# A Robust 2400 bps Subband LPC Vocoder

P.A Laurent, P. de La Noue
Speech Group
Thomson CSF-RGS
BP156
92231 Gennevilliers France
PIERRE.DELANOUE@STS.RGS.THOMSON.FR

## ABSTRACT

This paper presents a new voice coder for applications in future low bit rate communication systems. The emphasis has been put on speech quality, noise robustness and complexity. The coder realizes a multiband+LPC spectral analysis and synthesis of speech.

The transmitted information consists in a LPC10 filter, a set of voicing rates, a pitch, energies, spectral density of excitation in five sub-bands, and information about stationarity of the signal in each half-frame. Depending upon this stationarity, the quantization process is adapted to provide more spectral information (stable speech) or more temporal information (transitory speech).

In order to be less sensitive to surrounding noise, pitch and voicing rates are first computed in each subband. The final values of these parameters are obtained from the values in the current frame and its neighbours.

The excitation signal used at the synthesis side consists in a mixture of isolated pulses, periodic and aperiodic signals of adjustable spectral composition. Tests results are provided.

## 1. INTRODUCTION

Lots of research is currently carried out in several laboratories to replace the LPC-10e algorithm in preparation of a new standard at 2400 bps [3]. Most of the current high quality 2400 bps vocoders are not sufficiently robust to background noise and distorsion, as required in military and mobile communication applications.

The problem addressed in this paper is the evaluation of speech parameters independently from the surrounding noise, especially in the presence of structured and periodic noises. Mixed excitation vocoders [1] seem well suited to this application, especially when a flexible architecture is used [2].We present the architecture of a new subband vocoder combining spectral and temporal informations computed by robust algorithms.

The first part of this paper overviews the general architecture of the coder and describes the evaluation of Pitch/Voicing and transition parameters.The quantization scheme is presented in the second part with a brief description of ellipsoidal quantization. The third part is dedicated to the synthesis of the excitation and the processing of transitional speech. Finally, we present an evaluation of robustness and performances.

## 2. ANALYSIS

### 2.1. Architecture

The architecture is presented on fig 1. It is based upon a LPC vocoder with 22.5 ms frames. Pitch and voicing evaluation are carried out from a semi-whitened residual signal. The signal energy

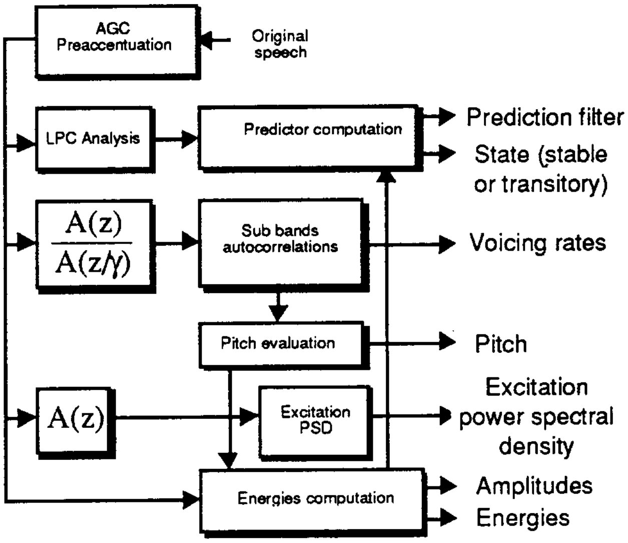

Figure 1 : Analysis architecture

0-7803-2431-5/95 $4.00 © 1995 IEEE

is computed pitch synchronously twice per frame. Other parameters are computed either once or twice per frame, and transmitted only once. For example, the prediction filter is computed in the middle of each frame and at the transition between frames. The filter at the transitions is used for computation of the residual signal and taken into account during quantization.

## 2.2. Pitch evaluation

In a noisy environment, the evaluation of the pitch and voicing parameters is the most difficult task for the vocoder. This evaluation is embedded in the general structure shown fig 1. The analysis of the residual speech from LPC10 is done through three fixed subbands from 300 to 3300Hz. This provides robustness against perturbations even though the analysis looses some accuracy in comparison with systems which use bands defined from instantaneous pitch fundamental frequency.

Several pitch extraction algorithms have been tested.We describe here the simplest one.

The novel approach consists in the evaluation of the voicing and pitch parameters.The pitch evaluation is carried out in two steps.

The first step consists in the computation of the normalized autocorrelation of the signal in each of the three subbands. Since these bands are only 1000 Hz wide the autocorrelation can be down sampled by a factor 4 which allows for a much shorter computation. Fig 2 shows an example of correlation in the three subbands for all values of possible pitch lags (20 to 160 samples) for voiced and unvoiced speech.

The second step consists in tracking the maxima of the subband autocorrelations.

