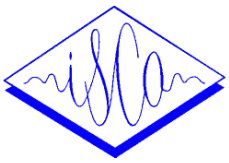

4th European Conference on
Speech Communication and Technology
EUROSPEECH '95
Madrid, Spain, September 18-21, 1995

ON THE USE OF THE DERIVATIVE OF THE POLE TRAJECTORIES OF THE LPC ANALYSIS PARAMETER SEQUENCE AS AN ALTERNATIVE TO DELTA PARAMETERS

F. Freitag, E. Monte, J. Hernando
Dpt.TSC.Universitat Politécnica de Catalunya Barcelona.Spain
E-mail:freitag@tsc.upc.es

ABSTRACT

In this paper a new approach for modelling time variations in the speech spectra is presented. We propose to approximate the trajectories of the frequency and amplitude of the poles of the LPC spectra with exponential functions. The obtained time constants of the exponential functions are incorporated in the observation vector and used for recognition in an HMM based recognition system. The performance of the new parameters is tested using a database of connected digits. The recognition rates obtained with the new parameters are compared to results obtained with delta parameters.

10.21437/Eurospeech.1995-218

1. INTRODUCTION

For some years it has been known that adding the knowledge of temporal changes in the spectra improves the recognition results in HMM based recognition systems. This information is usually incorporated in the form of the slope of the parameters, i.e. the linear regression coefficients [1], or can also be seen in the introduction of the temporal slope of the observation sequence of the delta cepstrum [2]. These parameters double the dimension of the observation vector as for each spectral parameter its corresponding delta parameter is incorporated. A higher number of parameters in the observation vector, however, increases the number of parameters of the HMMs that have to be estimated. Also the use of higher order differences like delta-delta parameters yield a slight improvement in recognition, but require a sequence of several speech frames for their calculation [3]. In this paper we propose an alternative way to model time-varying spectra. With the new parametrization the dimension of the slope information vector can be reduced.

2. MODELLING TIME VARIATIONS OF SPEECH SPECTRA.

2.1 Description

The parameters that we propose consist of the time constants (alfas) of exponential functions which model the trajectories of both the pole frequencies and amplitudes of the LPC spectrum. The LPC model is an all-pole model which can capture the resonant frequencies, or formants, but not the zeros of the spectrum. The pole frequencies correspond to peaks in the LPC spectrum. When the bandwidth of a peak is small then the pole frequency corresponds to a formant frequency. Spectral peaks are known to carry information valuable for speech recognition. In Fig.1 the trajectories of 4 pole frequencies of the word "one" are shown. It can be observed that slowly varying pole frequencies are obtained especially for the two poles in the lower part of the spectrum. Variations of the pole frequencies are smooth for voiced regions of speech. During unvoiced regions of speech or silence the pole frequencies of the LPC model can change rapidly between neighbouring frames. In this case the pole frequencies may not correspond to formants, but represent large-bandwidth peaks in the spectrum. Formant continuity also is not given in many cases for transitions between phonemes. During transitions significant changes between neighbouring pole frequencies can be observed. Nevertheless, it is believed that both smooth as well as changing trajectories of the pole frequencies carry information that can improve speech recognition results. The trajectories of the pole frequencies within a small number of frames can be approximated by exponential functions as shown in Fig.2. Besides of the change of the pole frequencies, the pole amplitudes are considered. The trajectory of the pole amplitude might contribute to recognize transitions between voiced and unvoiced regions. The trajectories of the pole amplitudes are also approximated by an exponential function.
Thus, an exponential function models explicitly spectral changes within a small number of frames and therefore might represent an alternative for the delta parameters. Since only 3 or 4 of the poles of the speech spectrum are considered to be important for recognition, the size of a parameter vector consisting of the time constants is limited and can be below the length of the vector of delta parameters. In this case

the number of parameters to be estimated for the HMM is smaller and the relation of training data to model parameters is better, which might lead to a better estimation of the parameters of the model.

