ROBUST COEFFICIENTS OF A HIGHER ORDER AR MODELLING IN A SPEECH ENHANCEMENT SYSTEM USING PARAMETERIZED WIENER FILTERING.

Josep M.SALAVEDRA*, Enrique MASGRAU**, Asunción MORENO*, Joan ESTARELLAS*, Xavier JOVE*

* Department of Signal Theory and Communications. Universitat Politècnica de Catalunya.
Apdo. 30002. 08080-BARCELONA. SPAIN. E-mail: mia@tsc.upc.es
** Department of Electrical Engineering and Computers. Universidad de Zaragoza.
María de Luna, 3. 50015-ZARAGOZA. SPAIN

ABSTRACT

We study some speech enhancement algorithms based on the iterative Wiener filtering method due to Lim-Oppenheim [2], where the AR spectral estimation of the speech is carried out using a second-order analysis. But in our algorithms we consider an AR estimation by means of cumulant analysis. This work extends some preceding papers due to the authors, providing a generalization of third- and fouth-order algorithms by means of adding two parameters in the general expression of Wiener filtering. These parameters allow a better control of their performance. Some results are presented considering AWGN but listening tests give similar performance when other noises (diesel engine) are considered.

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 in this paper concerns to a preprocessing front-end in order to enhance the speech quality by means of a speech parametric modelling insensitive to the noise. The use of HO cumulants for speech AR modelling calculation provides the desirable uncoupling between noise and speech. It is based on the property that for Gaussian processes only, all cumulants of order greater than two are identically zero. Moreover, the non-Gaussian processes presenting a symmetric p.d.f. have null odd-order cumulants. Considering a Gaussian or a symmetric p.d.f. noise (a good approximation of very real environments) and the non-Gaussian characteristic of the speech (principally for the voiced frames) it would be possible to obtain a spectral AR modelling of the speech more independent of the noise by using, e.g., third-order cumulants of noisy speech instead of common second-order cumulant. The problem arises of the higher spectral distortion presented by the AR modelling based on cumulant estimation when it is compared to autocorrelation case. It is due to the higher variance of cumulant estimation and the questionable "flatness" of the error sequence produced when the AR inverse filter works as a predictor over the speech signal. These drawbacks advise to make no more of two iterations using cumulant AR modelling.

2. THE PARAMETERIZED ALGORITHM

In preceding works, several modified approaches of the original Lim-Oppenheim algorithm have been considered. All of them use HOS to compute the AR modelling of speech signal. Thus AR modelling of the speech spectrum estimation is computed either from third-order cumulants in the third-order algorithm or from 4th-order cumulants when 4th-order algorithm is considered. For example, third-order cumulants are calculated using the covariance case:

$$C_k(i,j) = \sum_{n=p+1}^{N} x(n-k).x(n-i).x(n-j) \quad ,, \quad 0 \le k,i,j \le p \qquad (2)$$

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

where p=10 is the order of the filter. Then Wiener filter coefficients a_k are computed by solving the following equations [1]:

$$\sum_{k=0}^{p} a_k \cdot C_k(i,j) = 0 \quad ,, \quad 1 \le i \le p \; ; \; 0 \le j \le i \qquad (3)$$

At the beginning, third-order algorithm was compared to classical second-order one [5]. A twofold benefit is obtained by considering third-order AR modelling: Firstly, the convergence of the iterative algorithm is accelerated and therefore a reduction of both computational complexity and intelligibility loss is obtained; Secondly, a less polluted AR speech parameterization is achieved. So we get better performance but we must control the higher spectral distortion effect of third-order algorithm in comparison with second-order one, since a higher "peaking" or "narrowness" effect of speech formants is brought about [4]. In [6] a trade-off between noise reduction effect and spectral distortion effect is discussed. Then the parameterized algorithm seems to be a good solution to control the spectral distortion effect. It can be seen as a generalization of Wiener filtering by adding two parameters α and β to the general expression of Wiener filter [3] :

$$W_i(w) = \left(\frac{P_y}{P_y + \beta.P_r}\right)^{\alpha} \qquad (4)$$

where P_r is the spectrum of the noise signal estimated in non-speech frames and P_y is a spectrum estimation of the unavailable clean speech signal. By varying these parameters α and β, filters with different characteristics can be obtained. Thus, if $\alpha=\beta=1.0$ then expression (4) corresponds to the general Wiener filter equation and if $\alpha=0.5$, $\beta=1.0$ it corresponds to power spectrum filtering. Then, high values of α and β lead to a more aggressive Wiener filter and therefore noise suppression is increased but distortion may increase too.

