Sciences & Technologie B – **N°27**, Juin (2008), 89-94.

QUADRATIC NONLINEARITY DETECTION IN SPEECH SOUND

Reçu le 21/12/2006 – Accepté le 27/05/2008

Résumé

Ce papier présente l'analyse bispectrale du signal de parole comme détecteur du couplage quadratique de phase (CQP) entre les harmoniques du conduit vocal ; cette propriété nous fait la base de développer un modèle non linéaire quadratique de production de parole, appelé Auto-regréssif quadratique (ARQ), qui peut être une alternative intéressante à la modélisation linéaire classique.

Mots clés *: Phonèmes, Formants, Cumulants, Modèle autorégressif (AR), bispectre, CQP, autorégressif quadratique.*

Abstract

The purpose of this paper is to describe the bispectrum analysis application for detecting quadratic phase coupling (QPC) in speech signal. This information may provide clues about how to construct new models of speech production including quadratic nonlinearities, called autoregressive quadratic (ARQ), which can replace existing linear models.

Keywords *: phonemes, formants, cumulants, bispectrum, QPC, autoregressive model, autoregressive quadratic.*

H. MAALEM,
F. MARIR.
Département d'Electronique
Faculté des Sciences de
l'Ingénieur –
Université Mentouri
Constantine - Algérie

ملخص

في هده المقالة نقوم بالتحليل الطيفي من الدرجة الثانية لإنتاج إشارة الكلام و الدي يمكن استعماله ككاشف الربط الثنائي الموجود بين توفيقات القناة الصوتية, مما يساعدنا على إنجاز نموذج ثنائي غير خطي لإنتاج إشارة الكلام.

***الكلمات المفتاحية*:** ترددات التناغم, الإحصائيات من الدرجة العليا, التحليل الطيفي من الدرجة الثانية, الربط الثنائي

The subject of higher order statistics (H.O.S) has received much attention in recent years. H.O.S techniques are applied to many signal-processing problems, such as speech analysis concentrating on the bispectrum in particular.

The bispectrum is a third order frequency domain measure, which contains information that conventional spectral analysis techniques cannot provide. The bispectrum is developed as a signal processing tool over a period of 30 years [1], and has been used to provide information about signals which are non gaussian, non stationary and non-linear in a wide variety of signals [1], such as speech sounds.

But it raises the question as to whether simple non-linear speech model may be better at capturing some of the naturalness in real speech.

This paper investigates the use of the bispectrum as a detector of quadratic phase coupling (QPC) in voiced speech. It also investigates model with quadratic nonlinearity, to see whether it may be good for speech signals.

The performance of a new proposed speech production model (autoregressive quadratic (ARQ)) in the detection of Arabic vowels \a\, \i\ and \u\ formants is studied; the results show the usefulness of this model.

1. THEORIES

The higher-order statistics are the higher-order moments and certain non-linear combinations of these moments, called **cumulants** [2].

- Let x(n) denote a discrete-time, real-valued stationary zero mean random process. The third order cumulant is given by:

$$C_3(i,j) = E[x(n)x(n+i)x(n+j)] \qquad (1)$$

Where E denotes the expected value

If x (n) is a Gaussian process, then its statistics are completely characterized by its autocorrelation function, therefore, if x (n) is non-Gaussian then its higher-order statistics of carry information, which is not contained in the autocorrelation function.

- Two ways exist for determining the bispectrum, namely **:** non-parametric (indirect and direct method) and parametric method.

