ARTIFICIAL INTELLIGENCE APPROACHES TO COMPUTER-BASED ENGLISH TEACHING: A TUTORIAL SPEECH RECOGNITION

YELENA Y. DASHKO

Dept. of Foreign Languages
Mining Institute
Leningrad, USSR, 199026

ABSTRACT

With the increasing role of computers in teaching, there is little doubt that we will eventually want them to talk to us and allow us to speak to them. This review presents an experimental approach of voice I/O techniques to computer-based teaching English on the basis of Expert Type System. Advanced applications of the speech processing technology and some special linguistic/information problems are discussed.

INTRODUCTION

The (micro)computer in education has both stimulated research on linguistic database and has provided a more precise experimental vehicle for controlling the presentation of instruction and measuring responses /1/.

The basic hypothesis is that the computer offers an opportunity:

(i) to provide the teacher with a powerful resource to manage individual learning within the terminal room;
(ii) to enable the student to follow learning procedures which incorporate step-by-step feedback and to stimulate individual attention and to gain assistance discreetly;
(iii) to make it possible to the teacher to observe and monitor the progress of the student in detail.

The objectives of the preliminary study are, firstly, to investigate the feasibility of the approach and the appropriateness of the software facilities being employed.

AUTOMATED LEARNING SYSTEMS

Evidence indicates that it is most productive to teach grammatical and lexical bases in context. In order to carry out the exercise the student must thus draw upon reading (knowing) English letters, vocabulary.

In computer-based English teaching the student receives all of his training from the display device including tests and performance feedback. Variations and combinations of instructional arrangements are not uncommon. The computer-based study management model usually employs existing materials and the student spends only a part of his study time interacting with the computer; the material used in the initial study provides 20-40 minutes work for students. Testing and teaching may be done on the computer via keyboard input by students. Responses may be saved on diskette and at the end of testing the student's score can be displayed on the screen immediately, or a printed evaluation can be reproduced depending on the computer programme and the test construction. Questions to be printed are chosen from a master list and their descriptors (up to 1000 items).

Each Automated Learning System (ALS) comprises several learning volumes: Training volume; Control volume with its priority scoring due to three levels of complexity, grammatical and lexical databases; Reference volume with its grammar and lexicon retrieval; Encouragement volume.

Potential linguistic problems are worked out before they are translated into hardware and software. Needs, goals, constraints are described first, thus helping a complex problem to be divided into a hierarchy of simpler problems to understand/control the whole process of learning. Each training step is related to previous and subsequent stages, and misunderstandings among students are avoided. Questions/answers from the packet of exercises in Control volume should be typed with a 'minimum-energy' solution.

ALS AND DIALOG INTELLECTUAL EXPERT SYSTEM (DIS-332)

Due to the technological development of voice recognition systems voice input can be embodied in ALS. Voice input gives the chance to relieve the overloaded visual/manual channel and may achieve a natural form of human-computer communication.

A major problem is how to integrate the different sources of knowledge in such a way as to exploit their interaction. One can identify the following conceptually

distinct sources of knowledge as important in determining the interpretation of a spoken utterance/2/:

1. Segmentation, Feature Extraction and Labelling - processes of detecting acoustic-phonetic events in the speech signal and characterizing the nature of individual segments of the signal.

2. Lexical Retrieval - a process of retrieving candidate words from the lexicon base that are acoustically similar to the labelled segments.

3. Word Matching - a process of determining some measure of the goodness of a word hypothesis at a given point in the speech signal.

4. Syntax - the ability to determine if a given sequence of words is a possible subpart of a grammatical sentence and to predict possible continuations for such sentence fragments.

5. Semantics - the ability to determine if a sentence is appropriate to the context in which it is uttered, and what has been said previously in the discourse.

A variety of different approaches have been explored on the basis of the Dialog Intellectual Expert System (DIS-332)/3/. Generally, they fall into top-down and bottom-up strategies, where a single network parser combined syntactic, semantic and pragmatic components on the basis of Hidden Markov Models, is presented.
In these syntactically constrained tasks performance results ought to be reported with the following information: a) complete description of the domain grammar including full specification of the vocabulary in the form of scripts, b) frequencies of lexical units transitions from each task state to successive states and c) average branching factor.

It is important to distinguish between the total vocabulary capacity and the branching factor, the average number of words which must actually be discriminated at each stage of the task (sometimes referred to as the size of the average active subvocabulary). DIS-332 includes a study of extensive database collections of both isolated and connected utterances, spoken by ten Russian speakers.

Vocabulary size = 500 words, performance of the recognizer = 98-99%, recognition time = real, branching factor $\leq$ 10.

