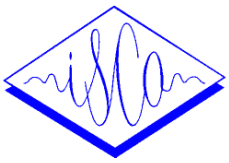

3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

A STOCHASTIC SPEECH CODER WITH MULTI-BAND LONG-TERM PREDICTION

C. García-Mateo*, J.L. Alba-Castro*, L.A. Hernández-Gómez**

* E.T.S.I. Telecomunicación. Dpto. Tecnologías Comunicaciones. VIGO (Spain). carmen@dtc.uvigo.es
** E.T.S.I. Telecomunicación. Dpto. SSR-UPM. Madrid (Spain). luis@gaps.ssr.upm.es

ABSTRACT

In this contribution we present a CELP coder using a multi-band strategy to obtain its adaptive contribution. This procedure shows the advantage of a better representation of the quasi-periodic nature of the short-time spectrum. The frequency splitting is made using two-band QMF filters in order to avoid an excessive computational complexity. The proposed coder can be employed at a bit rate between 3 and 7 Kbps depending on the quantization techniques.

Keywords: *CELP coding, subband analysis, QMF banks.*

1. INTRODUCTION

Many speech coders working at bit rates below 13 Kbps use a long-term predictor to remove the periodic structure in the short-time spectrum of the speech signal. Coders like CELP (Code-Excited Linear Prediction coder) or Multipulse (MP) are examples of such structures.

Several configurations have been employed in the past to implement this predictor. Most of them belong to two categories, both working in a close-loop form or an open-loop one:

- multi-tap pitch predictors
- fractionally spaced pitch predictors.

One example of the first class is the single-tap predictor which shows a nice compromise between computational complexity and performance, thus it is presently the preferred choice. Nevertheless, in order to achieve better quality more sophisticated algorithms have been developed.

Both of above mentioned categories try to deal with the quasi-periodic nature of the speech signal, modeling

a) the pitch time-variant behavior. Pitch period changes slightly from frame to frame and can be a non-integer value.

b) the pitch frequency variation. Most speech signals show less periodicity at high frequencies than a low frequencies.

The first topic has been well-solved by both techniques: multi-tap predictors like fractionally spaced ones are able to work with non-integer delays [1] [2]. The counter balanced part is the required extra computational load and the extra amount of bits. Increasing the bit rate is not a problem, if at the same time better quality is obtained. Nevertheless, both techniques do not be able to take into account the quasi-harmonic structure of the short-time spectrum. A "roughness" effect in the reconstructed speech [3] appears when the quantized parameters of the long-term predictor try to represent the whole periodic spectral information.

In the short-time spectrum of some voiced speech frames, a mixed unvoiced-voiced nature of the signal can be noticed. In order to represent this fact in an efficient way, a unique long-term factor is not enough. A more suitable approach would be to explore the spectrum at different frequency bands like sinusoidal coders do [4].

In [5] we have presented a multi-band structure for the long-term selection in a frequency domain stochastic coder. The main drawback was its large computational complexity, but the results achieved then, impelled us to move most of the computation to the time-domain. Thus, we simplify the algorithm, but maintaining the frequency-based model.

In this contribution, we will present a new structure for a CELP coder suitable for bit rates around 3-7 Kbps, feasible to be implemented in a single chip. The stochastic section stays as in the original coder, but the long term-prediction is built by frequency analysis, obtaining a set of a delay and a gain factor for each frequency band. The rest of the paper is organized as follows: in section 2, we briefly describe the structure of the proposed coder, showing the differences with the original CELP coder. In section 3, we will show the splitting procedure. In section 4, we will present the results we have achieved, and finally, some conclusions will be given at the end of the paper.

2. CODER STRUCTURE

We have applied a multi-band strategy to the adaptive stage of the CELP coder described in [6] by splitting the whole spectrum into several frequency bands and then looking for the best long-term predictor for each subband. Although the algorithm basis lies in the frequency domain, the procedures to obtain the long-term parameters work in the time domain using multi-rate digital signal processing techniques. This means an important computational saving compared with our first approach of [5].