1.1. Non parametric methods

1.1.1. Indirect method

The estimation of the third order cumulant is first computed, and then second fourier transform of this cumulant is computed after appropriate windowing [3, 4].

$$B(f_1, f_2) = \sum_m \sum_n c_3(m,n) e^{-j2\pi(mf_1+nf_2)} \qquad (2)$$

1.1.2. Direct method

The measured signal $x(n),\ n = 0,1,..,N-1$ is divided into p segments,b $x_i(n)\ \ i = 0,1,..,p-1$. Each segment can be first multiplied by a window and then its M-point discrete Fourier transforms

(DFT) $X_i(k),\ k = 0,1,...,M-1$,

is computed. The bispectrum is given by :

$$B(k,l) = \frac{1}{p}\sum_{i=0}^{p-1} X_i(k)X_i(l)X_i^*(k+l) \qquad (3)$$

1.2. Parametric method

If $x(n)$ is a linear process :

$$x(n) = \sum_i h(i)u(n-i)$$

and u(i) is some white noise, then the bispectrum is given by [2, 5] :

$$B(f_1, f_2) = c_{3u}(0,0).H(f_1).H(f_2).H^*(f_1+f_2) \qquad (4)$$

Where : H(.) is the Fourier transform of the filter impulse response.

- ***Quadratic phase coupling***

When a signal consisting of two harmonics, say at frequencies f1 and f2, with starting phases $\phi(1)$ and $\phi(2)$, is passed through a square-law device (a quadratic non-linearity), the output consists of harmonics at frequencies f1, f2, f1+f2 and f1-f2, with starting phases $\phi(1)$, $\phi(2)$, $\phi(1)+\ \phi(2)$ and $\phi(1)-\phi(2)$. These harmonics are said to be quadratically phase coupled. This phenomenon is observed in several situations, such as plasma physics, oceanography, EEG signals, signals generated by rotating machinery, etc. [6]

The bispectrum is a complex number, contains Fourier magnitude and phase information; in particular, the following relations are obtained from equation (3).

$$X(k) = |X(k)| e^{j\phi(k)} \qquad (5)$$

$$B(k,l) = |B(k,l)| e^{j\phi(k,l)} \qquad (6)$$

$$|B(k,l)| = |X(k)||X(l)||X^*(k+l)| \qquad (7)$$

$$\phi(k,l) = \phi(k) + \phi(l) - \phi(k+l) \qquad (8)$$

The phase of the bispectrum $\phi(k,l)$ called biphase

So, the quadratic phase coupling (QPC) detector is described, which comprises a magnitude test and a phase test [7]. The following figure shows a simple test for QPC (figure 1).

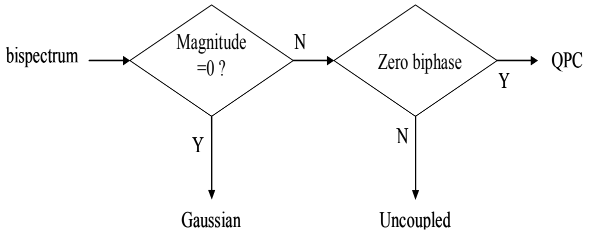

Figure 1 : Two part test for QPC

2. SPEECH PRODUCTION MODEL

2.1. Linear speech production model

Many different models have been postulated for quantitatively describing certain factors involved in the speech process. It can be stated with certainty that no single model has been developed, that can account for all of the observed characteristics in human speech [8].

One of the most used models of acoustical speech behavior is the linear speech production model developed by Fant [8]. The speech production model (AR) is show in figure 2.

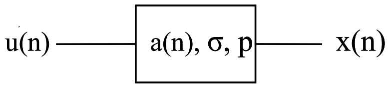

Figure 2 : Linear speech production model

The output $x(n)$ is related to the input $u(n)$ by :

$$x(n) = -\sum_{i=1}^{p} a(i)x(n-i) + \sigma u(n) \qquad (9)$$

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

The transfer function H (z) is then:

$$H(z) = \frac{\sigma}{A(z)} = \frac{\sigma}{1+\sum_{i=1}^{p} a(i)z^{-i}} \qquad (10)$$

The p.s.d. of the AR model is given by :

$$S_{AR}(f) = \frac{\sigma^2}{\left|\sum_{i=1}^{p} a(i)\exp(-j2\pi fi)\right|^2} \qquad (11)$$

2.2 Nonlinear speech production model

The quadratic phase coupling (QPC) can occur in a speech processing as a consequence of nonlinearities in the signal production mechanism. As it has been done with the speech signal produced by the linear model (AR), we propose a simple new nonlinear model of speech production called nonlinear prediction model or autoregressive quadratic (ARQ) (figure 3).

