Image Cover Sheet

CLASSIFICATION UNCLASSIFIED	**SYSTEM NUMBER** 516710

TITLE

Feeback requirements for a direct voice input system

System Number:

Patron Number:

Requester:

Notes:

DSIS Use only:

Deliver to: DK

This page is left blank

This page is left blank

DCIEM SL 2001-155

Feedback Requirements for a Direct Voice Input System

Philip S E Farrell, Mike Perlin, and Ghee W. Ho

Defence and Civil Institute of Environmental Medicine, 1133 Sheppard Avenue West, Toronto, Canada, M3M 3B9

ABSTRACT

Direct Voice Input (DVI) has been identified as an alternative control method for computer-based systems within military aircraft. Past studies revealed that DVI affords heads up time, which positively influences safety and situational awareness. DVI also requires new crew coordination procedures that effectively redistribute workload. In this study, DVI feedback requirements were investigated for helicopter operations Perceptual Control Theory was used to determine information requirements, and an experiment was performed to determine the effectiveness of various feedback forms. Data trends show that the heads up and heads down visual displays perform similarly, and both are slightly better than audio feedback in terms of the speed and accuracy of assessing the feedback.

1 INTRODUCTION

Advances in cockpit technologies will continue to impact operations within military aircraft and affect procurement strategies. Moving map displays, helmet mounted symbology, and flight management systems are three examples of technologies that may require alternative interfaces in order to ensure safe and effective human-computer interaction. Direct Voice Input (DVI) is being considered as an alternate control interface in the Canadian Forces Griffon Helicopter cockpit (tactical and transport missions) to allow pilots to interact with new aircraft systems while keeping their hands on the cyclic and collective.

Voice recognition software converts sound patterns into a form that the computer recognizes Kobierski and Swail (1997) resolved one technical challenge and reported 99% recognition rates in the noisy Griffon Helicopter. Integrating DVI in a busy workspace generates human factors challenges as well, such as grammar development, multi-task interference, attention demands, and crew coordination procedures, which have been addressed in several studies (Farrell et al., 2001).

This paper investigates the form and content of feedback for operating the communication functions of the Control Display Unit (CDU) with DVI in the Griffon multi-task environment. The CDU provides an interface to the navigation and communication systems in the Griffon Helicopter where radio frequencies and way-point details are entered. The CDU has a 5x5" CRT display with soft keys and a customized keyboard Information is organized in a nested menu structure. This nested menu structure and the CDU location in the cockpit (aft of the center console) requires heads down time to read and enter data, causing attention to be shifted from outside to inside the cockpit. DVI may possibly reduce workload by minimizing head and eye movements, and increase situational awareness by maintaining the point of regard out of the window.

Feedback information requirements for the proposed DVI system were generated using a Perceptual Control Theory (PCT) framework of human-computer interaction. The effects of type of feedback were explored experimentally using Signal Detection Theory (SDT). If DVI were the only task, then the optimal form for the feedback could be predicted from what is known about how interface modalities interfere with each other (Herdman and Beckett, 1996), however, the DVI load is added to existing sensory, cognitive, and psychomotor loads in the helicopter environment.

2. FEEDBACK INFORMATION REQUIREMENTS

A PCT approach (Farrell & Chéry, 1998) was used to derive the potential feedback information requirements. In this case, the approach begins with a Griffon Helicopter composite scenario A section of the scenario involves communication with an outside agent using the aircraft radios The pilot sets up the communication link by choosing the proper radio and frequency.

