AUTOMATIC SPEECH RECOGNITION AND COMPUTER SCIENCE

J.S.BRIDLE

1.0 ALGORITHMS FOR AUTOMATIC SPEECH RECOGNITION.

1.1 From waveform to sound pattern

(A brief review of relevant signal processing)

1.1.1 Digitisation. - Bandwidth, signal-to-noise ratio. Sampling rate, number of bits. Remarks on real-time response requirements of computers for digitisation.

1.1.2 Spectrum analysis. - Time and frequency structure of speech. Time and frequency resolution of analysis. Discrete Fourier transforms, recursive filters, all-pole analysis.

1.1.3 Spectrum distance measures. Computational requirements. -

1.2 Time Warps, String Edits and Hidden States.

1.2.1 The speech timescale problem. - A related mathematical problem and a Dynamic Programming solution.

1.2.2 Relationship to string matching with insertiuons, deletons and substitusions. -
Generalised Lenvenshtein distance,
Wagner-Fischer algorithm.

1.2.3 Connected word recognition using whole word templates. - Algorithm
Architectures for real-time connected word recognition

1.2.4 Probabalistic finite state automata - Relationship to regular languages, Markov sources.
Maximum likelihood explanation of a given output sequence.
Error-correcting parsers
Use in ASR: Word models Phrase models Hierarchies of PFSA's. (Integration of multiple knowledge sources)
5000 word continuous natural language transcription (IBM)
Speaker-independent isolated word recognition over the telephone (Bell Labs)

1.3 Use of Higher-level Knowledge.

(A brief review of the ARPA Speech Understanding Research project.)
Aims
Techniques used(Hearsay II, Harpy)
Achievements.

2.0 USING ASR AS A MAN-COMPUTER INTERFACE

2.1 Possible Reasons for Using Simple ASR for Data Entry

Speed, Accuracy, Size and cost of equipment;
Operational cost saving, parallel tasks, no alternative;
Easy learning, fantasticity value, gimmic value.

2.2 Voice Interaction Styles

2.2.1 Voice Buttons - Selection of single item from previously defined set ("Newcastle") (IWR).

2.2.2 Hierarchies of voice buttons - Vocabulary sub-set depends on context. User must know alternatives at each point.

2.2.3 Commands with arguments (numeric or alphabetic) - ("Frequency 347 KHz") (Simple CWR)

2.2.4 Simple artificial languages - For trained operators ("severe dent right offside wing") (CWR with simple syntax control).

2.2.5 Psuedo-natural, limited domain - Attempts to cope with naive users ("When is the next train to Cheltenham") (Graceful interaction).

2.2.6 Unlimited vocabulary, limited discrimination - (Information retrieval from existing large data base)

2.2.7 Natural language, limited domain - Large vocabulary, use of language statistics to guide ASR (business letter dictation).

2.3 Human Factors

The user's model of the machine: The machine's model of the user.
Difficulties with speaking style and consistent pronunciation: some people will have great difficulty using ASR.
User's assumptions about the computer: under- and over-estimation of ASR capabilities and the 'intelligence' that lies behind it.
Effects of feedback to the user- danger of runaway double adaptation.
Warning: are large vocabularies and complex artificial syntaxes too difficult to learn?

2.4 Interaction with other technologies

Displays, Pointing devices, Expert Systems, VLSI...

DISCUSSION

Professor Randell recalled that there were two research projects on Automatic Speech Recognition (ASR) at Carnegie-Mellon University and wondered which one had been more successful.

Mr. Bridle replied that out of the two projects, one was based on the use of AI techniques while the other utilised statistical techniques and was the more successful.

Dr. Card enquired about Man-Machine Interface aspects of ASR.

Mr. Bridle replied that there are many issues, depending upon what kind of use a system is put to. Many people will have great difficulty in using an ASR system; some reasons might be their particular way of speech or 'speaking to machine' hang-ups (some people feel awkward speaking to a machine). Also different users could have different assumptions about the capabilities of a given machine, in which case a user's model of a machine will not correspond to the machine's model of a user.

Professor van Rijsbergen asked about the error rate of the system developed at the speaker's Laboratory (the LOGOS system).

Mr. Bridle replied that they had not yet undertaken any proper measurements. However, he did have a figure for a similar system which was used experimentally by the U.S. Postal Service for Mail Room operators. This system had digit sequences recorded in it for matching against spoken sequences, and the error rate was a fraction of one percent.

Professor Randell asked whether any ASR system is in regular practical use.

Mr. Bridle replied that the Admiralty Hydrographic Survey Organization in the U.K. have been using a system regularly for the last five years. A simple isolated word recognizer is being used to record digitized map reference coordinates. Another example is one of the experimental versions of the Mirage 3 jet fighter plane built in France, whose avionics system used an ASR system for recognizing certain simple commands.