COMPARISON OF DIFFERENT ORDER CUMULANTS IN A SPEECH ENHANCEMENT SYSTEM BY ADAPTIVE WIENER FILTERING

J.M.SALAVEDRA*,E. MASGRAU**, A. MORENO*, X. 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 estudy 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 a cumulant (third- and fourth-order) analysis. This work extends some preceding papers due to the authors, providing a behavior comparison between the cumulant algorithms and the classical autocorrelation one. Some results are presented considering the noise (Additive White Gaussian Noise) that allows the best improvement and those noises (diesel engine and reactor noises) that leads to the worst one. And exhaustive empirical test shows that cumulant algorithms outperform the original autocorrelation algorithm, specially at low SNR.

1. INTRODUCTION

The use of higher order statistics for signal processing applications has become very popular during 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 [1]. As it is well known, 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.

Recently, the iterative speech enhancement method based in a sequential MAP estimation of the speech originally formulated by Lim-Oppenheim [2] has been object of interest [3] and its perfomance highly improved. This method consists of an iterative Wiener filtering of the noisy speech based on spectral estimation of the noise (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 intelligibility loss of the speech.

The use of the higher order cumulants for the speech AR modelling calculation provides the desirable uncoupling between the noise and the 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 probability density function 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 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. The problem arises of the higher spectral distortion presented by the AR modelling based on cumulants estimation when it is compared with the autocorrelation case. It is due to the higher variance of the cumulant estimation and the questionable "flatness" of the error sequence produced when the obtained 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.

In this paper an AR modelling of the speech based on the third- and fourth-order cumulants is used in the Wiener filter design and therefore a less contaminated AR parameterization of the speech is directly obtained. This results very useful, e.g., in recognition system based on speech parametrization [4]. In section 2 four different approaches to this speech enhancement system are described. Section 3

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

0-7803-1238-4/93 $3.00

contains the evaluation of these implementations under the test conditions and some comparisons among their performance are made. Finally, some conclusions are discussed in section 4.

2. ITERATIVE WIENER ALGORITHMS

Four different implementations of the classical iterative Wiener filtering Method based on an AR modelling of the speech signal have been considered. They have been tested under the same algorithm features:

1) segment the noisy speech by using a 50% overlapping and a frame length of N=256 samples (32ms at 8kHz sampling frequency).
2) window every frame by Hanning windowing.
3) estimate the noise spectrum inside of non-speech frames by means of a smoothing periodogram.
4) estimate the coefficients of the tenth-order AR modelling of the clean speech from the noisy speech signal.
5) dessign the non-causal Wiener filter from the above estimation of the speech and noise spectra.
6) filter the noisy speech frame through the previously designed Wiener filter. We consider a suitable FFT length in order to avoid aliasing effects caused by circular convolution (L=512 points FFT).
7) iterate until maximum number of iterations: GO TO step 4, by using the filtered speech signal instead of the noisy speech signal to estimate the clean speech spectrum.

At first sight, an improvement of the performance can be expected after every iteration since this current AR estimation is carried out from a cleaner speech signal than the preceding iteration estimation. However, other factors sidetrack this iterative algorithm, specially in speech signal disturbed by non-Gaussian noises, as it is discussed below.

2.1 Second-order algorithm

In the original Lim-Oppenheim Method [2] the Wiener filter is defined as

$$H(\omega) = \frac{Px}{Px+Pd} \qquad (1)$$

where Pd is the spectrum of the noise signal d(n), estimated in non-speech frames, and Px is a spectrum estimation of the unavailable clean speech signal. This spectrum estimation is computed by means of a second-order AR modelling from the available noisy speech signal x(n)=s(n)+d(n). To get a better estimation this AR modelling is updated every new iteration from the filtered speech signal obtained in the preceding iteration.