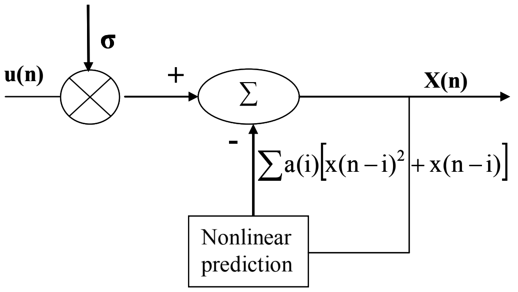

Figure 3 : Nonlinear speech prediction model

The output x (n) is related to the input u(n) by :

$$x(n) = -\sum_{i=1}^{p} a(i)\left[x(n-i)^2 + x(n-i)\right] + \sigma u(n) \qquad (12)$$

The coefficients a(i) can be determined using the autocorrelation function and the third order cumulant.

2. RESULTS AND DISCUSSION

Data acquisition is provided exclusively by arab speakers, with an audio-digital (pc) recording in a quiet room with microphone. Phonemes, words and sentences were recorded (three time or more). Signals were sampled at 11025Hz and quantified on 16 bits using an audio-media card. Therefore, we have chosen for this paper, segments of three Arabic vowels (\a\, \u\ and \i\) pronounced by mature students. Each frame includes 512 samples which correspond to duration of 46ms.

Firstly, Bispectra and their phases of the vowels \a\, \i\ and \u\ are estimated, where obtained by parametric methods. A 14th order AR model, of unity gain, based on the third order cumulant is used for QPC detection (figures 4, 5, 6, 7, 8 and 9) respectively; the bispectra show a single sharp peak and almost zero biphase at around (f_1, f_2) = (687, 601), (279, 150) and (300, 279) respectively. We note that QPC exists in the triplet frequencies (687, 601, 1288) (279, 150, 429) and (300, 279, 579). Notice that only the $0 \le f_1 \le f_2$ portion of the bispectrum is displayed.