3. PERFORMANCE OF THIRD-ORDER PARAMETERIZED ALGORITHM

Performance of parameterized third-order algorithm has been evaluated in terms of standard spectral measures such as Itakura, Cosh and Cepstrum distances. However, the performance evaluation, discussed in this section,considers only Cepstrum distance because it looks at the overall spectrum in a more uniform way than the other two distances. Therefore, Cepstrum distance is more sensitive to distortion in valleys and flat zones of the speech spectrum since the known "peaking" effect introduced by the iterative Wiener filtering algorithms causes higher spectral distortion in these zones [4]. The following speech enhancement experience has been considered: noise-free utterances are disturbed by additive noises (AWGN,diesel engine). All figures in this section refers to AWG noise at three different levels of noise: low , middle and high. SNR. We range parameter α from 0.5 to 1.4 (step size=0.1) and parameter β from 0.2 to 2.0 (step size=0.2). Every pair of values corresponding to a value of every parameter is noted as (α,β), belonging to the plane $0.5 \le \alpha \le 1.4$; $0.2 \le \beta \le 2.0$; and its associated Cepstrum distance is noted as $C_i(\alpha,\beta)$ (i:iteration number) whose maximum and minimum values are noted as MAX and MIN respectively.

0-7803-1772-6/94/$3.00

3.1. High level of noise (SNR=0dB)

After first iteration, level of noise decreases uniformly from MAX=C_1(0.5,0.2)=11.22 dB to MIN=C_1(1.5,2)=9.31dB (Initial value of Cepstrum distance before processing is C_0=12.02dB). Noise suppression effect is much higher than spectral distortion effect, therefore C_1 (α,β) decreases when parameter α and β increase because parametric Wiener filter becomes more aggressive and furthermore first iteration of the algorithm introduces a very small spectral distortion effect (see [4]). It must be remarked that high values of (α,β) get a noise reduction 1.9dB better than low values of them. So, we have a good control of Wiener filtering by using these two parameters: a high value of α and low β is equivalent to a low α and high β and then a curve representing a fixed Cepstrum distance has a tendency to a diagonal line on the representation plane.

After second iteration level of noise decreases from MAX=C_2(0.5,0.2)=10.62dB to MIN=C_2(1.4,2)=8.13dB but not so uniformly as before. Low values of (α,β) have a high variance of values in Cepstrum distance (curves very close) but difference of level on high values of (α,β) is very small. So, we have a zone of the plane with a similar performance: $39 \le 20\alpha+12\beta$. When high (α,β) most part of noise suppression has been done and spectral distortion effect begins to be similar to noise reduction effect. However in the other corner (low values of α and β) noise reduction effect copletely overshadows spectral distortion effect. It can be noted that high (α,β) have already arrived to the saturation level where both spectral distortion and noise reduction effects have the same magnitude.

After third iteration (see fig.1.a), Cepstrum distance of these high values of (α,β) begin to deteriorate (C_3(1.4,2.0)=8.49dB) because the level of residual noise that can be eliMINated is lower than level of added distortion. On the other, low values (α,β) achieve better performance than previous iteration but still higher than 9dB and therefore they need some iterations to get a saturation level where both effects are similar. Better results are obtained when values (α,β) are near to the following equation : $5\alpha+2\beta = 8$, where $0.8 \le \alpha \le 1.3$. Level of Cepstrum distance varies from MAX=C_3(0.5,0.2)=10.13dB to MIN=C_3(1.2,1.0)=8.02dB. Values of (α,β) belonging to this valley of the best performance are those that get the saturation level of noise reduction after 3 iterations.

When 4 or 5 iterations are processed (see fig.1.b) we appreciate the same features as previous iteration but the valley of the best performance moves to left side, corresponding to a less aggressive zone. This valley may be represented by $2\alpha+\beta = 3.4$ after 5 iterations. Furthermore, some local maximum and local minimum distances appear along this valley. Thus, Cepstrum distance takes values from MAX=C_4(0.5,0.2)=9.72dB to MIN=C_4(0.9,1.2)= 8.13dB (4 iterations) and from MAX=C_5(0.5,0.2)=9.45dB to MIN=C_5(0.8,1.8)=8.15dB .

