Proceedings of the

TENTH INTERNATIONAL SYMPOSIUM ON AVIATION PSYCHOLOGY

May 3 – 6, 1999

Sponsored by

The Ohio State University
Department of Aviation
Aviation Psychology Laboratory
Columbus, OH 43210

The International Journal of Aviation Psychology

The Association of Aviation Psychologists

Edited by

Dr. Richard S. Jensen
Brian Cox
Dr. Joseph D. Callister
Robyn Lavis

Dr. Richard S. Jensen – *General Chair*
Dr. Joseph D. Callister – *Program Chair*
Neysa Huber, Angela Weir & Robyn Lavis – *Technical Chair*

VOLUME 1

Proceedings of the

TENTH INTERNATIONAL SYMPOSIUM ON AVIATION PSYCHOLOGY

May 3 – 6, 1999
Columbus, Ohio USA

Sponsored By:

The Ohio State University Department of Aerospace Engineering, Applied Mechanics and Aviation

The International Journal of Aviation Psychology

and

The Association of Aviation Psychologists

Edited by:

Dr. Richard Jensen
Brian Cox
Dr. Joseph Callister
Robyn Lavis

Dr. Richard Jensen, *General Chair*

Dr. Joseph Callister, *Program Chair*

Neysa Huber, Angela Weir and Robyn Lavis, *Technical Chair*

DCIEM SL 1999-054

DIRECT VOICE VERSUS MANUAL INPUT TECHNOLOGY FOR AIRCRAFT CDU OPERATIONS

Philip S. E. Farrell, Brad Cain, Paul Sarkar, and Ronald Ebbers[1]
Defence and Civil Institute of Environmental Medicine
Toronto, Canada

[1]Bundesamt für Wehrtechnik und Beschaffung
Munich, Germany

ABSTRACT

Direct Voice Input (DVI) is often regarded as a replacement technology for keyboard entry activities in the cockpit. DVI may reduce workload and increase performance by exploiting verbal psychomotor resources. For this study it was assumed that DVI and manual input (MI) operations might interfere with other flying activities because these tasks require some combination of four limited resources (visual, auditory, cognitive, and psychomotor; VACP). The experiment was performed in a helicopter simulation environment. A mission scenario was developed wherein Canadian Forces Griffon helicopter pilots were asked to set radio frequencies (or radio links) using a Control Display Unit (CDU) either with MI or DVI, and to perform flying tasks that manipulated the demand on the individual VACP resources. The radio link completion times and errors were recorded along with flight, subjective workload, and video data. An analysis of variance was performed on a small subset of data, and the results suggested that DVI was superior to manual input during the psychomotor and visual phases due to the increased *heads up* time afforded by DVI (as reported in the literature). However, high psychomotor activity reduced the frequency of verbalization. Interestingly, MI was superior to DVI during the cognitive phase. Unexpectedly, performance did not differ between MI and DVI during the audio phase. The study showed that DVI did not outperform MI under all conditions. The implication of this study is that DVI should be utilized to augment rather than replace manual input activities in the cockpit.

INTRODUCTION

The CH146 Griffon helicopter is now part of the Canadian Forces fleet. The helicopter is equipped with a flight management system and the Control Display Unit (CDU) is the system's interface. The CDU allows the crew to manipulate and monitor various aircraft systems such as communication and navigational settings. The CDU has been nick-named "the head magnet" because it draws and holds visual attention. Most of the CDU entry is done before departure, or by the pilot not flying while in the air. During an in-flight re-tasking, the Pilot Flying often elects to land the aircraft before new settings are typed into the CDU for reasons of safety.

Direct Voice Input (DVI) has been suggested as an alternative to Manual Input (MI) so pilots can maintain a *heads up* posture while still entering data. However, there may be certain circumstances where DVI breaks down. Anecdotally, speech begins to trail off when a driver merges onto a highway. Talking to the passenger is distracting and becomes a casualty during this phase of high load. Ironically, DVI is targeted to reduce workload during periods of high activity in the helicopter.

A literature review was performed to determine the extent of the DVI research (CMC, 1999). The technical issues include recognition rates under various conditions like male/female or multiple accent recognition in noisy or G environments, DVI recognition rate improvements with lip motion detector, and word versus intent recognition rates. A few studies looked at human factors issues such as the application of DVI to a battlefield helicopter, control of flare deployment using DVI, feedback for DVI, and DVI and workload in a laboratory setting.

