SPEECH ENHANCEMENT BY ADAPTIVE WIENER FILTERING BASED ON CUMULANT AR MODELLING

E. Masgrau*, J. Salavedra**, A. Moreno**, A. Ardanuy**.

* Department of Electrical Engineering and Computers.
Centro Politécnico Superior de Ingenieros. University of Zaragoza.
María de Luna, 3. 50015-ZARAGOZA. SPAIN.

* * Department of Signal Theory and Communications.
E.T.S.I. de Telecomunicación. Politechnic University of Cataluña.
Apdo. 30002. 08080 BARCELONA. SPAIN.

In this paper we study some speech enhancement algorithms based on the iterative Wiener method due to Lim and Oppenheim [1], but where the AR spectral estimation of the speech is carried out using a third order cumulant analysis. This work extends some preceding papers due to the authors, providing a detailed analytical study of the convergence of the iterative algorithms. This analysis allows to understand the behaviour of the original [1] and the proposed cumulant algorithms, providing some insigths and suggestions to develop new and improved algorithms. An exhaustive empirical analysis establishes that the hybrid algorithm presented in [3] by the authors outperforms the original correlation algorithm, specially at low SNR largerly .

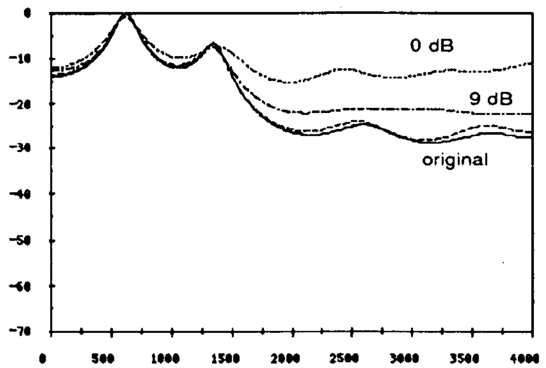
1. INTRODUCTION

The number of signal processing applications based on higher order statistics (HOS) is continuously stepping up in the last years. The principal and more esteemed properties of the so called higher order cumulants are their ability to estimate the phase of the non-Gaussian parametric signals and to distinguish between Gaussian and non-Gaussian processes [2].

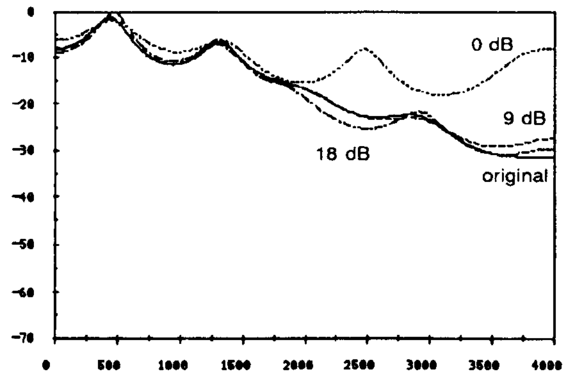
A promising application field of the HOS is that of speech processing. Many applications of speech processing (e.g., speech recognition, source speech coding) that show very high perfomance in laboratory conditions degrade dramatically when working in real environments. It is due principally to the low robustness in front of the noise offered by the standard signal processing algorithms. A possible solution of this problem consists of a preprocessing front-end in order to enhance the speech quality by means a noise reduction or a speech parametric modelling insensitive to the noise. In this paper we analyse a family of preprocessing approaches based in HOS analysis.

A preliminar study of these algorithms was presented in [3]. They are derived from the iterative speech enhancement method based in a sequential MAP estimation of the speech originally formulated by Lim-Oppenheim [1]. The original method consists of an iterative Wiener filtering of the noisy speech based on spectral estimation of the noise (obtained in non-speech frames) and an AR modelling of the speech. This speech model is continuosly improved by using the filtered speech obtained in the preceding iteration. The convergence of the algorithm is very impaired by the residual noise influence in the speech AR modelling. Also, this noise-speech coupling causes a spectral distortion and a subsequent loss of the speech intelligibility .

The use of the higher order cumulants for speech AR modelling provides the desirable uncoupling between the noise and the speech. 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 an spectral AR modelling of the speech more independent of the noise by using, e.g., the third order cumulants of the noisy speech instead of the common second order cumulant or autocorrelation. In Figure 1 is illustrated this property for several noise levels.



Correlation (second order)



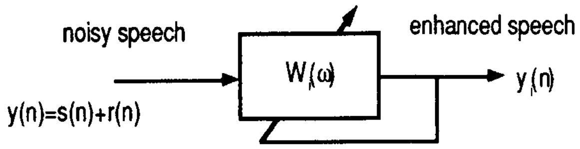
Cumulants (third order)

Figure 1.Noise influence in the AR spectral estimation of the speech using second and third order cumulants.

