

THE FOURTH ORDER CUMULANT OF SPEECH SIGNALS APPLIED TO PITCH ESTIMATION

Reçu le 04/10/2004 – Accepté le 12/11/2005

Résumé

Dans certain nombre d'application de la parole, tel que le codage, la synthèse ou l'identification, il est crucial de faire une discrimination fiable entre les sons voisé et non voisé et de déterminer la fréquence fondamentale (Pitch). Dans ce papier on utilise le cumulants d'ordre 4 et le modèle (Auto-régressif AR) pour concevoir un nouveau algorithme pour la détection de pitch des sons voisé même dans le cas de présence d'un bruit Gaussien, et de montrer la supériorité de cet algorithme par rapport autre méthodes classiques (méthode cepstral).

Mots Clés : Pitch, Formants, Cumulants, Modèle autoregressif (AR), Cepstre

Abstract

In a number of speech applications, such as coding, synthesis or recognition, it is crucial to make a reliable discrimination between voiced/unvoiced segments and accurately determine the pitch period. The problem of an accurate estimation and decision in noisy condition remains open; Higher-order statistics (H.O.S) have inherent properties that make them well suited when dealing with a mixture of Gaussian and non-Gaussian processes.

This paper explores the fourth order cumulant using autoregressive model (AR(p)) and presents a new algorithm for pitch detection of voiced sounds with and without colored Gaussian noise and shows the superiority of the novel method over the classical methods such as cepstral method.

Keywords : Pitch, Formants, Cumulants, Autoregressive model (AR), Cepstre

H. MAALEM

F. MARIR

Department of electronic,
Mentouri University of
Constantine
Algeria.

ملخص:

المركبتان الأساسيتان للصوتيات الفيزيائية لإشارة الكلام هما ترددات التناغم و التواتر الأساسي. هذه الأخيرة التي تمكننا من معرفة و التمييز بين الصوت فوزي (voisé) و غير فوزي (non voisé).

في هذه المقالة نعطي طريقة جديدة لحساب قيمة التواتر الأساسي في حالة وجود و عدم وجود الضوضاء الغوسية و ذلك با استعمال الإحصائيات من الدرجة الرابعة.

الكلمات المفتاحية: ترددات التناغم ، التواتر الأساسي ، الإحصائيات من الدرجة العليا.

It is well known that the performance of speech recognition algorithms is influenced by environmental conditions in which speech is produced; factors which influenced recognition performance include the noise. Most of the methods developed in this area have been based on the second order statistics or the power information [1].

Higher-order statistics (H.O.S) are a class of tools that have taken a potential in such fields as radar, image processing and others; they are of particular value when dealing with a mixture of Gaussian and non-Gaussian processes in non-linear system [1]. In addition, the H.O.S of speech signals have distinctive features that may be exploited to lead to a better estimation and a more accurate discrimination between various speech types.

In this paper we explore the characteristics of the fourth order cumulant of the speech signals using the linear model (autoregressive AR) and to develop a new algorithm for pitch estimation in the two cases presence and absence of the Gaussian noise. The results are compared with those obtained by the classical methods (autocorrelation and cepstral methods).

1 DEFINITIONS

1.1 Fourth order statistics:

If $x(n)$, $n=0, 1, 2, 3, \dots$ is a real stationary discrete time signal and its moments up to order p exist, then its p^{th} order moment function is given by [2, 3].

$$m_p(\tau_1, \tau_2, \tau_3, \dots, \tau_{p-1}) \\ = E\{x(n)x(n+\tau_1)x(n+\tau_2)x(n+\tau_3)\dots x(n+\tau_{p-1})\} \quad (1)$$

and depends only on the time differences τ_i for all i , here $E\{\cdot\}$ denotes statistical expectation, if in addition the signal has zero mean, then its 2nd and 4th order cumulant function are given by :

2nd order cumulant :

$$C_{2,x}(\tau) = m_{2,x}(\tau) \quad (2)$$

4th order cumulant :

$$C_{4,x}(\tau_1, \tau_2, \tau_3) \\ = E[x(n)x(n+\tau_1)x(n+\tau_2)x(n+\tau_3)] - C_{2,x}(\tau_1)C_{2,x}(\tau_2 - \tau_3) \\ - C_{2,x}(\tau_2)C_{2,x}(\tau_3 - \tau_1) - C_{2,x}(\tau_3)C_{2,x}(\tau_1 - \tau_2) \quad (3)$$

For non zero mean signal, it is enough to replace $x(n)$ by $x(n) - E[x(n)]$.

1.2 Cepstrum

The cepstrum is defined by the opposite transform of the logarithm of the absolute value of the Fourier transform of the signal [4].

$$C_r = \int_{-\infty}^{\infty} \text{Ln} |X(f)| e^{-j2\pi fr} df \quad (4)$$

1.3 Linear speech production model

Many different models have been postulated for quantitatively describing certain factors involved in the speech process, one of the most used models of acoustical speech behaviour is the linear speech production model developed by Fant [5]. The speech production model (AR) is shown in figure 1.

