ENVELOPE AND INSTANTANEOUS PHASE CONSIDERATIONS IN SPEECH MODELLING

Asunción MORENO, Miguel A. LAGUNAS

Dpto. Teoría de la Señal y Comunicaciones
E.T.S.I. Telecomunicación. Apdo. 30.002
08080 Barcelona, Spain

ABSTRACT

This paper presents a low bit rate coding system where envelope and instantaneous phase of the residual are used. A time varying filter (short delay filter) is excited by a signal composed by a parametric version of the residual, multiplied by a sequence from a codebook. Two alternatives are studied for the design of the codebook: sequences formed by random pulses and sequences formed by random phases to simulate the instantaneous phase of the residual.

INTRODUCTION

Hight quality synthetic speech at low bit rates has been a topic of extensive research in the last few years.

Multi pulse and stochastically excited coders have been used successfully to obtain a good quality speech signal at low bit rates.

The multipulse model /1/,/2/, uses a short delay filter excited by a sequence of pulses. Each pulse is calculated sequentially in position and amplitude in order to reduce a perceptually weighted error between original and synthetic speech. Amplitude and position of each pulse is coded and transmitted to the receiver.

Stochastically excited coders (CELP) /2/,/3/, represent speech by means of a time varying filter excited by a Gaussian sequence. This sequence is selected from a codebook of white Gaussian sequences by minimizing a perceptual error measurement. 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.

Vector quantized multipulse coder (VQMP) /4/, has been proposed to obtain a multipulse coder so efficient as stochastic coder from the point of view of coding bit rate. 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 5 ms and discontinuities between frames affects the quality of the synthesized signal.

Concerning CELP coders, long delay filter is useful in voiced regions but during unvoiced regions has small effect. In this paper authors propose to use a function which preserves pitch and temporal energy evolution in order to keep pitch information in voiced regions and energy evolution in all kind of regions.

The envelope of the residual signal, is an associated function which preserves the above mentioned time evolution 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.

PROPOSED MODEL

This paper presents a low bit rate coding system as is shown in Fig. 1 where a time varying filter (short delay filter) is excited by a signal composed by a parametric envelope version of the residual, 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 (5 ms).

The envelope of the residual is taken from the original prediction error and a parametric version of this envelope is obtained in the following way:

The analytic signal associated with a real signal x(n) is defined as:

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

ISCAS'88

CH2458-8/88/0000-1557$1.00 © 1988 IEEE

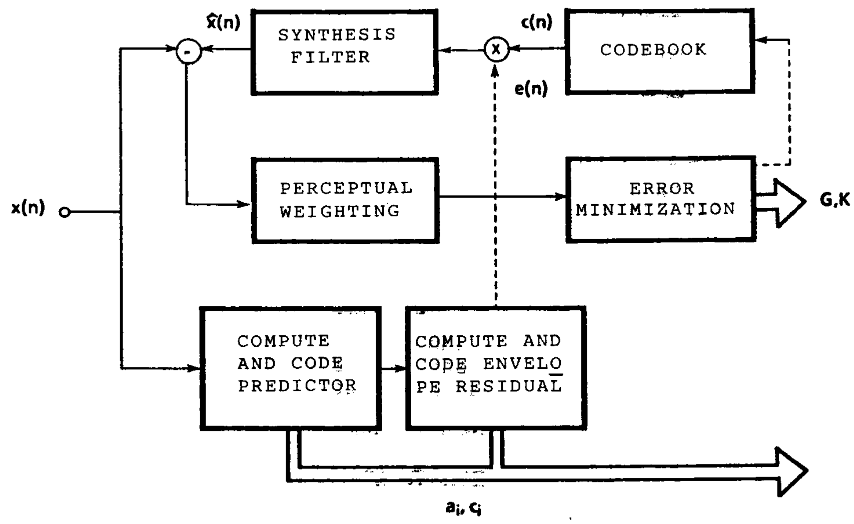

FIG 1. Block diagram of the proposed coder to determine the optimum excitation sequence.

Where $h_x(n)$ is the Hilbert Transform of x(n). The envelope of the signal x(n) is the module of its analytic signal. This envelope can be parametrized in the following way. We can compute samples of the analytic signal Fourier Transform $A_x(k)$ $(0 \leq k \leq N-1)$ obtained from a windowed signal x(n). $(0 \leq n \leq N-1)$ 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.

$$E = \sum_{k=0}^{N-1} A_x(k) - \sum_{q=1}^{Q} c(q)\, A_x(k-q) \qquad (2)$$

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

$$\hat{e}_x^{\,2}(n) = k_o / \left| N\left(1+\sum_{q=1}^{Q} c(q)\exp(j(2\pi/N)nq)\right)\right|^2 \qquad (3)$$

Other spectral estimation methods can be applied, but this method has been chosen due to practical considerations in the implementation of the hardware processor. Avoiding discontinuities in the parametric envelope among frames is specially important in this scheme. This is achieved by properly choosing a trapezoid window to process the signal. A good smoothed residual envelope is obtained with 10 parcors over frames of 32 ms.

CODEBOOK

Two different alternatives have been used when generating the codebook.

The first one uses the same philosophy as VQMC. Sequences are generated with a limited number of pulses distributed in position by means a pdf uniform.

Their amplitude is +1,-1 fixed at random. The length of the codebook is 1024 different sequences of 40 points each.

Next alternative implies to consider the residual as the product of its envelope by the cosine of its instantaneous phase. In this case, the codebook acts as an instantaneous phase generator and the cosine of this phase is taken before it is multiplied by the envelope. In this way, instantaneous phase is chosen from a codebook of 1024 sequences of 40 samples each in an analysis by synthesis loop.

