PITCH DETERMINATION OF NOISY SPEECH USING HIGHER ORDER STATISTICS

Asunción Moreno and José A. R. Fonollosa

Universitat Politècnica de Catalunya.
E.T.S.E.Telecomunicació, Apdo 30.002, 08080 Barcelona, Spain
E-Mail: amoreno@tsc.upc.es

ABSTRACT

In this paper it is shown that the use of third order statistics are useful to determine the pitch of a speech signal and how they can eliminate the effect of a wide kind of noises, including those generated by periodic sources. The proposed algorithm is based on the property that higher-order statistics can extract useful information about the statistics of voiced frames, and they can separate speech from noise. Third-order statistics are quite insensitive to most noises (gaussian, sinusoidal, car noise, ...) because these noises have a symmetric probability density function and, therefore, their third-order cumulants are zero. The algorithm has been tested in noise corrupted speech, at different levels of signal to noise ratio, and with different kinds of noise. The results show that this new algorithm gives in all the cases a much better estimation of the pitch than the conventional autocorrelation method.

1. INTRODUCTION

Most of the speech analysis methods developed up to date have been based on the autocorrelation function or power spectrum, and it is well known that second-order statistics completely characterize a gaussian process. However, in many applications where non gaussian processes or non-linearities are present, analysis based on autocorrelation (and hence, on power spectrum) fails to provide all the useful information about the process. Cumulants [1,2], and their Fourier Transform, Poly-spectrum, have information about the presence of non gaussian signals or non linearities, and for this reason there is an increasing interest in their application to signal processing.

There are few articles where higher-order statistics are applied to speech signals. Bispectrum of several English phonemes is studied in [3] and a method is proposed to decide if a given segment of speech is voiced or unvoiced.

This work was supported by the Spanish Government under grand PRONTIC 105/88

The most important conclusion of the results presented in [3] is that voiced phonemes have a non-gaussian distribution and third-order cumulants permit to extract additional information to the provided by the autocorrelation.

Higher-order statistics are also interesting when the speech to process has been recorded in a noisy environment, since an analysis based on cumulants can separate both processes (speech and noise). For example if the noise has a gaussian distribution but the signal does not, it is possible to obtain a non-biased estimation of the LPC parameters. This property is applied in [4] to obtain a robust speech recognition system in noisy environments and it is also the base of the pitch estimator proposed in this paper.

Pitch information is useful in many applications as coding, recognition, synthesis of speech, speaker identification, aids to the handicapped, etc.

Although many pitch detection algorithms have been developed up to now, the problem of a correct detection still remains open [5]. Methods based on second-order statistics, autocorrelation or its Fourier Transform, have been shown to be efficient. A comparative analysis of the more significant algorithms developed until 1975 can be found in [6]. However, in noisy environments, pitch detection algorithms fail, and robust systems against noise has been usually developed and tested only with white gaussian noise [7,8]. Few articles consider other kind of noises [9].

The algorithm presented in this paper has been developed to obtain a correct pitch estimation even if the signal is corrupted with a periodic noise as those produced by car or aircraft engines. In order to achieve our objective we take into account that most of these noises have a symmetric pdf and their third-order cumulants is zero. On the other hand, voiced frames of speech signals has a bispectrum that permits to estimate the pitch period from these statistics.

2. ALGORITHM

Most of the pitch detection algorithms are based on the autocorrelation (AC) of the frame of the signal under analysis.The autocorrelation is determined by:

0-7803-0532-9/92 $3.00

$$r[k] = \sum_{i=0}^{L-k-1} s[i]\, s[i+k] \qquad k = 0, \ldots, L-1$$

Where s[i] is the frame to analyze, and L its length. The autocorrelation presents a maximum in the signal period and their multiples and pitch is determined from the index where r[k] takes its maximum value:

$$AC = \text{argMax}(r[k]) \qquad P_m \le k \le P_M$$

where P_m and P_M are the minimum and maximum permitted pitch values.

