



VOICE RECOGNITION - WHERE ARE THE END-USERS ?

J.M. Noyes* and C.R. Frankish*

ABSTRACT

After a brief review charting the progress of voice recognition applications over the last 15 years, successful present-day uses of voice input in the UK are described. Although a vast amount of experimental work has been carried out into automatic speech recognition (ASR), there are surprisingly few users who are working in a non-research context. Factors which explain why there are such a small number of speech input users in the UK are outlined.

INTRODUCTION

Although work on automatic speech recognition has been carried out since the 1950's (ref 1), it was not until the early 1970's that such systems became operational. Since then, identification of a wide range of potential uses for voice input has stimulated a vigorous research effort. Despite this, a survey of currently successful applications reveals a relatively small number of end-users, engaged in a limited range of tasks. The aim of the present paper is to address some of the issues raised by this apparent slowness in the growth of what many regard as an attractive technology. The first section presents a brief outline of established and projected uses of voice input in four major areas; industry, avionics, office, and aids for the disabled. This is followed by a discussion of some of the factors that presently restrict the range of successful applications.

VOICE RECOGNITION APPLICATIONS

Industrial

The Owens-Illinois Corporation in the USA was one of the first companies to use voice recognition for inspection and quality control. Early in 1973 they began using voice data entry for the inspection of colour television faceplates (ref 2). Other successful applications in the USA include the quality control of pull-ring can lids, textiles, oven parts, automobiles, jet engines at GEC and Lockheed-Georgia, compressors at Tecumseh Products Company, and the inspection of microelectronic circuits at Allied Corporation and Lockheed Missile Systems Division. There are far fewer European industrial users of automatic speech recognition, although some car manufacturers are beginning to investigate the introduction of speech technology, and have trained groups of users to carry out inspection tasks using voice. Another successful industrial application of ASR has been concerned with the directing of materials such as cartons, parcels, luggage, etc. In the USA, speech recognition systems have been introduced at airports, where they are used for the routing of baggage (ref 3).

*University of Bristol, Department of Psychology, 8-10 Berkeley Square, Bristol, BS8 1HH

Avionics

The majority of the research work on automatic speech recognition has been carried out in the field of avionics. Due to technological advancements, the complexity of avionic systems is ever increasing, and speech technology has been considered as a viable means of reducing the demands made on the crew of modern aircraft. A large amount of experimental work on ASR has been carried out into many aspects of avionics, from fixed wing aircraft and helicopters, to space stations and air traffic control, with the American, British and French aerospace industries all engaged in active research. Despite this concentration of research effort, the installation of speech recognisers into the cockpits of operational aircraft is not imminent. Much of the research in this area has involved simulation experiments, and it was not until 1984 that Coler remedied this situation by carrying out some in-flight testing of ASR systems (ref 4). A variety of reasons account for this slowness in the practical application of ASR systems. In avionics, the speech recogniser must have a high recognition accuracy, as misrecognitions may have serious consequences for safety and efficiency, and to date the performance of recognisers has not been sufficiently reliable. Few current commercial recognisers are robust enough to cope with the G-forces, high noise and vibration levels, and turbulence of the airborne environment (ref 5), and changes in the voice characteristics resulting from psychological and physiological stresses (ref 6).

Office

Another area which has attracted a large amount of research interest has been the application of ASR in the office. The concept of a "listening typewriter" was investigated as early as the late 1950's (ref 7), and since 1972, IBM in the USA have been working on the development of a voice-activated typewriter (ref 8). Others working in this field include AT & T Conversant Systems in the USA, the Nippon Electric Corporation (NEC) in Japan, and Olivetti in Italy. In the UK, Plessey is working on this application as part of an Alvey demonstrator project, in collaboration with three academic partners. Office applications such as information retrieval, word processing and voice messaging systems all require large vocabularies of at least 5,000 words. A vocabulary of this size demands the introduction of syntax, and in order to make the user's communication with the ASR device as natural as possible a continuous as opposed to a connected or isolated speech recogniser would be needed. Mangione (ref 9) hypothesised that voice input tasks would be carried out by a large number of non-expert users, so a speaker-independent system may be preferable, with minimal constraints on syntax. In 1987 there are no large vocabulary continuous speech recognisers available which meet these requirements. As in the case of avionics applications, limitations of the technology largely account for the lack of office applications of ASR.

