



## **A FULL SPEED SPEECH SIMULATION OF SPEECH RECOGNITION MACHINES**

A F Newell, J L Arnott, R Dye

### **ABSTRACT**

A speech driven word-processor has been devised which is based on Palantype machine shorthand transcription. Although the system uses a trained human operator, it does provide a very high performance simulation of a speech recognition system. This particular simulation does not suffer from any unrealistic restrictions on speed of dictation, such as occur with typewriter-keyboard based simulations. The simulation has been set in the context of a 'speech-driven' office, This simulation, and the associated experiments, are being used to produce a realistic assessment of users' response to such a system, and measures of performance criteria required. It is also enabling the development of appropriate dialogue design strategies for speech operation of word processors.

### **INTRODUCTION**

Recently there has been an upsurge of interest in speech recognition technology; a number of systems with various performance characteristics have been launched onto the market (see Speech Tech 1987) and there are confident predictions by many that speech will become a very popular method of inputting data into computers. Nevertheless, to date, there have been few systems which have been used consistently in real situations. Frankish (1987) reviewed the current and projected applications of speech recognition, dividing the potential market into 'avionics', 'office systems', 'aids for the disabled', and 'industrial applications'. He reported that only the industrial applications area had produced a significant number of working systems, these being in the areas of quality control, inspection and material handling. These particular applications are well matched to the performance characteristics of current recognisers: the vocabulary size is small, the syntax of the dialogue is very constrained and also the tasks have 'hands-busy and eyes busy' characteristics.

### **LISTENING TYPEWRITER - PERFORMANCE AND EXPECTATIONS**

In contrast to the above applications, the 'listening typewriter' has long been a 'holy grail' of speech recognition technology, and working systems have confidently been predicted to be 'almost available' for some time. Nevertheless, Frankish reported no large-scale use of speech recognition technology in the office. Newell (1984) has discussed some of the functional problems of the use of speech but, in addition, a major problem is the need for a clear specification of the performance required of a speech recognition machine before it will be accepted in such an environment.

Unfortunately, however, there is little or no reliable data on what the user is actually looking for in a 'listening typewriter'. The task of obtaining appropriate data is made difficult by current systems having

Microcomputer Centre, University of Dundee, Dundee DD1 4HN, Scotland.

inadequate performance, and also that it is not easy to specify the future performance (particularly the limitations) of speech recognition machines, clearly and accurately to potential users. It is not unlikely that, when some people say they would use a 'listening typewriter', they are envisaging a machine which would act in a similar way to their 'ideal' shorthand typist: that is it would produce a letter which was not only visually pleasing on the page, but was also what the author intended to say, rather than what he actually dictated.

A major problem is thus to ascertain the relationships between what will be achievable by automatic recognisers in the future and what a user may be prepared to accept. This is a very important question for the speech recognition research community. The answers to it should enable design effort to be targetted in appropriate directions, and towards goals which are acceptable to the user population.

### **DESIGN GUIDELINES FOR LISTENING TYPEWRITERS**

Johnson (1987) has been examining this question by a technique of making a video in which a person appears to be operating a 'listening typewriter'. The video was shown to groups of potential customers who were asked for their comments. One of his more disturbing findings was that the perceived value of the system to members of the group was much greater immediately after having seen the video than it was following a group discussion in which the potential users exchanged their views. Johnson has also conducted live experiments in which a typist acted as speech recogniser. These were in the same mould as the well known experiments reported by Gould in 1983, although with different protocols and objectives.

A major drawback of the simulation experiments is the limited speed capability of traditional typists. A good typists can perform at 60 to 70 words per minute, as opposed to commercial shorthand speeds of up to 120 words per minute, and unconstrained speech rates which are usually in the range 150 to 200 words per minute. Any simulation using a typist, therefore, severely restricts the speed at which a user can dictate free text, and this may bias the results of the simulation.

### **A FULL-SPEED SPEECH DRIVEN OFFICE**

We have thus set up a system where the speech recognition task is performed by a trained verbatim reporter using a Palantype shorthand machine keyboard. The output from the shorthand keyboard is automatically transcribed into orthography using a commercially available speech transcription system, (Newell et al, Possum Controls Ltd.). It operates in real-time with insignificant processing delays. The orthographic output from the transcription system is used to drive a word processing package.

