RAPID PROTOTYPING OF HUMAN INTERFACE TECHNOLOGIES USING SIMULATION

MARIA F. GRABOVAC, DAVID A. CRAVEN and JAMES W. MEEHAN

Defence Science and Technology Organisation, Australia
maria.grabovac@dsto.defence.gov.au

Abstract: This paper reports the use of simulation to perform rapid prototyping of advanced control and display technologies (direct voice input, 3-dimensional sound and helmet mounted display) for fast-jet aircraft. By integrating the new prototypic systems into a research simulator and making rapid changes to the direct voice input, experienced pilots were able to use the systems in simulated flight, gaining insights into, and providing expert comments on their future application. The rapid-prototyping approach was shown to be useful in establishing how the technologies might be employed in future cockpit systems.

keywords: simulation; rapid-prototyping; direct voice input; human interface technologies

Introduction

A research simulator was used to perform rapid prototyping of advanced control and display technologies for future fast-jet aircraft cockpits. Rapid prototyping is a process by which the prototypes are developed by making successive changes to models in a simulation environment. The technique is widely used in the design, construction and even testing of engineered systems, including automobiles and aircraft [Boeing, Caltech 1997], [Hardtke, 2001] and [Lind et al, 2000].

The general purpose of the study was to investigate principles for the interoperability of advanced aircraft control and display systems. The three systems involved in this study were: a voice-recognition system referred to here as direct voice input (DVI); a spatially-encoded auditory signal generator, incorporating 3-dimensional sound (3DS); and a visual helmet-mounted display (HMD). The DVI prototypic voice-recognition system enabled the pilot to use voice commands to select and control aircraft systems. The 3DS was a spatially-encoded sound profiling system that enabled localisation of sound played in the simulator. The visual symbology was projected on the HMD. The monocular display projected symbology in the line of sight of the pilot's right eye.

The objective of the study was aimed primarily at assessing the interoperability of the systems in simulated flight using rapid prototyping. This required integration of the systems into a flight simulator. The prototypic equipment provided for the simulation consisted of individual closed-system technologies. The core of the ensemble was non-modifiable, while the modifiable sections governed the display of the systems including aspects of HMD symbology, 3DS and DVI; new commands could be created or the functionality of existing commands could be changed. The DVI was tightly coupled with both the HMD and 3DS; however, this paper focuses on the DVI system. Voice commands enabled 3DS and HMD displays to be executed on demand, hence the description of the rapid prototyping of DVI will also reveal aspects of its interoperability with the other prototypic systems.

The general objective of an input system operated by the pilot's voice is to enable the pilot to control the aircraft more intuitively. There are putative conceptual advantages to this form of control such as time efficiency, i.e. it can be faster to call up a command than it is to scroll through pages of menus using switchology. The second is the ability to control the aircraft or its subsystems while the pilot's hands (or eyes) are otherwise occupied.

Direct Voice Input

The systems were integrated in the simulator as is shown in Figure 1. Since the purpose of the exercise was to explore the possibilities of the technologies and make modifications in response to expert comments, the systems had to be integrated so that making changes "on the fly" was acceptable. The new systems were a stand-alone avionics set, having a single point of contact with the simulator via a TCP/IP connection over ethernet [Robbie, 2001]. The communications protocol was fixed before the system was flown in the simulator; however this was flexible enough to permit incorporation of new DVI commands as required. On the simulator side, the "HMI Suite Interface" module received commands from the DVI, and fed back information to drive the HMD and 3DS displays. The "HMI Suite Controller" was at the other end of this connection, where all of the processing for the new systems took place. The HMI suite could be taken off line, modified, and re-connected while the simulator was still being flown.