Aids for the Disabled

A fourth area is that of ASR devices for the disabled. Work sponsored by the New York University Institute of Rehabilitation Medicine, the Palo Alto Veterans Administration, California, and in the UK by the DHSS and Nuffield Foundation has investigated the use of voice input systems

in aids for the disabled. Applications have included the use of ASR for educational and therapeutic purposes, and as a mode of communication. Specific examples include voice-controlled environmental control units (ECUs), wheelchairs and robotic arms. Research is still very much in an evaluative phase, although interest in this area has increased during the 1980's (ref 10). Due to the diverse nature of the handicaps of the population it is not possible to generalise about the disabled group of users. However, certain hypotheses could be devised about the requirements of the speech recogniser. The majority of applications require small vocabularies (less than 50 words) with no syntax. For example, Haigh (ref 11) commented that she found it difficult for her disabled population to identify 16 objects which could be switched on and off by the VADAS (Voice Activated Domestic Appliance System) environmental control system. A speaker-dependent, isolated word recognition system would also seem appropriate, since some groups of handicapped individuals have problems with the production of fluent speech.

ASR USERS IN THE UK

In 1986, an attempt by our own research group to locate everyday users of ASR systems in the UK, revealed only one group of individuals. This established and successful application of ASR technology is concerned with the making of maps and charts. Hydrographers at the Ministry of Defence, Taunton, have been using voice input in the process of making charts, while the cartographers at Clyde Surveys Ltd., Maidenhead, have employed ASR in mapmaking. Both groups have been using speech technology since the late 70's, and voice input is fully integrated into their systems.

Since that survey was conducted, a further application has been announced by the Austin-Rover Group, where voice input equipment has recently been installed for use in vehicle inspection. It is anticipated that other motor manufacturers will investigate this option in the near future, and interest is also being shown by some companies in the possible use of voice input for warehousing and stock control functions. Despite this, the number of developed UK applications of speech recognition technology remains extremely small.

CONCLUSIONS

It becomes apparent that one of the main reasons restricting the more widespread use of speech recognisers is the limitations inherent in the existing technology. This is particularly true for avionics, where high performance levels for the recogniser are necessary for reasons of safety and efficiency. A similar situation exists, although perhaps to a lesser extent, for the disabled group of users. The situation is likely to arise when it is important to the well-being of the handicapped individual that their commands are understood and carried out by the speech recognition system, for example, when maneuvering a wheelchair or calling for assistance. A different situation arises in the office, where it is not the accuracy of the recogniser which is the limiting factor, but the inability of present speech technology to match the demands of office applications. These are specific reasons relating to the constraints of the technology, but there are many general factors

which can also help explain the lack of users. The myth that voice is the most 'natural' means of communicating with a machine is repeatedly cited as an advantage of ASR technology. Although talking may be the most natural way for humans to communicate, this may not be true for a human and a machine. A further problem concerns the expectations of the user, who often has misconceptions of the system's capabilities and imagines that using a speech recogniser will be similar to holding a conversation with a fellow human. In some applications, the implementation of speech technology replaces jobs normally carried out by the workforce, and hence there is a certain resistance to the new technology. This was found by United Airlines when introducing ASR to the baggage handlers (ref 3). Finally, the inadequate human factors research in this area is frequently acknowledged as one of the reasons explaining the small number of voice input users. Crucial research is needed into the areas of training and enrollment, dialogue design, feedback, error detection and correction, before more individuals will benefit from speech recognition technology.

REFERENCES

1. W A Lea (ed.), Trends in Speech Recognition (Prentice-Hall, Englewood Cliffs, NJ, 1980).
2. T B Martin and J C Collins, ASQC Technical Conference Transactions, San Diego, CA, 8-15 (1975).
3. A J Lind, Proceedings of Speech Tech '86, Voice I/O Applications Show and Conference, New York, 66-67 (1986).
4. C R Coler, Proceedings of Speech Tech '84, Voice I/O Applications Show and Conference, New York, 95-98 (1984).
5. C A Simpson, C A Moore and J C Ruth, IEEE, 790-798 (1986).
6. A VanBronkhorst and R Abraczinskas, IEEE, 332-338 (1982).
7. J L Flanagan, Proceedings of the IEEE, 64(4),405-415 (1976).
8. F Jelinek, Proceedings of the IEEE, 73(11), 1616-1624 (1985).
9. P A Mangione, Proceedings of Speech Tech '86, Voice I/O Applications Show and Conference, New York, 154-156 (1986).
10. R I Damper, International Journal of Man-Machine Studies, 21, 541-553 (1984)
11. R Haigh, Personal Communication.