The algorithm developed in this communication is based on the cumulants of order 3, in spite of the autocorrelation. In general, the value of this cumulant depends on two indexes. Nevertheless it is not necessary to compute all the cumulants to obtain an estimation of the pitch. We have come to the conclusion that a good choice is to use the diagonal cumulant slice c[0, k] = c[-k, -k] that is calculated as

$$c[0, k] = \sum_{i=0}^{L-k-1} s[i]\, s[i]\, s[i+k] \qquad k = -(L-1), \ldots, L-1$$

The cumulant of a periodic signal is also periodic, but, in general, we do not have a maximum at the origin (k=0). For this reason we cannot use the position of the maximum to compute the periodicity. It is necessary to develop a method to extract this periodicity of the cumulant slice.

We have found that a robust method to estimate the pitch from the cumulants consist on finding when the autocorrelation of the sequence c[0,k] takes its maximum value. That is, applying to the cumulants c[0,k] the AC method of pitch determination as if it were a frame of the speech signal.

$$R[n] = \sum_{k=-(L-1)}^{L-n-1} c[0,k]\, c[0,k+n] \qquad P_m \le n \le P_M$$

$$MR = \text{argMax}(R[n]) \qquad P_m \le n \le P_M$$

We call the resulting method MR.

3. RESULTS

The pitch determination algorithm proposed in this communication (MR) has been compared against the classical method of pitch detection based on autocorrelation (AC). The objective of this comparison has been to show if the MR method based on third-order cumulants permits to obtain better results in corrupted speech, specially when periodic noise with a period close to the pitch is present.

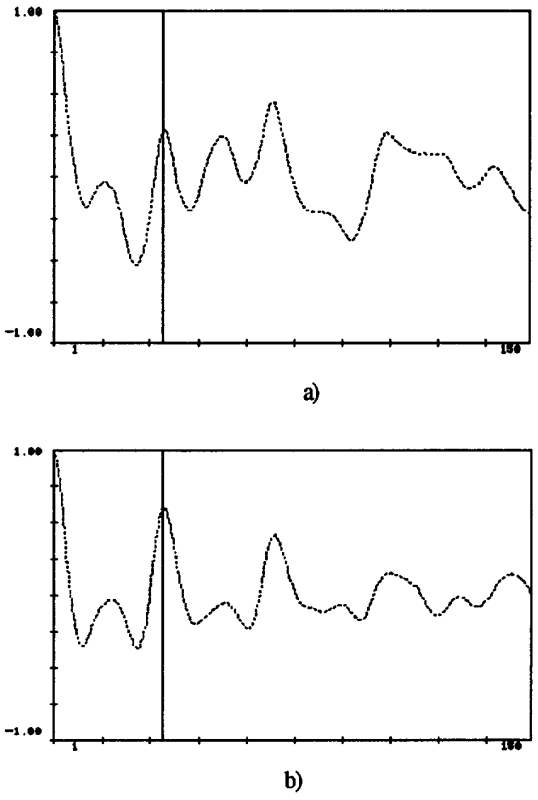


Fig. 1.a) Autocorrelation of a frame of the signal AMB with car engine noise (C4) and SNR = 0 dB. b) Autocorrelation of the cumulants of the same frame. The vertical line indicates the correct pitch value.

Neither preprocessing, as center clipping or inverse filtering; nor postprocessing, as smoothing or tracking, were applied to either the AC or MR methods. In this way we compare the characteristics of the basic extractor.

To test the system, we selected several phonetically balanced utterances by four male and four female speakers that cover a pitch range between 70 and 300 Hz. Recording was made into a silent room.

A pattern pitch was established manually for each utterance by a semiautomatic system. Pitch was evaluated in frames of 40 ms, every 10 ms. This pattern pitch was used as the reference to evaluate the tested methods.