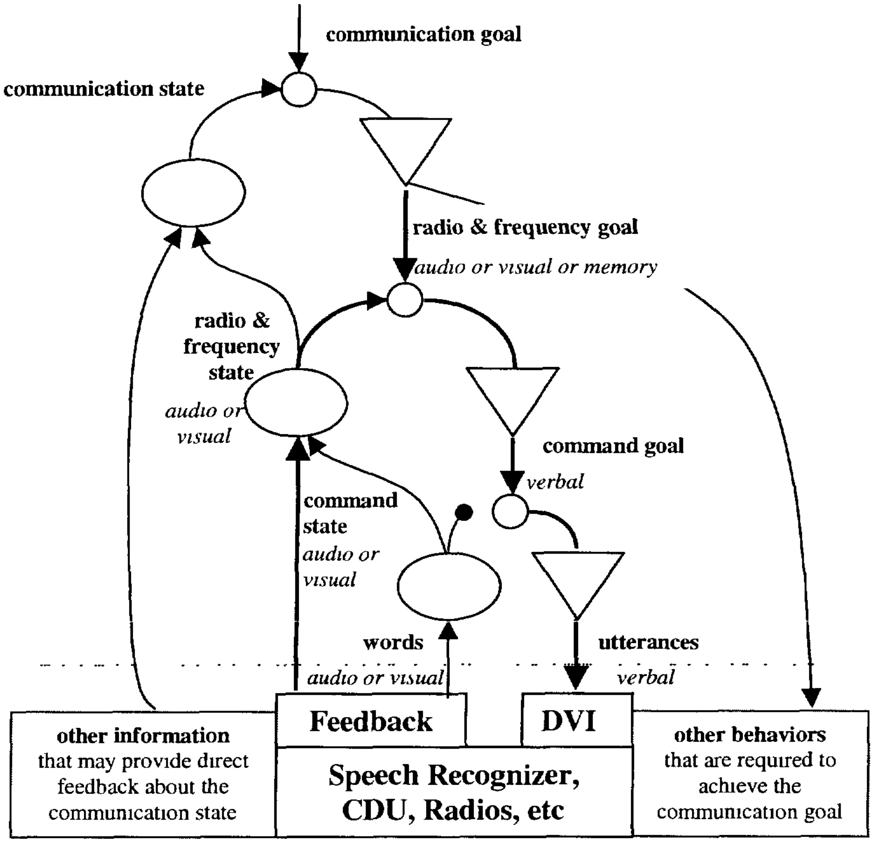

Figure 1. A PCT model of the proposed DVI system showing the feedback information requirements

Next, a goal decomposition of the scenario section is performed as illustrated in Figure 1. The highest level goal is to be in a desired communication state One of many sub-tasks or behaviors needed to achieve the goal is to set the desired radio and frequency. Behaviors become sub-goals at the next level down in the PCT paradigm. Therefore, the desired radio and frequency becomes a sub-goal that is further decomposed into the command goal and utterance behavior. The world then changes state and sensory information is generated about the words, command, and radio and frequency, and this feedback is used to generate internal perceptions about the command, radio and frequency, and communication states, respectively. Each perception is compared to its goal, thus closing the control loop.

The PCT model provides design guidance. For example, all abstraction levels may be controlled when manual input is integrated with DVI, which may be necessary as the vocabulary increases. Also, the model suggests that explicit feedback for all levels will reduce the cognitive integration loads. The interface designer must decide whether to display each aspect of the feedback information explicitly, or leave it up to the human to integrate the information and generate the higher level perceptions. A design decision was made to provide feedback for the command state only for the following experiment. No direct control over the commanded words was possible (indicated by the broken line), unlike manual input with delete and insert capability. Thus, the entire command was repeated if an utterance error was made or was not recognized. Also, participants did not control for the communication goal, and so no other information was given to help generate a perception about the communication state. The proposed feedback display was based on a DVI Action-Target-Parameter structure (Churchill & Herdman, 2000). For example, if the command was "set radio 1 to 118.5" then the feedback display would be "radio 1 is set to 118.5", which contained information about the command and radio and frequency state.

3. FEEDBACK FORM REQUIREMENTS

3.1 Purpose

The italicized text in Figure 1 indicates control and display modalities of the DVI system. Other aviation, navigation, and communication tasks will also use audio, visual, memory (cognitive), and verbal processes, and potentially interfere with the DVI task. The purpose of the experiment is to determine any performance differences due to interference effects between the DVI feedback forms in a Griffon Helicopter multi-task simulated environment. The results will be used as advice when integrating DVI in a fielded system.