Obviously the filtered speech signal contains a smaller residual noise but it presents a larger spectral distortion. Therefore, increasing the number of iterations doesn't always involve a better speech estimation. It is well known that this algorithm, after processing some iterations, leads to a narrowness and a shifting of the speech formants [3], providing an unnatural sounding speech. In [5] a detailed convergence analysis of this algorithm is carried out. It is proved that this estimated Wiener filter converges to a more selective filter (higher slopes) than the optimum one. Thus it tends to cancel all the signal frequencies with signal-to-noise ratios lower than 4.77dB, and an additional attenuation, proporcionally to the noise level, affects signal frequencies with higher SNR, in comparison to the optimum Wiener filter. Only the non-contaminated speech frequencies undergo a null attenuation.

2.2 Third-order algorithm

The Wiener filtering is computed by means of expression (1), but now the AR modelling is computed from third-order cumulants to get Px. Third order cumulants of every speech frame are computed by using the covariance case:

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

$$0 \le k,i,j \le p$$

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

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

$$1 \le i \le p \; ; \; 0 \le j \le i$$

Considering this third-order AR modelling we hope a twofold benefit: Firstly, the convergence speed of the iterative algorithm is highly accelerated and therefore both the computacional complexity and the intelligibility loss of speech can be greatly reduced;

Secondly, a non-polluted AR parameterization of the speech signal is directly obtained. It is proved in [5] that this third-order Wiener filter tends to cancel more frequencies than the filter estimated by using the classical correlation method, depending on the grade of speech-noise independence provided by the cumulant analysis. Thus a higher "peaking" or "narrowness" effect of the speech formants is brought about.

2.3 Hybrid algorithm

We have seen that the number of iterations is hardly limited when we use the third order algorithm. Fortunately this algorithm provides an important enhancement with only one or two iterations. Therefore an hybrid algorithm was proposed in [5] consisting of one up to three iterations using autocorrelation AR modelling following the first iteration based on a cumulant AR modelling. This method tries to get advantage of the favourable features of the two previous methods: good convergence speed and a low distortion effect of the speech signal. This method obtains good results when the speech is disturbed by Additive Gaussian White Noise [5].

2.4 Fourth-order algorithm

Sometimes the hybrid algorithm working at specific environments has a worse performance than usual because even the first iteration of the third order algorithm gets no improvement since distortion effect overpowers suppresed noise effect. It must be noted that third-order algorithm doesn't reproduce the symmetrical components of speech signal and the distortion increases. Therefore, we have considered a fourth-order AR estimation because it preserves these symmetrical components. We compute the coefficients in the same way represented by expression (3) but using the fourth-order cumulants instead of third-order ones. We expect to have a fast convergence speed and a low distortion effect.

3 EMPIRICAL EVALUATION

In order to obtain a comparison of the different approaches described in the previous section, we present an exhaustive evaluation of the correlation, cumulant (third- and fourth-order) and hybrid algorithms. We consider the following speech enhancement experience: noise-free utterances are disturbed by additive noises. The results using some sentences from female and male speakers (including an utterance provided by ESCA society), and different kinds of noises: AWG noise and several real noises (diesel engine and reactor noises) are shown in this section, where different global SNR ranging from 0dB to 18dB have been considered.

The performance of these four algorithms is evaluated in terms of the standard spectral measures such as Itakura, Cosh and Cepstrum distances. We can see in Table 1 that the improvement obtained over the second order algorithm is very considerable for any number of iterations, when the additive noise is AWGN at a level of SNR=0dB. Because of the properties of cumulant estimation we get the best achievement (in comparison to the second-order algorithm) when the noise is AWGN. In the second-order approach the improvement increases gradually, but slowly, iteration by iteration. On the contrary, the third- and fourth-order methods obtain a good improvement (about 3 dB) after two and three iterations respectively, obtaining a faster convergence speed. The fourth-order algorithm enhances the noisy speech at the same convergence speed as the hybrid one and a little bit slower than the third-order one, however its speech distortion results less important when a listening test is made, since the symmetrical components of the speech are preserved in a better form.