Also mentioned in the literature review is DVI syntax for helicopter operations, which we are currently investigating. The research is aimed to determine whether improvements to syntax alone will significantly reduce workload, as well as to discover the efficacy of new interface design techniques. Good syntax should allow the user to navigate the system naturally and smoothly. However, a CDU's manual interface (i.e., its *syntax*) may be improved by applying the design techniques of Ecological Interface Design (Vicente & Rasmussen, 1996) or Perceptual Control Theory Task Analysis (Farrell & Chéry, 1998), and be made just as effective as DVI for a single task scenario.

The studies mentioned above essentially look at DVI in a single task environment. However, a helicopter (and most other applications) presents a multi-task regime, and so there is potential for the flying task to interfere with the task of operating the DVI system. This experiment attempts to determine the types of flying tasks that might interfere with DVI, and determine whether MI may be a better input mode during those instances.

The study's objectives are to determine the conditions under which DVI might do better than MI, and the degree of DVI performance degradation under increasing loading conditions. DVI and MI are compared under four conditions that load the subject in the Visual, Audio, Cognitive, and Psychomotor (VACP) domains. An attempt was made to manipulate the domains independently. Within each condition the load was varied from low to medium to high.

This paper reports on the experimental design, proposed analysis, sample results, and implications of the study. It is hoped that the results will provide insight for making recommendations for the tasks and mission segments best suited for DVI, as well as the procedures for operating a DVI system within the helicopter.

EXPERIMENTAL DESIGN

The general purpose of the experiment is to explore the human factors issues related to the application of speech recognition technology in military aircraft. A very gross description of a helicopter mission may consist of four basic tasks: conduct of flight, spatial orientation, calculation, and communication. VACP resources can be identified for the four basic tasks as well as the CDU task. The general hypothesis is that at some high load level, the operator will reach a VACP resource limit or processing capacity limit and the flying and CDU tasks will interfere with each other. A performance decrement will follow either with the flying task, the CDU task, or both depending on the subjects' strategies. Therefore matrices of hypotheses may be identified as shown in Figure 1. The degree of interference might be determined by examining the performance decrement as a function of load level.

The matrix entries are ideal. The number "1" represents complete interference and the tasks are done serially with respect to that resource. A "0" means no sharing of resources and both tasks can be performed in parallel. For example, the V-V entry is 1 for the CDU MI task because the pilot can not look down for key entry and Out-The-Window (OTW) for the flying task at the same time, so the hypothesis is complete interference. The V-V entry is 0 for the CDU DVI task because the pilot may maintain his or her gaze OTW while operating the CDU. Thus, the hypothesis is no interference.

		CDU MI Task				CDU DVI Task			
		V	A	C	P_{hand}	V	A	C	P_{voice}
Flying Task	V	1	0	1	1	0	1	1	0
	A	0	0	1	0	0	1	1	0
	C	0	0	1	0	0	1	1	0
	P	1	0	1	1	0	1	1	0

Figure 1. Matrices of Hypotheses

The VACP resources were loaded by manipulating the flying task only. The visual condition involved a dense fog while tracking an undulating road (random changes of direction by 30° every nautical mile, see Figure 2). In the audio condition, numbers were presented aurally at a constant rate from which subjects needed to respond to a target number by pressing a button on the cyclic. In the cognitive condition subjects were asked to memorize a string of numbers and letters (similar to a call sign) while performing the flying and CDU tasks, and then recall the call sign. The psychomotor condition involved random cross-track wind gusts causing the aircraft to drift from a straight road. At the end of the experiment, subjects performed a fifth condition where they were loaded in all domains while performing a Search and Rescue task. The condition was not included in the analysis due to the task complexity.

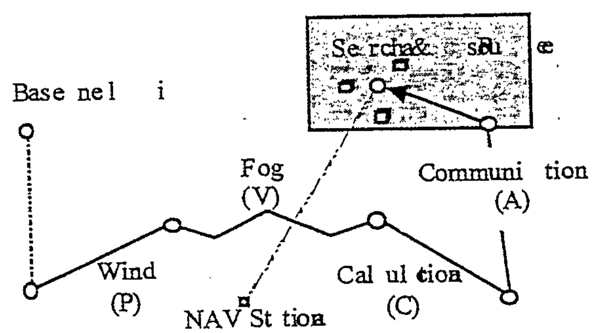

Figure 2. Plan View of Experimental Scenario

Subjects were asked to set a radio link as quickly and as accurately as possible for the CDU task, and to maintain 200 ft and 90 knots for the flying task. Baseline data was collected separately for both tasks at the beginning and end of the experiment. Also, flying task baseline data was collected at the beginning of each load condition.