3.2 Experimental Design

In the PCT model, goal achievement is possible only after the feedback information is perceived correctly. The user must pay attention in order to detect any errors in the feedback. If the user is occupied with other tasks then there is the potential to perceive *no error* when one exists (miss), or perceive an *error* when none exists (false alarm) "Signal Detection Theory (SDT) is applicable in any situation in which there are two discrete states of the world (call these signal and noise) that cannot easily be discriminated" (Wickens, 1992). In this case, an *error* and *no error* are the two discrete states of the feedback Since error detection is the variable under investigation, the *error* has been defined as the signal and *no error* as the noise. Table 1 outlines the four possible outcomes using SDT terminology. The performance measures for comparing different feedback forms are the detection accuracy - the user's ability to detect an error - and the detection time.

Participants used DVI to set communication links, and were given correct or erroneous radio and frequency feedback. Probability data were generated for a Receiver Operating Characteristic (ROC) curve which is a plot of the Hit probability, $P_H = f_H/(f_H + f_M)$ (f - frequency), versus the False Alarm probability, $P_{FA} = f_{FA}/(f_{FA} + f_{CR})$ over a range of theoretical biases to the Signal-to-Noise Ratio (SNR). That is, if SNR is zero (no signal) then subjects will tend to respond with "no error" and $(P_H, P_{FA}) \rightarrow (0,0)$. Similarly, if SNR is infinite then $(P_H, P_{FA}) \rightarrow (1,1)$.

Table 1. SDT Terminology for experiment.

Participant's Assessment	**Signal (error)**	**Noise (no error)**
"error"	Hit (H)	False Alarm (FA)
"no error"	Miss (M)	Correct Rejection (CR)

Table 2. Feedback form conditions for each flight and SNR

	Condition 1	**Condition 2**	**Condition 3**	**Condition 4**
PNF	HUD/CDU	VO/CDU	HUD/VO	CDU
PF	VO	HUD	CDU	HUD/VO

The ROC curve shows whether there are feedback form sensitivity (d′) differences. That is, a feedback form will be more sensitive to a correct response if its curve is closer the upper left corner of the ROC curve (i e., high P_H and low P_{FA}) than another feedback form's curve. In addition to the SDT analysis, the detection time is used to determine performance differences. The detection time is defined as the time difference between the onset of the feedback and the participant's response.

The experimental design involved four manipulations: feedback form, SNR, crew position, and mission segment. Combinations of three feedback forms were investigated. a visual heads up display (HUD), an audio voice output (VO), and a heads down display (CDU). The CDU was included since it is likely to remain in the cockpit regardless of the study's conclusions. The SNR values of 1:3, 1:1, and 3:1 were chosen so that empirical data might be compared to the theoretical ROC curve. Since the Pilot Flying (PF) and the Pilot Not Flying (PNF) perform different tasks, the crew position may affect the feedback form performance. Similarly, different tasks were executed in different flight phases, and so hover and cruise were chosen for the mission segment manipulation.

The experimental conditions are given in Table 2. Note that the CDU appears once per condition due to computer processing limitations. The PNF was given the CDU whenever possible, which reflects operating procedures in the Griffon Helicopter. Participants worked as a crew and completed a total of 96 DVI commands during the hover and cruise segments of each flight. For each of the four display conditions, participants switched crew position and repeated the flight. The crew repeated these eight flights at the three different SNRs for a total of 24 flights.

3.3 Apparatus

The experiment is performed in the Aircraft Crewstation Demonstrator (ACD), which is a fixed based, low–fidelity, rapidly reconfigurable simulator for investigating new cockpit technologies with the human-in-the-loop. The ACD was configured to represent a Griffon Helicopter. Drivers for the DVI system, feedback displays, and desired command displays were written as well as a means for introducing errors and a data collection application. The PF occupied the right seat and PNF occupied the left seat in the ACD. Both positions had virtual instrument panels with a window containing the desired command. The CDU display was in the center floor console position, the HUD was a Liquid Crystal Display mounted on the simulator's glare shield, and VO was heard through the DVI headset.

