VADAS, A SIMPLE VOICE RECOGNITION SYSTEM FOR ENVIRONMENTAL CONTROL

R HAIGH*, A K CLARKE*

ABSTRACT

Severely disabled people frequently require environmental control equipment to operate electrical appliances. The possibility of using voice recognition rather than microswitches to interface with the equipment is currently under investigation.

Such a system has been evaluated to establish the equipment's potential for improving the quality of life for disabled people.

Trials have been carried out both in the laboratory and in the homes of disabled people, to determine the reliability, usefulness and acceptability of the systems.

INTRODUCTION

VADAS, Voice Activated Domestic Appliance System, is a very simple word recogniser, developed by Voice Control Ltd of Cambridge. It is a speaker dependent, language independent, isolated word recogniser (ref 2). Each word in the vocabulary must be trained twice during the training process and an average of the two utterances is stored as the template.

VADAS, is a unit measuring 18 x 14 x 7 cm with a built-in microphone, a screen to give visual feedback, push buttons to enable different functions to be carried out by the helper and operates on the ring main. The system has a set vocabulary of seven words but these can be changed, added to or erased as required. Up to 16 appliances can be operated by the system and a mains signal receiver switch is required for each. A code is set on the back of each receiver switch to indicate to VADAS which appliance is which. Once VADAS's vocabulary has been trained with the user's voice household appliances, such as, a light, fan, radio or television, can be operated by voice command.

The potential users for such equipment are those with severely impaired upper and lower limb function. These people are frequently housebound, often wheelchair bound, require constant attention and special equipment to aid daily living.

AIMS OF STUDY

The aims of the study were to:

a) establish the length of time taken and ease with which the system could be trained with the user's voice and the number of times retraining was required;

b) assess whether the set vocabulary is suitable both for the user and for VADAS and to obtain data on substitution and rejection error rates;

* Royal National Hospital for Rheumatic Diseases, Bath BA1 1RL
c) ascertain the effects of background noise;
d) assess the effects of voice fatigue and deterioration;
e) collect information on which people with which types and degrees of disability would gain most benefit from a voice activated system.

METHODOLOGY

Laboratory trials were carried out in order to collect data in a controlled setting which could then be compared with results from the field trials.

Twelve subjects (six male, six female), age range 21 to 57 tested the system for 15 minutes a day on 10 working days to establish whether recognition performance remained consistent over a given period of time. Comparisons were made between training and use in quiet and noisy backgrounds and rejection and substitution error rates were monitored. If a command had to be repeated five times the word was retrained.

VADAS was then tested in the homes of disabled people for approximately four weeks to determine reliability, usefulness and acceptability of the system. A pilot trial involving 6 people was conducted and subsequently 25 people (16 male, 9 female) age range 20 to 78 participated in the main trial. The following disabilities were represented; Multiple Sclerosis, Motor Neurone Disease, Polio, severe Rheumatoid Arthritis, Cerebral Palsy and Spinal Injury. Only people with the ability to speak words with a degree of consistency were included.

Data was collected using structured interviews and event monitors. These were fitted into each VADAS for the purpose of this study and counted the number of times each appliance was switched on and off to establish how frequently the system was used.

LABORATORY TRIAL RESULTS

The results showed marked variation in performance between different users.

Retraining requirements varied between those for whom this was not necessary to those who had to retrain daily.

Confusion matrices were generated for the set vocabulary both for individual subjects and between subjects. These showed that there is no apparent consistency in which words are rejected or substituted either for individual subjects or between subjects.

Background noise did not significantly affect rejection and substitution error rates.

FIELD TRIAL RESULTS

On the whole the subjects were very positive about the principles of voice recognition and many preferred it as a way to interact with their environment over the use of suck/blow or head pointers. Though most were concerned that reliability should be improved.

There were few problems with training the system, since in most cases the training session took only a few minutes. Most subjects trained
the system daily for the first few days while they experimented with different vocabularies and familiarised themselves with the system. Later most found it necessary only to retrain approximately once a week.

Some subjects experimented with alternative vocabularies. Occasionally the words chosen did not resemble the appliance. The maximum number of commands with this system is 16 and subjects did not report forgetting the command words. However, it was more difficult to choose words which would not cause substitution errors (ref 3). There was no connection between words which sound the same to humans and those which are substituted by the system, which some subjects found frustrating.

Frequent rejection errors were also reported; most users had to repeat themselves often in order to activate the system and then found that the command word had to be repeated frequently, often with the result that a substitution error occurred. This was frustrating and fatiguing.

Background noise caused two main problems; one was that the system was sometimes unable to detect the user's voice above the sounds of the radio or television, the other was that the noise from the radio or television activated the system.

Although the tolerance level of the system can be altered this did not appear to be sensitive enough to accommodate people whose voices deteriorated throughout the day.

There did not appear to be any correlation between those with a strong confident, authoritative voice and those who found the system to be reliable.

It appears that it is individual, personality characteristics rather than types of disability which affect the performance of the system.

**DISCUSSION**

Environmental Control Units (ECU's) enable simple tasks to be performed, such as turning on/off a light, radio, television, by those with limited movement. ECU's can, therefore, improve quality of life, increase independence and reduce the nursing requirements of a person with severe disabilities (ref 4). However, the present interfaces require physical contact with the equipment. Therefore being able to command the environment by simply speaking to appliances could have many advantages for a disabled person.

The subjects who participated in this study may have demonstrated the 'sheep and goats' phenomena recognised by Doddington and Schalk (ref 1) which are at present being researched further at Bristol University. Their suggestion is that word recognition systems can be reliable for the majority of the population (the sheep), but that reliability can never be attained by a small section of the population (the goats). It is unclear what causes 'goats' though there are two possible reasons; during the template training session 'goats' may not produce a speech template representative of their usual speech pattern for that particular word, or they may be generally inconsistent in the speech they produce.
It was observed that some subjects trained the system with a different voice than that with which they commanded it. This was attributed to the fact that during training the subject reads the words to be trained from the screen, during use a command tone is used. Also, once the system had failed to recognise the command or repeatedly mis-recognised it the subjects frequently became irritated and their voice patterns changed. This occurred both in the laboratory and during the field trials.

Some subjects took advantage of the fact that the system vocabulary could be altered. Words were chosen which the user could pronounce most easily, clearly and consistently, which enabled them to feel more confident. For example, some users trained the system with numbers or colours or with a word which represented the appliance but not the word in the set vocabulary, eg, 'tele', 'TV', or 'box', for television, one user even gave appliances Welsh and Biblical names. Therefore it is difficult to recommend a set vocabulary which will be appropriate for all users as different people give the same appliances different names and some people will find some words easier to pronounce than others.

Substitution and rejection errors may be potentially dangerous. For example; if the user gives the command 'fan' and the alarm is activated the helper may cease to respond to the alarm. Conversely if help is required and any appliance but 'help' is activated there may be serious consequences. One aspect of training the system may be at fault here. In training the subject simply repeats the words which appear on the screen, in a 'reading' voice. In a situation where help is required the voice may be substantially altered. Another reason why recognition rates may be low is the fact that the system is trained with the user sitting directly in front of and close to the microphone, partly to be able to read the screen, during use the user may be further away from the microphone and at a different orientation. To reduce recognition error rates the user could wear a headband microphone but this would defeat the object of designing a system to which the user does not have to be attached. Trials may be conducted with a radio microphone to establish whether this will be satisfactory for some users and whether this will improve performance.

The principle of interacting with equipment via voice recognition has so far been found to be generally acceptable and useful to disabled people, however, further studies are required.

REFERENCES