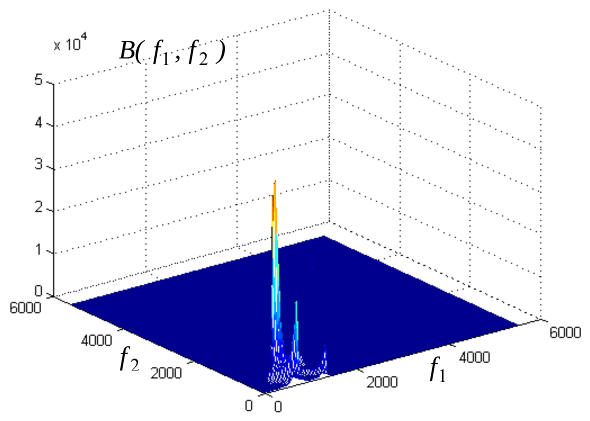

Figure 4 : Bispectrum of the vowel \a\

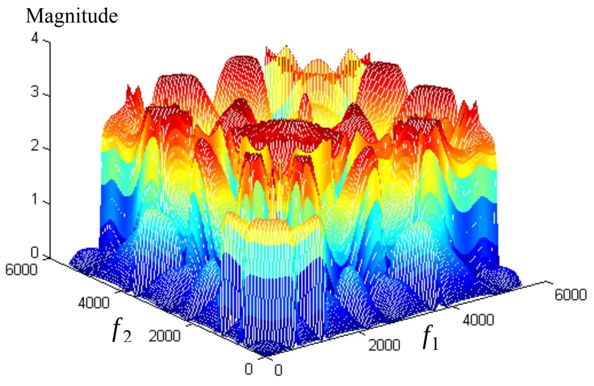

Figure 7: Bispectrum phase of the vowel \a\

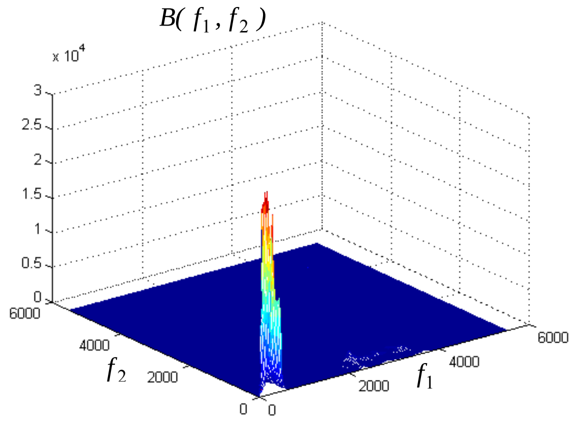

Figure 5 : Bispectrum of the vowel \i\

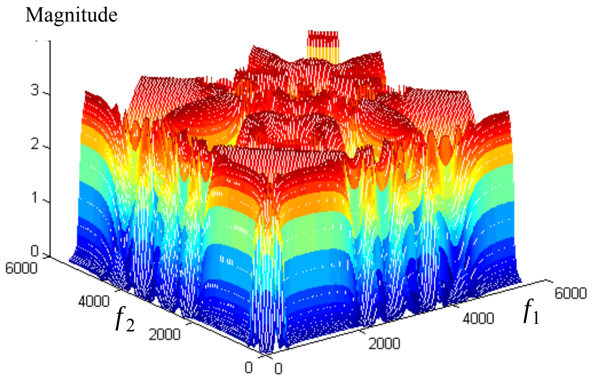

Figure 8 : Bispectrum phase of the vowel \i\

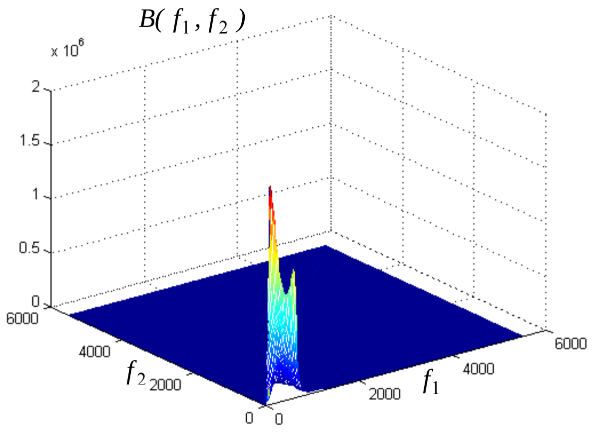

Figure 6 : Bispectrum of the vowel \u \

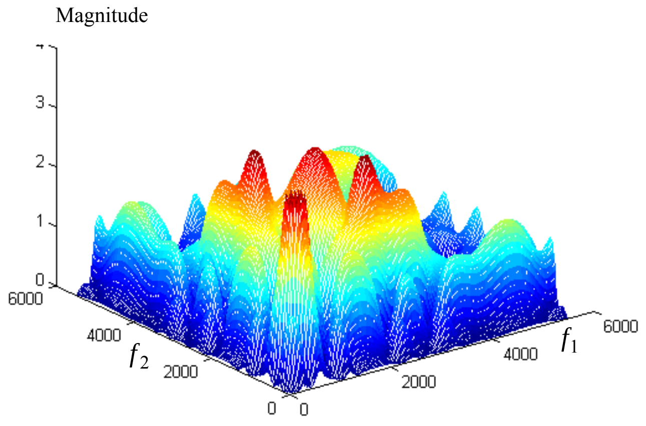

Figure 9 : Bispectrum phase of the vowel \u\

The second estimation computes the spectrum using a 14th order linear AR model (figures 10, 12 and 14).