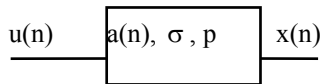


Figure 1 : Linear speech production model

The transfer function $H(z)$ is then:

$$H(z) = \frac{\sigma}{A(z)} = \frac{\sigma}{1 + \sum_{i=1}^p a(i)z^{-i}} \quad (5)$$

Where σ represents the system gain and $a(i)$ the AR coefficients.

Many ways exist for determining the AR coefficients, one of which is based on the following normal equation [3]:

$$\sum_{k=1}^p a(k)r_y(k) = r_y(0) \quad (6)$$

Where r_y is the autocorrelation function.

It has been shown [3] that the AR coefficient can also be determined using 4th order cumulant, giving the following equation:

$$\sum_{k=1}^p a(k)C_{4,y}(\tau - k, k_0, 0) = C_{4,y}(\tau, k_0, 0) \quad (7)$$

With :

$$\tau = 1, 2, \dots, p + n, \quad n \geq 0$$

and

$$k_0 = -p, -p + 1, \dots, 0$$

2. PITCH ESTIMATION

The analysis tries to capture the fundamental frequency of the sound source by analysing the final speech utterance. The fundamental frequency is the dominating frequency of the sound produced by the vocal chords. This analysis is quite difficult to perform. There are several problems in trying to decide which speech signal parts are voiced or not, several algorithms have been developed, but no algorithm has been found which is efficient and correct for all situations; such as the cepstral method or SIFT method (simplified inverse filtering) which is based on the estimation of the maximum of the autocorrelation.

Then, we present in our work the simple novel technique based on the 2nd and 4th order cumulants and the autoregressive model by using the two linear equations (6) and (7), whereas it's diagram of detection is represented by figure 2.

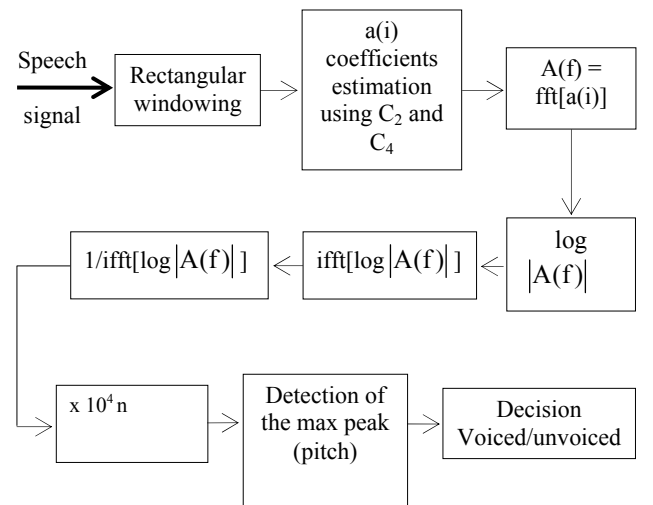


Figure 2 : Diagram of the estimate of the pitch using the 2nd and 4th order cumulants.

3. EXPERIMENTAL RESULTS

The signals used in the applications, consist of 512 recorded sample sections of several voiced sound (Example: Arabic vowels \a, \u and \i of a girl). First, estimation computes the spectrum of these vowels using the autoregressive model (AR) based on the second and fourth order cumulant (figures 3, 4 and 5).