The problem arise from the higher spectral distortion presented by the AR modelling based on cumulant estimation when it is compared with the autocorrelation case. It is due to the higher variance of the cumulant estimation and the questinable "flatness" of the error sequence when the obtained AR inverse filter works as a predictor over the speech signal. These drawbacks advise to make no more of two iterations when using cumulant AR modelling. In [3] we proposed a hybrid method consisting in 1 up 3 iterations using autocorrelation AR modelling following to 1 (or 2) iteration based in a cumulant AR modelling. Thus, an AR modelling of the speech based on the third-order cumulants is used for the Wiener filter design in the first iteration of the Lim-Oppenheim algorithm. We hoped a twofold benefit: firstly, the convergence of the iterative algorithm is highly accelerated and the number of iterations and, therefore, the computational complexity can be greatly reduced (ideally, an iteration is sufficient; in practice, the noise is not exactly Gaussian and the speech is not exactly non-Gaussian, and, therefore, more than one iteration is necessary); secondly, a non-contaminated AR parameterization of the speech is directly obtained. This latter results very useful, e.g., in recognition system based on speech parameterization. Following, the 1 up 3 iterations using second order AR modelling are carried out in order to improve the overall algorithm perfomance. The benefices of these aditional second order iterations are favoured by the strong noise reduction provided by the first cumulant iteration.

2. CONVERGENCE ANALYSIS OF THE ITERATIVE WIENER ALGORITHMS

In this paper we extended the previous works and we present a detailed analysis of the convergence characteristics of the iterative Wiener algorithms. In the Figure 2 a scheme of the iterative Wiener algorithm is shown, where the enhanced speech of the preceding iteration, y_i, is used for the speech spectral estimation in the current one.



$$W_i(\omega) = \frac{P_{y_{i-1}}(\omega)}{P_{y_{i-1}}(\omega) + P_r(\omega)} \qquad \text{where} \quad P_{y_{i-1}}(\omega) = \frac{g^2}{\left|1 + \sum_{k=1}^{p} a_k e^{-j\omega k}\right|^2}$$

Figure 2. Scheme of the iterative Wiener algorithm.

The AR spectral estimation of the speech obtained from y_i is:

$$P_{y_{i-1}} = (P_s + \alpha\, P_d) \cdot |W_{i-1}|^2 \qquad (1)$$

where the parameter α represents the uncoupling noise-speech effect due to the cumulant analysis. Thus, the parameter α is 1 in the correlation case (null uncoupling) and 0 in the ideal third order cumulant case. The parameter α is really frequency-dependent and the spectral estimation of the speech, P_s, and the noise, P_d, are not exact. Both facts are not considered in (1) for simplicity.

The estimation recursion of the consecutive Wiener filters results:

$$W_i(\omega) = \frac{1}{1 + W^{c'}_{opt}(\omega) \cdot [W^{-1}_{i-1}(\omega)]^2}$$

being

$$W^{c'}_{opt}(\omega) = \frac{P_r}{P_s + \alpha . P_r}$$

This recursion converges to:

$$W_\infty(\omega) \le W_{opt}(\omega) \quad \text{si } W_{opt}(\omega) \ge C \ ; \quad C = \frac{4-\alpha}{5-\alpha}$$

$$W_\infty(\omega) = 0 \quad \text{si } W_{opt}(\omega) \le C \ ; \quad " \quad "$$

Thus, the value of the variable C ranges in the interval:

$$0.75\ (\alpha=1;\text{corr.}) \le C \le 0.8\ (\alpha=1;\text{cum.})$$

corresponding the limits of this interval to SNR of 4.77 dB and 6 dB, respectively.

This convergence behaviour of the algorithm causes a spectral distortion in both the cumulant and the correlation cases, such as is illustrated in Figure 3.

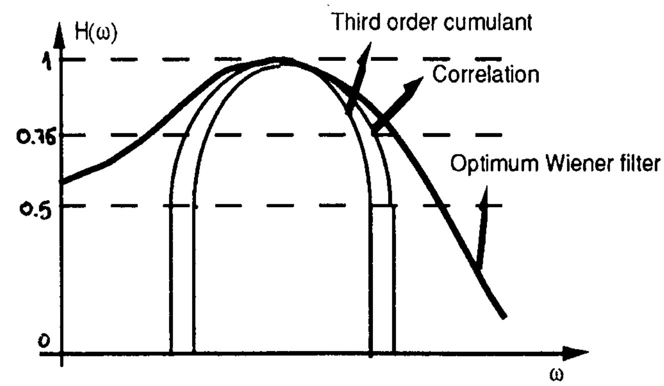


Figure 3. Spectral "peaking" distortion of the Iterative Algorithm.