As in the pattern pitch, the pitch detection methods under analysis, AC and MR, work with an analysis window of 40 ms delayed 10 ms. Sampling frequency was 8 KHz. Pitch values where searched in a range between 53 and 400 Hz, that means L=320, P_m=20 and P_M=150.

		20 dB		10 dB		5 dB		0 dB	
		AC	MR	AC	MR	AC	MR	AC	MR
TOTAL	Males	1.89	1.67	4.01	3.22	8.97	6.45	22.89	16.84
	Females	0.00	0.15	3.35	1.03	20.28	7.23	44.13	29.40
	Total	0.94	0.91	3.68	2.12	14.62	6.84	33.51	23.12
WN	Males	1.93	1.58	1.93	1.93	2.81	2.98	5.09	5.61
	Females	0.00	0.20	0.00	0.00	0.60	0.20	5.40	1.00
	Total	0.96	0.89	0.96	0.96	1.70	1.59	5.24	3.31
CAR	Males	1.71	1.45	3.82	2.98	8.90	5.92	25.48	17.98
	Females	0.00	0.10	3.25	1.25	25.15	7.55	50.50	34.50
	Total	0.86	0.77	3.53	2.22	17.03	6.74	37.99	26.24

Table I. Pitch gross errors in %.Results obtained at different SNR. TOTAL includes the results for all the noises tested, WN includes the results for white gaussian noise and CAR shows the results for C2, C3, C4 and C5.

		20 dB		10 dB		5 dB		0 dB	
		AC	MR	AC	MR	AC	MR	AC	MR
TOTAL	Males	1.42	1.61	1.37	1.55	1.37	1.56	1.42	1.65
	Females	0.81	0.85	0.85	0.87	0.90	0.92	1.11	1.11
WN	Males	1.40	1.62	1.40	1.54	1.43	1.59	1.42	1.69
	Females	0.80	0.86	0.85	0.89	0.96	0.98	1.10	1.09
CAR	Males	1.43	1.61	1.39	1.55	1.42	1.59	1.47	1.69
	Females	0.81	0.85	0.85	0.88	0.91	0.94	1.21	1.19

Table II. Average of the variances of fine pitch errors .Results obtained at different SNR. TOTAL includes the results for all the noises tested, WN includes the results for white gaussian noise and CAR shows the results for C2, C3, C4 and C5.

We considered the following noises: white gaussian noise (WN), aircraft take off noise inside the cabin (TK), aircraft take off noise recorded from the airport (RD), noise from a diesel engine (MD), noise inside a car at 2000 r.p.m. (C2), 3000 r.p.m. (C3), 4000 r.p.m. (C4) and 5000 r.p.m. (C5). Noise was added to obtain SNR of 20, 10, 5 and 0 dB.

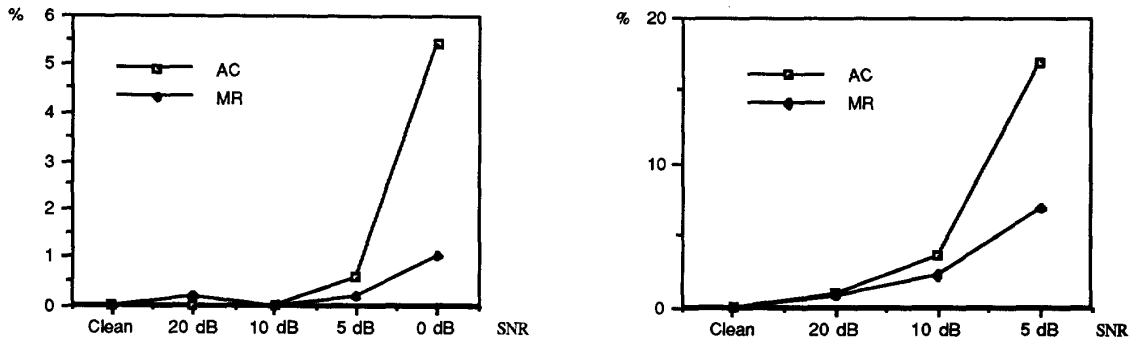
The comparison has been based on the gross pitch errors. We consider a gross pitch error when the difference between the estimated pitch and the pitch from the pattern is greater than 1 ms [6].