The synthesis stage of our coder, we call it sub-band CELP coder, is shown in Figure 1. It can be noticed that by using subband analysis-synthesis a set of pitch lag and gain factor is attached to every frequency band. This procedure increases the number of parameters to send to the receiver, but using efficient quantization techniques (there is a strong correlation among the pitch lags and gain factors) the bit rate of the full-quantized coders can be between 3 Kbps and 7 Kbps without important perceptually distortion.

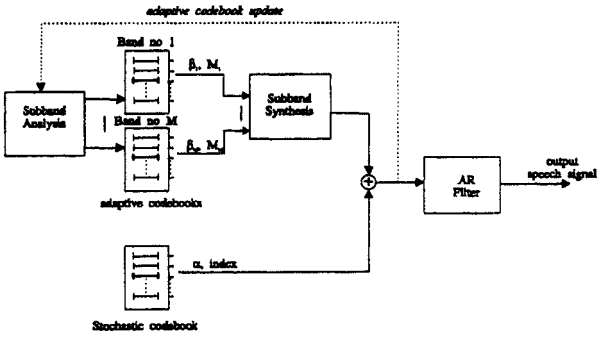

Figure 1: Synthesis stage of the sub-band CELP coder.

The main points to describe are: first, how the frequency band signals are obtained; second, the procedures to obtain the long term parameters for every band.

We have employed QMF filters with and without decimators to split the spectrum. This method imposes some restrictions to the overall performance of the sub-band CELP coder. In the next section, we will address this problem.

Once the time-domain signal for every band is obtained, we have used the same procedure as in the original CELP coder to determine the pitch lag and the gain factor. The block diagram of this task is described in Figure 2.

The complexity of the long-term search procedure for each band is reduced by a factor equal to the number of bands, if decimation is used to down-sample the sequences at the output of the subband analysis. Thus, the computational load slightly increases (only due to the band splitting procedures) compared with the original CELP coder.

If the subband signals are not maximally decimated, the computational complexity of the adaptive contribution increases by a factor approximately equal to the number of bands.

3. SUBBAND ANALYSIS STRATEGY

There are many techniques to split the spectrum into a set of non-overlapped frequency bands. QMF banks is a time-domain approach widely used in speech coding due to its nice compromise between performance and complexity.

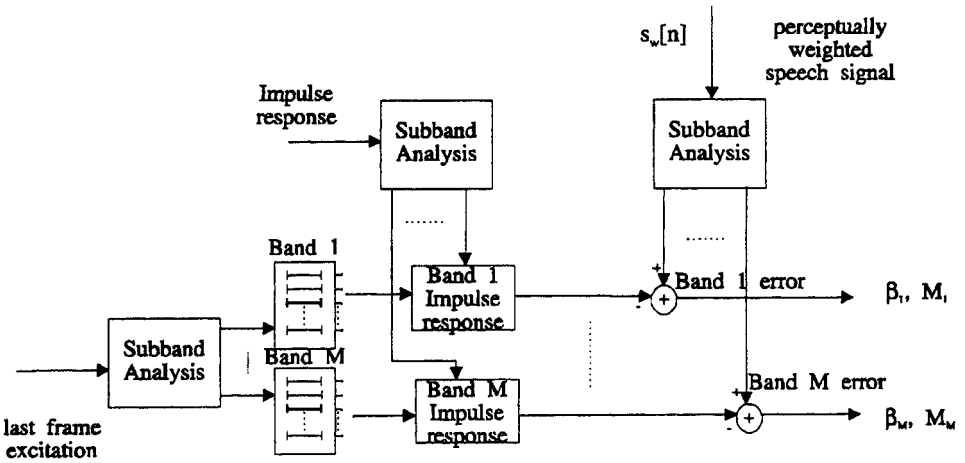

Figure 2: Adaptive contribution search procedure in the sub-band CELP coder.

We have also chosen it to obtain two equally spaced bands but using new sharp cut-off basic filters. A multiple band splitting can be obtained by a tree structure. This imposes more coding delay than a parallel approach, but it is difficult to build a polyphasic M-band structure with the required performance and thus, at this early stage of the work, we have only used two-band splitting, halving the spectrum in cascade so many times as necessary.

