SIGNAL PROCESSING IV: Theories and Applications
J.L. Lacoume, A. Chehikian, N. Martin, and J. Malbos (Editors)
Elsevier Science Publishers B.V. (North-Holland)

ENVELOPE AND INSTANTANEOUS PHASE IN RESIDUAL REPRESENTATION

A. Moreno, M.A. Lagunas

Dpto. Teoría de la Señal y Comunicaciones. U.P.C.
E.T.S.I. Telecomunicacion
Apdo. 30.002, 08080 Barcelona, Spain.

Envelope and instantaneous phase of a real signal are two concepts mainly associated with modulation systems. In this paper is shown how envelope and instantaneous phase can be applied in coding systems. In particular,the envelope of the residual provides information about pitch and energy of the signal and can be used to properly codify it. A parametric version of the envelope is taken from the residual of a signal and this information is used in two different coders: an APC coder and a stochastically excited coder. In APC coder the residual envelope is used to quantize the prediction error providing step size and dynamic bit assignment information. In stochastically excited coders, the input to the synthesis filter is composed by the product of the parametric envelope of the residual multiplied by an instantaneous phase taken from a codebook of random phases.

INTRODUCTION

The characterization of the prediction error in speech coding is important to get a good synthetic speech quality. In waveform coders, as ADPCM /1/, the prediction error is quantized and send to the receiver. In this case, a priori estimation of the waveform of the prediction error is useful to quantize it properly.

Stochastically excited coders /2/, by the other hand, represents speech by means of a time varying filter excited by a Gaussian sequence. A time varying filter is built using a short delay filter which provides the formants structure and a long delay filter, (pitch filter),which provides the fine structure for the spectrum. Excitation sequence is selected from a codebook of white Gaussian sequences by minimizing a perceptual error measurement.

Long delay filter is useful in voiced regions but during unvoiced regions has small effect. Therefore it is useful to have a function which preserves pitch and temporal energy evolution of the residual in order to keep pitch information in voiced regions and energy evolution in all kind of regions.

Some efforts have been made in this sense as the system Vector Quantized Multipulse Coder VQMP /3/. This coder uses a model where a short delay filter is excited by a signal obtained from a codebook of multipulse_like sequences, that is, only few values of these sequences are different from zero. Each sequence is multiplied by a polynomial weighting function whose coefficients are estimated in order to minimize a quadratic error between the prediction error and the function itself. This function is calculated every 5ms and discontinuities between frames affects the quality of the synthesized signal.

The envelope of the residual signal, is an associated function which preserves the above mentioned time evolution features of the residual: pitch and energy. This function possesses another important advantage: the positive character of the envelope allows to obtain a parametric version of the envelope, with a considerable data reduction in its representation.

In this paper the information of the envelope is used in two different applications in speech coding. First, is presented how the envelope of the residual can help to quantize properly the prediction error in an ADPCM system. Energy information derived from the envelope are used to control the step size of the quantizer and to assign bits dynamically. Second, the envelope of the residual is applied in a low bit rate coding system where a time varying filter (short delay filter) is excited by a signal composed by a parametric version of the envelope of the residual, multiplied by the cosine of an instantaneous phase. Instantaneous phase is chosen from a codebook in an analysis by synthesis loop in order to minimize a perceptual weighting error.

This work has been partially supported by the CAICYT grant 2196/84 of the Spanish Government.

ENVELOPE AND INSTANTANEOUS PHASE

The extraction of information of a signal and to find alternative versions of representation are two important goals in signal processing. The decomposition of a signal in envelope and instantaneous phase is an alternative representation widely used in the literature for band-pass signals. The envelope and the instantaneous phase of a real signal x(n) are defined from its analytic signal ax(n) defined as:

$$a_x(n) = x(n)+j\ h_x(n) \qquad (1)$$

Being $h_x(n)$ the Hilbert Transform of x(n). $a_x(n)$ is a complex signal whose modulus argument representation is:

$$a_x(n) = e_x(n)\ \exp(\emptyset x(n)) \qquad (2)$$

Where $e_x(n)$ and $\emptyset_x(n)$ are the envelope and instantaneous phase of x(n).

