SPEECH RECOGNITION SYSTEM ON A MICROCUTTER

V.G. Lebedev
Lab. of Technical Cybernetics, Novosibirsk State University
Novosibirsk, USSR 630090

INTRODUCTION

A system is discussed which is designed to understand phrases of discrete speech. The system is based on connected responses and recognizes phrases in the form of connected words. It was developed by the Novosibirsk Research Institute of Automatics and Cybernetics and is the first in the USSR to be used in a practical application.

ARCHITECTURE AND ALGORITHM OF SYSTEM

The system is based on a microcomputer "Electronika-60" incorporating a preprocessor for the initial speech signal processing and a complex of programs written in the languages MACROCOM/MANS AND FORTRAN in the "RASPA" operation system. Through a microphone a speech signal is fed to the preprocessor which every 16 ms measures the values of intensities of several 25 ms segments covering a band of 400 to 5000 Hz.

The value of the total intensity which is in excess of the assigned threshold determines the start of signal input. The final decision on the start of input is taken if several successive input segments meet this condition. The decision to end the input is taken if several successive segments have a total intensity below the threshold. Otherwise the segments of low intensity correspond to speech pauses.

The threshold values and the required number of segments in the first and the second case are assigned by the user. Performed parallel with the process of input signal processing is the work by the preprocessor which generates the following:

- the average intensity value for each segment,
- the average intensity of the speech envelope for each segment,
- determining the average closed contiguous vowel segments and the subsequent averaging segments.

The secondary averaging is performed for a group of segments which are located between the first vector and all the subsequent ones less than the segment added by the user. After the segmentation the second index of the secondary averaging segment is assigned by the user. The user may assign a segment following the initial segment or any segment along with the selected segment.

The processing of the speech recognition system is carried out by a second program which is capable of doing the following:

- determining the total number of connected words that were recognized by the preprocessor,
- determining the number of false recognitions, cases in which a phrase is not contained in the dictionary,
- determining the number of correct recognitions that were not contained in the dictionary.

ASSIGNMENT FOR LINGUISTIC RESTRICTIONS

Admissible sequences of words in phrases are processed by a preprocessor. If the value of the word stresses marks the selection of a node of the reduced vocabulary, the node is placed in the dictionary.

The tree is built in the form of a 2ndimensional array with a sequence of the words in each of which is a root of a tree, which is a sequence of admissible nodes where each branch reflects a continuation of the process, the selection of a word is made while the program is running in a waste of the value of the total intensity which is in excess of the assigned threshold determines the start of signal input. The final decision on the start of input is taken if several successive input segments meet this condition. The decision to end the input is taken if several successive segments have a total intensity below the threshold. Otherwise the segments of low intensity correspond to speech pauses.

The threshold values and the required number of segments in the first and the second case are assigned by the user. Performed parallel with the process of input signal processing is the work by the preprocessor which generates the following:

- the average intensity value for each segment,
- the average intensity of the speech envelope for each segment,
- determining the average closed contiguous vowel segments and the subsequent averaging segments.

The secondary averaging is performed for a group of segments which are located between the first vector and all the subsequent ones less than the segment added by the user. After the segmentation the second index of the secondary averaging segment is assigned by the user. The user may assign a segment following the initial segment or any segment along with the selected segment.

The processing of the speech recognition system is carried out by a second program which is capable of doing the following:

- determining the total number of connected words that were recognized by the preprocessor,
- determining the number of false recognitions, cases in which a phrase is not contained in the dictionary,
- determining the number of correct recognitions that were not contained in the dictionary.

ASSIGNMENT FOR LINGUISTIC RESTRICTIONS

Admissible sequences of words in phrases are processed by a preprocessor. If the value of the word stresses marks the selection of a node of the reduced vocabulary, the node is placed in the dictionary.

The tree is built in the form of a 2ndimensional array with a sequence of the words in each of which is a root of a tree, which is a sequence of admissible nodes where each branch reflects a continuation of the process, the selection of a word is made while the program is running in a waste of the value of the total intensity which is in excess of the assigned threshold determines the start of signal input. The final decision on the start of input is taken if several successive input segments meet this condition. The decision to end the input is taken if several successive segments have a total intensity below the threshold. Otherwise the segments of low intensity correspond to speech pauses.

The threshold values and the required number of segments in the first and the second case are assigned by the user. Performed parallel with the process of input signal processing is the work by the preprocessor which generates the following:

- the average intensity value for each segment,
- the average intensity of the speech envelope for each segment,
- determining the average closed contiguous vowel segments and the subsequent averaging segments.

The secondary averaging is performed for a group of segments which are located between the first vector and all the subsequent ones less than the segment added by the user. After the segmentation the second index of the secondary averaging segment is assigned by the user. The user may assign a segment following the initial segment or any segment along with the selected segment.

The processing of the speech recognition system is carried out by a second program which is capable of doing the following:

- determining the total number of connected words that were recognized by the preprocessor,
- determining the number of false recognitions, cases in which a phrase is not contained in the dictionary,
- determining the number of correct recognitions that were not contained in the dictionary.
to the units of the subsequent level (the second line). The reference addresses equal to zero determines the end of the phrase.

The input data describing the phrases are represented as a sequence of lines each of which describes a particular phrase (a group of phrases) or part of it and has the form: \(L_{j}A_{1},A_{2},...,A_{i},...,A_{k}[\ast]\) where \(j\) is the current phrase number, \(A_{i}\) — a number or a set of numbers for words that could stand in the \(i\)-th place of a phrase, \(\ast\) is the symbol for the phrase continuation in the next line if its description fails to fit into one line.

Example:

1. \(10,12,11(1,2,3,4,5,6,7,8,9), (13,14)\)
2. \(25(23,21)(13,14)\)
3. \((15)(29,92)\)

The first line of this file describes 36 phrases in which the first place may be occupied by the 10-th or 12-th word, the second by the 11-th, the third by the 1-st, 2-nd,..., or 9-th; the fourth by the 13-th or 14-th word of the fixed vocabulary. The second line describes 4 phrases in which the first place may be occupied by the 25-th word, the second by the 24-th, the third by the 23-d or 21-st, the fourth by the 15-th or 14-th. The program of forming a tree of word compatibility in phrases operates in the dialogue mode and makes it possible to introduce initial data determining the sequence of words in a phrase from the terminal keyboard or from an earlier prepared external file. Taking into account the large variety of identical branches the program eliminates repeated branches which allows the required volume of memory to be reduced 5 to 6 fold. The array of phrases constructed according to the above example has the form

1 2 3 4 5 6 7 8 9 10 11 12 13 14
28 10 12 25 15 -2 1 12 3 4 5 6
4 6 6 21 26 1 8 99 18 18 18 18 18
15 16 17 18 19 20 21 22 23 24 25 26 27 28
7 8 9 -2 13 14 -1 24 -1 23 21 -1 29 92
18 18 18 2 0 0 0 1 23 2 13 18 18 2 0 0

The initially constructed array describing the tree of word compatibility in phrases would contain 306 words instead of 56 as is the case after the optimization.

EXPERIMENTAL RESULTS

The system was tested on phrases of a problem-oriented vocabulary belonging to the language of an air-traffic-dispatcher. The vocabulary contained 120 words. On the material of 140 phrases made up of 3 to 7 words with speaker adaptation the recognition reliability obtained amounted to 93%. The branching factor varied from 1 to 48 and on the average was equal to 13. The system worked in the real time scale. At present the system is in experimental operation.

REFERENCES


Fig. 1. Example of the tree of permissible phrases.