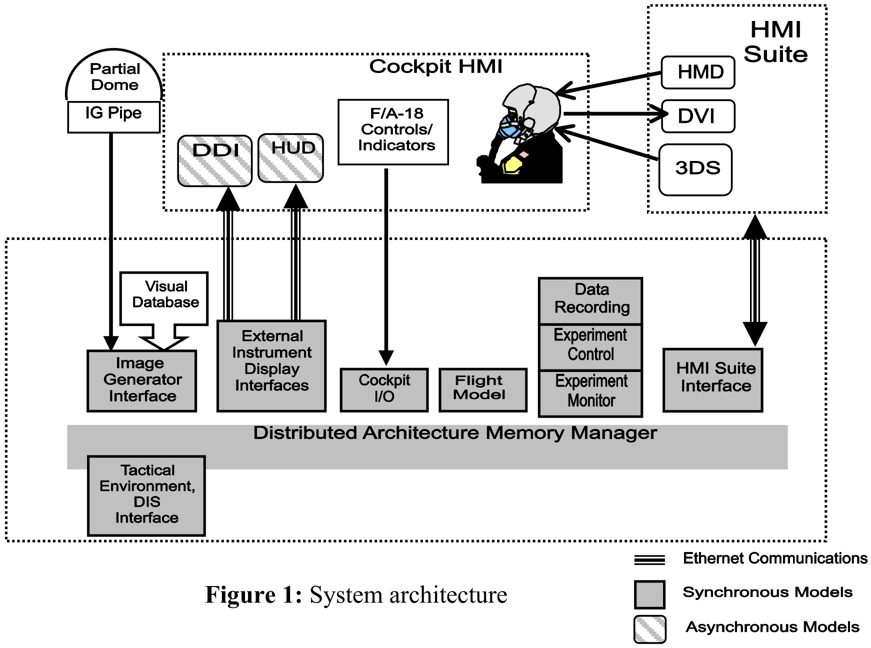

Figure 1: System architecture

DVI was tested in the simulator both as a stand-alone system and as part of the integrated suite in order to verify its functionality and test its compliance with performance specification.

Testing the system in the simulator revealed reduced reliability of the DVI compared with its performance in voice training. Among the problems identified were consistency in switchology, the voice-training environment, and software. A computer joystick was used in training the DVI, whereas a cockpit throttle communications switch was used in the simulation environment. The voice training was conducted at a computer console in a quiet location in the simulator facility, and the oxygen mask incorporating the microphone was held at the pilot's mouth. There was more background noise in the simulation environment, and the oxygen mask was mounted on the helmet, positioning it at a slightly different distance from the mouth. This may have resulted in a marginal change in the acoustical environment. The software was also tested to determine whether it performed better with or without noise-reduction. These potential causes of reduced recognition were tested systematically both in and out of the simulator.

It was noted that the voice-recognition software using noise reduction performed better overall than without noise reduction in both the simulation and training acoustic environments. It also seemed that the use of the cockpit button yielded more consistency in recognition rates than the use of the joystick, but this may have been due to the pilots' familiarity with the cockpit controls. In spite of these changes it was found that the recognition rate still did not reach the reliability level required. This suggested that the execution of the algorithm in the software may have been involved. Recognising the link between the algorithm's pattern recognition phase and the need for a phonetically diverse syntax, the syntax structure was identified as a possible cause of errors. In order to understand the types of errors that were being made, it was necessary to examine the algorithm used in the system.

Three main algorithms used in voice recognition systems are: Dynamic Time Warping (DTW), Hidden Markov Models, and Neural Networks. The algorithm that was used in the DVI system was a speaker-dependent version of DTW. This type of voice recognition algorithm uses pattern-matching techniques to achieve voice recognition, and hence requires a syntax. The system needs to be trained in the syntax with the user's voice to create templates for recognition [Rabiner and Juang, 1993]. It was therefore necessary to assess the syntax, make changes and then examine the effect of the changes on the performance of the DVI system. This approach to "training" DVI fitted in very well with the intended rapid-prototyping approach of the study whereby rapid changes would be made to the systems involved, allowing pilots to assess their interoperability and potential during simulated flight. By enabling pilots to modify the functionality of some DVI commands, they were able to interact with the systems during flight, develop ideas and suggest new functions to implement and test "on the fly".