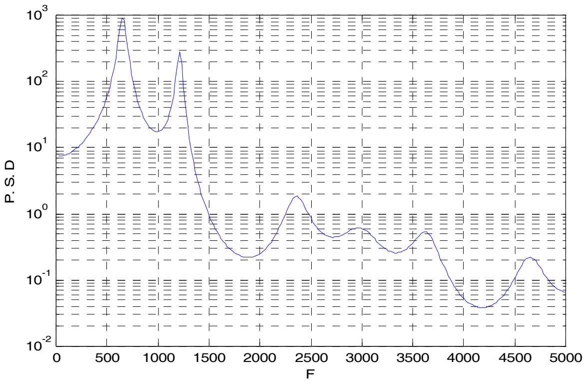

Figure 10 : Spectrum of \a\ using AR model

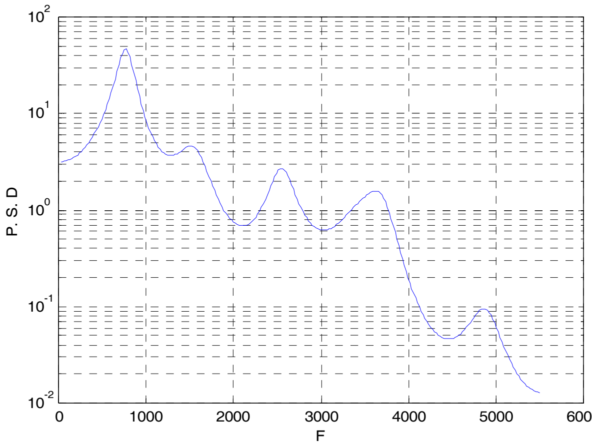

Figure 11 : Spectrum of \a\ using ARQ model

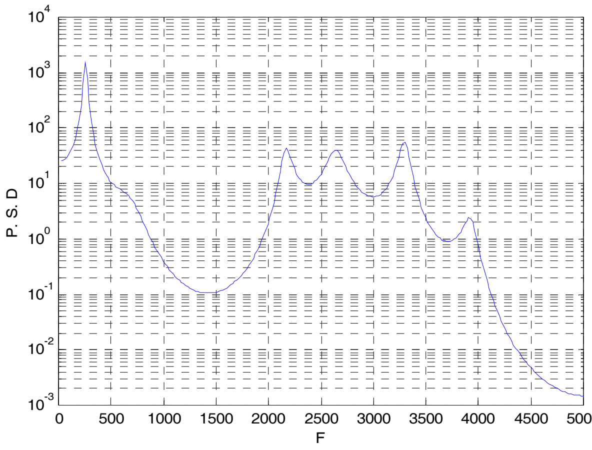

Figure 12 : Spectrum of \i\ using AR model

The same estimation is then repeated with a 14th order nonlinear model (ARQ) (fig 11, 13 and 15)