The pitch candidates correspond to autocorrelation maxima. Hence a list of those maxima is drawn starting from the short pitches. In order to avoid pitch multiples, entries in the list are limited to candidates the correlation of which is slightly higher then the correlation of previous entries. This list in then pruned from artefacts in order to get rid of small maxima resulting from noise. The next step consists in setting an interval of decreasing confidence around each maximum by building a Model consisting of normalized pulses lying at the pitch candidates' positions, with a 10% relative width, as shown Fig 3 :

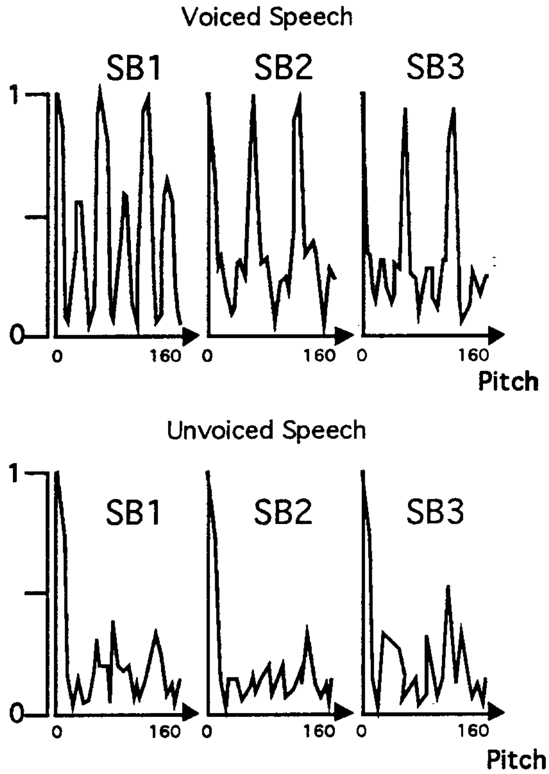

Figure 2 : Sub bands autocorrelations

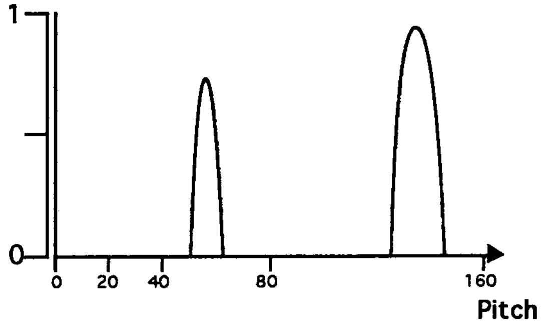

Figure 3 : Autocorrelation Model

In order to automatically track pitch variations, a smoothing algorithm is used, namely:

SmoothModel (frame n) =
0.5 SmoothModel (frame n-1) + 0.5 Model (frame n)

The energy of the current frame is compared with the energies of its two neighbours and an average running energy. If it is too weak, the SmoothModel(frame n) is replaced by SmoothModel(frame n-1 or n+1). Then (the half or the double of) the positions of the maxima of the SmoothModel are compared to the previous pitch, with some tolerance (5 %) in order to find a new pitch value. If there is no clear continuity the pitch is simply taken as the position of the maximum of the smoothed model provided the frame is of sufficient energy.

### 2.3 Voicing rates

The voicing rate in each subband is a function of the value of the corresponding autocorrelation. However the uncertainty interval on the voicing rate increases for low voicing rates. In addition, due to the relative instability of the pitch, the average voicing rate decreases as frequency increases. Conversion of autocorrelations to voicing rates takes this fact into account.

### 2.4. Special case of Transitory Speech

Some segments of speech do not exhibit a very stable structure : the energy of the signal changes very rapidly. The usual coding scheme is not suited to those cases. An alternative approach has been used in those cases where spectral information is of less importance than in stable sounds.

Since the transitionality of speech is connected with unstability, the following criteria are used to decide that a half frame of speech is transitory :

1) the LPC filter gain must be less than some threshold (4.0)

2) and the (pitch-synchronous) energies computed in three 3.75 ms long sub-frames must vary sufficiently (max/min ratio > 3.0).

The filter is then quantized with less accuracy so that a few bits are available for quantizing the relative energies of all sub frames.

The steadiness information is transmitted for each frame. If both half-frames are transitory the voicing information is dropped and the frame is assumed to be fully unvoiced.

On the average about 10% of clean speech frames are declared transitory.

## 3. QUANTIZATION

A new quantization scheme has been implemented for multidimensional data (voicings, energies). It basically consists in considering only the volume which contains most of the observations. This is usually a mutidimensional ellipsoid, not a parallelepiped the corners of which are almost never reached by the data. The axes of the ellipsoid are the M first eigen vectors of the correlation matrix of the observation vectors ; their lengths are proportional to the square root of the corresponding eigen values.The quantized values lie inside the ellipsoid on the vertices of a given lattice : integer coordinates, the sum of which being possibly odd or even or unconstrained (Ellipsoidal Vector Quantization, EVQ).