The envelope is a positive function whose energy is related with the energy of x(n) in the following way:

$$Ex = \sum x^2(n) = (1/2) \sum e_x{}^2(n) \quad (3)$$

The envelope can be parametrized in the following way: we can compute samples of the analytic signal Fourier Transform Ax(k) ($0 \leqslant k \leqslant N-1$) obtained by a windowed signal x(n) ($0 \leqslant n \leqslant N-1$). Applying linear prediction over the sequence $A_x(k)$ we estimate the predictor coefficients c(q) (q=1...Q) according to minimize the mean square error.

$$E = \left| A_x(k) - \sum_{q=1}^{Q} c(q)\ A_x(k-q) \right|^2 \qquad (4)$$

Solving this problem by correlation method, a smoothed version $\hat{e}_x(n)$ of the original envelope is obtained.

$$\hat{e}_x{}^2(n) = ko/\left| N(1+\sum_{q=1}^{Q} c(q)\ e^{j(2\pi/N)nq}) \right|^2 \qquad (5)$$

From equation (3), it is shown that the total energy of the signal coincides with half the energy of the envelope. We can't say that the envelope is a time energy temporal density. However, it can be proved that short time average energy of the residual is very close to short time average energy of the envelope. In this sense, $\hat{e}_x(n)$ is a smoothed version of the envelope and can be used to estimate directly the short time energy of the signal.

Pitch information that remains in the residual, also holds in the envelope. Our parametric envelope keeps this information because of the way it has been obtained. Minimizing (4) means to minimize the sum between the quotient of the envelope $e_x(n)$ and its approximation $\hat{e}_x(n)$. As a result, appears that the contribution to the total error is larger when $e_x(n) > \hat{e}_x(n)$ than when $e_x(n) < \hat{e}_x(n)$. As a consequence, the smoothed envelope follows the pitch variations of the envelope because in this points the envelope of the residual takes more significants values.

APPLICATIONS

a) Waveform coding

Adaptive predictive coding is a useful way to encode speech waveforms. SNR can be increased by using a quantizer based in the waveform of the prediction residual. This can be achieved by a parametric version of the envelope of the residual controlling the range of the quantizer and assigning quantification bits dynamically. the basic configuration of the transmission system is depicted in Fig. 1.

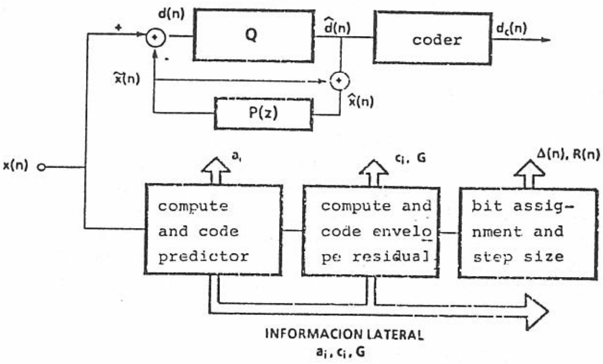

Fig.1 Coder system. Residual envelope is used in the quantizer.

For each input sample to the quantizer, the step size is given by the expression:

$$\Delta(n) = G((R(n))\sqrt{E(n)} \qquad (6)$$

Where R(n) is the number of bits/sample, E(n) is the estimation of the energy of the nth sample and G is the optimal step size for a Gaussian uniform quantizer with unit variance, which is evaluated as a function of the number of bits.

The average number of bits/sample for the input signal is given by:

$$R = \frac{1}{N} \sum_{n=1}^{N} R(n) \qquad (7)$$

And the waveform distortion, defined as the mean square value of the difference between the input and the output of the coder is:

$$D = \frac{1}{2N} \sum_{n=1}^{N} K \, 2^{-2R(n)} \, \hat{e}_x^{\,2}(n) \quad (8)$$

Where K is a term which depends on the probability density function of the residual and we'll assume constant for simplicity. $\hat{e}_x^{\,2}(n)$ is the estimation of the energy.