Three load levels (low, medium, and high) used during the wind, fog, calculation, and communication conditions. The visual load was manipulated by increasing the fog density. The audio load was changed as the presentation rate was set at a higher value. The cognitive load was manipulated by providing different string lengths. Random wind gusts of up to 10, 20, and 30 knots were used for the three psychomotor loads.

Subjects performed the low-load condition during which they set a radio link upon hearing a start tone. Six tones per level were presented separated by 40 seconds for a total of six frequency settings per load level. A NASA TLX type questionnaire was administered after the load level. Subjects repeated the protocol for the medium and high-load conditions. The total number of links is 240 ((4 + 2 baseline) conditions × 2 input modes × 3 levels × 6 links (only 5 in the baseline conditions)). Video data is collected from an over-the-shoulder view, a view looking into the subject's face, and a bird's eye view of the CDU.

Ten male subjects performed the experiment. All subjects were CH146 Griffon helicopter pilots with a range of hours on type from 500 to 3000. Each subject was briefed on the experimental procedure and was given time to practice with the systems. The input mode was counter-balanced amongst subjects where five subjects performed the experiment with M1 (2.5 hours) and then with DVI (2.5 hours). Subjects were given an hour break between the sessions and short breaks at their own request.

PROPOSED ANALYSIS

The intent of the analysis is to generate statistical descriptions that show the effect of DVI and MI on flying and CDU performance variables. For example, if the mean values of cross-track error and link completion time during the wind condition and DVI manipulation are significantly lower than their MI counterpart, then one might conclude that DVI performs better than MI under these conditions.

It will be difficult to make such straightforward conclusions due to the complex nature and interaction of this experimental design. The independent or manipulated variables are load condition (VACP and baseline), input mode (MI and DVI), load level (low, medium, and high), and number of links (one through six). There are other independent variables that are not included initially in the analysis such as the subject's flying experience. The recorded data included key presses, utterances, flight parameters, and video. These data were translated into the dependent variables for radio links (completion times and error rates) and flight performance (altitude deviation, speed deviation, cross-track error, and heading error). Other dependent variables that are not considered initially include the ratio of *head up* versus *head down* time retrieved from the video data analysis as well as subjective workload.

The general hypothesis assumes that several tasks are competing for the same resource, a Multiple Analysis of Variance (MANOVA) is the suggested statistical test. The MANOVA should indicate the significance of the independent variables (and their combinations) in affecting the dependent variables.

However, as a prelude to the MANOVA, an Analysis of Variance (ANOVA) is performed on each condition separately so that any trends in the data may be identified, making the interpretation of the results more tractable. For this paper, a single flight variable and a single link variable are selected that most likely draw on a single resource more than other measures for that condition. This will provide a partial set of results from which general comments can be made. The full analysis and results will be available as a DCIEM report (Cain et al., forthcoming).

The completion time (CT) is the chosen link variable for each of the four conditions. CT is defined as the interval between the start tone and the "enter" key press indicating the end of a setting. The flight variables for the visual, audio, cognitive and psychomotor conditions are the root mean square (rms) values for altitude deviation, heading deviation, heading deviation, and cross-track error, respectively.

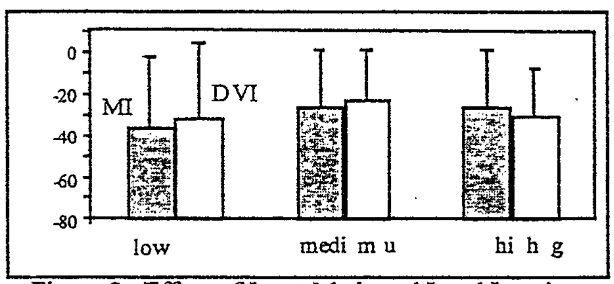

Figure 3. Effect of Input Mode and Load Level on Altitude Deviation during the Visual Condition.