Itakura distance weighs the spectral resonances and these spectrum frequencies are well preserved by all these cumulant approaches. Then a distortion in the remaining frequencies is less notorious for this spectral distance measure. Therefore we have represented Cepstrum distance in the figures to support the remarks because it looks at the overall spectrum in a more uniform way and it is more sensitive to the distortion in valleys and flat zones of the spectrum, since the known peaky effect of the iterative Wiener filtering methods [3] causes higher distortion in these zones. So it is interesting to use a measure that considers this effect to evaluate the performance of cumulant techniques that goes towards a less accurate spectral convergence [5].

Figure 1 shows the values of cepstrum distance versus number of iterations when a high noise level (SNR=0dB) is added to ESCA utterance. Figure 1a shows the above remarks: third-order approach has a faster convergence but its distortion is greater. So a good trade-off between convergence speed and distortion was obtained considering the hybrid

algorithm [5]. Fourth-order algorithm gets the same performance than the hybrid one and better than the correlation one. When the noise is not Gaussian the behavior is quite different: the performance decreases after first iteration for all of the algorithms. In figures 1b and 1c the cumulant algorithms overcome the second-order one, specially the fourth-order one, when only one iteration is processed.

These results are speaker dependent since third-order algorithm performance decreases when diesel engine and reactor noises are considered, while the second- and fourth-order approaches get similar performance. Considering a medium noise level of SNR=9dB (figure 2) we obtain an enhancement of 3 dB in the cumulant approach and the performance is greatly overcome in comparison to autocorrelation approach, when we take ESCA utterance disturbed by AWGN (fig.2a). We assess similar conclusions when the noises are not Gaussian: the performance decreases after the first iteration; third-order performance decreases when we disturb with diesel engine noise while the remaining approaches have similar behavior (fig.2b) and fourth order algorithm assess better results when we add reactor noise (fig.2c).

Another kind of results are shown in figure 3. Cepstrum distance is represented versus SNR ranging from 0 to 18 dB considering the first iteration of the three algorithms that have been tested before. Second-order approach obtains an uniform improvement independently of the noise level and the kind of noise we add to the noise-free speech sentence. Fourth-order algorithm assess better improvement than second-order one at low and medium SNR for all of different noise natures. The improvement of third-order algorithm decreases when SNR increases because spectral distortion effect is dominant and sometimes it leads to worse results than either second-order algorithm or no-filtering case in a reduced subset of speakers, when diesel engine noise is considered. So in this case Wiener filtering based on third order cumulants produces distortion in the noisy signal without considerable noise reduction because of the spectrum of this noise.

4 CONCLUSIONS

A speech enhancement method based on an iterative Wiener filtering have been proposed in this paper. Spectral estimation of speech is made by means of an AR modelling based on third- and fourth-order cumulant analysis to provide the desirable uncoupling between noise and speech. Some different approaches of the Lim-Oppenheim algorithm using cumulant AR estimation have been compared to the classical autocorrelation algorithm. Cumulant based algorithms assess better results when noise is AWGN. So the hybrid algorithm represents a good trade-off among convergence speed, distortion effect and computational complexity. However the performance of the third-order approach decreases when other kind of noises (diesel engine and reactor noises) have been evaluated, whereas fourth-order algorith has the best performance in most part of experiments.Finally, the convergence of the iterative algorithms based on cumulant AR estimation is strongly accelerated. Therefore, fourth-order algorithm needs only the first iteration to assess the same improvement as the classical autocorrelation method after more than three iterations, and sometimes the implicit distortion of the iterative fitering leads to lower improvement for any number of iterations by using the latter method.

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

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.	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

c)	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

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

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.H.L. Hansen and M.A. Clements."Constrained Iterative Speech Enhancement with Apllications to Spech Recognition". IEEE Trans on Signal Processing, pp 795-805. April 1991.

[4] J.M. Tapia and E. Masgrau."Reconocimiento del Habla en Ambientes Ruidosos". Proc. Congreso URSI'91. Septiembre 1991. Cáceres.Spain.

