PITCH ANALYSIS OF NOISY SPEECH

Asunción Moreno, José A. R. Fonollosa

Dep. of Signal Theory and Communications
Universidad Politécnica de Cataluña
E.T.S.I.T., Apdo 30002, 08080 Barcelona, Spain

ABSTRACT. In this paper it is shown that third order statistics of speech provides useful information about these signals. The addressed problem is the determination of the pitch in signals corrupted by noise. A wide kind of noises, including those generated by periodic sources have been considered. Results show that the proposed algorithm gives a better estimation of the pitch than a second order statistics based algorithm.

1. INTRODUCTION

Pitch information is useful in many applications as coding, recognition, synthesis of speech, speaker identification, aids to handicaps etc. Most of the robust pitch detection algorithms are based on the autocorrelation of the signal or in the power spectrum [1,2,3,4]. Results are given, generally, over corrupted speech in white gaussian noise but many algorithms fail when some kind of periodicity appears in the noise, specially if it is close to the speaker pitch.

Second order statistics completely characterise a gaussian process. However, in many applications where non gaussian processes or non-linearities are present, analysis based on the autocorrelation (and hence, the power spectrum) fails to provide all the useful information about the process. Cumulants [5,6], 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 [7] and a method is proposed to decide if a given segment of speech is voiced or unvoiced. The most important conclusion of the results presented in [7] 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 have been recorded in a noisy environment, because an analysis based on cumulants can separate both processes (speech and noise). This property is the base of the pitch estimator proposed in this paper.

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 engines or plain engines. In order to achieve our objective we take into account that most of these noises have a symmetric pdf and their third order cumulant 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 analysis methods can be divided into three blocks: a preprocessing block, a basic extractor and a postprocessing block.
Preprocessing of the speech signal as low pass filtering, inverse filtering or clipping are used to decrease the influence of the noise or the influence of the first formant in the determination of the pitch. The basic extractor is usually based or related with second order statistics: Autocorrelation or Power Spectrum (AC, EP, Cepstrum, AMDF, ML). The postprocessing block consist in a smoothing or tracking of the pitch values delivered by the basic extractor to obtain a more reliable pitch curve.

The most common method of basic extractor is based in the autocorrelation of the frame under analysis and is defined as:

\[ r[k] = \sum_{i=0}^{L-k-1} s[i]s[i+k] \quad k=0,\ldots,L-1 \]  \hspace{1cm} (1)

where \( s[i] \) is the frame under analysis and \( L \) its length.

A simple method to determine the pitch is to find where the autocorrelation takes the maximum value.

\[ AC = \arg \max \, r[k] \quad PM \leq k \leq PM \]  \hspace{1cm} (2)

Being \( PM \) and \( PM \) the minimum and maximum pitch values permitted.

The developed algorithm is based on the cumulant of order three. The value of the cumulant depend on two indexes and is defined as:

\[ c[j,k] = E \{ s[i] s[i+j] s[i+k] \} \]  \hspace{1cm} (3)

We have found that it is not necessary to calculate all the values of the cumulant. It is enough to calculate for each frame the one slice cumulant \( c[0,k] \):

\[ c[0,k] = \sum s[i] s[i] s[i+k] \]  \hspace{1cm} (4)

We have obtained that a robust method to estimate the pitch from the cumulants is to find where the autocorrelation of the function \( c[0,k] \) has the maximum value, that is, applying to the cumulants \( c[0,k] \) the autocorrelation method AC as it were an speech signal. We call this method MR.

In order to determine the periodicity of \( c[0,k] \), two periods of the cumulant, at least, have to be calculated. We have chosen these points as \( k \leq PM \). It is possible to calculate more points but will be worse estimated due to the finite frame length.

Fig 1a) shows an example of a frame of a speech signal from a female speaker and Fig 1 b) shows the same signal contaminated with noise from a car running at 3000 r.p.m. The signal to noise ratio for the all utterance
was SNR=0dB and the sampling frequency 8 KHz. In this frame, the correct pitch value was 43 samples, and the periodicity of the noise was 80 samples.

Fig 2 a) shows the autocorrelation obtained from the noisy signal. It can be observed that there is no peak at the correct pitch value, and the first peak that appears is located between the noise periodicity and the double of the pitch period. In this case, the noise affects severely the autocorrelation function. Fig. 2 b) shows the autocorrelation of the cumulant. In this Figure it can be observed that peaks appear in positions related with the pitch period, and the influence of the noise is very low.

3. RESULTS

The developed algorithm to determine the pitch (MR) has been compared against the autocorrelation method (AC). The comparison has been based on the gross pitch errors (>1ms) \[1\].

The objective of the comparison was to determine if the MR method based in third order cumulants permits to obtain better results in noisy corrupted speech signals, specially in periodic noise with period next to the pitch. No preprocessing or postprocessing were applied.

4 male speakers and 4 female speakers were used to test the system. Speakers were chosen to cover a pitch range between 70 and 300 Hz. Utterances were sampled at 8KHz. Patterns of the pitch of the utterances were manually established by a semiautomatic system. Pitch was estimated in frames of 40 ms delayed 10 ms.

We considered the following noises: White gaussian noise (WN), plane take off noise inside the cabin (TK), plane take off noise recorded from the airport (RD), noise from a diesel engine (MD), noise inside a car to 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.

Fig. 3 shows the global results for all the speakers and all the noises. (1) and (4) were calculated in frames of L= 320 samples. It can be seen that cumulant based method always is better than autocorrelation method. The error rate of the autocorrelation method at SNR ~5 dB has been reduced 43%.

Two kinds of noises are specially significant: white gaussian noise, because the autocorrelation method is robust against this noise, and car noises because the periodicity of them is in the allowed pitch range and the power is concentrated around these frequencies.

Comparing the results for gaussian noise, we see in Fig 4 that MR method also reduces significantly the error rate.

We studied the effect of a low pass filtering over the signals. A low pass filter with a
cutoff frequency of 600 Hz was designed. Results for gaussian noise are represented in Fig 4. It can be seen as both methods, AC and MR are improved. However, low pass filtering does not improve the results for periodic noise because an important part of the energy is concentrated in the frequency range of the filter.

In order to improve the results with a better estimation of the cumulant, we enlarged the window analysis to L=470 samples for AC and MR methods. In this case a significant improvement in gaussian noise has been achieved for females speakers as it is shown in Fig 5. Improvements for male speakers were less significants and results very similar for AC and MR methods.

Enlarging the frame length to calculate the pitch in signals corrupted by car noise, also improves the results for MR method while AC method remains the same. A relative reduction of the gross pitch error rate of 63% at SNR=5dB is accomplished. Following the idea of better estimate the cumulant a smoothing of three frames was implemented previous to the pitch period estimation. By this technique, a relative additional reduction of 16% in the error rate was accomplished. Fig. 6 shows the results for all the speakers and car noises.

4. CONCLUSIONS

From the results we can conclude that third order statistical analysis of speech contains important information about the speech signal. Results show that the proposed pitch determination algorithm based on third order statistics provides better results than autocorrelation method in all the tested cases, even in white gaussian noise corrupted speech where autocorrelation is a robust method.

REFERENCES