The subband signals are often maximally decimated in order to adjust the digital spectrum. We consider also the possibility of non decimation achieving then a better resolution for the pitch predictor. This is equivalent to implement a fractionally-spaced pitch predictor including also frequency discrimination.

Some considerations must be taken into account for the design of the basic low-pass filter for this application:

1) The type of the filter. The filter must have an odd number of taps and, in case of maximally decimated outputs, the order must also be a multiple of 4. The reason can be noticed in the Figure 1 at the point where the contributions of the long term predictor and the stochastic codebook are added to build the reconstruct speech signal (the same occurs at the analysis stage when the long term contribution must be subtracted from the original signal to obtain the residual error). For these operations both signals must be in phase. If the delay between both signals is an integer number of samples no extra processing is required, in other case an interpolator is required which increases the computational complexity of the system.

2) The filter length must be fixed as a compromise between frame length and spectral resolution, avoiding as much as possible the aliasing effect between adjacent bands.

Taking into account both considerations, we have designed two QMF banks following [7] for 9 and 17 taps respectively. The last one splits better the spectrum but the resolution loss in the estimation of the long-term predictors is bigger than for the first bank. In the Figure 3 the frequency response of both QMF banks is showed.

4. CODER PERFORMANCE

The coder we have tested uses a subframe length of 7.5 ms (60 samples at 8 KHz). The implementation of the basic structure is detailed in [6]. The parameters of the adaptive contribution are:

- two band splitting using filters of 9 and 17 taps.
- pitch lags between 2.5 ms and 18 ms.
- with and without decimation.

In Figure 4, it is shown the dispersion between the values of the gain factors for the lower band and the upper band respectively for a sentence with many voiced sounds and pronounced by a male speaker. The values are concentrated along the diagonal. The dispersion area reflects the flexibility of our coder distributing the energy with the frequency.

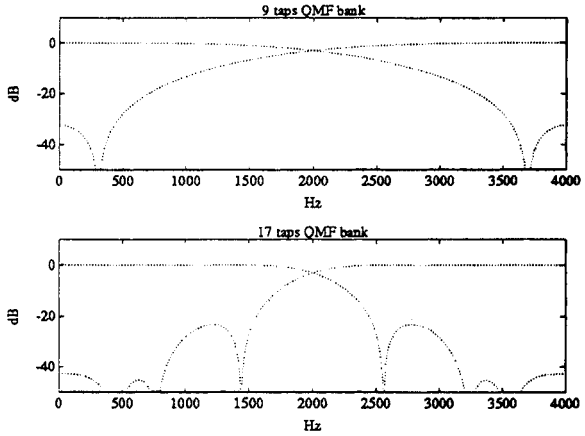

Figure 3: QMF banks for 9 taps filter (above), and 17 taps (below).

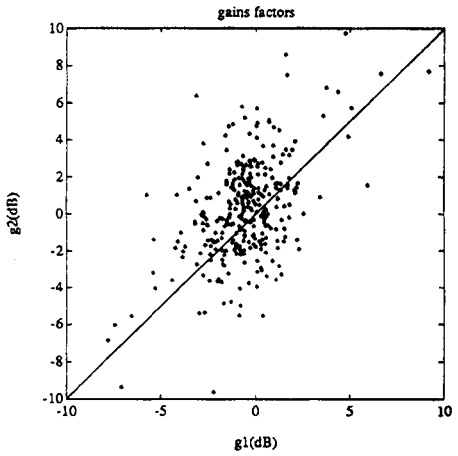

Figure 4: Relationship between the gain factor of the lower band (g1) and the gain factor of the upper band (g2).

In Figure 5, it is shown the differences between the pitch lags (in number of samples) of the different bands. The values are concentrated around the mean pitch value. The little dispersion, very important from a perceptual point of view, allows to use efficient quantization schemes for the pitch lags without increasing too much the final bit rate. It can be also noticed the effect of pitch halving. The dispersion from the diagonal around the mean pitch value is mainly

caused by the unvoiced part of the sentence and has minor influence in the final quality.