3.4 Participants and Procedure

Participants were 18 years and older and were recruited from a pool of trained "simulator pilots" (> 10 hours simulator flying time) that included DCIEM employees, students, military personnel, and the general public. The results reported in this paper include data from three of the investigators, since the data from the expected 16 participants had not been fully collected. Participants were briefed on the experiment and individually trained the DVI recognizer with the grammar developed for this experiment. During a flight, a DVI command was displayed on the instrument panel and the participant uttered the command. The feedback forms were displayed, and the participant decided whether or not the feedback was correct by saying "right" or "wrong". Participants were asked to make the assessment as quickly and as accurately as possible. All utterances and displays were recorded and time stamped. The total time to conduct the experiment was 14 hours spread over six days.

4. RESULTS AND DISCUSSION

Eight subjects participated in this initial test, and 3 of them were investigators. Thus, the results presented below are data trends only. Up to sixteen subjects will participate in the full experiment. The ROC curves were plotted, and it was immediately observed that the data points were clustered in the upper left corner. This has two implications. First, the mean values of the calculated biases for each SNR condition ($c_{25} = -.38$, $c_{50} = -.13$, and $c_{75} = .02$) did not differ significantly from $c = 0$ (no bias). Thus, the data from the separate SNR conditions may be grouped together. Second, the mean accuracy for all manipulations was 90% compared to the baseline accuracy of 100%. Regardless of any statistical difference, any display combination would most likely suffice in practice.

Table 3 shows the sensitivity values for each of the conditions, however there are no significant differences. The visual displays performed equally well and better than VO, which moderated the performance. The feedback forms performed better in the hover than during the cruise, and there was little difference between crew positions

Table 3. Mean sensitivity values, d′

PNF			PF		
Hover	**Feedback**	**Cruise**	**Hover**	**Feedback**	**Cruise**
3.37	**HUD/CDU**	2.96	3.04	**HUD**	2.92
3.43	**CDU**	2.96	3.43	**CDU**	2.73
3.22	**VO/HUD**	2.94	3.32	**VO/HUD**	2.81
3.36	**VO/CDU**	2.65	2.94	**VO**	2.54

Table 4. Mean probability ratio, κ

PNF			PF		
Hover	**Feedback**	**Cruise**	**Hover**	**Feedback**	**Cruise**
.86	**HUD/CDU**	.79	.81	**HUD**	.79
.86	**CDU**	.80	.87	**CDU**	.76
.85	**VO/HUD**	.79	.84	**VO/HUD**	.76
.86	**VO/CDU**	.75	.79	**VO**	.69

The data was further analyzed as a single probability space ($P_H+P_M+P_{CR}+P_{FA} = 1$) as opposed to the dual probability space of SDT ($P_H+P_M = 1$ and $P_{CR}+P_{FA} = 1$). The probability ratio, kappa (κ) is defined as $\kappa = (P_o-P_c)/(1-P_c)$ where $P_o = (f_H + f_{CR})/(f_H + f_M + f_{FA} + f_{CR})$ is the observed probability of a correct detection and P_c is the expected chance probability (Reynold, 1977). When κ is zero, the observed probability is equal to chance, or $P_o = P_c$. For example, flipping a coin 100 times is likely to yield $P_o \approx P_c = 0.5$. If the participant predicts (detects) the coin face correctly every time, then $P_o = 1$ and $\kappa = 1$. $\kappa > 0$ means that the probability of a correct detection is greater than chance and must be weighted or influenced, in this case, by the four manipulations. The κ trends in Table 4 are identical to the d′ trends. Moreover, κ provides a normalized number of the observed probability for a given feedback form.