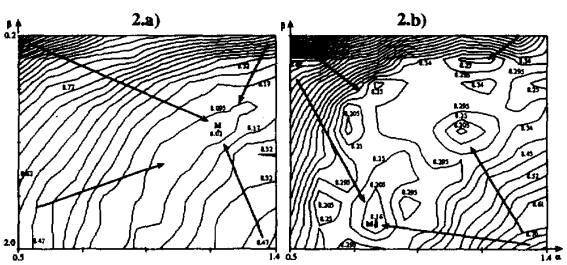

Fig.2. Level lines representation corresponding to parameterized third-order algorithm (SNR=0dB) : a) 3 iterations; b) 5 iterations (M: minima ; $\rightarrow$: Lower distance zone)

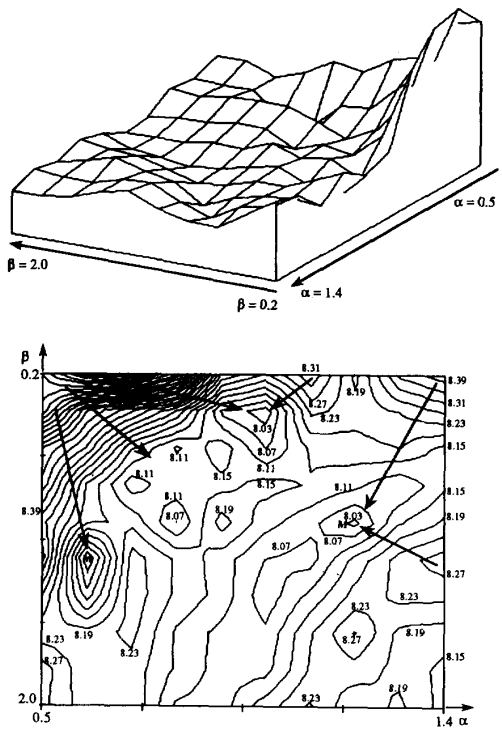

Fig.1. Level lines representation corresponding to parameterized third-order algorithm (SNR=0dB) at the best iteration (M: minima ; $\rightarrow$: Lower distance zone)

Fig.2 shows the best Cepstrum distance value for every pair (α,β) independently of the number of iterations that has been necessary. Table.1 contains these values and the number of iterations (in brackets) necessary to get the minimum Cepstrum distance. Distances are between MAX=C_{13}(0.6,0.2)=9.11dB and MIN=C_{17}(0.6,1.2)= 7.95dB . Performance of low values (α,β) never becomes good, even if more than 10 iterations are processed and then computational complexity increases too much. minimum MIN is obtained after 17 iterations and because of computational complexity it must be discarded. Listening tests at this MIN shows a residual musical noise that is maintained at a constant level after 6 iterations and distortion effect added by every additional iteration is very small. Therefore, a better selection is MIN=C_3(1.2,1)=8.02dB where only 3 iterations are needed and so distortion effect is not important yet. Expression $7\alpha+3\beta = 11.6$ seems to be a good choice of the best performance zone. Values on the left side ($12\alpha+5\beta \le 16.6$) must be discarded because of its higher computational complexity and sometimes bad quality. High values of (α,β) must also be discarded ($5\alpha+2\beta > 9$) because they present residual musical noise after 2 iterations and intelligibility loss is too important if more iterations are processed.

In short, best results are obtained when Cepstrum distance is lower than 8.1dB after 3 or 4 iterations. Listening tests show a slightly better performance 1 iteration after saturation level because intelligibility loss of this additional iteration is small and then parametric Wiener filter removes the residual musical noise that noisy speech signal still has.