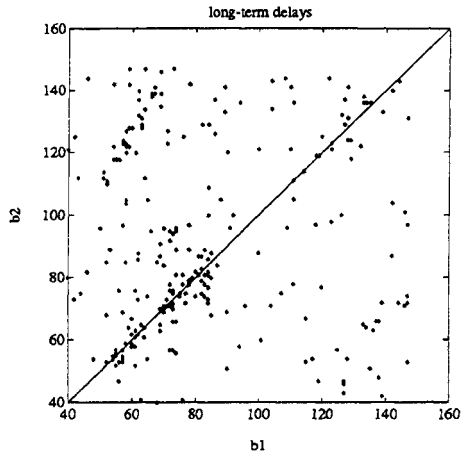

Figure 5: Relationship between the pitch lag of the lower band (b1) and the pitch lag of the upper band (b2). The axis units are in number of samples at fs=8 KHz

The coder shows better segmental SNR which is corroborated by a better speech quality specially for female speakers. The improvement is higher for the non-decimated coder than for the maximally decimated one.

We have also tested a three-band CELP coder using two stages of QMF filtering. The bands were: 0-π/4, π/4-π/2, and π/2-π. The results were poored than expected, mainly due to the acumulating effect of aliasing among frequency bands and loss of temporal resolution caused by the filtering delays.

5. CONCLUSIONS AND FURTHER WORK

We have applied the subband technique to a CELP coder to improve the long-term representation specially for quasi-periodic sounds. In spite of being at an early stage of the work (we have only tested a two-band splitting), the preliminary results show the nice possibilities of this approach, providing subjectively better speech quality than the original CELP coder.

The complexity of the proposed coder with decimated outputs has been slightly increased due to the filtering tasks but this is not a restriction for the real time implementation using a single chip.

The performance would improve by a multi-band approach. In that sense we are designing a parallel structure following [8] but with lower order filters.

6. REFERENCES

[1] Lupini P., Hassanein H., Cuperman V. "A 2.4 Kb/s Celp Speech Codec with Class-Dependent Structure". Proc. IEEE ICASSP93. Minneapolis (USA) 1993. Vol II. pp. 143-146.

[2] Marques J.S., Trancoso I.M., Tribolet J.M., Almeida L.B. "Improved Pitch Prediction with Fractional Delays in CELP Coding". Proc. IEEE ICASSP90. Albuquerque (USA) 1990. pp. 665-668.

[3] Granzow W., Atal B.S, Paliwal K.K., and Schroeter J. "Speech Coding at 4Kb/s and Lower Using Single-Pulse and Stochastic Models of LPC Excitation". Proc. IEEE ICASSP. Toronto (Canada) 1991.pp. 217-220.

[4] García Mateo C., Rodríguez Banga E., Alba J.L., Hernández Gómez L., "Analysis, Synthesis and Quantization Procedures for a 2.5 Kbps Voice Coder Obtained by Combining LP and Harmonic Coding". Signal Processing VI: Theories and Applications. Ed. Elsevier. 1992. pp. 471-474.

[5] García Mateo C., Casajús F.J., Hernández Gómez L. "Multi-Band Vector Excitation Coding of Speech at 4.8 Kbps". Proc. IEEE ICASSP90. Albuquerque (USA) 1990. pp. 13-16.

[6] Campbell J.P., Welch V.C., Tremain T.E. "CELP Documentation Version 3.2". 17 September 1990. U.S. Government. Department of Defense, R2, Fort Meade, MD 20755-6000.

[7] Galand C.R., Nussbaumer H.J. "New Quadrature Mirror Filters Structures". IEEE Transactions on ASSP. Vol. 32, June 1984. pp. 522-531.

[8] Nguyen T.Q., Vaidyanathan P.P. "Structures for M-Channel Perfect-Reconstruction FIR QMF Banks which Yield Linear-Phase Analysis Filters". IEEE Transactions on ASSP. Vol. 38, No. 3, March 1990. pp. 433-446.