Proceedings of The Institute of Acoustics 2A8

DEVELOPMENT AND APPLICATION OF AUTOMATIC SPEECH RECOGNITION

JEAN PAUL HATON

CENTRE DE RECHERCHE EN INFORMATIQUE DE NANCY, FRANCE

Introduction

Several projects have been developed in our laboratory during the past ten years in various areas of speech recognition and understanding. All these different projects are related to the general problem of man-machine communication by voice. We have first designed an automatic isolated word recognition system using the principles of dynamic programming in the time normalization of the words. This system was then used for the recognition of sequences of digits (postal codes) and for speech training of deaf children in the framework of the SIRENE project. Since 1974 we have investigated the problem of connected sentences recognition. This is the most difficult problem in automatic speech recognition ; it makes it necessary to take into account syntactic and semantic-pragmatic information during the recognition process .
In a first time we restricted ourselves to artificial language recognition. The MYRTILLE I system was realized in 1975 for the recognition of an artificial language in the application of a telephone exchange. MYRTILLE I uses a top-down strategy which turns out to be efficient as far as artificial languages are concerned. For natural and pseudo-natural languages in which the syntactic constraints are far less strong it is necessary to use more sophisticated strategies. This is done in the MYRTILLE II project which started in 1977. MYRTILLE II will be able to understand sentences of spoken French with few restrictions on syntax but in a well-defined application.
This paper will briefly describe these various projects, the methods used and the results obtained. It will also outline the techniques involved at the various processing levels : speech signal processing, pattern recognition and speech understanding.

Isolated Word Recognition

1) Basic algorithm

The most important problem in the recognition of isolated word for a single speaker is the time normalization of the words. The technique that we use is based on the computation by dynamic programming (DP) of a time warping function which makes it easy to compensate for the non-linear variations in duration which can affect a given word during several pronunciations. Similar techniques were first used in USSR (1) and Japan (2). Our algorithm is more simple since it only computes a local optimum for the warping function instead of a global optimum like DP algorithms usually do. This simplified algorithm is efficient enough for the problem of word matching. Moreover it is less time consuming : a version of the system was implemented on a Z80 8-bit microprocessor, it recognizes a ten digit vocabulary with a recognition time of about 1 second. The speech signal is analyzed by a bank of band pass filters (15 in the general system and 3 in microprocessor based system).

Several experiments have given recognition rates from 95 % to 99 % for different vocabularies up to 100 words.

2) Extension to digit sequence recognition (3)
The isolated word recognizer is used in a system for the recognition of french postal codes pronounced continuously (i.e. sequences of 2 or 3 digits). The approach we have chosen is a semi-global one ; it consists of two steps :
- segmentation of an utterance into digits. This segmentation is carried out using various parameters as energy, voicing and spectral variations computed from the outputs of a spectral analyzer,
- recognition of the digits by the dynamic matching algorithm.
Work is presently in progress in order to test this system on a great amount of data.

3) Application to speech training of the deaf

Since 1974 we have developed the SIRENE system for speech training of deaf children [4]. The basic idea of this system is to compensate for the lack of auditory feed back by use of visual displays. SIRENE is presently based on a minicomputer but some parts have already been implemented on microprocessors. The system is intended to be used by speech teachers and therapists, but is can also be used by children for self-training. Several speech parameters are hardware extracted and displayed on a screen. In fact three categories of training procedures are related to these parameters :
- prosodic parameters : pitch, intensity, rhythm and correlation between these parameters,
- frequency parameters : formants, and so on. This makes it possible to train the child to utter isolated sounds,
- words and phrases : these procedures are intended to train the child to learn a vocabulary of isolated words and/or phrases. The matching procedure is the DP procedure described previously. Dynamic programming is very interesting for the deaf since the variations of rhythm and duration may be very important.
SIRENE has been used for one year by therapists with deaf children and also with hearing children who have typical pronunciation defects. The results obtained are quite good.

Recognition of sentences

The first attempt to sentence recognition in our laboratory was concerned with the recognition of isolated word sentences of a machine-tool control language (5). We have then developed the MYRTILLE I system for connected sentence recognition. This system was designed for artificial language processing ; i.e. languages with strong syntactic constraints (the language used in MYRTILLE I is a small subset of spoken French corresponding to the dialog with a telephone exchange). The constraints on syntax made it possible to build a top-down system in which the recognition process is driven by a parser (5).
This strategy cannot be implemented for natural language. In the MYRTILLE II project we have therefore designed a new strategy which is both to-down and bottom-up according to the context. In the case of natural language the role of syntax becomes less preponderant and it is impossible to completely distinguish between syntax and semantics. The representation we have choosen incorporates both syntax and semantics in a unified way. It consists of a net work with procedural nodes. This representation is in some sense similar to an ATN grammar (6) but the procedures which are at the nodes of the network make it possible to dynamically modify the parsing according to semantic information.

MYRTILLE II is presently completely defined and its implementation is in progress.

References

1- V.M. VELICHKO, N.G.ZAGORUJKO, 1970 "Automatic Recognition of 200 words" Int. J. Man-Machine Stud., 2, pp. 223-235.

2- H. SAKOE, S. CHIBA "A Dynamic Programming Approach to Continuous Speech Recognition" Proc. Int. Cong. Acoustics, Budepest, 1971.

3- J.P. HATON, D. MOREL "Automatic Recognition of Connected Digits Sequences By Means of Segmentation and Dynamic Programming" Int. Conf. Conf. ASSP, Washington, D.C. April 1979.

4- M.C. HATON, J.P. HATON "SIRENE, A System For Speech Training of Deaf People" Int. Conf. ASSP, Washington, DC, April 1979.

5- J.P. HATON, J.M. PIERREL "Organization and Operation of MYRTILLE I Speech Understanding System" Int. Conf. ASSP, Philadelphia, April 1976.

6- W. A. WOODS, 1970 "Transition Network Grammars for Natural Language Processing" Comm. ACM, 13, pp. 591-606.