is very likely to be short and not suitable for DVI commands.

Figure 6: 3DS activity count with regards to pilots and workload level.

3DS use needs to be associated to DVI use, because, though it is imposed on pilots most of the time, they could also use it as a request. For instance, the "wingman" 3DS was on demand via DVI.

Figure 6 shows the total of 3DS events per pilots with regards to workload. It reveals two peaks of exposure, first at low WL figuring the 3DS from friendly parties and second at the exposure to enemy 3DS for the highest level of WL.
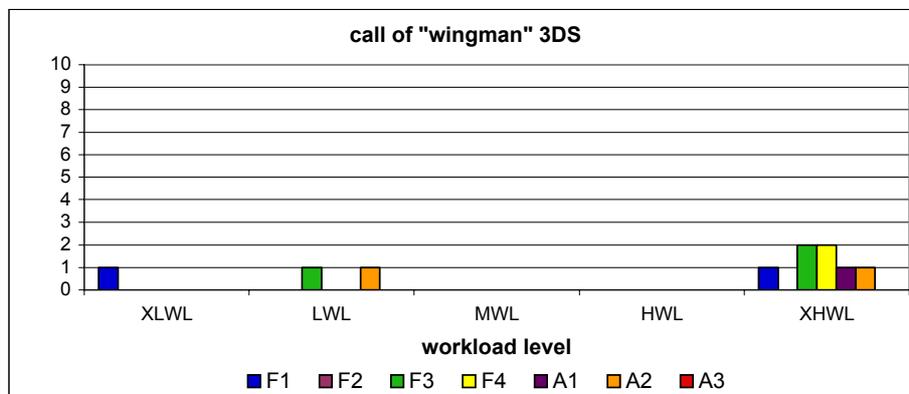
Figure 7: number of DVI "call wingman" the pilots used to localize wingman with 3DS.

3DS contribution to performance remains low whatever the workload, signifying low mastery (<45% success) besides low involvement due to poor expected gain. Such a low effectiveness needs to be commented; the only pilot, who localized well, used it with great advantage for his spatial awareness. However, the majority of the pilots were only able to localize moderately to poorly and 3DS performed for them accordingly. The reasons for this are unclear, as a number of factors appear to be involved. For example, some compromises needed to be made in the implementation of the 3DS in the simulator that impoverished the 3D stimulus. However this was the case for all pilots and hence cannot account for the variability witnessed. Possible factors are multiple but can be grouped with regards to technology, sound physics and human psychophysiology

On the other hand, DVI comes as a surprise as its performance (both activity and event) increases with workload; most pilots use it at the highest level. In fact, the increase of performance with workload happens for pilots that know they are efficient with DVI. Figure 7 shows the on demand ("wingman") 3DS and that they concentrate at the highest level.

Anecdotal reports suggested that it would be impossible to use it at higher workload levels. However pilots used it both for routine commands at low workload, and cost effective commands at higher workloads. Implications are that such strategies of use should be taken into account for training. At higher WL it gave a clear advantage to pilots in selecting functions immediately and directly with no regards to its hierarchical status.

5.0 CONCLUSION

In realistic simulated air combats, operational pilots had a relative easy appropriation of technology they had never used before. Their strategy of use was workload dependant, yet still efficient at the highest workload levels especially for DVI. Throughout missions, pilots displayed an appropriate cognitive resource management strategy mostly based on an effective "trade-off-guided" resource allocation of inputs.
Interestingly, new and resource hungry interface systems are not especially discarded at highest levels of workload when pilots think their use brings a fair trade-off. Such high level resource management strategy should be implemented in novel HMI design.

[1]   Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63-102). New York, NY: Academic Press.

[2]   Flanagan, J. C. (1954). The critical incident technique. *Psychological bulletin., 51*(4), 327-359.

[3]   Dix, A., Finlay, J., Abowd, G., & Beale, R. (1998). *Human-Computer Interaction, second edition*: Prentice Hall.