Above mentioned configuration provides a large scale of opportunities while using in ALS:

1. Russian lexical vocabulary and corresponding English items Input;

2. Printed text Visualization;

3. Impartial Control upon English words learning;

4. **Sounding for each input word (speech synthesis);**

5. Voice input ($\leq$500 words) for English spelling correction of phonemic baseforms in training and recognition mode;

6. Voice input of English sentences (sentence length $\leq$15) in training and recognition mode;

7. Impartial Control upon learned grammatical/lexical English level;

8. Reference information output due to the error rate or to inquiry.

LINGUISTIC AND INFORMATION PROBLEMS OF VOICE I/O

Finally, prediction models of recognizer performance can be used in determining optimal operating conditions for voice input /4/. A major problem is the identification of those factors having significant influence on the performance of the speech recognizer. W.A.Lea (1982) has compiled an extensive list of more than 80 variables including language and task factors (number of training passes, reject threshold, size of the active vocabulary, inter-word confusability), human factors (sex of the speaker), algoritmic factors, channel and environmental factors (microphone type and position), performance factors (type of feedback, error correction).

Ergonomic aspects for improving recogni3 tion performance should include:

a) a short speaker training of several hours is necessary;

b) if possible, a 3-5 repetitions in system training should be carried out;

c) equally-positioned phonemes of the vocabulary should be out of different articulation types;

d) vocabularies should be splitted down even in smaller subvocabularies than specified;

e) system training must be performed with the operational noise at minimum.

There was found an improvement for DIS-332 from 96% to 98-99% when three instead of one word repetition was chosen. Summing up, it provides a sufficient variety of coarticulatory environments. 5 - 6 word repetitions brought no further improvements which is not surprising in view of the high level of the recognition rate. Generally, it is suspected that the necessary number of word repetition during system training mode between about 3 to 6 depends positively on vocabulary size, complexity and confusability. The results show that equally-positioned phonemes can be better distinguished from/among one another when different articulation types are used (e.g. plosive - fricative). Such features as voicing, nasality, affrication, duration and place of articulation are the primary channels of the intelligibility-relevant information. Different vocals tend to be good features of words to be recognized if they do not belong simultaneously to the high vowels "e" and "i" and to the deep vowels "o" and "u". We believe that for each doubling of the vocabulary size, the recognition accuracy tends to decrease by a fixed amount, which is different for each talker.

Yet there can hardly be any more important task in speech recognition than determining now well algorithms or devices work. Thus the error rate as a performance measure conveys no information about performance except the relative number of errors made on a given task. It tells nothing about the distribution of errors and the costs of making particular errors; depends on vocabulary size and doesn't reflect large vocabulary difficulty, the inherent acoustic confusability, the difficulty of the speaker, or the environment. A new information-theoretic performance measure is based, in part, on the idea that automatic as well as human speech recognition systems can be modelled as communication channels. A more meaningful measure, called the Relative Information Loss (RIL) would normalize the amount of information lost in a recognition process with the amount transmitted/5/. Woodard and Nelson /6/ propose combining the 'Human Equivalent Noise Reference' (HENR) method with a RIL method. HENR is based on the confusions between speech sounds by humans listening in noise. The model predicts the percentage word recognition rate, and the confusions at any signal to noise ratio for any vocabulary which has been defined in phonetic terms. This combined approach may be used to relate device performance to task difficulty.

Grammatical constraints whether they will be stochastic or deterministic, have the effect of decreasing entropy, increasing redundancy and hence decreasing error rate (entropy is, of course, a statistical property). Each natural language requires that some assumption be made about the likelihood of occurence of trained difficulty at a given point in sentence.

Two reasonable assumptions are that the difficulties are equiprobable or distributed to maximize entropy. Under the medium entropy assumption,

$$H_{eq} = \frac{\log_2 |L(G)|}{E_1(G)\{W\}}$$

so that entropy in bits/word is the base-two logarithm of the size of the language divided by average sentence length. For ALS redundancy is to be increased up to 20% against existing Expert Systems.

CONCLUSIONS

The development of faster microprocessors, larger memories, better printers and storage devices, together with pricing competition, will play roles, too. But the factor likely to be judged most significant in the academic microcomputer revolution will probably be the rate at which these recognition systems have gained widespread acceptance by humans in serving their diverse educational needs on the basis of ALS.

REFERENCES

/1/W.D.Fattig, "Microcomputers in Academia", Journal of the Alabama Academy of Science, vol.55, No.1, Jan.1984, pp.31-37

/2/"Computer Speech Processing" ed.by F.Fallside, W.A.Woods (UK) LTD, London, 1985

/3/Петров А.Н., Туркин В.Н. "Система речевого диалога", Автоматическое распознавание речевых образов, ч.II, Каунас, 1986, с.97-99

/4/H.Mutschler, "Ergonomic aspects for improving recognition performance of voice input systems", IFAC Analysis, Design and Evaluation of Man-Machine Systems, Baden-Baden, 1982, pp.261-267

/5/Th.M.Spine, B.H.Williges, J.F.Mainard, "An economical approach to modeling speech recognition accuracy", Int.J. Man-Machine Studies, 1984, 21, pp.191-202

/6/J.M.Baker, D.S.Pallett, J.S.Bridle, "Speech recognition performance assessments and available databases", Proceedings ICASSP-83, Boston, pp.527-532