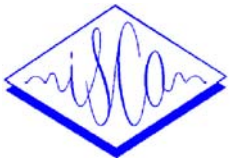


EUROSPEECH '97
5th European Conference on Speech Communication and Technology
Rhodes, Greece, September 22-25, 1997

NON-QUADRATIC CRITERION ALGORITHMS FOR SPEECH ENHANCEMENT

Enrique Masgrau, Eduardo Lleida, Luis Vicente

Communication Technologies Group (GTC).
Department of Electronic Engineering & Communications
Centro Politécnico Superior. C/María de Luna 3, 50015-Zaragoza. Spain
Universidad de Zaragoza
Tel: +34-976-761930, FAX: +34-976-762111, E-mail: masgrau@posta.unizar.es

ABSTRACT

A new algorithm for speech enhancement based on the iterative Wiener filtering method due to Lim-Oppenheim [1] is presented. We propose the use of a generalized non-quadratic cost function in addition to the classical MSE term (quadratic term). The proposed cost function includes two signal-error cross-correlation terms and a L2 norm term of the filter weights. The signal-error cross-correlation terms reduce both the residual noise and the signal distortion in the enhanced speech. The L2 norm term of the filter weights reduces the overall gain of the filter, decreasing the weight noise variance and removing the side lobe of the filter response. Two solutions to the new cost function are presented: the classical non-causal type (ideal Wiener), working in the frequency domain; and a causal finite length in the time domain. In both cases, as Lim's algorithm, the filter output of each iteration is used as "noiseless" speech signal for the following one. Simulation results demonstrate the effectiveness of these algorithms.

1. INTRODUCTION

As it is well known, many applications of speech processing that show very high performance in laboratory conditions degrade dramatically when working in real environments because of low robustness. In the more common case (which is addressed here) where only the corrupted signal source is available, the noise reduction must be only carried out by exploiting the statistical differences between noise and signal sources. One of the more popular algorithms is the iterative speech enhancement method, originally formulated by Lim-Oppenheim [1] and based in a sequential MAP estimation of the speech. This method consists of an iterative Wiener filtering of the noisy speech based on spectral estimation of the noise (obtained in a non-speech frames) and an AR modeling of the speech. This speech model is continuously 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 modeling. Also, this noise-speech coupling causes a spectral distortion ("peaking" or "narrowness" formant effect) and a subsequent intelligibility loss of the speech. The authors proposed in previous works [2-4] some solutions to these drawbacks based in the use of HOS (Higher Order Statistics) to overcome the mentioned drawbacks of the Lim algorithm. They are based on the HOS uncoupling properties between noise (gaussian supposed) and speech. The obtained results were very good but the computational load was too great due to HOS estimations mainly. In this paper, we propose the use of generalized non-quadratic cost function to design the filter response. Section 2 gives a short review of the iterative Wiener algorithm, the new algorithm is presented in section 3. Preliminary results are discussed in section 4, and finally, conclusions are presented in section 5.

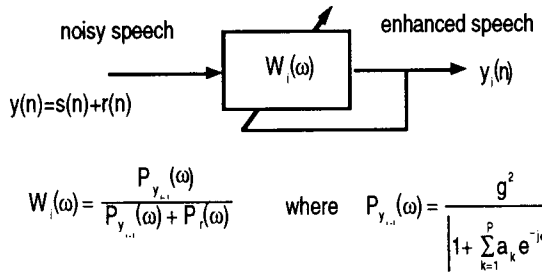
2. ORIGINAL ITERATIVE WIENER FILTERING

In the original Lim-Oppenheim Method [1], noisy speech is enhanced by means of an iterative Wiener filtering that is defined as:

$$W(f) = \frac{Ps(f)}{Ps(f) + Pr(f)} \qquad (1)$$

where Pr(f) is the spectrum of the noise signal r(n) estimated in non-speech frames, and Ps(f) is a spectrum estimation of the unavailable clean speech signal. An iterative Wiener filtering is used to obtain a better estimation of

the AR speech modeling as shown in figure 1.



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

Figure 1. Scheme of the classical iterative Wiener algorithm.

At first glace, an improvement of performance can be expected after every iteration since this current AR speech estimation is carried out from a cleaner speech signal than filter estimation of the preceding iteration. But other factors sidetrack this iterative algorithm and a limitation in the number of iterations must be taken in account. Clearly 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 leads to a narrowness ("peaking" effects of the formants) and a shifting of the speech formants, providing an unnatural sounding speech. These effects can be observed in Figures 2.

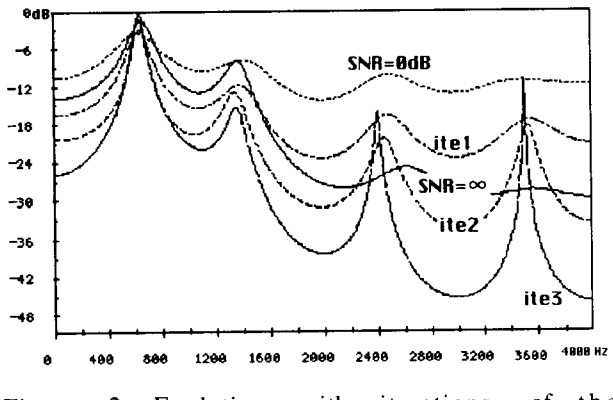


Figure 2. Evolution with iterations of the formant "peaking" effect. SNR=0 dB

The formant peaking effect appears although the exact Wiener filter was used in each iteration, since the MMSE solution cause a spectral envelope distortion. In [2] a detailed convergence analysis of this algorithm is carried out. It is proved that this estimated Wiener filter tends to cancel all signal frequencies with SNR lower than 4.77dB, and an additional attenuation, proportionally 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.