2.2 Implementation

For each speech frame we can obtain the spectral envelope using LPC analysis. The poles of the LPC spectrum are obtained numerically as roots of the equation

$$1+\sum_{i=1}^{p} a_i z^{-i} = 0 \qquad (1)$$

We used a LPC transfer function of order p=12. The pole frequency F_i of the ith pole z_i is:

$$F_i = \frac{1}{2\pi T} * \arg(z_i) \qquad (2)$$

For each frame the poles are ordered from the lowest frequency to the highest. Thus the gaps are reduced when determining the trajectories of the poles. Large changes between neighbouring pole frequencies can indicate unvoiced regions or phoneme transitions. We have restricted the minimum pole frequency to 300 KHz, since the poles found below 300 Hz most probably represent energy and not formants. We model the trajectories of the pole frequencies within a window of 5 frames by exponential functions of
the form k*exp(alfa*n) n=-2, ..,2,
and alfa representing the slope of the trajectory. In the same way the trajectory of the pole amplitude is modelled.The pole amplitude is represented by the radius of the complex pole. Both alfa parameters (which represent the variation of frequency and amplitude) are appended to the observation vector that is used for recognition.

3. DATABASE AND PREPROCESSING

The database used was the TI. Sequences of isolated digits and of 7 connected digits spoken by 50 male talkers were used for training the HMMs. For testing sequences of 7 connected digits spoken by 50 different male talkers were used. Speech data was sampled at 8 KHz, pre-emphasized and Hamming windowed. The analysis window was 25 ms, the analysis window shift 10 ms. The observation vector consisted of 12 mel-frequency cepstral coefficients (MFCCs) to which delta parameters or alfa parameters or both were appended. The recognition system was the HTK. Each digit was represented by a CDHMM of 10 states with 1 mixture for the mel-frequency cepstral coefficients and delta coefficients and 1 and 3 mixtures, respectively, for the alfas. Each parametrization was represented by a separated stream. Only transitions between neighbouring states in a HMM were allowed. The grammar applied for recognition defines the start and end of each recognition sequence as "silence", but does not restrict the number of digits contained in the sequence.

4. EXPERIMENTAL RESULTS

Experiments with various observation vectors were performed to evaluate the influence of the alfas on the recognition results. The recognition rates are given in % correct. A preliminary experiment was done in order to ascertain if the parameters that we propose are capable of extracting enough information from the speech signal. The experiment was done using as observation vector for the HMMs only the alfa parameters (i.e. variation of the pole frequency and amplitude). In this experiment the observation vector consisted of one stream of frequency parameters (af) followed by the amplitude parameters (ar). The recognition results with 3 mixtureswere 2af-2ar: 66,48%; 3af-3ar: 70,05%; 4af-4ar: 71,34%, from which we can conclude that these parameters offer pertinent information for the recognition process. It can be observed that the recognition rates improve consistently as the order of the analysis grows.
The next experiment was done in order to ascertain that the use of these parameters can give improvements in comparison to the use of MFCC parameters only. The results are shown in Table I. The recognition was done using an HMM with two streams, one stream consisted of 12 MFCC coefficients and the other of the alfa coefficients of different length. The experiment was performed twice using one mixture for the two streams in one case and in the other case, one mixture was used for the MFCC coefficients and three mixtures were used for modelling the stream of alfa parameters. The modelling of the MFCC stream was left unchanged in order to compare the influence of different modelling of the alfa stream.