Pronunciation and Articulation

Other factors that influence a DVI syntax are articulation and pronunciation. Three different accents were encountered during the study: French, New Zealand and Australian. There are differences in Australian and New Zealand pronunciation, particularly in the articulation of vowels, and there are other differences between English spoken by native and non-native speakers. This added to the complexity in establishing a syntax whose commands differed enough linguistically and phonetically that the commands would not be confused.

Example of syntactic development

In navigating an aircraft, the pilot flies in straight lines between predetermined points called waypoints. A waypoint has map co-ordinates that correspond with longitude and latitude. The aim is to turn the aircraft on a waypoint and fly on a new bearing to the next. In existing layouts the pilot is often required to select and update waypoints manually, but doing this by voice command to the flight computer would allow the pilot's hands to remain on the throttle and stick controls. As navigation is not normally a time-critical function, voice commands are potentially well adapted to this function.

A typical command to navigate the aircraft could be "waypoint five". In the DVI system, this call would display appropriate symbology on the HMD, such as an arrow pointing to the location of the waypoint. In the course of a mission, there would be many waypoints and the pilot would have to navigate from one to the next. The syntax for the DVI included navigational commands of the type "waypoint XX" (where XX represented numerals). This command would bring up the symbology for the nominated waypoint. The numerals were entered into the DVI syntax as English text and pronounced by the pilot accordingly ("one", "two", "three", etc.). However, some commands ("waypoint nine" and "waypoint five") were confused by the voice recognition system. These numerals were therefore changed to NATO-compliant phonetic translations, ("wun", "too", "tree", etc.). This change was made to ensure uniformity in pronunciation during training, and again during flight. Hence, the numeral "five" changed to "fife", and "nine" changed to "niner". This transformation increased the number of syllables in "nine" and shortened the syllable length in "five". It was observed that this change enabled the pattern-matching algorithm to use the length of the command during its pattern-matching stage as well as the spectral content of the commands. This increased the rate of recognition for waypoint calls. Pilots were then more confident to experiment with the calls and explore the potential of the systems more fully.

In addition to changes in the syntax, changes were made in display of functions resulting from DVI commands. An example was a command to display a heading tape, pitch, roll and altitude-above-ground in the HMD. Initially the heading tape displayed the heading of the aircraft regardless of the direction of the pilot's head. Pilots suggested that the heading tape could be changed to take account of the direction the pilot was looking. The associated voice command, "call attitude" remained, but the functionality of the HMD symbology was altered accordingly.

DTW is a spectral-content-over-time pattern-recognition algorithm. An inherent problem in such algorithms is the incorrect recognition of words due to differing pronunciation, for example, when under stress, high g, or with a dry mouth. Depending on where emphasis is placed on a word, the spectral content can differ from one utterance to the next. Therefore, if a command were emphasised differently during training than when the command was used in flight, the next closest match may be returned instead of the correct one. The voice-recognition algorithm (speaker-dependent DTW) required the syntax to be trained to the individual's voice prior to recognition. Pilots sometimes found it difficult to remember the exact intonation used in training. In some cases this led to mispronunciation and hence misinterpretation. To improve this, each pilot was asked to use a more "robot-like" manner of speaking when training, and during runs. An associated change made to the syntax involved the addition of the primer "call" to every DVI command. The primer served two purposes: the first was to give the pilot a constant word to use to initiate the command; the second was to lengthen the command so that the pilot could more easily adopt the same rhythm used in recording the training commands. Hence, commands such as "stores" became "call stores", and "wingman" became "call wingman".

Numerous such changes were made to the syntax to improve the performance of the DVI system during rapid prototyping. Figure 2 shows examples of the structure of the syntax before and after the changes.

