NEW HOS-BASED PARAMETER ESTIMATION METHODS FOR SPEECH RECOGNITION IN NOISY ENVIRONMENTS

Asunción Moreno, Sergio Tortola, Josep Vidal, José A. R. Fonollosa

Dpt. Signal Theory and Communications
Universitat Politècnica de Catalunya
Barcelona, Spain

ABSTRACT

In this paper the problem of recognition in noisy environments is addressed. Often, a recognition system is used in a noisy environment and there is no possibility of training it with noisy samples. Classical speech analysis techniques are based on second-order statistics and their performance dramatically decreases when noise is present in the signal under analysis. In this paper new methods based on Higher-Order Statistics (HOS) are applied in a recognition system and compared against the autocorrelation method. Cumulant-based methods show better performance than autocorrelation-based methods for low SNR.

1.INTRODUCTION

In the last few years, there has been an increasing interest in the application of Higher-Order Statistics in signal processing. Signal analysis systems based on cumulants and their Fourier Transform are a powerful tool since they have very useful properties. Detection of non linearity, identification of non minimum phase systems and immunity to white or colored Gaussian noise are some examples [1]. Voiced-unvoiced classification, phoneme segmentation or pitch estimation are problems that have been addressed using HOS. Results show that speech signal can be characterized not only by its autocorrelation but also by its third- and fourth-order cumulants.

Immunity to noise is an important feature in any signal processing system. Classical speech analysis techniques are based on second-order statistics and their performance dramatically decrease when noise is present in the signal under analysis. Cumulants of order greater than two are zero for white and colored Gaussian noise. Analysis of noisy speech signals based on HOS permits to separate the speech from noise and in this paper this property is used.

We address the problem of recognition in noisy environments. Often, a recognition system is used in a noisy environment and there is no possibility of training it with noisy samples. Paliwal [4] made use of HOS in a recognition system and he showed that results remain constant under a great variability of SNR.

However, in high SNR conditions, the performance of the autocorrelation method was clearly better. This fact can be related to the following points:

*Cumulants-based normal equations can give a non minimum-phase filter. The estimation of the AR parameters can arise a non stable solution in some frames of speech signals.

*The variance of the estimation is greater than in the case of autocorrelation.

To avoid those two problems we have developed new methods to estimate the AR parameters that give stable solutions and have less variance. These methods use a linear combination of cumulant slices (WS) [2] or an unique slice (DS) [3]. In this paper those methods are compared in a speech recognition task against order three and four Yule-Walker based equations and autocorrelation method.

2. HOS BASED PARAMETER ESTIMATION METHODS

Consider the speech signal generated by a causal and stable AR (p) model with added noise :

$$y(n) = \sum_{i=1}^{p} a(i)\, y(n-i) + v(n) \qquad (1)$$

$$z(n) = y(n) + w(n)$$

The input process v(n) is a zero mean non-Gaussian i.i.d. sequence, with k-order cumulant $\gamma_{k,v} \neq 0$. The additive noise w(n) is independent of v(n), zero mean, and either Gaussian with unknown power spectrum or non-Gaussian with $\gamma_{k,w}=0$. The filter H(z) is exponentially stable. The above conditions imposed over w(n) guarantee $C_{k,w}=0$ and $C_{k,z}=C_{k,y}$.

In this section we consider three algorithms named Yule-Walker, W-slice and 1-D slice.

This paper has been supported by Spanish Government grant TIC 92-0800-C05/04

0-7803-2431-5/95 $4.00 © 1995 IEEE

Higher-order Yule- Walker algorithm.

From the well known equation [5]:

$$\sum_{l=0}^{p} a(l)\, C_{k,y}(m-l\,,\, k_o,\, 0,\, ..0) = \gamma_{k,y} h^{k-2}(-m)\, h(-m+k_o) \qquad (2)$$

We obtain

$$\sum_{l=0}^{p} a(l)\, C_{k,y}(m-l\,,\, k_o,\, 0,\, ...0) = 0 \quad \text{if } m>0, \text{ or } m>k_o \qquad (3)$$

To solve those equations is necessary to concatenate p+1 slices, k_o=-p..., 0 in (3) because a single slice does not guarantee a full rank system of equations [5].

Equations (3) must be solved using cumulants estimates and the solution using LS or TLS may yield a unstable solution. In this paper we test two alternatives to improve the stability: to increase the number of slices (k_o=-p-M,...0 and M>0), or to increase the number of equations per slice. In the later case we consider three alternatives: (S1) p equations per slice (minimum), m=0...,p; (S2) the negative slices (k_o<0) have p+1 equations, m=0,...,p; (S3) the negative slices have p-k_o equations, m=l+k_o,...p.

W-slice algorithm.