[5] 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.November 92.

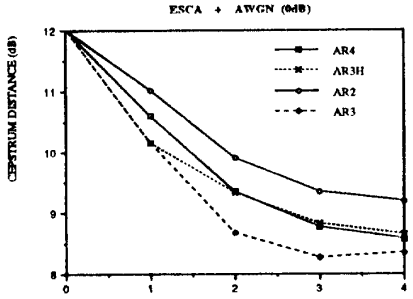

figure 1.a

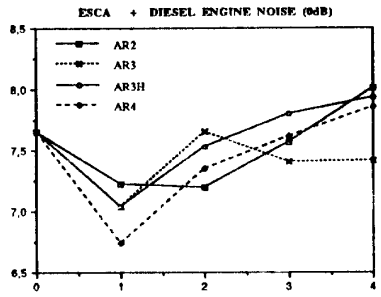

figure 1.b

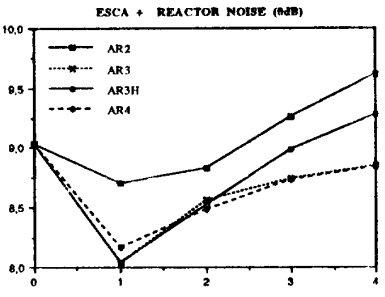

figure 1.c

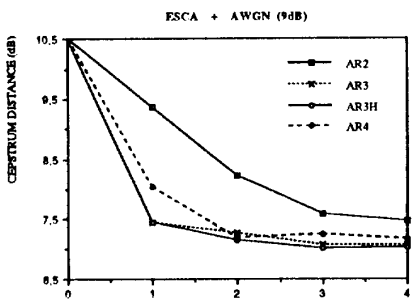

figure 2.a

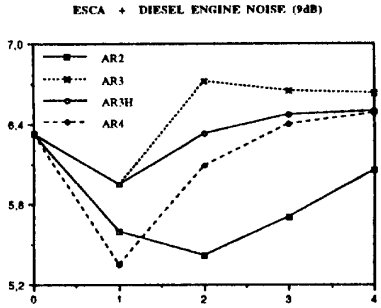

figure 2.b

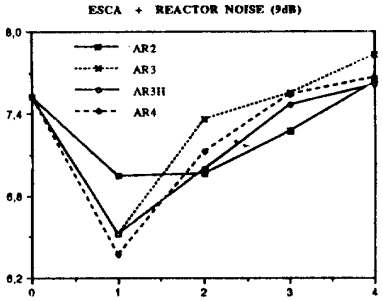

figure 2.c

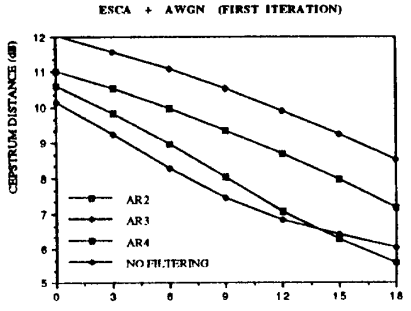

figure 3.a

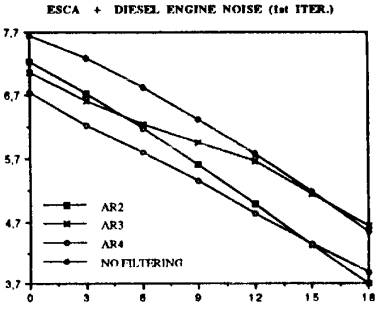

figure 3.b

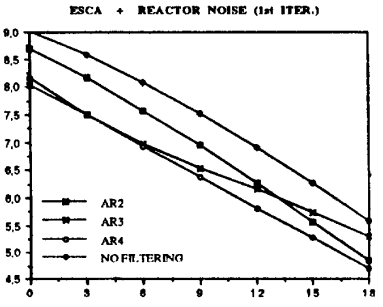

figure 3.c