Initial syntax	Final syntax
wingman	call wingman
waypoint one	waypoint wun
next waypoint	call next waypoint
82X	call 82X-ray

Figure 2: Examples of the syntax before and after changes

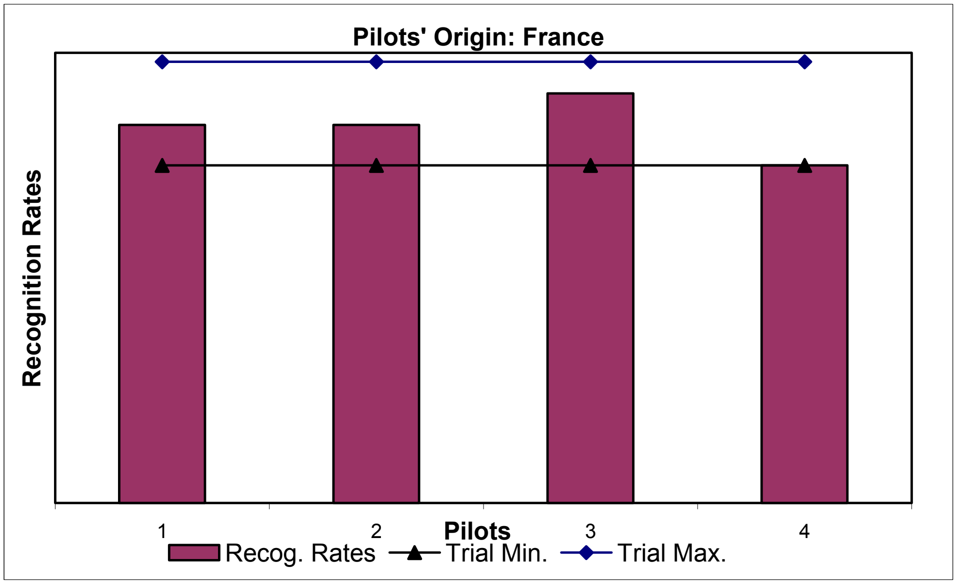

Figure 3: Recognition rates for French pilots, original syntax

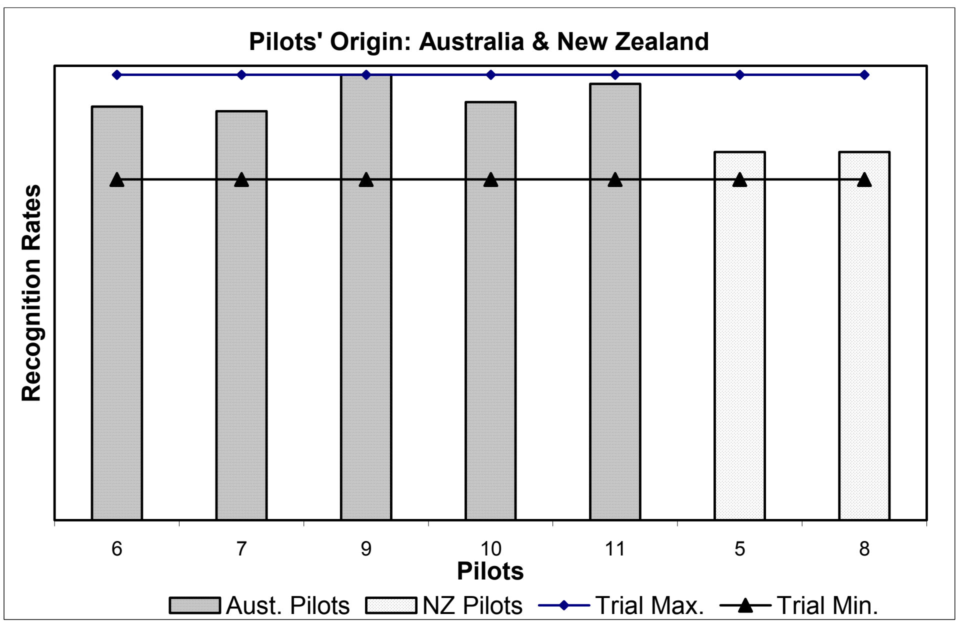

Figure 4: Recognition rates for New Zealand and Australian pilots, revised syntax

Pilot Nationality

The changes to the syntax may have impacted differently on pilots with varying nationalities. Although the study was not designed to investigate this aspect of DVI, some data were gathered that are interesting to compare. French pilots were required to use English throughout, and although fluency varied, all were proficient in English and experienced in the use of English in operational flight. Figure 3 shows recognition rates for the French pilots using the original syntax. The lower than expected recognition rates prompted restructure of the syntax and investigation of the differences between the training and simulation conditions.