The w-slice [2] algorithm is based on the following weighted sum of cumulant slices:

$$C_w(i) = w_2 C_{2,y}(i) + \sum_{j=-L}^{N} w_3(j)\, C_{3,y}(i,j) + \sum_{j=-l}^{N}\sum_{k=-L}^{N} w_4(j,k)\, C_{4,y}(i,j,k) + \ldots \qquad (4)$$

and it is developed in three steps:

a). Choose w_2, $w_3(j)$, $w_4(j,k)$, such that:

$$C_w(i) = 0, \qquad i=-P,\ldots,-1$$

$$C_w(0) = 1 \qquad (5)$$

being P ≥ p , N ≥ 0 and L≥ p+M, where M is the over determination.

b). Estimate the first P terms of the impulse response from the weighted cumulant $C_w(i)$.

$$h(i) = C_w(i) \qquad i=1,\ldots P.$$

c). Solve the filter coefficients from the following equation:

$$\sum_{l=0}^{p} a(l)\, h(i-l) = 0, \quad i=1\ldots P \qquad (6)$$

Step a) is solved by LS. To solve (6) is preferable to use LS or TLS than backsustitution to minimize the variance of the solution and obtain stable solutions (P=p+M, M>0). We have also considered the autocorrelation method:

$$\sum_{k=0}^{p} a(k)\, R_{hh}(k-l) = 0, \quad k=1\ldots P \qquad (7)$$

that assures a stable solution

1-D slice algorithm

This algorithm obtains the AR coefficients from a single cumulant slice $C_{k,y}(m, k_o,\ldots,0)$. An one-dimensional slice does not guarantee a full rank system of equations and for this reason is not reasonable to solve (3) directly; the solution may not be unique or stable. Instead, we consider the (deterministic) autocorrelation of the cumulants to form a Toeplitz matrix with a stable solution.

If we multiply (2) by the one-dimensional slice $C_{k,y}(m-l', k_o, 0\ldots,0)$ and we sum in an interval with m>0, we obtain:

$$\sum_{l=0}^{p} a(l) \sum_{m>0} C_{k,y}(m-l;k_o, 0\ldots,0)\, C_{k,y}(m-l';k_o, 0\ldots,0) = 0$$

This equation can be expresed as:

$$\sum_{l=0}^{p} a(l)\, \phi_c(l,l', k_o, 0,\ldots,0) = 0 \qquad (8)$$

where

$$\phi_c(l,l', k_o, 0,\ldots, 0) = \sum_{m>0} C_{k,y}(m-l; k_o, 0\ldots,0)\, C_{k,y}(m-l'; k_o, 0\ldots,0) \qquad (9)$$

Instead of $\phi_c(l,l', k_o, 0,\ldots, 0)$ we use the following aproximation:

$$\phi_c(l,l', k_o, 0,\ldots, 0) \approx R_c(l-l', k_o, 0,\ldots, 0)$$

where $R_c(i)$ is the autocorrelation of the causal part of the one slice cumulant. Substituing this aproximation in (8) gives:

		∞dB		20dB		10dB		0dB	
	M	sinus	ramp	sinus	ramp	sinus	ramp	sinus	ramp
YW3	0	99.2	99.6	95.6	96.2	67.6	76.8	21.2	25.6
	16	99.4	99.2	94.8	96.4	65.4	72.4	28.2	24.8
YW4	0	99.2	100	95.4	96.6	68.2	69.0	19.4	24.4
	16	99.4	99.0	96.4	96.4	64.8	79.4	19.0	25.4

Table I. Recognition rates in % for different SNR. obtained by methods Yule-Walker order 3 (YW3) and Yule-Walker order 4 (YW4). M: is the over-determination., andwindows: sinus and ramp are tested.

		∞ dB		20 dB		10dB		0 dB	
	Method	sinus	ramp	sinus	ramp	sinus	ramp	sinus	ramp
YW3	LS	98.6	98.6	96.4	95.6	80.0	82.6	40.0	44.8
	COR	99.4	99.0	94.6	96.0	81.2	82.8	29.8	48.8
YW4	LS	98.8	97.8	95.6	93.6	82.2	83.8	39.8	46.0
	COR	98.4	99.4	96.6	97.2	85.0	88.0	40.0	51.6

Table II. Recognition rates (%) obtained for different SNR., using W-Slice order 3 (WS3) and W-Slice order 4 (WS4). Equations are solved by methods LS and Correlation. and two windows: sinus and ramp are tested.

$$\sum_{l=0}^{p} a(l)\, R_c(l-l') = 0 \qquad l' = 1...p \qquad (10)$$

And can be solved forming a Toeplitz matrix with a stable solution.

The coefficients obtained using (10) differ from the true AR coefficients but the results confirm their usefulness in speech recognition tasks. In order to improve the aproximation and to reduce the variance of the computed autocorrelation we used a large number of samples of the cumulants (>>p) and k_o=0.