The number of eigen vectors to keep depends upon the values of the axes of the elllipsoids and the number of bits used to quantize the data. In the current implementation, all multidimensional parameters are quantized with this scheme, except the prediction filter, which is temporarily quantized using scalar quantization.

The transmitted frame consists in 54 bits as shown in the following table :

| Parameter | Frame State | | |
|---|---|---|---|
| | Stable | Partly Transit. | Transit. |
| Pitch<br>Total power<br>Sub Band PSD | 6<br>5+3<br>6 | | |
| Mode | 1 | 1+2 | |
| LPC Filter | 28 | 23 | |
| Voicing rates | 5 | 3 | 0 |
| Sub Frame rel. powers | 0 | 5 | 4+4 |
| TOTAL | 54 | | |

## 4. SYNTHESIS ALGORITHM

The quality of the synthesis depends upon the richness of the excitation. An improvement comes from the introduction of a stationarity dependent excitation. Fig 4 shows different excitation waveforms according to steadiness and voicing.

Transitory speech is synthesized by non periodic pulses. This seems simpler than pitch jitter [1] and allows an accurate time positioning of this excitation in segments where energetic information is more important than spectral information.

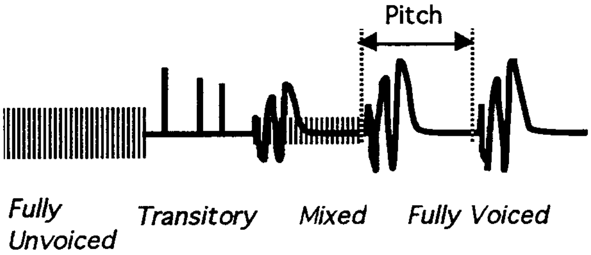

Figure 4 : Composite excitation

When a mixture of unvoiced and voiced excitations is used, each of them is subband filtered in order to obtain the right level of voicing rate and energy as a function of frequency. Energy scaling corresponds to a correction of the LPC synthesis (the latter being an all pole filtering). It permits a better reconstitution of anti-resonances, hence a better synthesis of nasal sounds. In order to take full advantage of this energy description there are six energy subbands.

## 5. PERFORMANCES

### 5.1. Coder Quality

Simplified DRT tests have been carried on. In its current implementation, the new Subband Vocoder performs half way between the LPC10e and 4800 ACELP. With added noise, the vocoder remains intelligible down to a SNR of 10 dB while the LPC10e and other protypes of 2400 bps vocoders are no longer intelligible. The complexity is reduced so that it can run in real time on a TMS320C30 based DSP board.

| Vocoder | DRT Scores |
|---|---|
| 2400 LPC 10e | 94.8 |
| 2400 SubBand | 95.8 |
| 4800 ACELP | 96.7 |

### 5.2 Planned Improvements

The voicing and energy evaluations are computed in evenly spaced bands. The use of a logarithmic scale (MEL) could provide a better match of the ear response. This MEL scaling could also be used to modify LPC analysis itself (spectral weighting).

Depending upon the available computing power, interpolation of synthesis parameters could be performed more frequenly than just once a for each pitch period. However some method of pitch interpolation [4] are under study to decrease the amont of data to transmit concerning this parameter.

The filter quantization has not been optimized and it will be done using the EVQ scheme.

## 6. CONCLUSION

In this paper we describe some of the main points of the vocoder architecture. We show that with a new voicing evaluation and pitch measurement we get a vocoder robust to noise. An additional feature, namely the processing of transitions, allows, when needed, to put the emphasis on either time or frequency description. A new quantization scheme of moderate complexity has also been developped (EVQ). These features make this vocoder a valuable candidate for new 2400 bps normalization.

[1] A. McCree, T. Barnwell "A New Mixed Excitation LPC Vocoder", ICASSP 1991 S9.6 pp 593-596

[2] A. McCree, B. Kleijn "Mixed Excitation Prototype Waveform Interpolation for low bit rate speech coding", IEEE Workshop on Speech Coding for Telecom. Oct 1993 pp 51-52

[3] V. Welch, T. Tremain "A New Government-Standard 2400 bps Speech Coder", IEEE Workshop on Speech Coding for Telecom. Oct 1993 pp 41-42

[4] B. Kleijn, R. Ramachandran, P. Kroon "Interpolation of the Pitch-Predictor Parameters in Analysis-by-Synthesis Speech Coders", IEEE Trans. on Speech and Audio Processing, Jan 1994, Vol 2, n° 1, pp 42-54