Table 5. Baseline Detection Times (sec)

Feedback	Mean	Median	Std Dev	Skew
HUD	2.00	1.81	0.564	1.08
CDU	2.30	2.00	1.06	1.34
VO	3.65	3.57	0.41	1.02

Table 6. Experimental Median Detection Times (sec)

PNF			PF		
Hover	**Feedback**	**Cruise**	**Hover**	**Feedback**	**Cruise**
1.89	**HUD/CDU**	2.32	1.98	**HUD**	2.22
2.09	**CDU**	2.44	2.42	**CDU**	2.71
2.59	**VO/HUD**	3.01	2.75	**VO/HUD**	3.09
2.90	**VO/CDU**	3.18	3.72	**VO**	3.82

The detection time data was analyzed and the results are shown in Table 5 Baseline data was collected for the three separate feedback forms only since the baseline for a combined feedback form is likely to be the fastest individual time. Note that the data is skewed and so the median value best represents the distribution. The CDU takes up to 0.5 seconds before the feedback begins, and VO takes about 3 seconds to complete the command. When an error was heard the participant could respond as soon as possible, but for a correct response the entire feedback needed to be listened to before responding. The HUD and VO displays required little or no head movement, while the CDU display required some time to look down. These times are included in the detection time results. The detection times from fastest to slowest are HUD, CDU, and VO.

The experimental times in Table 6 were compared to the baseline. All times were greater than baseline as expected under increased load Again, the visual displays were faster than the audio display. VO moderates the visual display times when in combination. PNF and hover were slightly faster than PF and cruise, respectively, in all conditions.

5. CONCLUSIONS

A PCT analysis provided feedback information requirements for the DVI system. An Action-Target-Parameter structure was used to develop the vocabulary and grammar, and integrate the feedback into a single display. This structure made the feedback salient and contributed to the high accuracy probabilities. The experimental trends showed that all feedback forms were 90% accurate on average and about 0.5 seconds slower while multi-tasking. The feedback forms performed better in the hover segment than in the cruise segment with little difference between the PNF and PF performance. The audio display performed consistently poorer than the visual displays although these differences are not statistically significant. Interestingly, the CDU is similar to the HUD, which would not have been predicted at the beginning of the study

For implementation in a Griffon Helicopter, the feedback requirements will then be dictated by the operational impact of DVI. As the number of systems operated by DVI increase, a consistent design of the DVI vocabulary and grammar will become even more critical. It is likely that manual input would be integrated with DVI in order to ensure control at all abstraction levels. Also, explicit information for all levels would be desirable so to reduce the cognitive load of information integration. The sensitivity of the feedback will be degraded as threats, additional communications, poor visibility, and system failures are added to the scenario The differences in feedback form most likely will be pronounced in a full operational environment.

REFERENCES

Churchill, L.L. & Herdman, C (2000). *CH146 Griffon Direct Voice Input Control For The Canadian Forces Utility Tactical Transport Helicopter. Final Project Report.* Prepared for DCIEM under contract no. W8477-8-AC40/003/SV. BSC Document Number 1000-1029, BAE Systems Canada, Kanata, Ontario, Canada. August, 2000.

Farrell, P. S. E. Chéry S. (1998). PAT: Perceptual Control Theory Analysis Technique. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting.* Chicago, Illinois. 5 – 9 October, 1998

Farrell, P S. E., Churchill, L. L., Wellwood, M. & Herdman, C. M. (2001) Integrating Direct Voice Input in a Helicopter Crew Environment *Proceedings of the Eleventh International Symposium on Aviation Psychology* Columbus, Ohio. 5 – 8 March, 2001.

Herdman C. M., & Beckett, L (1996) Code-Specific Processes in Word Naming: Evidence Supporting a Dual-Route Model of Word Recognition. *Journal of Experimental Psychology Human Perception and Performance,* 1996 Vol. 22, No. 5, 1149-1165

Powers, W.T Clark, R. K., & McFarland, R. L. (1960) A general feedback theory of human behavior: Part 1, *Perceptual and Motor* Skills, 1960, 11, 71-88.

Reynolds H.T. (1977). *The analysis of cross-correlations.* Freeions, Collier, Macmillan Publishers. New York, USA.

Swail, C. & Kobierski, R. (1997) Direct Voice Input for Control of an Avionics Management System. In *American Helicopter Society Avionics and Crew Systems Technical Specialists' Meeting.* Philadelphia.

Wickens, C (1992) *Engineering Psychology and Human Performance, Second Edition.* HarperCollins Publishers, New York, NY, 1992.

#516710

CA020059