ISCA Archive
http://www.isca-speech.org/archive

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

SOME FAST HIGHER ORDER AR ESTIMATION TECHNIQUES APPLIED TO PARAMETRIC WIENER FILTERING

Josep SALAVEDRA*, Enrique MASGRAU**, Asunción MORENO*, Joan ESTARELLAS*, Javier HERNANDO*

* Department of Signal Theory and Communications. Universitat Politècnica de Catalunya. Apartat 30002. 08080-BARCELONA. SPAIN. Tel/Fax: +34-3- 4017404 / 4016447 . E-mail: mia@tsc.upc.es

** Department of Electrical Engineering and Computers. Universidad de Zaragoza. María de Luna, 3. 50015-ZARAGOZA. SPAIN

10.21437/ICSLP.1994-428

ABSTRACT

Some Speech Enhancement algorithms based on the iterative Wiener filtering Method due to Lim-Oppenheim [2] are presented. In the original Lim-Oppenheim algorithm, speech AR estimation is carried out using classic second-order analysis, but our algorithms consider a more robust AR modelling. Two different strategies of speech AR estimation are presented and both estimators are trying to see as less amount of noise as possible. First one uses a previous One-Sided Autocorrelation computation, that is a pole-preserving function, and the actual SNR in the second-order LPC analysis is increased. Second one combines advantages of Higher-Order Statistics [1] with a linear combination of AR coefficients, belonging to two consecutive overlapped frames, to assess a less disturbed speech estimation.

1. INTRODUCTION.

It is well known, that many applications of speech processing that show very high perfomance in laboratory conditions degrade dramatically when working in real environments because of low robustness. The solution we propose here concerns to a preprocessing front-end in order to enhance the speech quality by means of a speech parametric modelling insensitive to the noise. Original Lim-Oppenheim algorithm [2] enhances noisy speech signal by means of an iterative Wiener filtering using speech AR modelling coming from second-order statistics estimation. But, its performance degrades when low SNR environments are considered, because second-order speech AR estimation is too sensitive to the presence of noise. In this paper we propose some robust approaches working in very noisy environments, where main part of Single Microphone Techniques of Speech Enhancement sidetrack their objectives and, therefore, some authors propose a solution based on a Multimicrophone Strategy [3].

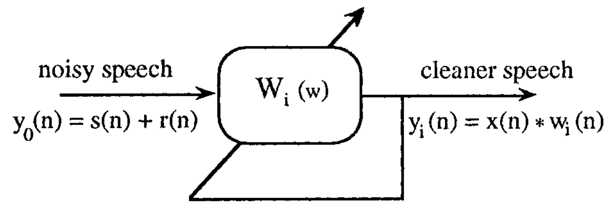
2. SPEECH AR MODELLING IN THE AUTOCORRELATION DOMAIN.

An iterative Wiener Filtering is considered to enhance noisy speech. This filter is dessigned at every frame by means of the following expression :

$$W_i(w) = \left(\frac{P_y}{P_y + ß.P_r} \right)^{\partial} \qquad (1)$$

where P_r represents the power spectrum of noise signal, that is estimated inside of non-speech activity frames by using a smoothing periodogram; P_y is the power spectrum of the unavailable clean speech signal; and parameters ∂ and ß allow a better control over filtering features but, in this section they are set to ß=∂=1.0 . Clean speech power spectrum must be estimated from the noisy speech signal because just single microphone techniques of Speech Enhancement are considered to remove the additive noise signal. This speech spectrum is obtained by means of an all-pole modelling of speech signal. Therefore, after first speech AR modelling, noisy speech is filtered and a cleaner noisy speech signal is available at Wiener Filtering Output. It seems successful to obtain a more accurate speech AR modelling coming from this cleaner speech signal. Thus, an iterative Wiener algorithm is considered, where an improvement of performance may be expected after every iteration since current AR speech estimation is carried out from a cleaner speech signal than filter estimation of the preceding iteration. The performance of this algorithm basically depends on the fidelity of the AR coefficients inside of the system shown in Fig.1 .

