Editorial

The annual "arrival" of speech recognition

The idea of the eventual development of a HAL-type machine capable of recognising continuous speech from any speaker, prompted a 1970 ARPA report recommending a research effort. After 5 years work from several prestigious groups the "winner" was adjudged to be Carnegie-Mellon's HARPY. It was concluded that "cost-effective speech input is an attainable goal".

The following quotations are taken from several different 1990 publications:

"The most widely used and most financially successful speech recognition application is the Japanese ANSER system. Bank customers, having provided an ID to the bank's computer over the telephone, are informed which of their bills have been paid, how much was paid, and what their remaining balance is... The vocabulary is limited to six isolated control words and ten digits".

"We have demonstrated that large-vocabulary speaker-independent continuous speech recognition is feasible".

"A voice activated 'typewriter' that can recognise 30,000 words and adapts itself to individual speakers has been unveiled... It costs $9,000 and consists of software and a speech-recognition circuit board for personal computers".

"The present generation of speech recognition devices has made little, if any, impact on society... It is possible that we are really on the right track... it is unfortunately also possible that this strategy will run afoul of the 'first step fallacy' which warns that one cannot reach the moon by climbing a tree though to do so is to move in the right direction".

"PC voice card for speaker-independent continuous-speech with a maximum active vocabulary of 40,000 words... acceptable accuracy and speed require syntactic language model constraints on active vocabulary. Price $33,900".

For the ultimate claim I must go back to a 1989 advertisement in *Byte* for "The Voice" which will recognise anybody's speech regardless of dialect and which accesses files "100,000 times faster than the IBM PC". A modest price of $2,000 is asked for this remarkable device.

You will gather from this collection—representative of the shades of opinion expressed about current activities—that it is not easy to assess how far we have progressed.

ANSER deals with half a million transactions a month and does its job very well. But although its ability to recognise 16 isolated words from any speaker could be bettered today, its performance is not untypical of in-the-field systems which work and fulfill a real market need.

On the other hand, *something* must exist to prompt the claim about the voice activated typewriter. Similarly the $39,000 continuous speech 40,000 word voice card, taken from a table listing a number of voice recognition products, must exist—I don't know how severely its performance is limited by "the syntactic language model constraint".

It's a pity that the hype obscures limited but very useful achievements in this field. For example the Cherry Voicescribe PC board enables James Godfrey, an 18-year-old paralysed by a rugby accident in the UK, to produce limited but useable text on his Word Processor.

When the first affordable voice-recognising "typewriter" worth having appears it will provide just acceptable results from a person speaking slowly and clearly once it has been trained to that person's voice. I would be very surprised to see it in the shops for at least another five years and quite surprised to see it in ten.,

A.E. Cawkell