In our speech driven office we have used the decor and acoustic conditions typical of an office environment, but on the desk there is a large, high resolution, visual display unit (without a keyboard) and an overt, boom-mounted, microphone. The microphone is connected to a digital speech recorder and the speech is also fed into earphones worn by a palantypist who is located in another room. (Although the microphone is currently the only input device available to the subject, other input devices will be added at a later stage of the research.)

The palantypist is instructed to encode on her keyboard the words spoken by the subject and thus to produce a full and accurate verbatim record of the speech. The orthographic output from the speech transcription system is fed into a separate computer which acts as a 'dialogue processor', and also runs a specially developed 'speech-driven' word processor. This package has many of the characteristics of a standard word processor, but has been designed to accept the orthographic version of spoken commands as well as text entry. Because of the geographic layout of the equipment, the subjects need never see the palantypist. This assists in creating the illusion of a completely automatic system if this is required.

The performance of the simulation is clearly dependent upon the skill of the operator, and some errors will occur due to keying mistakes. Nevertheless, the performance of the transcription system is substantially better than can be expected of fully automatic speech recognisers which may be built within the foreseeable future. In order to simulate the performance of fully automatic systems more accurately, we will provide the facility for degrading the performance of the transcription system by introducing errors of various types.

#### **RAISON D'ETRE OF THE SIMULATION**

The simulation is being used to investigate users' reponse to a speech driven text processor, and for the development of appropriate human interface characteristics, and dialogues for such a system. In particular we will:

- 1) Investigate the effects of simultaneous feedback of speech on dictation. This will include not only changes in the words which the talker might use, but also changes in timing and prosody which may occur,
- 2) Investigate and develop error correction and editing facilities which are appropriate to a speech driven system,
- 3) Determine the performance characteristics (e.g. text feedback methods, error rate, error correction capabilities) which users find acceptable and those which are unacceptable, and
- 4) Develop dialogue structures which are appropriate for a continuous speech recogniser with the potential of immediate feedback.

#### **RESULTS FROM PILOT TRIALS AND FUTURE WORK**

Early experimental results have underlined some of the problems which arise when using a speech driven word processor. Examples include subjects finding difficulty in:

Formatting text properly, giving appropriate typographic commands, (punctuation, capitalisation, etc), and particularly, recovering appropriately from errors of dictation, or recognition errors.

We are also planning to conduct a series of experiments where the operator is overtly present. We will thus be able to compare the results of our other tests with the performance of subjects, when they are deliberately made aware that there is a human operator within the system. These experiments will also help to indicate the potential benefit of machine shorthand systems in the electronic office.

### CONCLUSIONS

Although experimentation is at an early stage, the results from full-speed simulation experiments on a speech driven word processor have confirmed the suspicion of many that there is much more to designing a useful speech driven word processor than simply attaching a speech recognition system as a 'front-end' to a standard word processor. For such a system to be acceptable, the performance requirements are very high. The human factors of the interface, and the dialogue design are a vital part of the overall system, and it is essential that these are fully integrated with the speech recognition technology. There is also evidence that the goal of a 'natural interface' which requires no training is unlikely to be achieved in the foreseeable future. The simulation experiments which are possible using machine shorthand will thus provide many useful design guidelines for those teams who are developing fully automatic speech recognition systems.

In addition, a word processor driven by a palantype shorthand machine provides a speech driven office, which is technically feasible today, and which is appropriate in those situations where the economic advantages of such a system are sufficient to justify the employment of a trained operator.

### REFERENCES

- (1) Speech Tech. New York, April 1987.
- (2) C R Frankish, "Users, Applications and Human Factors". Proc. Intl. Speech Tech. (London), June 1987.
- (3) C Johnson, "Speech Technology and Word Processing", *ibid.*
- (4) A F Newell, "Speech the Natural Method of Man-Machine Communication", Proc. 1st IFIP Conf. on Human-Computer Interaction, Imperial College, London, Sept. 1984, (North-Holland), pp.231-238.
- (5) A C Downton and C P Brooks, "Automated Machine Shorthand Transcription in Commercial Applications", *ibid.* pp.151-156.
- (6) J D Gould, J Conti and T Hovanyecz, "Composing Letters with a Simulated Listening Typewriter", Comm. of A.C.M. Vol.26, No.4, April 1983.
- (7) Possum Controls Limited, Middlegreen Trading Estate, Langley, Slough.

### ACKNOWLEDGEMENTS

This work is supported by an SERC/ALVEY grant, and Possum Controls Ltd.