Interestingly, there was also a difference in DVI recognition rates between Australian and New Zealand "native" English speakers. Figure 4 shows the recognition rates for the New Zealand pilots after the syntax had been changed. Figure 4 also shows recognition rates for the Australian pilots that were consistently better than those for the

French and New Zealand pilots. Although the causes of these variable patterns in DVI performance are likely to be multi-factorial, it is worth noting that Australian and New Zealand pronunciation of vowels is the most noticeable difference in spoken English in the region. Hence, it is possible that both structure and content of a syntax in speaker-dependent voice recognition algorithms of this type can influence recognition rates.

Conclusion

This paper has described just one element of an extensive and complex prototyping activity. The DVI example presented shows that this was not a straightforward case of engineering, but that it drew on a number of different disciplines to perform a virtual test of physical prototypes, and a test of their potential use in future aircraft control and information display systems.

A critical element in such an endeavour is how the prototypic system is integrated in the test environment. The interplay between the test system characteristics, the test environment and the method of evaluation together impose limitations on what can be done. Hence, an optimum evaluation using simulation should be based on thorough understanding of the limitations of the engineered components and the scientific objectives of the evaluation itself. Failing to take full account of both will limit the ability to fully exploit the potential of this approach.

Simulation is being used much more extensively today in the design, prototyping, and testing of technologically advanced systems, such as those encountered in aeronautics. Simulation is less expensive than testing in the field. It is also safer and can even permit systems to be tested to destruction – which is neither possible nor desirable in the real world.

Acknowledgement

The prototypic DVI, 3DS and HMD systems employed in the work reported here were provided by Thales Avionics, Bordeaux, France, under a Technical Arrangement between the Délégation Générale pour l'Armement of France and the Defence Science and Technology Organisation of Australia.

References

1. Boeing, Caltech 1997, http://www.cds.caltech.edu/conferences/1997/vecs/tutorial/Examples/Cases/777.htm; *the use of dynamic simulations of aircraft flight.*
2. Hardtke F. 2001, "Rapid Prototyping for User-Friendly and Useful Human Machine Interfaces". Proceedings of SIMTECT 2001, Simulation Industry Association of Australia, Canberra, Australia, pp. 239-242.
3. Lind A, Hjorth CG, Hakansson U, Lorenzetto P. 2000, "Evaluation of mock-ups before manufacture of a shield block prototype by powder HIP". [Conference Paper] Elsevier. Fusion Engineering & Design, vol.49-50, Nov. 2000, pp.599-604. Switzerland.
4. Robbie, Andrew. 2001, "Design of an Architecture For Reconfigurable Real-Time Simulation". Proceedings of SIMTECT 2001, Simulation Industry Association of Australia, May 2001, Canberra, Australia, pp. 251-258.
5. Rabiner L, Juang BH. 1993, "Fundamentals of Speech Recognition", Englewood Cliffs, N.J. PTR Prentice Hall, c1993.

Biography

Maria Frances Grabovac obtained a Bachelor of Engineering (Electrical, Hons) and a Bachelor of Arts (French) in 2002 from The University of Melbourne, Australia.

Maria joined the Defence Science and Technology Organisation (DSTO), Air Operations Division as a Simulation Engineer in 2002. Her main areas of interest are head-mounted display symbology and speech recognition in simulated flight. She has worked on 'human-in-the-loop' simulation experiments such as the rapid-prototyping of human-machine interface technologies. Other areas of research include investigation of the effects of visual cues on height perception during simulated flight.

Maria is a Member of the Institute of Engineers Australia and a committee member of the Victorian Branch of the Institute of Electrical Engineers. Currently she is pursuing a Diploma in Modern Languages at Macquarie University.