3											
	0dB	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4
	0.2	8.91 (14)	9.11 (13)	9.06 (18)	8.83 (8)	8.30 (19)	8.22 (14)	8.34 (7)	8.18 (17)	8.36 (7)	8.41 (5)
B	0.4	8.86 (19)	8.55 (16)	8.44 (19)	8.31 (17)	8.06 (14)	8.01 (13)	8.24 (6)	8.21 (12)	8.23 (4)	8.27 (3)
	0.6	8.55 (12)	8.43 (19)	8.19 (16)	8.10 (14)	8.17 (7)	8.07 (12)	8.16 (4)	8.17 (3)	8.19 (3)	8.13 (3)
E	0.8	8.50 (13)	8.23 (15)	8.09 (13)	8.14 (13)	8.13 (4)	8.17 (4)	8.15 (3)	8.12 (3)	8.07 (3)	8.14 (3)
	1.0	8.36 (16)	8.20 (14)	8.18 (13)	8.07 (16)	8.21 (4)	8.12 (3)	8.08 (3)	8.02 (3)	8.15 (3)	8.20 (3)
T	1.2	8.40 (12)	7.95 (17)	8.25 (4)	8.18 (3)	8.13 (4)	8.07 (3)	8.05 (3)	8.16 (3)	8.25 (3)	8.29 (2)
	1.4	8.27 (12)	8.04 (11)	8.25 (4)	8.18 (4)	8.13 (3)	8.04 (3)	8.12 (3)	8.20 (3)	8.24 (2)	8.24 (2)
A	1.6	8.20 (12)	8.20 (5)	8.24 (4)	8.14 (3)	8.10 (3)	8.08 (3)	8.18 (3)	8.27 (2)	8.18 (2)	8.17 (2)
	1.8	8.30 (7)	8.20 (4)	8.21 (4)	8.14 (4)	8.09 (3)	8.10 (3)	8.19 (3)	8.33 (2)	8.18 (2)	8.14 (2)
	2.0	8.28 (7)	8.20 (4)	8.23 (3)	8.10 (3)	8.09 (3)	8.17 (3)	8.23 (2)	8.16 (2)	8.20 (2)	8.13 (2)

Table.1 Performance of parameterized third-order algorithm at SNR=0dB : Cepstrum distance in the best iteration.

3											
	9dB	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4
	0.2	7.28 (3)	7.35 (3)	7.40 (2)	7.38 (2)	7.33 (2)	7.35 (2)	7.34 (2)	7.36 (11)	7.39 (6)	7.29 (6)
B	0.4	7.22 (3)	7.25 (2)	7.22 (2)	7.26 (2)	7.30 (4)	7.18 (5)	7.07 (4)	7.15 (4)	7.03 (4)	7.08 (3)
	0.6	7.18 (3)	7.19 (2)	7.21 (2)	7.17 (12)	7.05 (5)	7.06 (4)	7.00 (4)	7.00 (3)	6.97 (3)	6.90 (3)
E	0.8	7.15 (2)	7.16 (2)	7.11 (11)	6.99 (4)	7.00 (4)	7.03 (4)	6.95 (3)	6.84 (3)	7.02 (3)	7.00 (3)
	1.0	7.13 (2)	7.14 (12)	7.09 (15)	6.93 (5)	7.02 (4)	6.91 (3)	6.92 (3)	7.02 (2)	6.99 (2)	6.96 (2)
T	1.2	7.11 (2)	7.08 (12)	6.95 (4)	6.97 (4)	6.98 (3)	6.88 (3)	6.98 (3)	6.96 (2)	6.98 (2)	7.05 (3)
	1.4	7.10 (2)	7.12 (6)	7.00 (4)	7.00 (3)	6.94 (3)	6.93 (3)	6.97 (2)	6.94 (2)	6.96 (2)	6.96 (2)
A	1.6	7.08 (2)	6.98 (9)	6.96 (5)	6.93 (3)	6.98 (3)	6.98 (2)	6.94 (2)	6.96 (2)	6.97 (2)	6.97 (2)
	1.8	7.07 (2)	7.03 (5)	6.92 (4)	6.95 (3)	6.94 (3)	6.94 (3)	6.94 (2)	6.94 (2)	6.95 (2)	6.98 (2)
	2.0	7.08 (2)	7.05 (4)	6.95 (4)	6.97 (3)	6.96 (3)	6.93 (2)	6.94 (2)	6.95 (2)	6.98 (2)	6.98 (2)

Table.2 Performance of parameterized third-order algorithm at SNR=9dB : Cepstrum distance in the best iteration.