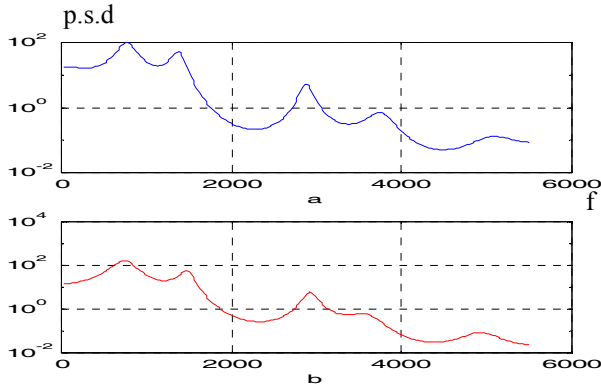


Figure 3 : Spectrum of vowel \a -b- based C_2 -r- based C_4 respectively

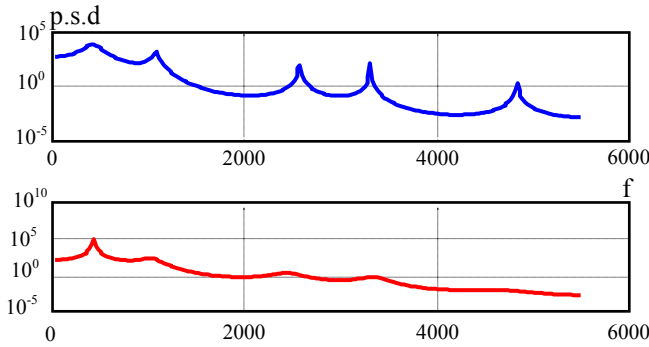


Figure 4 : Spectrum of vowel \u -b- based C_2 -r- based C_4 respectively

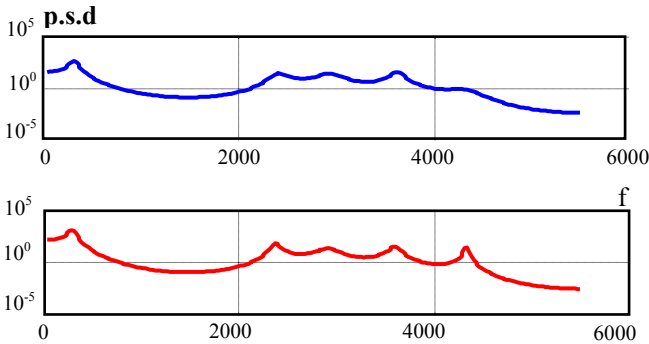


Figure 5 : Spectrum of vowel \i -b- based C_2 -r- based C_4 respectively

The second estimation concerning the detection of the pitch of the vowels by cepstral method and by the proposed technique using the 2nd and 4th order cumulants, the results are given respectively by the figures (6, 7 and 8).

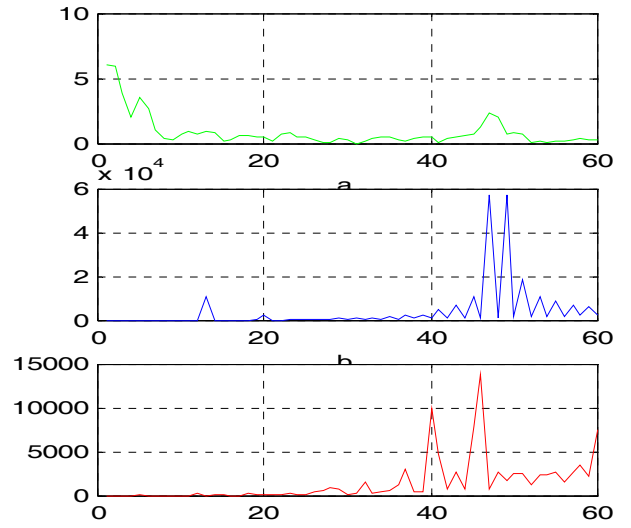


Figure 6 : Pitch of vowel \a -g- based Cepstre -b- based C_2 -r- based C_4 respectively

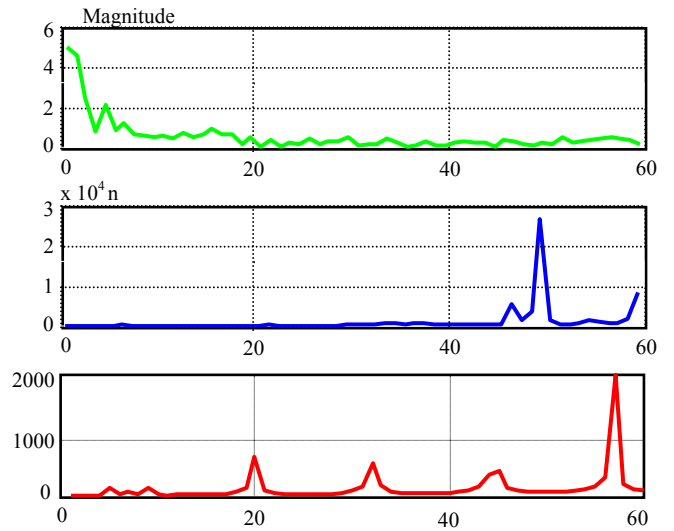


Figure 7 : Pitch of vowel \u -g- based Cepstre -b- based C_2 -r- based C_4 respectively

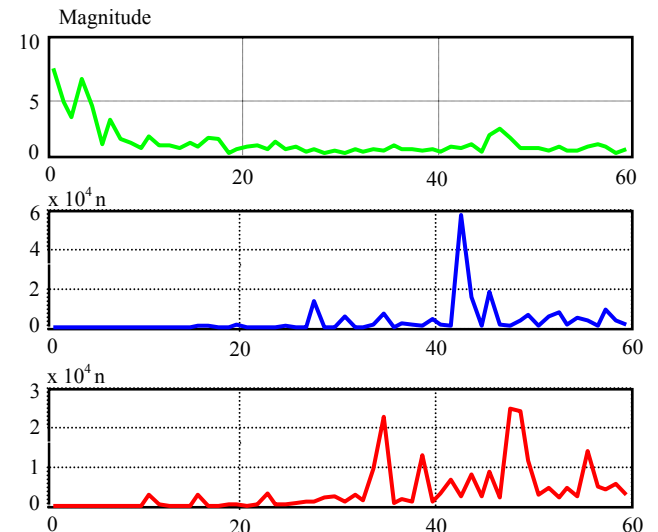


Figure 8 : Pitch of vowel \i -g- based Cepstre -b- based C_2 -r- based C_4 respectively