As can be seen, the cumulant algorithm causes a slightly greater peaking distortion than the correlation one. But the convergence of the cumulant algorithm is faster due the noise-speech uncoupling property, and it allows a reduction in the number of the iterations and, therefore, a smaller spectral distortion than the correlation case.

It is clear from the preceding results that a hybrid algorithm profiting the faster convergence speed of the cumulant algorithm and the more accurate spectral convergence of the correlation one is a good solution.

	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

a)

	SNR	SEGSN	ITAKU	COSH	CEPST
0 iter.	0.00	0.79	9.57	11.67	12.02
1 iter.	7.96	4.87	8.73	10.23	10.15
2 iter.	7.81	5.41	6.25	8.44	8.67
3 iter.	7.85	5.73	5.63	7.91	8.27
4 iter.	7.62	5.75	5.46	7.83	8.35

b)

	SNR	SEGSN	ITAKU	COSH	CEPST
0 iter.	0.00	0.79	9.57	11.67	12.02
1 iter.	7.96	4.87	8.73	10.23	10.15
2 iter.	8.79	5.97	7.31	9.15	9.33
3 iter.	9.00	6.29	6.01	8.15	8.82
4 iter.	8.88	6.33	5.62	7.87	8.65

c)

Table I. Distance measures using the algorithms based on a) second order statistic; b) third order cumulants; c) hybrid at **SNR = 0 dB**.

	SNR	SEGSN	ITAKU	COSH	CEPST
0 iter.	18.00	13.43	6.33	7.89	8.52
1 iter.	21.84	16.83	5.26	6.75	7.17
2 iter.	22.55	17.68	4.32	5.91	6.11
3 iter.	22.55	17.77	3.02	5.03	5.63
4 iter.	22.33	17.60	2.50	4.86	5.69

a)

	SNR	SEGSN	ITAKU	COSH	CEPST
0 iter.	18.00	13.43	6.33	7.89	8.52
1 iter.	21.47	17.01	2.93	5.15	6.03
2 iter.	20.23	16.21	2.82	5.06	6.35
3 iter.	19.46	15.70	2.75	4.96	6.45
4 iter.	19.37	15.60	2.82	5.05	6.56

b)

	SNR	SEGSN	ITAKU	COSH	CESTR
0 iter.	18.00	13.41	6.33	7.89	8.52
1 iter.	21.47	17.01	2.93	5.15	6.03
2 iter.	21.66	17.13	2.82	5.04	6.33
3 iter.	21.47	16.92	2.47	4.85	6.56
4 iter.	19.62	16.28	2.53	4.81	6.17

c)

Table II. Distance measures using the algorithms based on a) second order statistic; b) third order cumulants; c) hybrid at **SNR = 18 dB**.

3. RESULTS

In this section we present an exhaustive evaluation of the correlation, cumulant (third order) and hybrid algorithms. We use as test signal the noise-free utterance provided by ESCA. It is disturbed by additive Gaussian white noise with different global signal-to-noise ratios ranging from 24 dB to 0 dB. The perfomances of the three algorithms are evaluated as in SNR and SEGSNR terms as with Itakura, Cosh and Cepstrum distances. Tables I, II shows the results for two noise levels, SNR=0 db and SNR=18dB, respectively.

As can be seen, the correlation algorithm improves its perfomance continuously but slowly. On the contrary, the cumulant algorithm improves its perfomance quickly but it degrades from the second or third iteration, specially in terms of SNR distances and for high SNR. For low SNR, the cumulant algorithm is more robust, and it degrades from the fourth or fifth iteration. In both cases, the hybrid algorithm behaves very good, and it represents a good tradeoff between the correlation and the cumulant algorithms, specially when the iterations number is limited (2 or 3 iterations) and/or the noise level is variant or unknown.

REFERENCES

[1]J.S.Lim and A.V. Oppenheim."All-Pole Modeling of Degraded Speech". IEEE Trans. on ASSP,pp197-210. June 1978.
[2]C.L. Nikias, M.R. Raghuveer ."Bispectrum Estimation: A Digital Signal Processing Framework". Proc. of IEEE, pp 869-891. July 1987.
[3]E.Masgrau,J.A.R.Fonollosa,A.Ardanuy. "Enhancement of Speech by Using Higher-Order spectral Modelling. Proc. EUSIPCO'92.Brussels.Belgium.August 92.