In order to keep the envelope in the synthesized residual, signals of constant envelope must be multiplied by the residual envelope, that is, instantaneous phases are generated from random analytic signals of constant modulus. When generating the codebook it is also important to keep in mind 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 the sequence is spectrally flat and has constant envelope.

As a simplification of this codebook also was used an instantaneous phase generator uniformly distributed over-π and π

RESULTS

In Fig.2 is shown different waveforms obtained with the proposed coder and the codebook of pulses.

Envelope is parametrized with 10 parcors every 32 ms. From this figure it can be appreciated the evolution of the envelope following the waveform of the residual without discontinuities.

The length of the codebook is 1024 sequences of 40 points each. The number of pulses different from zero in each sequence is fixed randomly at tree or four pulses every 20 points.

Fig. 2 Waveforms obtained in the proposed coder using a codebook of pulses. a) Sinthesized signal. b) Original reidual. c) Sinthesized residual. d) Parametric envelope obtained with 10 parcors every 32 ms.

Fig. 3 shows the energie of the signal to be coded in frames of 16 ms.

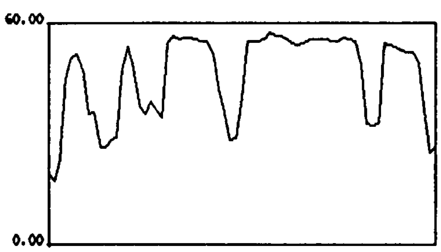

Fig. 3 Energie of the signal to code calculated in frames of 16 ms.

Segmented signal to noise ratio is shown in Fig.4, and compared with VQMC using the same codebook and being the weighting function updated every 5 ms. Continous line is VQMC system and dashed line is the result for the proposed coder. This coder really improves quality respect VQMC and also listening test agree with this. Main improvement in quality is due in avoiding discontinuities in the residual envelope.

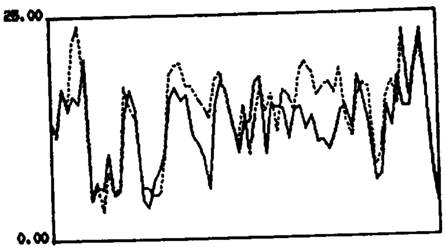

Fig 4. Segmented signal to noise ratio obtained with the proposed coder and a codebook of pulses. Continous line: system VQMP (weighting function updated every 5 ms). Dashed line: proposed method. (the envelope has been represented with 10 parcors every 32 ms).

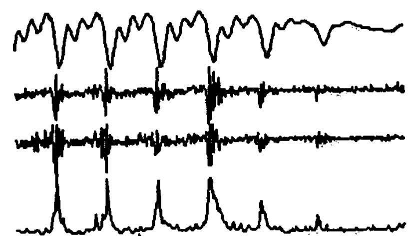

Fig 5. Waveforms obtained in the proposed coder using a codebook of random phases uniformly distributed between -π and π .a) Sinthesized signal. b) Original residual. c) Sinthesized residual. d) Parametric envelope obtained with 20 parcors every 16 ms.

Fig 5 shows the waveforms obtained with the proposed coder and sequences formed with random instantaneous phases uniformly distributed between-πandπ. No meaninghful different results are observed between gaussian generation with flat spectrum and constant envelope constraints and phases uniformly distributed. In this Figure envelope is represented with 20 parcors every 16 ms, to compare with the envelope obtained in Fig.2.

Fig 6 shows SNRseg obtained with this system and the codebook formed with random instantaneous phase. The codebook has a lenght of 1024 sequences and each sequence has 40 points. Envelope is represented with 10 parcors every 32 ms, the same as in the previous example. Continous line shows SNRseg for VQMP system and dashed line shows SNRseg for this system. Segmented signal to noise ratio is about 1dB better with stochastic instantaneous phase than using a codebook of pulses.

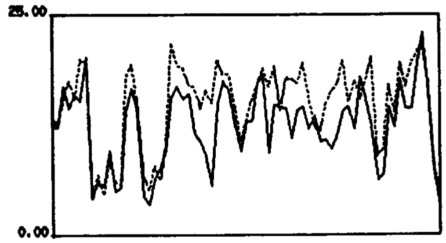

Fig 6. Segmented signal to noise ratio obtained with the proposed coder and a codebook of random phases uniformly distributed between $-\pi$ and π. Continous line: system VQMP (weighting function updated every 5 ms). Dashed line: proposed method. (the envelope has been represented with 10 parcors every 32 ms).

CONCLUSIONS

In this paper has been shown how envelope and instantaneous phase concepts can be used in a coding system at low bit rates. A parametric version of the envelope has been obtained by linear prediction methods. Sthocastic instantaneous phase has been introduced and very promisisng results have been obtained.

REFERENCES

/1/ B.S. Atal, J.R. Remde. "A New Model of LPC Excitation for Producing Natural Sounding Speech at Low Bit Rates". Proc. Int. Conf. on Acoustic, Speech and Signal Processing. Paris. France. 1982.

/2/ B.S. Atal. "Hight Quality Speech at Low Bit Rates: Multipulse and Stochastically Excited Linear Predictive Coders". Proc. Int. Conf. on Acoustic, Speech and Signal Processing. Tokyo. Japan. 1986.

/3/ R. García Gómez, F. J. Casajús Quirós, L. Hernández Gómez. "Vector Quantized Multipulse LPC". Proc. Int. Conf. on Acoustic, Speech and Signal Processing. Dallas. USA. 1987.