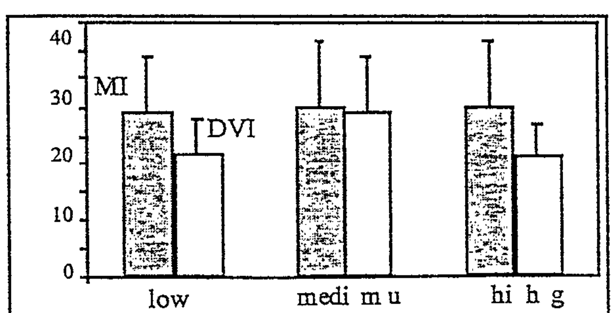

Figure 4. Effect of Input Mode and Load Level on Completion Time during the Visual Condition.

RESULTS

An ANOVA was performed independently for the flight variable and CT. The factors for the ANOVA were input mode (DVI and MI), load level (low, medium, and high), and link number (six plus one baseline). One subject did not complete the experiment due to his susceptibility to simulator sickness. His data were left out of the analysis.

Figure 3 is a plot of the mean value and standard deviation of the altitude deviation for MI and DVI at the three fog density levels during the visual condition. The ANOVA results show that there is no effect of input mode or load level on altitude deviation. Figure 4 is a similar plot for the variable, CT. It shows that DVI is significantly different (lower) than MI ($p = 0.0065$). However, there is an unexpected significant interaction between input mode and load level ($p = 0.0092$). That is, the DVI CT is clearly lower the MI CT in the low and high load levels but not for the medium level.

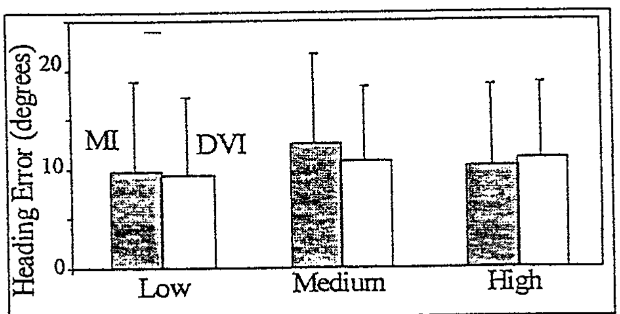

Figure 5. Effect of Input Mode and Load Level on Heading Deviation during the Audio Condition.

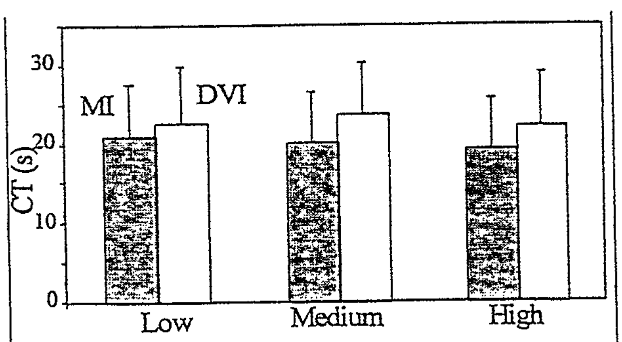

Figure 6. Effect of Input Mode and Load Level on Completion Time during the Audio Condition.

The results for the audio condition show that none of the factors or their interactions were significant for heading deviation as indicated in Figure 5. The same is true for CT as shown in Figure 6. Note that the MI CT mean is constantly lower than the DVI CT mean at each load level.

The cognitive condition results show that subjects maintained their heading an average of 1.5 degrees from the commanded value with no significant difference between input modes as shown in Figure 7. However, CT increases with load level significantly ($p < 0.0001$), while the MI CT is lower than DVI CT as shown in Figure 8.

The psychomotor condition also has expected and unexpected results. The results show that the factors have a significant effect on the cross-track error with the p-values being 0.0732 for the effect of the input mode, 0.0007 for the load level, and 0.0005 for link number. While it is expected that performance will DVI will differ from, the link number is significant for a different reason. The data shows the x-track error diminishing as the link number increases, and a downward trend exists as shown in Figure 9. In contrast, the CT results show no trend with link number as shown in Figure 10.