Fig.1 shows the results of applying AC y MR methods over the same frame of a female speech signal corrupted with noise of a car engine at 4000 r.p.m. and a signal to noise ratio of 0 dB. The vertical line indicates the correct value of the pitch. It can be observed that MR method cancels the periodic interference (Fig 1b) while the autocorrelation method gives an incorrect pitch estimation.

Table I shows the results of pitch gross errors in % obtained with both methods. For each tested SNR the percentage of gross pitch errors obtained is given. The rows corresponding to TOTAL indicate the average results for all the tested noises. Partial results are separated: WN rows indicate the results for gaussian noise, and CAR rows give the averaged results for all the car noises.

From the results obtained we can see that MR method gives better results than the AC method in almost all the cases. For gaussian noise, AC is a robust method and its results are similar to the obtained with MR, except for SNR=0 dB and female speakers where MR is clearly superior as can be seen in Fig 2 a).

Car engine noise has a clear periodicity and MR gives in this case significantly better results than AC. For female speakers and a SNR of 5 dB Table I shows an error rate of 25,15% for the AC method and of only 7,55% for the MR method. Furthermore, errors in the AC method are related with noise periodicity while errors in the MR method are



a) b)

Fig. 2. Percentage of gross pitch errors for AC and MR methods with a) white gaussian noise and female speakers and b) car engine noise (C1, C2, C3 and C4) and all the speakers.

mainly pitch doubling or pitch halving. Results obtained for car engine noise are shown in Fig 2b).
Table II shows the average of the variances of fine pitch errors. Results are very similar, being slighly higher MR method

4. CONCLUSIONS

From the results we can conclude that the statistical analysis of order three of speech signals provides useful information about the periodicity of this signal. Speech corrupted by noise can be better analyzed using these statistics because the third-order cumulants of most noises are lower than the third-order cumulants of speech signals in voiced frames. A robust procedure to extract the pitch has been developed and tested with good results in different kind of noises. A direct extension of the presented work is the development of a Voiced / Unvoiced detection algorithm.

5. REFERENCES

[1] C. L. Nikias, M. R. Raghuveer, "Bispectrum Estimation: A Digital Signal Processing Framework". *Proc. IEEE*, Vol. 75, No. 7, pp. 869-891, July 1987.

[2] J. M. Mendel, "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications", *Proc. IEEE*, vol. 79, no. 3, pp. 278-305, March 1991.

[3] B. B. Welss, "Voiced/Unvoiced Decision Based on the Bispectrum". in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 1589-1592, May 1985.

[4] K. K. Paliwal and M. M. Sondhi, "Recognition of Noisy Speech Using Cumulant-Based Linear Prediction Analysis", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 429-432, May 1991.

[5] W. Hess "Pitch Determination of Speech Signals". Springer - Verlag 1983

[6] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg and C. A. McGonegal, "A Comparative Performance Study of Several Pitch Detection Algorithms", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, pp. 399-417, Oct. 1976.

[7] D. A. Krubsack and R. J. Niederjohn, "An Autocorrelation Pitch Detector and Voicing Decision with Confidence Measures Developed for Noise-Corrupted Speech", *IEEE Trans. Signal Processing*, Vol. 39, No. 2, pp. 319-329, Feb. 1991.

[8] C. Nadeu, J. Pascual and Javier Hernando, "Pitch Determination Using the Cepstrum of the One-Sided Autocorrelation Sequence", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 3677-3680, May 1991.

[9] J. D. Wise, J. R. Carpio and T. W. Parks, "Maximum Likelihood Pitch Estimation" *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, pp. 418-423, Oct. 1976.