3.2 Middle level of noise (SNR=9dB)

After first iteration, Cepstrum distance decreases uniformly from MAX=$C_1(0.5,0.2)$=8.73dB to MIN=$C_1(1.4,2,0)$= 7.02dB (Initial Cepstrum distance is C_0=10.51dB). Now we have a lower level of noise, therefore saturation zone appears after first iteration at high values of (α,β) and the best performance is obtained in $22.6 \le 14\alpha+5\beta$. Furthermore, after two iterations it appears the valley of best performance in $5\alpha+2\beta=9$ (see fig.3.a), MAX=$C_2(0.6,0.2)$= 7.59dB , MIN=$C_2(1,2)$= 6.93dB and added distortion at high (α,β) is not important ($C_2(1.4,2)$= 6.98dB. Low values of (α,β) achieve to reduce most part of noise with only 2 iterations (noise reduction is almost 3dB). We may conclude that most part of (α,β) are in the saturation zone. When the iterative algorithm has processed 3 iterations (see fig.3.b), values of Cepstrum distance are between MAX= $C_3(1,0.2)$=7.62dB and MIN=$C_3(1.2,0.8)$=6.84dB. The best performance zone is $2\alpha+\beta = 3.2$,, $0.6 \le \beta \le 2.0$. Therefore this valley moves to the left side, as discussed before at low SNR. Distortion effect begins to be important because all distances of high (α,β) get worse about 0.25dB after 4 iterations have been processed because spectral distortion effect overshadows noise reduction effect: MAX= $C_4(0.7,0.2)$=7.76dB, MIN=$C_4(1.4,0.6)$=6.84dB.

Optimum number of iterations, for every pair (α,β) is represented in Table.2. Most part of (α,β) have good performance with no more than three iterations. Cepstrum distance varies from MAX= $C_2(0.7,0.2)$=7.40dB to MIN=$C_3(1.2,0.8)$=6.84dB. Listening tests show no residual noise is present after 3 iterations and spectral distortion is not important at this level of noise. In short, most part of values lead to a good performance, but parameter β must be greater or equal than 0.4.

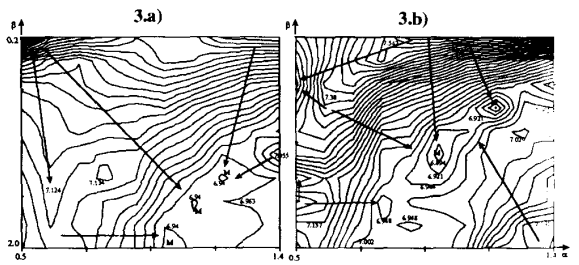

Fig.3. Level lines representation corresponding to parameterized third-order algorithm (SNR=9dB) : a) 2 iterations; b) 3 iterations (M: minima ; $\rightarrow$: Lower distance zone)

3.3 Low level of noise (SNR=18dB)

Only an iteration must be processed to eliMINate all of the noise. Third-order statistics algorithm seems to be too aggresive at this level of noise because in most part of middle and high (α,β) spectral distortion effect masks noise reduction effect and the valley of the best performance appear at low (α,β) : $7\alpha+\beta = 5.5$ where MAX= $C_1(0.5,0.2)$=6.12dB, MIN=$C_1(0.6,1)$=5.77dB (C_0=8.52dB). Listening tests show a very good quality with only an iteration of processing.

4. PERFORMANCE OF FOURTH-ORDER PARAMETERIZED ALGORITHM

All the test features are identical to previous section but now we consider the following zone :

$$0.5 \le \alpha \le 1.5$$

$$0.2 \le \beta \le 1.8$$

4.1 High level of noise (SNR=0dB)

The behaviour of the iterative Wiener filtering when fourth-order statistics are considered is much more conservative than third-order one. Therefore, noise suppression effect is much lower and more iterations are necessary. Thus saturation level appears only after 3 iterations at high values of (α,β), and the valley of the best performance appears after a processing containing a minimum of 4 iterations. This conservative behaviour makes this level of noise too high to be cancelled by this algorithm. Performance in the best iteration gets similar results to those obtained by means of third-order algorithm at middle and high (α,β), but more iterations are needed to get the same value of Cepstrum distance (see Table.3). Therefore, computational complexity increases too much and listening tests show that musical noise is never eliMINated. It must be remarked that an iteration using fouth-order statistics implies much more computational complexity than an iteration using third-order statistics.

