European Conference on
Speech Technology
Edinburgh, Scotland, UK
September 1987

ON ISOLATED WORDS RECOGNITION SYSTEMS USING TIME DEPENDANT LINEAR PREDICTION

A. Torices-Arguelles*, A. J. Rubio-Ayuso*

ABSTRACT

In this work we have try to use the time dependant linear prediction technique in order to automatically recognize complete isolated words. We have selected the autocorrelation method because it works faster than convariance since it gives a correlation matrix with a high redundancy. We consider a word as a point in a 72-dimensions space, where each dimension corresponds to one coefficient of the time varying linear prediction. In order to obtain the reference patterns we have try two methods. The first one is selecting among the versions of a word the one with minimal distance added among versions. The second method consists on computing the gravity center of the several versions of a word. Betttween these two methods the better has been the second one. As far as the base functions is concerned we have used potential functions, trigonometric functions, Walsh functions and Haar functions.

10.21437/ECST.1987-42

INTRODUCTION

In this work we have studied the application of the technique known as Linear Prediction with Time Depending Coefficients to the recognition of isolated words (Ref 1). The recognition system under testing consists on a data acquisition subsystem, a word delimiter, linear prediction processing, reference pattern elaboration (Ref 7) and true recognition phase. The data acquisition subsystem includes a microphone, an amplifier and a 12 bits analog to digital converter, the speech signal being samples at 10 kHz during 1.5 s. The word delimiter algorithm is an explicit one (Ref 4) and it is a modification of the proposed by Rabiner and Sambur (Ref 6). A detailed description of this algorithm can be found in Ref 5.

In the linear prediction processing we have implemented a linear prediction algorithm with time depending coefficients and we tested several alternatives (base functions, lenght of the developments, etc). The reference pattern elaboration is carried out in two ways: one of them by using the pattern with a minimum sum of distances (MSDRP) and the other by computing the gravity center of the various utterances for each word (GCRP), (Ref 7).

BRIEF DESCRIPTION OF LINEAR PREDICTION WITH TIME DEPENDING COEFFICIENTS

The basic idea is to express each prediction coefficient as a time function (Ref 1):

$$s(n) = -\sum_{i=1}^{p} a_i(n)\, s(n-i) + G\, u(n) \qquad (1)$$

*University of Granada, Spain.

where s(n) is the sample n-th, u(n) is the excitation to the system, G is the gain of the system and $a_i(n)$ are the time depending coefficients. The way in which the coefficients depend on time is expressed as a linear combination of a set of base time functions:

$$a_i(n) = \sum_{k=0}^{q} a_{ik}\, b_k(n) \qquad (2)$$

where $b_k(n)$ is the base function of order k.

Two important methods in the implementation of linear prediction are autocorrelation and covariance methods (Ref 1). In both cases, a system of linear equations must be solved in order to determine the prediction coefficients. In matrix form, this equations sytem is $\Phi A = -\Psi$. As can be seen in the covariance method, Φ is a symmetrical matrix with (1/4)p(p+1)(q+1)(q+2) distinct elements (Ref 1). In the case of autocorrelation method we have a number of distinct elements to compute of $p(q^2+2q+1)-(1/2)(q+1)q$ (Ref 1). This number is quite lower then the one for the covariance method. This point is very important as most of the time spent in the computations is used to determine the elements of matrix Φ . Thus, the difference is enough to decide us to use the autocorrelation method.

DETERMINATION OF P AND Q

In the expansion (1) we must determine the value of p, by using our knowledge on linear prediction with constant coefficients. It must be an integer lightly greater than sampling frequency in kilohertz (Ref 2). As samplig frequency is 10 kHz, we have taken p=12. This choice base in linear prediction with constant coefficients is justified by the good results we obtain in linear prediction with time depending coefficientes. In the expansion (2) we must determine the value of q (number of base functions, $b_k(n)$). To do this, we use experimental criteria, taking into account that the value of q has a great influence on the computing time. In order to determine q we have studied the recognition system with a vocabulary consisting on 15 words with 4 versions per word, taking as base functions the trigonometric ones, euclidean distance and making an estimation of the minimum rate of recognition (i. e., without taking into account the influence of the method of forming the reference patterns). We have found that a good value for q (trade-off between short computing time and high recognition rate) is q=5.

DISTANCE BETWEEN WORDS. REFERENCE PATTERNS ELABORATION

We can consider each word as a point in a space of n=p(q+1) dimensions (each dimension coresponding to one of the q+1 coefficients of the development of every prediction coefficient). In order to evaluate the proximity among words we have to define a distance (Ref 7). If $X(x_1, x_2, \ldots, x_n)$ and $Y(y_1, y_2, \ldots, y_n)$ are two words, we have used the following definitions:

-Euclidean distance: $D_e(X,Y) = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2}$ (3)

-Maximum distance: $D_m(X,Y) = \max\{|x_1-y_1|, \ldots, |x_n-y_n|\}$ (4)

-Relative distance: $D_r(X,Y) = \sum_{i=1}^{n} (x_i-y_i)/(x_i+y_i)$ (5)

By using the same situation and procedure as when determining the optimum value of q, we have found that euclidean distance is better than relative distance and this one is better than maximum distance.

