SPEAKER RECOGNITION BY MEANS OF SHORT SPEECH SEGMENTS ANALYSIS USING TIME-VARYING LINEAR PREDICTION IN LATTICE FORMULATION

JANUSZ ZALEWSKI
Technical University of Wroclaw,50-370 Wroclaw,Poland

Abstract - This paper presents the method of speaker recognition.In this technique the reflection coefficients obtained from short speech segments by means of time-varying linear prediction in lattice formulation procedure was utilized as the identification parameters and the minimum of time-average spectral difference between the corresponding short speech segments was the recognition criterion. The results of the recognition task using this method has been compared with others.

INTRODUCTION

The procedure utilized in any approach to speaker identification could substantially influence the resulting level of the ultimate identification accuracy of the used technique.In this regard,two distinctly separate operational phases may be identified for any approach of this type. First,the identification parameters and associated measurement technique must be chosen.Secondly,statistical distance measurement and associatted decision criterion must be identified and evaluated.

In the research we have previously reported /1,2,3/ the speaker has been represented by some phonemes,or short segments of speech regarded as the reference samples. The minimum cumulated distance measure between corresponding test and reference samples was the decision criterion.The method we have presented may be succesfully used as a procedure for identifying individuals from their speech.- at last under laboratory conditions.The parameter sets,chosen for speech waveforms parametrisation was the predictor coefficients and the cepstrum obtained via parametric analysis of speech signals,using an autoregressive model.For linear predictive coding,it is asumed that the signal is stationary over the time of analysis,and therefore the coefficients given in this model are constants.However speech signal to be modeled,even in short segments as are the phonemes, is not stationary.Therefore it seems to be reasonable to use an autoregressive signal modelling in which the coefficients are time-varying i.e. each coefficient in the model is allowed to change in time,by assuming it is a

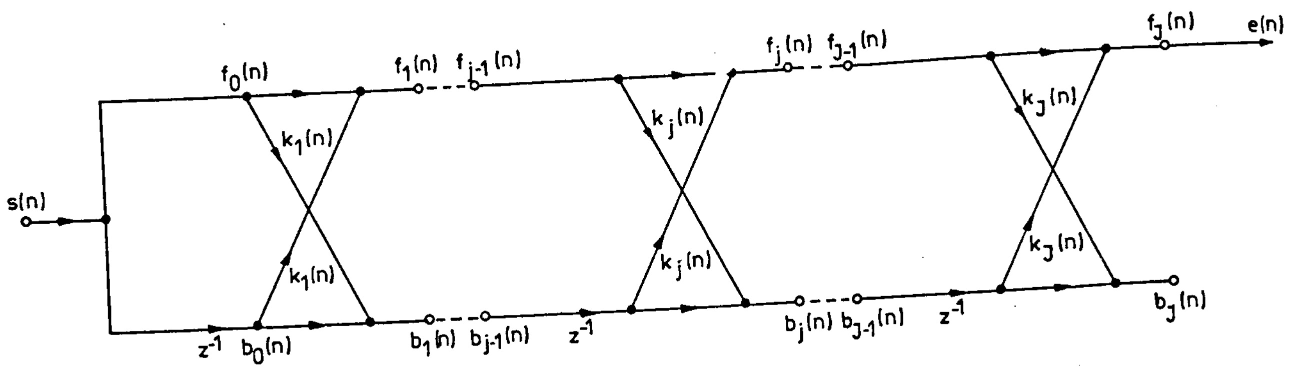

Fig. 1. The inverse filter in lattice form

linear combination of some set of known time functions. This model allows for continuously changing behavior of the signal, such propriety should enable the model to have possible better accuracy and allows for the analysis over longer data windows.