By contemplating these different figures one notices that according to the proposed technique analysis, the peaks which represent the fundamental frequencies, are very clear in the two cases (autocorrelation and the fourth order cumulant), in comparison to that obtained by the cepstral method.

The estimations of the pitch is then repeated when the vowel \u\ is corrupted by colored Gaussian noise, when the SNR equal 29db, (figure 9).

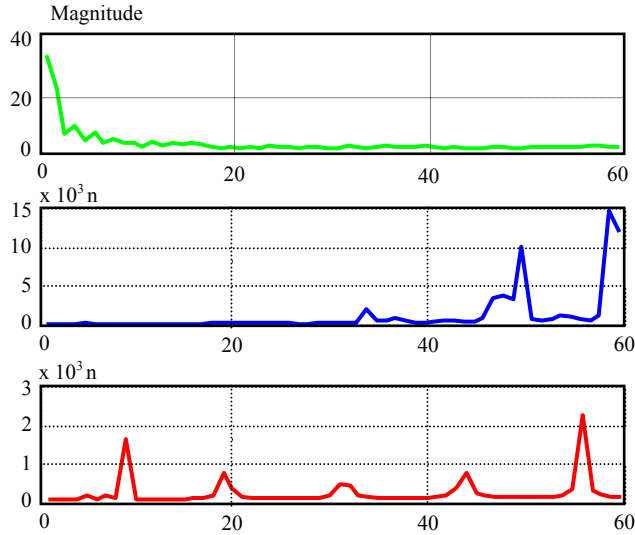


Figure 9 : Pitch of vowel \u\ -g- based Cepstre -b- based C₂ -r- based C₄ with additive Colored Gaussian noise

The values of the pitch corresponding to these peaks and the first three formants of these vowels are given in the table 1, 2 and 3.

Comparing the results in the absence and the presence of the noise, it is observed that the analysis based 4th order cumulants preserves its results.

Table 1 : Values of the first three formants and the pitch (in Hz) of the Arabic vowel \a\.

	Pitch	Formant 1	Formant 2	Formant 3
Using C ₂	225	775	1378	2885
Using C ₄	239	732	1464	2928
Cepstre	234			

Table 2 : Values of the first three formants and the pitch (in Hz) of the Arabic vowel \i\.

	Pitch	Formant 1	Formant 2	Formant 3
Using C ₂	256	301	2411.7	2928.5
Using C ₄	229	279.9	2368.7	2907
Cepstre	234			

Table 3 : Values of the first three formants and the pitch (in Hz) of the Arabic vowel \u\.

-1- without Gaussian noise,
-2- with additive colored Gaussian noise

	Pitch 1	Pitch 2	Formant 1	Formant 2	Formant 3
Using C ₂	220	186	409	1076.7	2584
Using C ₄	193	196	430.7	1012	2433.3
Cepstre	193	689.4			

CONCLUSION

It has been shown, in this paper, that the 4th order cumulant constitutes an efficient tool for the speech signal analysis, and it has also, been shown that the proposed algorithm for pitch estimation gives a better resolution in the two cases, based on the 2nd and 4th order cumulants over the cepstral method.

The blindness of the 4th order cumulant to additive Gaussian noise is an attribute that can be used further to enhance signal analysis in many other fields.

REFERENCES

- [1]- E. Nemer, R.A. Goubran and al "Voicing decision and pitch estimation using third order statistics" Proc. ISCPAT-97, San Diego, CA, USA, September 1997.
- [2]- Elias Nemer, and al "The fourth order cumulant of speech signals with application to voice activity detection" IEEE Trans. On signal Proc., vol. 42, N° 1, Jan. 1994, pp. 222-224.
- [3]- J. Mendel , "Tutorial on higher- order statistics" Proc. IEEE, vol 79, N°3, March 1991
- [4]- L. Kadri « Analyse spectrale du signal parole arabe application aux enfants à audition déficiente » Mémoire de magister, Université de Constantine 2000.
- [5]- J. Markel and al « Linear prediction of speech », Springer-Verlag, 1976.
- [6]- J.W.A. Fackrell and al, "quadratic phase coupling detection using H.O.S" IEE Coll. H.O.S. Savoy, Place, London, 22 May 1995