4.2 Low level of noise (SNR=18dB)

Its conservative behaviour makes parameterized fourth-order algorithm better than third-order one at low level noise conditions for any pair (α,β). Spectral distortion effect is very low and then even high (α,β) have a good performance. So, values of distances are better in the best iteration: $MAX=C_2(1.5,0.2)=5.70dB$, $MIN=C_2(0.5,1.2)=5.34dB$ ($C_0=8.52dB$). However, listening tests show a very similar quality to third-order algorithm

4		ALFA										
BETA	0dB	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
	0.2	10.6 (12)	10.3 (10)	10.0 (8)	9.83 (10)	9.65 (11)	9.63 (14)	9.22 (19)	9.01 (16)	8.79 (13)	8.74 (12)	8.69 (12)
	0.4	10.2 (10)	9.75 (11)	9.34 (12)	9.23 (11)	9.13 (12)	9.01 (14)	8.77 (15)	8.65 (12)	8.53 (10)	8.49 (10)	8.46 (10)
	0.6	9.69 (7)	9.17 (10)	8.66 (15)	8.63 (13)	8.61 (12)	8.40 (14)	8.31 (10)	8.29 (8)	8.27 (6)	8.25 (7)	8.23 (8)
	0.8	9.48 (6)	8.83 (11)	8.18 (15)	8.29 (12)	8.41 (10)	8.15 (8)	8.23 (6)	8.22 (8)	8.21 (11)	8.30 (7)	8.39 (4)
	1.0	9.29 (16)	8.85 (14)	8.40 (12)	8.32 (10)	8.24 (8)	8.30 (6)	8.21 (13)	8.18 (12)	8.15 (12)	8.30 (8)	8.46 (4)
	1.2	9.28 (5)	8.83 (8)	8.39 (11)	8.23 (9)	8.08 (8)	8.18 (6)	8.32 (7)	8.36 (6)	8.41 (5)	8.40 (4)	8.39 (3)
	1.4	9.09 (8)	8.66 (9)	8.23 (10)	8.20 (8)	8.16 (7)	8.26 (5)	8.27 (5)	8.34 (4)	8.40 (4)	8.35 (3)	8.37 (3)
	1.6	8.84 (13)	8.49 (11)	8.13 (9)	8.18 (7)	8.23 (5)	8.29 (4)	8.30 (4)	8.33 (3)	8.35 (3)	8.39 (3)	8.44 (3)
	1.8	8.53 (17)	8.31 (12)	8.09 (8)	8.19 (6)	8.29 (4)	8.29 (4)	8.40 (3)	8.33 (3)	8.26 (3)	8.44 (3)	8.61 (3)

Table.3 Performance of parameterized fourth-order algorithm at SNR=0dB : Cepstrum distance in the best iteration.

5 CONCLUSIONS

A speech enhancement method based on an iterative Wiener filtering have been proposed. Spectral estimation of speech is got by means of an AR modelling based on 3rd- and 4th-order cumulant analysis to provide the desirable noise-speech uncoupling. Some different approaches using parametric iterative Wiener filtering and HOS have been propossed. Parameterized third-order algorithms assess better results than fourth-order ones, specially at high and middle levels of noise. Their performance is similar only at low levels of noise. So, the parameterized third-order algorithms are a good trade-off among convergence speed, distortion effect and computational complexity.

REFERENCES

[1] C.L. Nikias, M.R. Raghuveer."Bispectrum Estimation: A Digital Signal Processing Framework". Proc. of IEEE, pp 869-891. July 1987.

[2] J.S.Lim and A.V. Oppenheim."All-Pole Modeling of Degraded Speech". IEEE Trans. on ASSP, pp197-210. June 1978.

[3] J.S.Lim and A.V. Oppenheim."Enhancement and Bandwidth Compression of noisy Speech". Proc. of the IEEE, pp1586-1604. December 1979.

[4]E.Masgrau,J.M.Salavedra,A.Moreno,A.Ardanuy. "Speech Enhancement by Adaptive Wiener Filtering based on Cumulant AR Modelling". Proc. ESCA Workshop on Speech Processing in Adverse Conditions, pp 143-146. Cannes, France.November92.

[5]J.M.Salavedra,E.Masgrau,A.Moreno,X.Jové. "Comparison of different order cumulants in a Speech Enhancement System by adaptive Wiener Filtering". Proc.IEEE Signal Processing Workshop on HOS, pp. 61-65. June'93. South Lake Tahoe, CA, USA.

[6]J.M.Salavedra,E.Masgrau,A.Moreno,X.Jové. "A speech enhancement system using higher-order AR estimation in real environments". Proc. EUROSPEECH'93, pp 223-226. Berlin, Germany. September 1993