THEORETICAL BASES

The fundamental works on linear prediction of time-varying signals was done by Liporace /4/, Hall/5/, Hall et al. /6/ Turner and Dickinson/7/, and Jurkiewicz/8/. In present research it was utilized time-varying linear predictor in lattice formulation done by Jurkiewicz, who has reformulated the linear predictive technique to estimate the variable parameters $k_j(n)$ of the inversef filter in lattice form, as depicted in Fig.1, rather than in direct form. That is the inverse filter is in lattice form, and its parameters $k_j(n)$ are estimated by minimizing the given (after Burg) MSE norm /8/.

$$D_j = \sum_{n=o}^{L} (f_j^2(n) + b_j^2(n)) \qquad (1)$$

where

$$f_j(n) = f_{j-1}(n) + k_j(n).b_{j-1}(n-1), \qquad (2)$$

$$b_j(n) = b_{j-1}(n-1) + k_j(n).f_{j-1}(n) \qquad (3)$$

$$f_o(n) = b_o(n) = s(n) \qquad (4)$$

$$k_j(n) = \sum_{l=o}^{N-1} q_{lj} \cdot u_l(n) \qquad (5)$$

and $u_l(n)$ are the time series (eg trigonometric functions as in Fourier series)

Denoting

$$f_j'(n) = f_{j-1}(n) \qquad (6)$$

$$b_j'(n) = b_{j-1}(n-1) \qquad (7)$$

and omitting the subscript j to simplificate notation, equations (1) - (5) become

$$D = \sum_{n=o}^{L} (f^2(n) + b^2(n)) \qquad (8)$$

$$f(n) = f'(n) + k(n).b'(n) \qquad (9)$$

$$b(n) = b'(n) + k(n).f'(n) \qquad (10)$$

$$k(n) = \sum_{l=o}^{N-1} q_l \cdot u_l(n) \qquad (11)$$

Minimizing the error D with respect to each coefficient q_l by setting

$$\frac{\partial D}{\partial q} = 0 \quad , l=o,1,..N-1 \qquad (12)$$

yelds the linear normal equations

$$q_l \cdot R_{il} = S_i \ , i= 0,1, ..N-1 \qquad (13)$$

where

$$R_{i,l} = \sum_{n=o}^{L} u_i(n).\ u_l(n).d(n) \qquad (14)$$

$$S_i = \sum_{n=o}^{L} u_i(n) \ . \ c(n) \qquad (15)$$

$$c(n) = -2 f'(n) \ . \ b'(n) \qquad (16)$$

$$d(n) = f'^2(n) + b'^2(n) \qquad (17)$$

The coefficients q are specified by the equation (13), or in matrix form

$$R \times Q = S \qquad (18)$$

Below is the complete algorithm for described time-varying linear prediction in lattice formulation:

In each j-th step of analysis (i.e. in j-th section of the filter:

- the matrix R and the vector S are computed from equations (14),(15),(16) and(17)
- the set of equations (13) or (17) are solved,
- the signals $f'(n)$ and $b'(n)$ are filtered according eq (9) and (1o) in the lattice system (Fig.1)

This set of operations is repeated in each succeding step j, for j=1,2 to J.

In the experiments described in this paper each of 1o reflection coefficients $k_j(n)$ was evaluated, according eq (9), as the linear combination of 3 or 5 time functions.

$$u_i(n) = \begin{cases} 1 & i=o \\ \cos(n(i+1)\pi/m), & i \text{ odd} \\ \sin(n\ i\ \pi / m), & i \text{ even} \end{cases} \qquad (19)$$

$i= o,1,..., N_T$,

$N_T = \Omega_c M / T$,

8 Ω_c = digital cut-off frequency of $k_i(n)$ spectrum,

M = period of the $u_i(n)$ functions set.

For each sample, from the filter parameters trajectories $k_i(n)$, the 10 sets of 40 cepstrum coefficients was evaluated; each set at one of 10 equidistant time instants. From the cepstrum coefficients of the reference sample c_i, and those of the test sample c_i, the time average spectral difference (i.e. the time-average Euclidean distance of c_i and c_i sets, multiplied by 10/ln 10) was computed. The time-average spectral difference is

$$d = 10 \cdot (\log e) \cdot (L^{-1} \sum_{l=1}^{L} 2 \sum_{k=1}^{K} (c_k - c_k)^2)^{1/2}$$

where $c_k = c_k(l)$ are the cepstral coefficients of the test sample, $c_k' = c_k\ l)$ are the cepstral coefficients of the reference sample, l - succeding time instant at which the cepstra are evaluated, L - number of time instants at which the cepstra are evaluated (here 1o), K number of cepstral coefficients representing the sample (here 4o)