This work was supported by TIC 92-0800-C05-04



$$W_i = \frac{P_{y_{i-1}}}{P_{y_{i-1}} + P_r} \qquad P_{y_{i-1}}(w) = \frac{g^2}{\left| 1 - \sum_{k=1}^{P} a_k \cdot e^{-jwk} \right|^2}$$

Fig.1: *Scheme of the Iterative Wiener Algorithm.*

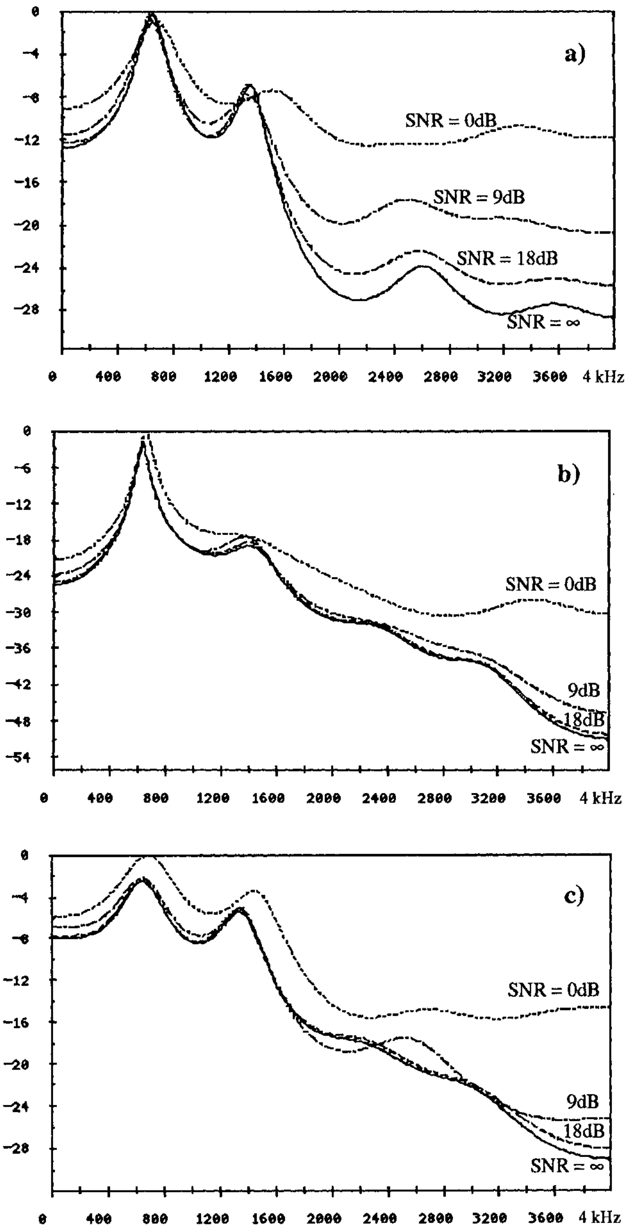


Fig.2 : *LPC spectra corresponding to: a) classic 2nd-order analysis; b) OSA function and 2nd-order analysis; c) ordinary 3rd-order cumulant analysis.*

Standard LPC technique is known to be very sensitive to the presence of additive noise and, so, a poor noise suppression is achieved after some iterations of this algorithm (see Table.1.a). The quality of the AR estimation may be evaluated in terms of two important statistical properties: the estimator bias and its covariance matrix. Some authors have shown that classic least squares estimator of AR coefficients (Yule-Walker equations), applied to noise degraded all-pole sequences, leads to biased estimates of these AR coefficients. This biased estimation may be avoided by considering Higher Order Yule-Walker Equations but, then, a serious problem is the large variance of this estimation.

The procedure that is presented here tries to be less sensitive to the additive noise and it increases the signal-to-noise-ratio while pole location of the speech signal model is preserved. If the result of applying a function to an all-pole sequence is a sequence that has the same poles as the original sequence, then, it is a pole-preserving function. In [4], autocorrelation function is introduced as a good pole-preserving function. Applying least squares estimation after this autocorrelation function, we obtain both a smaller bias and a smaller variance. Therefore, AR coefficients computed in the autocorrelation domain lead to better speech AR estimation. One-Sided Autocorrelation (OSA) sequence may be defined from the Autocorrelation sequence R(n) as follows :

$$R^+(n) = \begin{cases} R(n) & n > 0 \\ R(0)/2 & n = 0 \\ 0 & n < 0 \end{cases} \qquad (2)$$