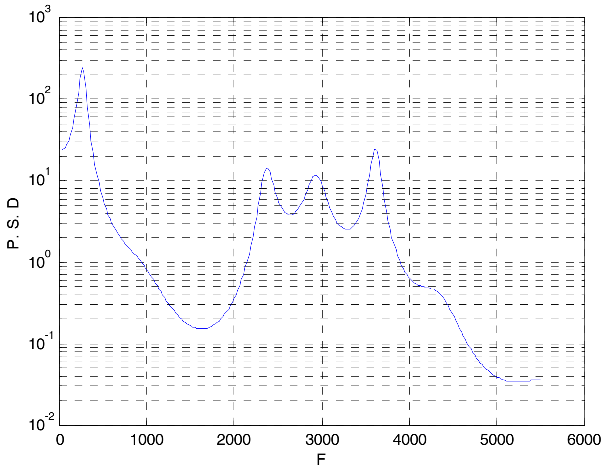

Figure 13 : Spectrum of \i\ using ARQ model

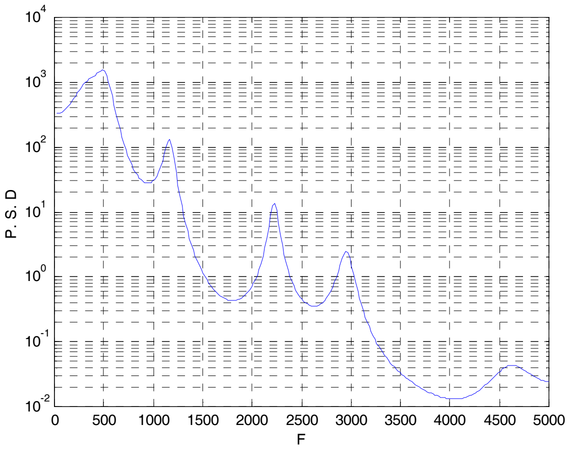

Figure 14 : Spectrum of \u\ using AR model

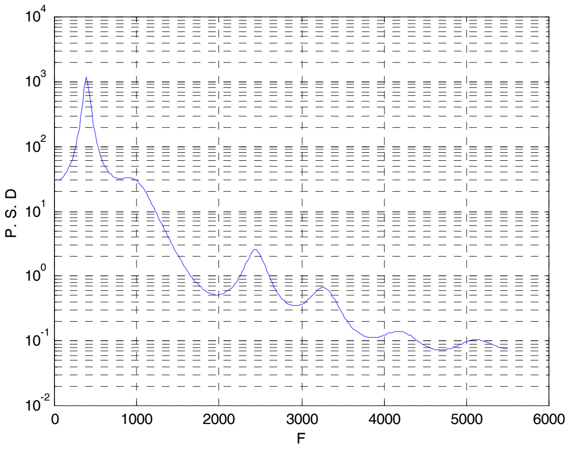

Figure 15 : Spectrum of \u\ using ARQ model

Analyzing the different figures, when comparing the different power spectra (p.s.d); we remark that the proposed nonlinear speech production model (ARQ) gives a good spectral resolution, which can be used to eventually detect formants.

The values of the first three formants are given in tables 1, 2 and 3 respectively :

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

	F1 (Hz)	F2 (Hz)	F3 (Hz)
AR model	709	1332	2599
ARQ model	773	1503	2556

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

	F1 (Hz)	F2 (Hz)	F3 (Hz)
AR model	279	2384	2900
ARQ model	257	2384	2921

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

	F1 (Hz)	F2 (Hz)	F3 (Hz)
AR model	537	1289	2449
ARQ model	386	902	2427

CONCLUSION

It has been shown, in this paper, that the bispectrum based on the third order cumulant constitutes an efficient tool for speech analysis, and can be used to effectively detect quadratic phase coupling.

It has, also, been shown that the proposed nonlinear model of speech production (autoregressive quadratic) can be used for formants detection for the Arabic vowels \a\, \i\ and \u\.

The idea behind our work was to test for quadratic phase coupling (QPC) between the formants. If such a QPC exists, then a nonlinear speech production model would give, in our opinion, better results than the existing linear model. For the considered vowels, a QPC was not found.

REFERENCES

[1] W. A. Fackrell "Bispectral analysis of speech signals" Thesis The University of Edinburgh, September 1996.

[2] J. Mendel "Tutorial on higher- order statistics" Proc. IEEE, vol 79, N°3, pp. 277-303, March 1991

[3] H. P. Rathy, S. Prasad and S. D. Joshi "Conditions for third order stationarity and ergodicity of a harmonic random process" IEEE Trans. On signal Proc., vol. 42, N° 1, pp. 222-224, Jan. 1994.

[4] V. Chandran and S. L. Elgar " A general procedure for derivation of Principal domains of higher-order spectra", IEEE Trans. On Signal Proc., vol. 42, N°1, pp. 229-233, Jan. 1994

[5] J. W. A Fackrell et al., "Non linearity detection for condition monitoring using h.o.s" Internal rep., Dept of Elect. and Elect. Eng., University of Edimburgh, Scotland, pp. 519-528, July 1996.

[6] Matlab version 5 & 6

[7] J. W. A. Fackrell and S. McLaughlin, "quadratic phase coupling detection using H.O.S" IEE Coll. H.O.S. Savoy, Place, London, 22 May 1995

[8] J. Markel and A. Gray " Linear prediction of speech " Springer-Verlag, 1976.