EXPERIMENTAL PROCEDURE

Subjects were the same 20 male speakers as in speaker recognition experiments /3/, where speakers have been represented by some phonemes, and the parameter set for speech waveforms parametrisation, was the predictor coefficients, obtained using autoregressive model with constant coefficients. The speech material consisted of 240 utterances, including 2 repetitions of 6 Polish vowels /a, o,e,i,u,y/ each spoken in two contexts. The speech signal was manually segmented, to detach the vowels,pre-emphasized 6 dB per octave,low -pass filtered with 5 kHz cut-off frequency,sampled at a rate of 10 ksamples per second and converted into digital form by means of 8-bit A/D converter.The segments of 100 ms duration was processed to obtain 10 time-varying reflection coefficients trajectories. To compare test and reference samples, the average spectral differences between them was computed. In the first speaker recognition experiment the speakers were represented by a single phoneme,in the second by pairs (15 combinations),in the third by three (20 comb.)and in the 4-th - by four phonemes (15 comb.). The minimum distance criterion was used as the decision rule,i.e. the m-th test sample was considered to be identical with the n-th reference,if for $j=1$ to 20 and $j \neq n, d_{mj} \; d_{mn}$ where d_{mn} denote the distance measure (average spectral difference) between the m-th test and the n-th reference sample.

RESULTS AND CONCLUSIONS

The detail results of all 112 recognition experiments will be presented at the Congress.Hereafter are presentedsome typical results obtained in two experiments,first where the speakers were represented by phoneme "i" and second where the representation included phonemes "i" and "a".The results are compared with results of experiments with parametrization obtained using constant model. In table 1, the average recognition errors are shown; subscript i denotes the first experiment,subscript a,i denotes the second,subscript v - variable model, subscript c - constant model.

TABLE 1. RECOGNITION ERRORS

$E_{i,v}$	$E_{i,c}$	$E_{a,i,v}$	$E_{a,i,c}$
0.050	0.183	0.000	0.008

Several conclusions can be drawn from the result of this research.First it may be stated that representation of speakers by short speech segments and comparison of corresponding segments may be succesfullly used in a procedure for identifying individuals from their speech. Second, the time-varying linear prediction procedure in lattice formulation is a convenient form of the parame trisation procedure.Finally,it is shown that augmenting the number of speech segments representing the speaker,could possibly result in an even more powerful identification process.

REFERENCES

/1/ J.Zalewski et al.,"Speaker recognition by means of linear predictor coefficients", Proc 9ICA,Madrid,1977,I-32,438

/2/ J.Zalewski et al.,"An application of the Itakura distance measure for the estimation of the predictive coded pattern similarity" /in Polish/,Proc XXIV Open Acoust. Seminar. ,Gdańsk,1077,Part 1 ,pp380,381

/3/ J.Zalewski,"A comparison of the effectiveness of some distance measures in spe speaker recognition experiments", Paper on the Speaker Recognition Working Group on theTenth International Congress of Phonetic Sciences, Utrecht, 1983 (also Reports of Techn.Univ.of Wroclaw,I-28/PRE-033/183 Wroclaw 1983)

/4/L.A. Liporace,"Linear Estimation of Nonstationary Signals", J.Acoust.Soc.Am., vol 58,no 6,December 1975,pp 1268-1295

/5/ M.G. Hall,"Time-Varying Linear Predictive Coding of Speech Signals,S.M. Thesis, Dept of Electrical Engineering and Computer Science, Mass.Inst.of Techn.,Cambridge, Massachusettes,August 1977

/6/ M.G.Hall et al. "Time-Varying parametric modelling of speech.", Proc. of the 1977 IEEE Conf. on Decision and Control, New Orlean, Dec 1977, pp 1095 - 1091

/7/ J.Turner E.Dickinson,"Linear prediction applied to time-varying all-pole signals",Proc.1977 ieee Int.Conf.on Acoust. Speech and Sign. Proc.,Hartford,Conn., 1977,pp 750 - 753

/8/ J.Jurkiewicz " Time-varying Linear Prediction in Lattice Formulation for Speech Analysis",Ph.D.Dissertation,Inst. of Telecommunication and Acoustics,Techn. University of Wroclaw,Wroclaw 1984.