For a given rate R, the optimal bit assignment that minimizes (8) is:

$$R(n)opt = R + \frac{1}{2} \log 2 \frac{\hat{e}_x^{\,2}(n)}{\prod_{m=1}^{N} (\hat{e}_x^{\,2}(n))^{1/N}} \quad (9)$$

In order to get R(n) an integer value, an iterative algorithm should be used.

This system has work at bit rates from 9.6 Kb/s to 32 Kb/s. The predictor coefficients and envelope coefficients have been calculated in frames of 32 ms (sample frecuency 8 KHz). Bit assignment for lateral information has been done as follows: for each frame of 32ms, 37 bits have been assigned to the parcors of the predictor and 34 bits have been assigned to the envelope parcors. Gain has been quantized with 5 bits. The bit rate for lateral information is 2.5 Kb/s and the others bits are assigned to the residual. This bit assignment has been chosen based in SNRseg measurements and informal subjective quality test. Short delay filter LAR parameters are quantized uniformly. Previously mean and variance had been calculated. The mean value of each parameter is substracted at the transmitter and is added in the receiver. After an envelope sensibility study, envelope parcors have been chosen to be quantized in a similar way. Real and imaginary part of each parcor is quantized independly.

At bit rates lower than 24 Kb/s this coder was compared with ATC /1/. ATC uses a cosinus transform and bit assignment based in a 12 poles LPC model. Quality is compared in SNRseg meausures and informal subjective test. At 9.6 Kb/s ATC gives 1.8 dB over this coder. At 12 Kb/s the SNRseg is the same with both coders and at 16 Kb/s the coder performance is 2 dB better than ATC. At 32 Kb/s this coder was compared with others ADPCM coders with better results.

b) Stochastically excited coders

In this application the information of the envelope of the residual is included in a stochastically excited coder to encode speech at low bit rate. The system is shown in Fig. 2. and consist in a time varying filter (short delay filter) excited by a signal composed by a parametric envelope version of the residual

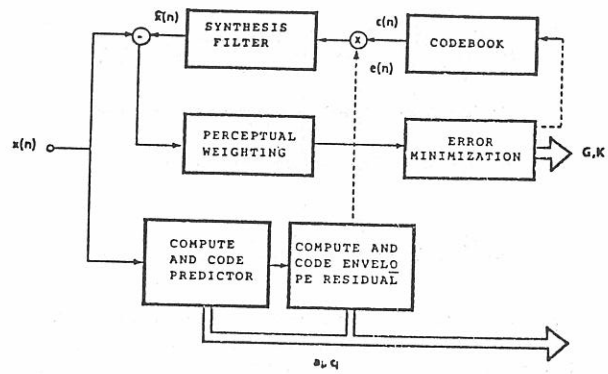

Fig.2. Block diagram of the proposed coder to determine the optimum excitation sequence.

multiplied by a sequence from a codebook.

Sequences are selected by minimizing a perceptually weighting error in an analysis by synthesis loop. The codebook has a total of 1024 positions and each sequence has 40 samples (5ms).

The envelope of the residual is taken from the original prediction error and a parametric version of this envelope is obtained.

The codebook acts as an instantaneous phase generator and the cosine of this phase is taken before it is multiplied by the envelope. In order to keep the envelope in the synthesized residual, signals of constant envelope must be multiplied by the residual envelope, that is, instantaneous phase are generated from random analytic signals of constant modulus. When generating the codebook it is also important to know that the cosine of these phases have to spread the spectrum of the synthesized residual respect the spectrum of the smoothed envelope. For these reasons it was used a Gaussian instantaneous phase generator with the following constraints: the cosine of this sequence is spectrally flat and has constant envelope.

Although these two constraints are very important to keep the desired envelope, the fact that each sequence is selected every a short number of samples makes that not meaningful differences appears between this codebook and a codebook of random phases uniformly distributed between $-\pi$ and π.