In order to determine the effect of electing one of the two types of reference patterns (MSDRP and GCRP), we have used the "standard" recognition system before described. In this case we have carried out a true recognition process. For MSDRP the recognition rate obtained was 83.3 %, while the recognition rate was of 96.6 % in the case of GCRP.

BASE FUNCTIONS. VOCABULARIES

We have tested the following set of base functions, studying their influence in the recognition rate of the standard system, for both MSDRP and GCRP.
-Trigonometric functions, defined as (Ref 1) $u_k(n)=\cos(2\omega kn)$ if k is even, and $u_k(n)=\sin(2\omega kn)$ if k is odd, with $\omega=\pi$. These functions give a recognition rate of 83.3 % for MSDRP and 96.6 for GCRP.
-Potential functions, defined as (Ref 1) $u_k(n)=n^k$. The results are 51.6 % and 65.0 % for MSDRP and GCRP, respectively.
-Walsh functions.
Defined as (Ref 3) $WAL(s,t)=(-1)^{\sum_{k=1}^{m}(s_k+s_{k-1})t_k}$ (6)

where $t=\sum_{k=1}^{m} t_k/2^k$ and the order $s=\sum_{k=0}^{m} s_k 2^k$ and $t=n/N$, where N is the number of samples in a word and n is the index on the sample. The use of Walsh functions has some advantages when they are electronically implemented as they have only two values (as binary digitasl circuits). The experience has shown us that best results are obtained for a range of orders from s=4 up to s=9. By using the "standard" recognition system, the rate of correct recognition was 81.6 % in the case of MSDRP and 99.3 for GCRP.
-Haar functions, defined as
$HAR(0,t)=1$

$$HAR(2^p+m,t) = \begin{cases} (\sqrt{2})^p & \text{for } m/2^p < t < (m+1/2)/2^p \\ -(\sqrt{2})^p & \text{for } (m+1/2^p)/2 < t < (m+1)/2^p \\ 0 & \text{in other case} \end{cases} \quad (7)$$

with p=0, 1, ... and m=0, 1, ..., 2^{p-1} and the order $s=2^p+m$. The results have been of 83.3 % for MSDRP and 100 % for GCRP.
Here we present the vocabularies we have tested and the results we have obtained with them.
-Numerical vocabulary, formed by the following spanish words: 'CERO', 'UNO', 'DOS', 'TRES', 'CUATRO', 'CINCO', 'SEIS', 'SIETE', 'OCHO', 'NUEVE', 'MAS', 'MENOS', 'IGUAL'.
-Provinces vocabulary, formed by the following spanish province names: 'ALICANTE', 'BALEARES', 'BURGOS', 'CACERES', 'CASTELLON', 'HUESCA', 'LOGRONO', 'LUGO', 'MURCIA', 'ORENSE', 'SEVILLA', 'SORIA', 'VALENCIA', 'VIZCAYA', 'ZARAGOZA'.
-Mixed vocabulary, as the union of the two preceding vocabularies.
The results are sumarized in the following table.

Functions	Patterns	Numerical V.	Provinces V.	Mixed V.
Trigonom.	MSDRP	59.6	83.3	68.0
	GCRP	90.4	96.6	92.0
Potential	MSDRP	48.0	51.6	39.0
	GCRP	50.0	65.0	58.0
Walsh	MSDRP	69.2	81.6	67.0
	GCRP	96.1	93.3	89.0
Haar	MSDRP	69.2	83.3	71.0
	GCRP	86.5	100.0	85.0

CONCLUSIONS

-In this work we have verified that linear prediction with time depending coefficients is a suitable approach to isolated words automatic recognition.

-We have selected autocorrelation method versus covariance method because it is faster.

-The parameters p and q have been fixed to 12 and 5, respectively, as they provide the best experimental results.

-The best distance measurement is euclidean distance, among the several we have tested.

-The set of reference patterns has been elaborated in two ways: minimal sum of distances and gravity center. This last method seems to be more suitable than the other one.

-As far as base functions is concerned, we have tested trigonometrics, potentials, Walsh functions and Haar functions. Only potentials gave a recognition rate lower than 90 %.

-Finally, in this work has been verified that the recognition rate of recognition system could depend on the vocabulary. Effectively, numerical vocabulary gives poorer results than provinces vocabulary.

REFERENCES

1. M. G. Hall, A. Openheim, A. S. Willsky "Time varying parametric modeling of speech" Signal Processing, 1983

2. L. R. Rabiner, R. W. Schafer "Digital processing of speech signal", Prentice-Hall, 1978.

3. R. G. Beauchamp "Walsh functions and their applications" Academic Press, 1975

4. M. C. Carriòn-Pèrez, A. J. Rubio-Ayuso "Un detector explìcito de principio y final para palabras aisladas" Informàtica y Automàtica, n 60, 1984.

5. L. F. Lamel, L. R. Rabiner, A. E. Rosemberg, J. G. Wilpon "An improved endpoint detector for isolated word recognition" IEEE Trans. ASSP, Aug. 1981.

6. L. R. Rabiner, M. R. Sambur "An algorithm for detecting the endpoints of isolated utterances" Bell Sys. Tec. vol. 54, Feb. 75

7. J. T. Tou and R. C. Gonzalez "Pattern Recognition Principles" Addison-Wesley Pub. Comp., 1975.