3. THE NON-QUADRATIC ALGORITHM

In the preceding section we commented the existence of a tradeoff between noise reduction level and signal spectral distortion in the classical Wiener algorithm. The speech estimation error obtained in the Wiener filter output consists of two different terms: a signal distortion, and a residual noise. The first term, correlated with the speech, results very more harmful to the listening than the second one, an incorrelated-signal distortion. The Wiener MMSE criterion treats both terms in a uniform way.

In order to earn control over this signal-noise distortions tradeoff we include two terms measuring both effects in the cost function to be minimized. So, we include a filter stress term such as is next explained. Thus, we define the following non-quadratic cost function:

$$L = \beta_1 E\{e^2(n)\} + \beta_2 E^2\left\{\sum_m y(n)r(n-m)\right\} + \\ + \beta_3 E^2\left\{\sum_m e(n)s(n-m)\right\} + \beta_4 E\left\{\sum_m w_m^2\right\} \qquad (2)$$

where e(n) is the global error signal, s(n) is the clean speech signal, r(n) is the noise signal, y(n) is the enhanced speech and w_m are the filter coefficients. The β_i are weighting parameters driving the relative importance given to the different terms. The first term presents a linear dependence with the signal power and the second and third ones present a quadratic dependence. For homogeneity, the β_2 and β_3 parameters are normalized by the input signal power estimation. The significance of all terms are the following:

- 1st term: the classic MSE term of the Wiener filtering.
- 2nd term: claims for the orthogonality between the noise r(n) and the enhanced signal y(n). It works in the same way of the MSE minimization (first term), strengthening the residual noise removal.
- 3rd term: claims for the orthogonality between the residual error e(n) (signal

s(n). It takes care for a minimum signal distortion. It works in a complementary way of the preceding term.

- 4th term: aims to the w_m coefficient solution of minimal L2-norm consistent with the minimization of the other cost terms. It tends to remove the filter weights with a smaller influence in the minimization process, decreasing the noise weight variance, and therefore, removing the side lobe of the filter response.

We points two solution types to the minimization problem of new L cost function. The classics non-causal Lim solution type (ideal Wiener) and a causal finite-length. In both cases, the consecutive iterations are carried out like the Lim algorithm: the filter output of each iteration is used as speech signal s(n) for the following one.

3.1.Non-causal solution type

In this case we consider a non-causal and unlimited length filter (ideal filter response) and the minimization of the proposed generalized cost function results directly in the following filter response:

$$W(f)=\frac{\beta_1 P_s(f)+\beta_3 P_s^2(f)}{\beta_1 [P_s(f)+P_r(f)]+\beta_2 P_r^2(f)+\beta_3 P_s^2(f)+\beta_4} \quad (3)$$

This expression reduces to the known ideal Wiener filter in the case of choosing the weighting parameters as $\beta_1=1$, rest of $\beta_i=0$. The numerator term $P_s^2(f)$ strengthens the high level signal frequencies preventing its distortion. On the contrary, the term denominator term $P_r^2(f)$ strengthens the noise level frequencies for an higher removal of these. Both spectra are estimated like in the Lim algorithm: $P_s(f)$ by means of an AR modeling and $P_r(f)$ with smoothing periodogram of the silence frames. The unrealizability of the ideal non-causal filter is overcome in the same way of the original Lim algorithm, by sampling (fine, N=256 points) of the ideal W(f) response and calculating the filter weights by inverse FFT. Also, the signal filtering is carried out by using 2N-FFT.

3.2.Causal solution type

The preceding solution uses an over-dimensioned filter length to prevent the aliasing and the ripple implied in the inverse FFT design method. A detailed study of the significant length of the design filter suggests that a filter length N=21 would be sufficient. Working in the time domain, the direct minimization of the L cost function (2) leads to a "special normal equation" whose solution is:

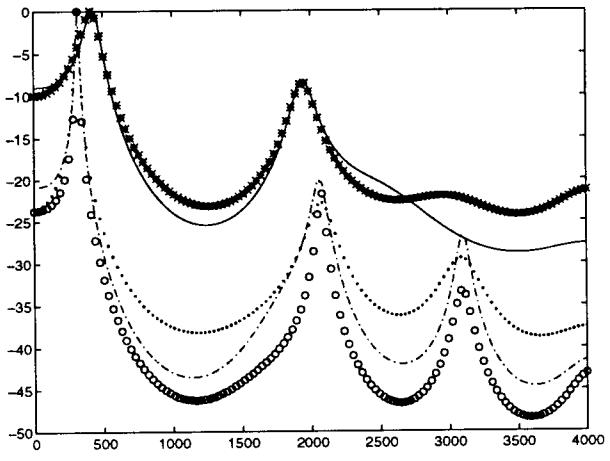
$$\underline{W}=[\beta_1 \underline{\underline{R}}_{xx}+\beta_2 \underline{\underline{R}}_{rr}^2+\beta_3 \underline{\underline{R}}_{ss}^2+\beta_4 \underline{\underline{I}}]^{-1}\,[\beta_1 \underline{P}_{ss}+\beta_3 \underline{\underline{R}}_{ss}\underline{P}_{ss}]$$
$$=\underline{\underline{R}}^{-1}\underline{P} \quad (4)$$

where double underlined indicates matrix and simple underlined indicates vector. The signal correlation coefficients are estimated from the frame data samples and the noise correlation ones from silence frames. This solution requires a great computational burden and, further, the inversion matrix can be produce serious numerical problems. Thus, we prefer to use the Steepest Descent (SD) algorithm, where the gradient estimation responds to the expression: $\nabla_w(n)= \underline{\underline{R}}\,\underline{W}(n)-\underline{P}$.

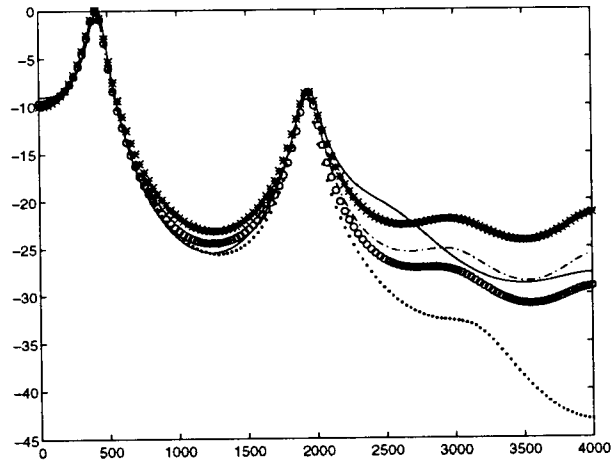
4. PRELIMINARY RESULTS

In this section we present some comparative results obtained with the proposed causal-time domain SD algorithm, which has proven to be the more efficient. Thus, in the figure 3 is shown the LPC envelope of a enhanced speech frame, by using: a) Classic Lim algorithm with up 3 iterations and b) proposed algorithm with two iterations and several combination of β_i parameters. The global SNR of the noisy speech is 9 dB. As can be seen, the peaking effect (increasing with the iteration number) is evident in the Lim results. On the contrary, the peaking effect is very low with the proposed algorithm (especially with $\beta_1=\beta_3=0.5$), and the spectral matching is very high including the valleys. Even the pure MMSE criterion outperforms the Lim results by using the SD algorithm. The listening quality of the enhanced signals confirms these results, resulting always a better quality with the proposed algorithm. With respect to objective measures, not always correlated with the subjective tests, we report an higher SNR improvement in the Wiener algorithm results, at least in the first iteration (after several iterations, the peaking effect becomes dominant and the SNR decreases quickly). On the

contrary, the proposed algorithm gives better results in terms of spectral distances (Cosh, Itakura and Cepstrum), more related with the spectral distortion and in agreement with the results in figure 4. Further results will be available in the paper presentation.



a)



b)

Figure 4. Spectral envelopes of the enhanced speech. SNR Noisy speech=9 dB.

a) Classic Lim algorithm.
Solid line: noiseless original speech.'*':noisy speech. '.','o' and '-.': enhanced speech after 1,2 and 3 iterations, respectively

b) Proposed algorithm.
Solid line: noiseless speech. '*': noisy speech. '.':β_1=1, β_3=0; 'o':β_1=0.7, β_3=0.3; '-.':β_1=0.5, β_3=0.5.
All cases $\beta_2=\beta_4=0$.

5. CONCLUSIONS

A new speech enhancement algorithm based on an iterative Wiener filtering have been proposed. We have defined a generalized non-quadratic cost function including two error-signal cross-correlation terms and a L2-norm of the filter weight in addition to the classical MSE term. The two cross-correlation terms control two components of estimation error: the signal distortion and the residual noise. The L2 weight term aims to remove the unnecessary filter weights reducing the noise weight variance and the side lobe of the filter response. We have proposed two algorithms solving the posed minimization problem: non causal-frequency domain algorithm and causal-time domain algorithm. Two approaches have been presented for the time domain algorithm: block and adaptive Steepest Descent estimation of the filter weights. In anycase, the proposed algorithms outperform the classical Wiener algorithm both in spectral distance measures and subjective listening. In the informal tests, the SD algorithm gives the best quality over all algorithms.

REFERENCES

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

[2] E.Masgrau et al ,"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.

[3] J.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 21-23, 1993.

[4]J.Salavedra,E.Masgrau,A.Moreno,J.Estarellas. "Some Robust Speech Enhancement Techniques using Higher-Order AR Estimation". Proc. EUSIPCO-94. Edinburgh, Scotland, U.K. September 13-16, 1994.