Fig. 3. shows different waveforms obtained with this coder and the codebook of random phases uniformly distributed between $-\pi$ and π.

The envelope is parametrized with 20 parcors every 16 ms. It can be apreciated how the envelope follows the waveform of the residual and has information about the energy. The length of the codebook is 1024 and each sequence has 40 samples.

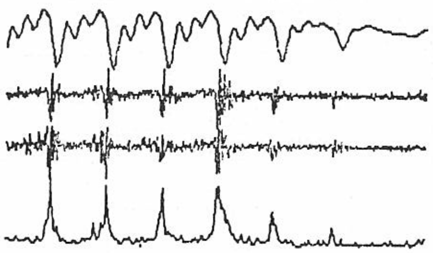

Fig.3. Waveforms obtained in this coder using a codebook of random phases uniformly distributed between $-\pi$ and π
From top to bottom are shown: Synthesized signal, Original residual, Synthesized residual, Parametric envelope obtained eith 20 parcors every 16 ms.

Fig.4 shows the energy of a signal to be coded. This energy has been calculated in frames of 16ms.

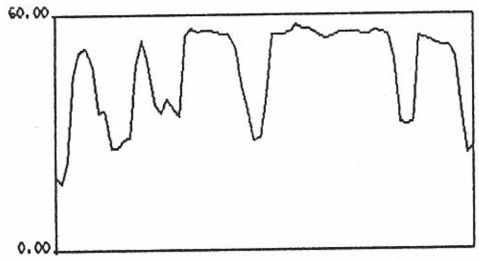

Fig.4. Energy of a signal to code in frames of 16 ms.

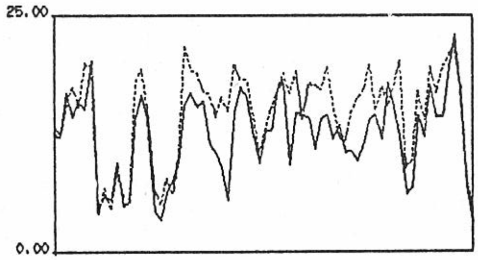

Fig.5 SNRseg with the proposed method and a codebook of random phases uniformly distributed between $-\pi$ and π. Continous line: sy.tem VQMP. Dashed line: proposed coder.

Finaly, Fig.5 shows the SNRseg obtained with this system where the envelope is represented with 10 parcors every 32 ms. The system is compared against VQMP with a codebook of the same lengh and updating the weighting function every 5ms. The proposed method obtains better results than VQMP both in SNRseg meausures and in listening test. This is mainly due to the fact that this system avoid discontinuities beetween frames in the calculus of the envelope.

CONCLUSIONS

In this paper has been shown the importance of the envelope and instantaneous phase of a real signal in speech coding problems.

Has been developed a method that obtains a parametric version of the envelope based in the positive character of this function. This parametric envelope has been used in different coders. First, has been used in a quantizer. The development has been made over an ADPCM system successfully. Has been proved that the envelope provides a valuable information in quantification problems i.e. step size control and variable bit assignment.

Second, the envelope is used in a low rate coding system. A time varying synthesis filter is excited by a function composed by the envelope of the residual multiplied by the cosine of a random instantaneous phase taken from a codebook in an analysis by synthesis loop.

Every system has been compared at the same bit rate with others coders.

REFERENCES

/1/ N.S. Jayant and Noll. 'Digital Coding of Waveform. Principles and Applications to Speech and Video'. Prentice Hall Inc. (1984).

/2/ M.R. Schroeder, B.S. Atal. 'Code excited Linear Prediction (CELP): High-Quality Speech at Very Low Bit Rates. Proc. Int. Conf. on Acoustics, Speech and Signal Proc. pp.937-940. 1985.

/3/ R. Garcia Gómez, F. J. Casajús Quirós, L.Hernández Gómez. 'Vector Quantized Multipulse LPC'. Proc. Int. Conf. on Acoustics, Speech and Signal Proc. Dallas. USA. 1987.