3. RESULTS

The experiment chosen to compare the different methods is the recognition of the ten digits. The database is composed by 10 speakers. Each speaker utters 10 repetitions of each digit. Speech signal is bandpass filtered between 100 and 3400 Hz and sampled at 8 KHz. 10 HMM states are used with an order 8 LPC analysis in frames of 37.5 ms (300 samples) delayed 150 samples. Cepstrum, delta cepstrum and delta energy are used in the recognition system. Five utterances of all the speakers are used for training and five for testing. Noisy signals are obtained adding white Gaussian noise at SNR = 0, 10, 20 dB. Results are compared against recognition rates obtained with the autocorrelation method (YW2) using a ramp window without preemphasis.

Higher-order Yule- Walker algorithm.

Table I shows the results obtained using the Higher-order Yule- Walker algorithm. YW3 and YW4 correspond to the third- and fourth- order cases respectively. When equations are solved by LS, the number of equations are chosen following S2 (the negative slices (k_o<0) have p+1 equations, m=0,...,p). The results are tabulated for different values of the over determination factor M and the window.

The number of slices chosen to solve (3) is not critical in high SNR conditions. The results are similar for the three tested conditions named S_1, S_2 and S_3. With a SNR of 10dB, S_2 is the best choice. TLS produces a great number

SNR	AC	YW3	YW4	WS3	WS4	DS3	DS4
Clean	99.6	99.6	99	99	99.4	98.4	98.0
20 dB	98.6	96.2	96.4	96	97.2	97.2	97.4
10 dB	82.2	76.8	79.4	82.8	88	90.4	92.4
0 dB	34.4	25.6	25.4	48.8	51.6	61.6	64.4

Table III Summary of recognition rates at different SNR.

of unstable frames and the best results are obtained with LS. Three methods were compared for processing unstable frames: E1: to eliminate unstable frames, E2: to invert poles inside the unit circle, E3 to work equally with stable and unstable frames. The best choice with LS was E3.

Results for ramp or sinus windows are similar at high SNR. When SNR=10 dB ramp window shows a slightly better performance.

With the YW3 method, over determination M=16 decreases the recognition rates at 10 dB. However for the YW4 method, over determination improves the results.

W-slice algorithm.

Table II shows the results obtained using the W-slice algorithm. WS3 and WS4 correspond to the third- and fourth- order cases respectively. The table compares the results obtained with the LS and the correlation method. TLS gives poorer results. If LS or TLS are chosen, the solution may be unstable. If LS is used the procedure named E2, (inverting the poles inside the unit circle) gives the best results which are shown in table II. Solving (6) using the correlation of the cumulants gives always a stable solution.

In noisy conditions, ramp window improves the recognition rate, specially with the correlation method.

1-D slice algorithm

This is the simplest method, it is always stable and gives the best results, using either third- or fourth- order cumulants. The window effect is not important and in Table III only the results obtained with a ramp window are presented.

Table III compares the best results obtained with each of the considered parameterization methods. At SNR=10dB, 1D3 and 1D4 clearly improve all the HOS-based methods and the conventional autocorrelation method (YW2)

4. CONCLUSIONS

The presented results show that the methods based on a weighting of third- or fourth-order cumulants are more robust than those based on the cumulant Yule-Walker equations as SNR is decreased. Compared with the autocorrelation method, DS is better for low SNR. Moreover, DS is the simplest method based on cumulants since only one slice has to be calculated.

The improvement in the recognition rate is due only to the changes in the parameter estimation stage. Other stages in the recognizer may also be modified to increase the recognition rate in noise, but our goal was only to compare different parameterization methods in the same standard HMM-based recognition task .

5. REFERENCES

[1] J. M, Mendel *"Tutorial on Higher Order Statistics (Spectra) in Signal Processing and System Theory: theoretical results and some applications"* Proc IEEE, vol 79, no 3 March 1991

[2] J. Vidal , J. A. R. Fonollosa, *Causal AR Modeling Using a Linear Combination of Cumulant Slices*, Elsevier Science: Signal Processing 36, Aug.1992

[3] A. Moreno, J. A. R. Fonollosa, J. Vidal *"HOS analysis of speech. A vocoder Application"*. 3rd European Conference on Speech Communication. EUROSPEECH 93. pp 519-522 Berlin. Germany.

[4] K.K. Paliwal, M.M. Sondhi. *"Recognition of noisy Speech Using Cumulant-Based Linear Prediction Analysis"*. Int Conf on Acoustics, Speech and Signal Processing. ICASSP'91. pp 429-432

[5] A. Swami, J.M.Mendel. *AR Identifiability Using Cumulant Slices*. Proceedings of the Workshop in Higher Order Spectral Analysis. CO pp 13-18 June 1989

[6] L.R.Rabiner. *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition.* IEEE Proceedings Vol 77. No2. Febr. 1989