	MFCC	M_2af_2ar	M_3af_3ar	M_4af_4ar
1Mix	96.07	96.32	96.24	96.22
3Mix (alfas)	96.07	96.44	96.44	96.40

Table I

From Table I, one can see that the use of the alfa parameters improve the recognition results in comparison to the use of the MFCC parameters alone. In can be seen that the use of an additional stream of length four (M_2af_2ar) with 3 mixtures gives an improvement of 0.37% in the error rate which in terms of number of errors is a reduction of 14 errors.
The other experiment was designed for determining the improvement given by the use of the alfa parameters over a system which already uses a derivative information. We decided to compare a baseline system consisting of the MFCC and the delta parameters with a version that in addition to the MFCC and delta parameters used the alfa parameters. In table II we present the results:

	M_D	MD_2af2ar	MD_3af3ar	MD_4af4ar
1Mix	98,10	98,12	98,16	98,24
3Mix (alfas)	98,10	98,16	98,20	98,22

Table II

It can be seen in Table II that the use of the delta parameter (M_D) gives an improvement on the recognition rate of about 2%. This improvement is better than the one obtained when using the alfa parameters (M_2af_2ar), nevertheless it must be emphasised that the delta parameters mean a stream with vectors of length 12 while the best improvement with the alfa parameters (M_2af_2ar) was obtained using a stream of length 4. In Table II the use of the alfa parameters also improves the recognition results compared to the use of the MFCC and delta parameters alone and this improvement is consistent in the sense that the use of the alfa parameters constantly yields an improvement.

5. CONCLUSIONS

In this paper a new approach for modelling spectral changes was presented. The experimental results showed that the time constants (of frequency and amplitude variations) used as parameters for recognition of connected digits carry information and that the recognition results can be improved in comparion to the baseline system. The size of the vector with the alfa parameters is small in comparison to the vector of delta parameters, so the increase in computation and memory requirements is small.

6. REFERENCES

[1]Furui, S. : *Speaker-Independent Isolated Word Recognition Using Dynamic Features of Speech Spectrum.* IEEE ASSP-34(1):52-59, February,1986.

[2]Rabiner, L.R. : *High Performance Connected Digit Recognition Using Hidden Markov Models.* IEEE ICASSP88. April, 1988.

[3]C.H.Lee, .Rabiner L, Pieraccini R. : *Speaker Independet Continuous Speech Recognition Using Continuous Density Hidden Markov Models.* Speech Recognition and Understanding, NATO ASI.Springer Verlag 1990

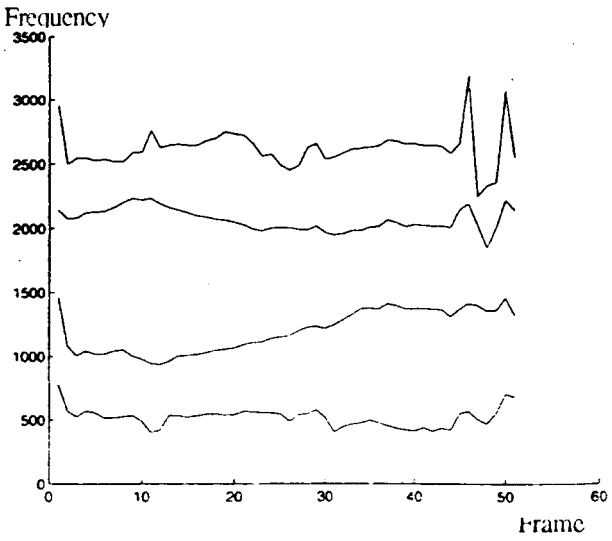

Fig. 1. Trajectories of four pole frequencies of the LPC analysis of the word "one".

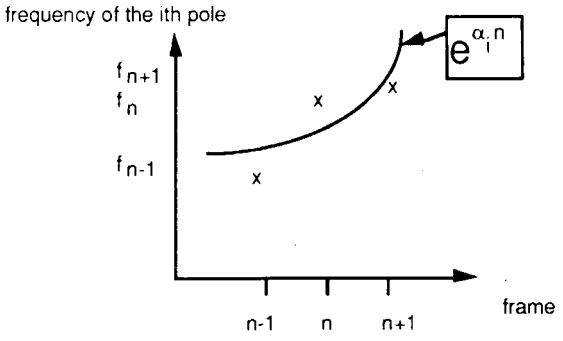

Fig. 2. The trajectory of the ith pole frequency is approximated by the parameter α_i.