AN INTERACTIVE SYSTEM FOR LANGUAGE IDENTIFICATION

V.B. KUZNETSOV

Department of Applied and Experimental Linguistics,
Moscow Linguistics University, Ostozhenka 38,
Moscow 119034, USSR

ABSTRACT

This paper describes a prototype language identification system LANIS based on hidden Markov modelling (HMM) of the way in which sounds combine together in a particular language. The results of LANIS performance in closed identification tests comprising four languages (English, French, German, Russian) are presented and discussed. The speech material used in the tests was of two kinds: transcribed texts and oral speech. An attempt is made to assess the effects of the length of the unknown speech sample and the structure of HMM (the number of hidden states and sound classes) on identification score.

1. INTRODUCTION

It is a well established fact that frequency distributions and patterns of occurence of phonemes vary from language to language. House and Neuburg !1! have shown that this information can form the basis of a powerful language recognition tool even if the phonetic inventory is reduced to a very small number of gross phonetic categories such as stop, fricative, nonvocalic sonorant, vowel, and silence. To model the phonotactics of the broad phonetic categories the authors applied an HMM technique. The devised procedure for automatic language identification goes as follows. During training session reference HMMs are generated by means of a maximization technique for the languages of interest. Then, using the constructed HMMs a set of maximum likelihood functions is calculated for the unknown input utterance. To cope with the inputs of different length the obtained values of the likelihood function are "normalized", i.e. the natural logorithms of the function values are divided by the number of elements in the utterance. According to the decision rule the unknown utterance is assigned to the language whose HMM has produced the highest score. It follows from the above that the identification takes place within the limits of a closed trial in which it is predetermined that the input utterance is spoken in one of the languages from the fixed set.

Due to the fact that the speech material was restricted to transcribed texts and the performance testing of the identification procedure was rather sketchy, the results obtained by House and Neuburg are primarily of methodological significance.

The present work is designed to develop and evaluate a prototype language identification system implementing the two main features of the approach suggested by House and Neuburg: broad phonetic classes and HMM technique. Two more issues are addressed: the determination of the optimal structure of the HMM and the minimum length of the unknown utterance sufficient for its reliable recognition.

2. SYSTEM DESCRIPTION

The basic structure of the LANIS system is shown in Fig. 1. The implementation strategy was strongly influenced by research needs. The system can deal with two types of input: a string of phonetic symbols or the speech signal. It consists of four program modules and a database.

DATA PREPROCESSING MODULE. The function of this module is to extract from the input data set the information needed for the generation of an HMM. Firstly, the frequency of occurrence of phonetic elements is counted and the average duration for each class of speech segment is determined. Secondly, the expert specifies the structure of HMM: there can be up to 5 states and 10 phonetic classes. Thirdly, the initial values of the HMM parameters are either computed using a standard formula or set by the expert. Then, the input data set and the information obtained become available either for HMM construction or language identification.

HMM GENERATION MODULE AND DATABASE The parameters for each of the language models are estimated from the training data set (maximum size - 2000 elements) using the Baum-Welch algorithm. There is a special mode of the algorithm application (chosen by the expert) where the parameter variances are evaluated as well. The resulting HMM with the associated value of the maximum likelihood function are stored in the database. For the present the storage capacity of the database is limited to 5 languages and 10 HMMs. If a language is represented by more than one HMM, then an average HMM can be calculated and stored. For any pair of languages marked by the expert the significantly different HMM parameters can be discovered.

LANGUAGE IDENTIFICATION MODULE. The identification program tests the unknown utterance against the reference HMMs. Then, judging from the computed values of the likelihood function the program decides the language of the test utterance. The segmentation module is described in a separate section below.

3. EVALUATION OF PERFORMANCE ON TRANSCRIBED TEXTS

To master the identification procedure and to evaluate its efficiency we used transcribed texts in four languages: English, German, French, and Russian. The length of the texts varied from 2.0 to 3.5 thousand phonetic symbols. Two types of transcription were used. One had a phonetic inventory of 6 classes: 1) vowels and sonorants, 2) voiced plosives and affricates, 3) voiced fricatives, 4) voiceless weak fricatives, 5) voiceless strong fricatives, and 6) voiceless plosives and affricates. The other represented the speech stream with 4 categories: they were the same as in the previous inventory, except for the first three classes that had fallen into one category.

For the generation of reference HMMs we used as training data either 1000 element extracts or the whole texts. The size of the test samples was 100 and 300 segments. Before reporting the results of the recognition tests, we want to discuss the data shown in Table 1. For the two types of the test samples Table 1 presents to the left of the sloping line (/) the average frequency of occurence of the phonetic classes (multiplied by 1000) , to the right - the corresponding variation coefficient (the ratio of the standard deviation of the frequency score to the mean, expressed in percent). Each mean was derived from 10 to 12 measurements. It is apparent from this data that the reduction in the size of the test utterance from 300 to 100 elements brings about on average a two-time increase in the dispersion of frequency of occurence. There were a few cases where phonetic classes with the lowest frequency of occurence were not present in the samples of 100 elements at all. Not suprisingly, the

results of language identification on these samples are rather poor. Table 2 displays the identification error rate for each of the languages and the averaged score. Several important characteristics of the tests are specified in Table 2: the structure of the reference HMMs; the total number of identifications (N), distributed more or less equally among the languages; the size of the unknown utterance (V) and its origin, i.e. whether it was taken from the test or training data sets. It should be remembered that in the latter case the training set covers the whole text of a particular language. From the analysis of the data presented in Table 2 two major points can be made. First, four phonetic categories are not enough to discriminate reliably between the languages in question. Second, the size of the test set should be about 300 elements.