and it verifies

$$R(n) = R^+(n) + R^+(-n) \quad ,, \quad -\infty \le n \le +\infty \qquad (3)$$

a)	SNR	SEGSN	ITAKU	COSH	CEPST
0 iter.	0.00	0.79	9.57	11.67	12.02
1 iter.	7.36	4.38	9.21	10.71	11.01
2 iter.	8.83	5.92	8.86	10.17	9.90
3 iter.	9.04	6.16	7.30	9.04	9.34
4 iter.	9.11	6.25	6.42	8.45	9.20

c)	SNR	SEGSN	ITAKU	COSH	CEPST
0 iter.	0.00	0.79	9.57	11.67	12.02
1 iter.	8.13	5.04	5.13	7.76	7.76
2 iter.	8.18	5.68	5.08	7.49	7.97
3 iter.	8.05	6.03	4.84	7.28	7.93
4 iter.	7.94	6.10	4.76	7.33	7.93

b)	SNR	SEGSN	ITAKU	COSH	CEPST
0 iter.	0.00	0.79	9.57	11.67	12.02
1 iter.	7.92	4.86	8.18	9.78	9.82
2 iter.	7.60	5.31	5.94	8.16	8.47
3 iter.	7.59	5.59	5.11	7.55	8.15
4 iter.	7.36	5.79	5.15	7.64	8.30

d)	SNR	SEGSN	ITAKU	COSH	CEPST
0 iter.	0.00	0.79	9.57	11.67	12.02
1 iter.	7.47	4.53	8.97	10.49	10.53
2 iter.	7.39	4.95	7.88	9.65	9.30
3 iter.	7.37	5.11	6.55	8.65	8.80
4 iter.	7.77	5.49	5.52	7.91	8.47

Table.1 : *Time (SNR, SEGSNR) and spectral (Itakura, Cosh, Cepstrum) distance Measures using algorithms based on: a) classic 2nd-order statistics; b) parameterized 3rd-order cumulants (∂=1.2;β=1.0); c) parameterized 3rd-order cumulants (∂=1.2;β=1) with IF=0.6 and PFI=5; d) 4th-order cumulants at SNR=0dB (AWGN).*

Furthermore, the spectral magnitude of its Fourier Transform may be seen as a spectral envelope [5] :

$$E(w) = \left|S^{+}(w)\right| \quad (4)$$

This envelope characteristic originates a strong enhancement in the highest power frequency bands and noise components lying outside these frequency bands are largely attenuated in comparison to LPC spectrum of noisy signal (see fig.2.a and fig.2.b)

3. PERFORMANCE OF THE OSA ALGORITHM

As it has been discussed in the previous section, One-Sided Autocorrelation (OSA) function is applied to the noisy speech signal and, then, its output is sent to Levinson-Durbin algorithm, where AR coefficients are calculated. This previous autocorrelation computation motivates that Levinson-Durbin algorithm is receiving a less contaminated sequence. This fact may be observed in Fig.2, where LPC spectrum using OSA algorithm is compared to those spectra obtained from classic second-order algorithm and third-order algorithm [6]. Clean speech signal (SNR=∞) and disturbed speech signal with different levels of additive noise have been processed. Fig.2.a shows that classic second-order statistics algorithm is too sensitive to the noise when middle and low SNR are considered. OSA algorithm gives us a good performance at low and medium levels of additive noise but, its performance begins to deteriorate at low SNR. Despite of this high level of noise, a good noise reduction is achieved after processing some iterations of the iterative Wiener filtering (see Table.2). In comparison to third-order algorithms, it seems to produce a higher distortion, specially inside of low energy frequency bands of the speech spectrum.