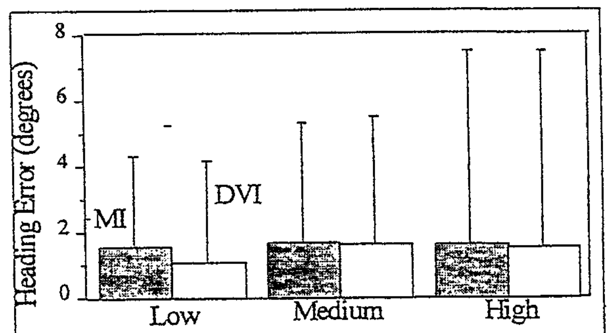

Figure 7. Effect of Input Mode and Load Level on Heading Deviation during the Cognitive Condition.

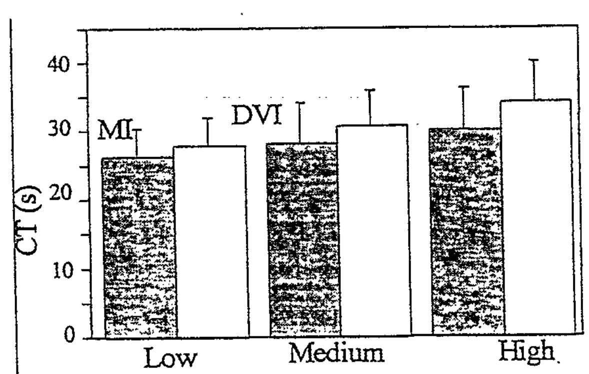

Figure 8. Effect of Input Mode and Load Level on Completion Time during the Cognitive Condition.

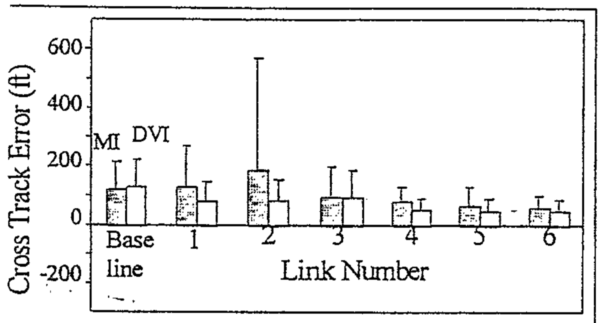

Figure 9. Effect of Input Mode and Link Number on Track Deviation during the Psychomotor Condition.

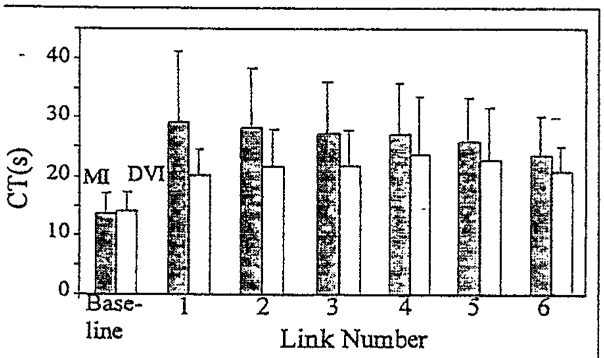

Figure 10. Effect of Input Mode and Link Number on Completion Time during the Psychomotor Condition.

DISCUSSION

The multi-task environment provided a realistic scenario for examining the impact of Direct Voice Input on helicopter operations. However, this environment yielded many degrees of freedom for the analysis. In addition, there were some variables that were not controlled for such as strategies. Thus, care must be taken when interpreting the results.

For instance, in the high-load wind condition, the mean value for the cross-track error was 170±370 ft during the MI manipulation, while all other conditions are in the order of 75±75 ft. Clearly, a subject strayed from the road for a long period of time during the manual input and link no. 2 condition, thus skewing the data.

Also, subjects chose from one of many strategies making it difficult to interpret the results. It is assumed that subjects considered the flying task as the primary task and the CDU task as secondary. However, in the audio and cognitive conditions where the flying task is relatively easy, subjects might have had a tendency to pay more attention to the CDU task rather than, say, holding the altitude exactly at 200 ft.

If a subject decided that the primary task was of utmost importance, then one would expect to see little change in the flight parameter and an increase in completion time as the load level increased. This trend was evident in the psychomotor condition where the cross track deviation was statistically the same while the completion time rose steadily as the load levels increased.