It was noticed that the choice of the initial HMM parameter values can play an important role in language identification. However our limited experience prevents us from making any definite suggestion on how to handle this problem.

4. AUTOMATIC SEGMENTATION AND LABELLING MODULE

The LANIS system incorporates a module performing speaker- and language-independent automatic segmentation and labelling (ASL) of continuous speech. Each extracted segment is identified as one of the following: 1) vocalic segment, 2) strong fricative, 3) weak fricative, 4) unidentified fricative, 5) stop, 6) silence. Segmentation and labelling are carried out on the basis of four parameters: fundamental frequency, zero-crossing rate, intensity and a parameter indicating sharp drops in intensity of the speech signal. The parameters are computed once every 16 ms.

The ASL program is a knowledge-based procedure. The rules implemented in the program were initially worked out by an experienced phonetician who analized the traces of the four parameters. In its present form the ASL module works as a two-pass routine. The first pass makes a provisional identification of each 16 ms frame as a member of our set of phonetic categories. Since this results in too many implausible identifications, the second pass attempts to group these labelled frames into likely phonetic segments, producing as output a string of phonetic symbols each of which has a duration value.

The ASL module is still at an experimental stage and no serious attempt has been made to test the degree of the accuracy obtained. None the less we decided to carry out a pilot experimental identification using the outputs of the ASL module. Speakers of the four languages in question were recorded while reading a piece of prose. 30 second passages of four male speakers were used to construct the refrence HMMs. For a given language the training data sets of the other languages served as test samples.

Table 3 lists the normalized values of the maximum likelihood function computed for each speech sample. Negative signs are omitted in the Table. Analysis of the figures in the rows reveals that in all cases the highest value was obtained where a speech sample had been tested against its "native" reference HMM. Evidently, this result does not lead to any general conclusion, and yet it is not felt to be discouraging.

5. REFERENCES

[1] HOUSE A.S. and NEUBURG E.P. (1977), *"Toward automatic identification of the language of an utterance:1: Preliminary methodological considerations"*, J. Acoust. Soc. Am., 62(3), 709-713.

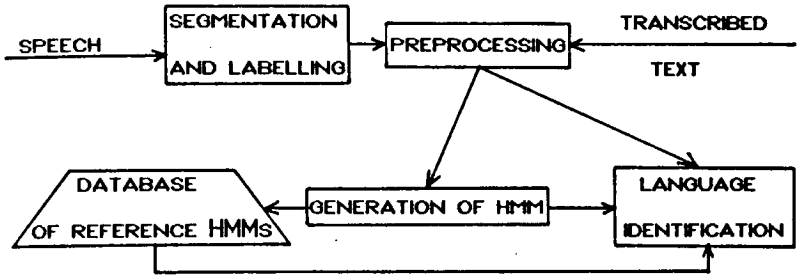

Figure 1. Basic structure of the LANIS system

Table 1. Frequency of occurence / Variation coefficient

LANGUAGE		PHONETIC CATEGORIES					
		1	2	3	4	5	6
ENGLISH	100	602/7	73/30	86/32	42/31	75/20	120/30
	300	610/5	78/22	82/33	50/14	68/27	112/8
FRENCH	100	638/7	75/36	62/36	59/38	62/17	104/24
	300	651/3	82/17	52/29	56/24	52/16	107/17
GERMAN	100	645/4	102/20	30/72	35/51	68/30	120/18
	300	659/3	88/16	35/37	39/25	62/23	116/8
RUSSIAN	100	665/6	48/42	68/39	20/78	59/28	140/22
	300	650/2	55/24	58/23	25/30	62/13	150/19

Table 2. Identification error rate (%)

	4 categories 3 states	6 categories 4 states			
	training V=300 N=42	training V=300 N=40	test V=300 N=37	training V=100 N=40	test V=100 N=54
GERMAN	46	10	30	20	29
ENGLISH	60	30	29	20	50
FRENCH	10	40	10	50	47
RUSSIAN	27	0	20	10	27
AVERAGE	36	20	22	25	38

Table 3. Normalized values of maximum likelihood function (negative signs are omitted)

Test sample of speech	Reference HMMs German	English	French	Russian
German	0.594	0.607	0.612	0.673
English	0.554	0.541	0.547	0.608
French	0.538	0.529	0.523	0.547
Russian	0.427	0.433	0.427	0.397