In preceding works we have found that AR estimation from third-order cumulants using an Interframe Factor (IF) is more reliable than the others (see Table.1). Third-order cumulants allow a desirable uncoupling between speech and noise signals because of its properties [1] : all cumulants of order greater than two are identically zero if Gaussian Processes are considered and all odd-order cumulants are null when non-Gaussian processes presenting a symmetric p.d.f. are evaluated.

a)	SNR	SEGSN	ITAKU	COSH	CEPST
0 iter.	0.00	0.79	9.57	11.67	12.02
1 iter.	9.18	5.99	6.28	8.34	7.94
2 iter.	9.06	6.14	4.79	7.47	6.58
3 iter.	8.77	6.05	4.76	7.36	6.65
4 iter.	8.58	5.97	4.64	7.25	6.60

Table.2 : *Time (SNR, SEGSNR) and spectral (Itakura, Cosh, Cepstrum) distance Measures using 2nd-order statistics and a previous one-sided autocorrelation function (L_R=80).*

Parameter IF weighs current frame AR coefficients with respect to previous frame AR coefficients. Parameter PFI (Previous Frame Iteration) corresponds to the iteration number of the previous frame that is considered to help first iteration of the current frame, because AR estimator is looking at a cleaner speech signal when paprameter PFI>1. As it is shown in Table.1, an improvement, higher than 4dB in Cepstrum distance, may be assessed when just first iteration of the iterative Wiener filtering has been processed. Therefore, a good reduction of computational complexity is achieved because the convergence of this algorithm is greatly accelerated, without any noticeable increase of distortion.

Table.2 shows that OSA algorithm also achieves a very fast convergence and it seems to be more aggressive than third-order one using Interframe Factor. A high noise reduction (more than 4dB in terms of Cepstrum distance) is assessed after processing just first iteration and, furthermore, this improvement increases to 5.5dB when two iterations are processed. The problem arises from its higher distortion effect appreciated in the listening tests. In short, both techniques seem to be two up-and-coming approaches when very noisy environments are evaluated.

4. CONCLUSIONS.

Two different approaches of speech AR estimation based on an iterative Wiener filtering have been proposed. Spectral estimation of speech signal is obtained by means of a robust AR modelling to provide a desirable noise-speech uncoupling. First technique calculates a previous OSA function to serve a less noisy speech signal to the LPC analysis system. Second one considers a third-order statistics analysis and two parameters, Interframe Factor (IF) and Previous Frame Iteration (PFI), have been introduced to take advantage of previous speech spectrum estimations to initiate 3rd-order AR modelling corresponding to first iteration of the current noisy speech frame. Both approaches are compared to classic 2nd-order analysis and ordinary 3rd- and 4th-order cumulant estimations. They achieve an important noise suppression (more than 4dB in terms of Cepstrum distance) after processing just first iteration of this algorithm, in a very noisy environment. Therefore, convergence of this iterative algorithm is strongly accelerated and, thus, a reduction of both computational complexity and processing delay are assessed, while no appreciable increase of distortion effect [7] is generated. All these features are specially esteemed when low and medium SNR are considered.

REFERENCES.

[1]C.L.Nikias,M.R.Raghuveer,"Bispectrum Estimation: A Digital Signal Processing Framework". Proc. of The IEEE, Vol. 75, No. 7, pp 869-891. July 1987.

[2]J.S.Lim, A.V.Oppenheim,"All-Pole Modeling of Degraded Speech". IEEE Trans. ASSP, Vol. ASSP-26, No. 3, pp197-210. June 1978.

[3]D.Van Compernolle. "DSP Techniques for Speech Enhancement". Proc. ESCA Workshop on Speech Processing in Adverse Conditions, pp. 21-30. Cannes, France. November 10-13, 1992.

[4]D.McGinn, D.H.Johnson. "Reduction of all-pole parameter estimator bias by successive autocorrelation". Proc.ICASSP, pp.1088-1091. Boston, USA. April 1983

[5]J.Hernando,C.Nadeu,E.Lleida. "On the AR Modelling of the One-sided Autocorrelation Sequence for noisy Speech Recognition". Proc. ICSLP, pp. 1593-1596. Banff, Alberta, Canada. October 1992.

[6]J.M.Salavedra, E.Masgrau, A.Moreno, X.Jove. "A speech enhancement system using higher-order AR estimation in real environments". Proc. EUROSPEECH, pp 223-226. Berlin, Germany. September 21-23, 1993.

[7]J.M.Salavedra, E.Masgrau, A.Moreno, J.Estarellas. "Some Robust Speech Enhancement Techniques using Higher-Order AR Estimation". Proc. EUSIPCO. Edinburgh, Scotland, U.K. September 13-16, 1994. To be published