Furthermore, MI CT was consistently higher than DVI CT. Given this limited set of results, one might conclude that DVI is better than MI during high psychomotor activity. The results tended toward the hypothesis given in Figure 1. That is, the psychomotor elements of the flying and MI CDU tasks interfered with each other while the flying and DVI CDU tasks will not interfere. It is clear from the data that there is some interference in the DVI task, otherwise the DVI completion time would be the same across all three-load levels. However, it is not clear presently how to calculate a number that would represent the amount of interference.

If a subject decides that the CDU task is more important, then one would expect to see a lot of effort in performing the task as fast and as accurately as possible at the decrement of the flying task. The audio condition shows no significant difference for the CT (as expected) as well as the heading (unexpected) across load level. Of course, there are other variables to consider, and initial results showed a trend of increasing tertiary task errors (recall that the tertiary task was to detect the target number out of a stream of numbers). Thus, none of the dependent variables can be looked at in isolation.

Note that MI CT mean value is consistently lower than DVI CT mean value during the audio condition. This effect might be comparable to the highway merger effect mentioned in the introduction. One might conclude that MI might be better than DVI in high audio load conditions.

This result is consistent with the audio hypothesis in Figure 1. Moreover, the mean values for MI CT across loads are statistically the same and so the CDU task seems to be de-coupled from the flying task. However, there is still some interference because the CT value is about 5 seconds longer than the baseline condition where the CDU is operated by itself. The additional time is attributed to maintaining a constant heading, altitude, and speed, and not changes in the tertiary task.

It is difficult to draw any conclusions from the visual condition data because the trends are not as expected and the data have the largest standard deviations of all conditions. The ANOVA showed that DVI may have a slight advantage (not significant) over MI in visually degraded conditions. One of the problems in making such an observation is that pilots would automatically revert to instrument flying or land immediately in such conditions and so the DVI/MI comparison may be not-applicable under this condition.

The only expected trend was that the altitude deviation is skewed negatively because subjects have a tendency to mover closer to the ground so they can see the road. Also, the completion times are longer in general in the visual condition than other conditions. During the experiment, subjects would begin the link only on the straight part of the road. They would stop in between a link or ignore the tone altogether to navigate a corner, and then complete the link once on the straight. One interesting analysis for the visual condition that is yet to be performed is the *heads down* versus *heads up* time.

The cognitive condition results are the easiest to interpret and they challenge the hypothesis in Figure 1. It is clear that the heading does not suffer as the load increases but the completion time does increase. Surprisingly, the MI performance is better than DVI in this condition. One might conclude that cognitive tasks may interfere more with DVI than MI.

The ANOVA results have provided strong indications of where to look for similarities and differences in the data. It has shown that sharing of resources amongst the tasks is not always as expected particularly in the audio and cognitive domains. The results have also reinforced the need to look at all the data rather than perform individual analysis and try to piece them together.

CONCLUSIONS

A multi-task experiment was performed to determine the efficacy of DVI compared to MI under various loading conditions. A matrix of hypotheses was generated and the experiment was designed to verify these hypotheses. Although a MANOVA is required to perform a full analysis, a pre-analysis was performed with multiple ANOVAs that yielded data trends. It is clear that DVI is not always superior, particularly in the audio and cognitive conditions. Any recommendations would be that DVI should not replace MI but complement it whenever necessary.

Further research must be done on the procedural aspects of a DVI and MI system. Given two pilots and two input modes, there is a great potential for mode confusion. At the very least, the switching from DVI to MI and back should be transparent.

Part of the integration of DVI into the cockpit will be the issue of direct voice output (DVO) or other forms of feedback (i.e., a visual display). The pilot must not only receive feedback to see if the system understood the commands, but also feedback about what the system did with the commands.

REFERENCES

Canadian Marconi Company (1999). DVI Literature Summary. Draft report prepared by L. L. Churchill.

Vicente, K.J. ,& Rasmussen, J. (1992). Ecological interface design: theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-22, 589-606.

Farrell, P.S.E. & Chéry, S. (1998). PTA: Perceptual Control Theory based task analysis. Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting. October 5-9, 1998: Chicago, IL, pp. 1314-1318.

Cain, B., Farrell, P.S.E., Sarkar, P. & Ebbers, R. (forthcoming). DVI Experiment Report. DCIEM Report No. 99-R-XX. Defence and Civil Institute of Environmental Medicine, Toronto, Canada.

#513092