Analysis-by-Synthesis Algorithms for Low Bit Rate Coding

G. Riccardi, G.A Mian
Dipartimento di Elettronica ed Informatica, Universitá di Padova
Via Gradenigo 6/A, 35131 Padova, Italy.
E-mail: dsp3@paola.dei.unipd.it
tel. +00-39-49-8287637,
fax +00-39-49-8287699

Keywords : Low bit-rate coding, Analysis-by-Synthesis, Prototype, Fractional Delay Pitch.

ABSTRACT

In speech coding at low bit-rates, the computation of a robust excitation signal is a key point, which characterizes the perceptive quality provided by a vocoder. Recently a lot of work has been done in this direction, namely for good quality and coding efficiency in the design of vocoders [4-6]. This work proposes a set of algorithms which apply the rationale of the Analysis-by-Synthesis together with High Resolution technique to low-bit rate coding. In the following we shall call this technique High Resolution Analysis-by-Synthesis (HRAS). As a whole this method provides a precise Analysis parameter estimation and a detailed excitation signal used in signal reconstruction.

1. THE HRAS TECHNIQUE

The high resolution analysis is based on the "fractional" pitch technique, introduced in [1,2]. The synthesis in HRAS, is provided by two types of excitation signals: a train of prototypes at pitch distance, $\pi(n)$ and a stochastic codebook excitation, $\sigma(n)$. The input $\sigma(n)$ is used as the second orthogonal component of the total excitation. The proposed work deals with algorithms for the computation of the component $\pi(n)$, which is related to the high resolution task in the signal reconstruction process. To this purpose an enhanced excitation $\pi(n)$ is provided by a train of prototypes which are adapted at each pitch period on a frame-by-frame basis:

$$\pi(n) = \sum_{i=0}^{\lfloor N/M \rfloor - 1} \sum_{j=-L}^{L} \gamma_{i,j}(n)\delta(n-n_0-iM-j) \quad (1)$$

where N is the length of the synthesis frame and M the pitch lag estimated in the analysis step. The synthesis parameters L, n_0, $\gamma_{i,j}(n)$ are, respectively, the temporal extension of each prototype waveform, the corresponding phase term and the long term excitation. The latter parameters should be computed through the minimization of the weighted reconstruction error energy over the synthesis frame:

$$\min_{n_0,i,j} \sum_{n} (\tilde{y}_w(n) - \sum_{i=0}^{\lfloor N/M \rfloor - 1} \sum_{j=-L}^{L} \gamma_{i,j}(n) h_w(n-n_0-iM-j))^2 \quad (2)$$

where $h_w(n)$ is the weighted pulse response of the all-pole model, $\tilde{y}_w(n) = y_w(n) - y_{w0}(n)$ is the weighted difference between the original signal,$y(n)$, and the term due to the free response $y_0(n)$ [3]. A weighting filter was chosen according to perceptive considerations given in [7]. No closed solution exists for the corresponding solution of equation (2), which results almost irresolvable for real time applications.Hence different sub-optimal procedures can be tried according to hypothesis applied to the signal reconstruction process. Three approaches will be presented.

1. The first method is based on the assumption that the signal in the analysis frame (usually in the range 20-30 ms) is not stationary. In this case each prototype is computed taking into account the contribution of the past prototypes.This way the computation of each prototype results in a sequence of minimization problems (and into the solution of the corresponding linear systems) associated to each pitch period.
2. As for the second approach, the stationary hypothesis within the analysis frame is used for efficient modeling of the prototype train. Thus the excitation $\pi(n)$ is obtained by means of a gain vector modulating a prototype associated to the processed frame. As

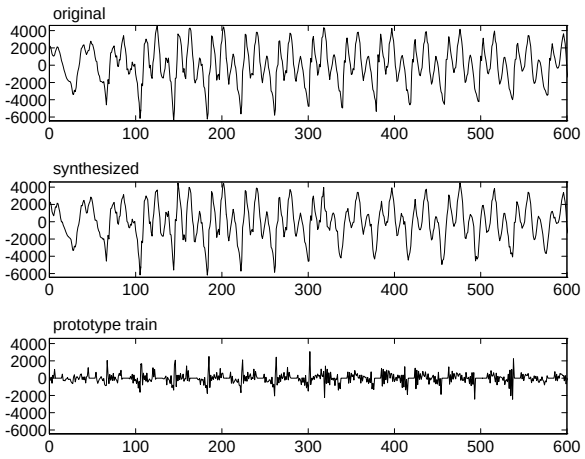

Figure 1: Plot of original , synthesized and prototype signal computed with algorithm 3 (L=15).

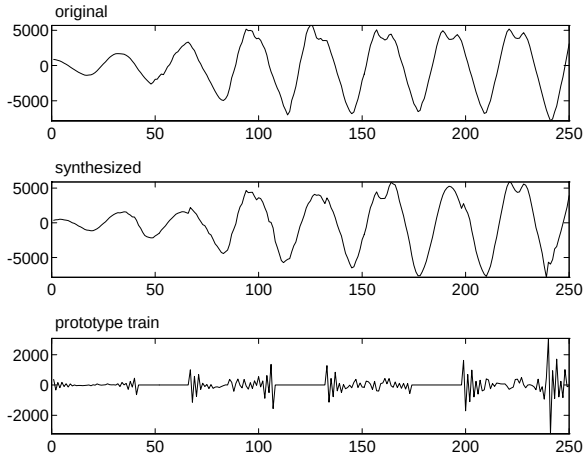

Figure 2: Plot of original , synthesized and prototype signal computed with algorithm 3 (L=15).

opposed to the first method the error energy minimization is carried out all along the analysis frame and the prototype is derived by the solution of only two linear systems.

3. In the third procedure no assumptions are made on the nature of the excitation signal $\pi(n)$. This approach turns out to be the best solution for the minimization problem stated in equation (2), compared to the other two methods, but it has a greater computational cost. The coefficients $\gamma_{i,j}$ are computed by means of a closed form, once the terms n_0 and L are derived with a suboptimal procedure. The adaptive structure of the signal $\pi(n)$ makes it possible to recover not only the matching errors between the estimated analysis parameters and the reference signal to be synthesized,but also the slowly time-varying parameters of voice signal.Fig. 1 shows the reconstructed signal by using signal $\pi(n)$ in case the voice signal is passing from a stationary interval into another. In fig. 2 an attack to a stationary piece of signal is processed and reconstructed with signal $\pi(n)$, as well.It is evident , in both cases, the good adaptation of the excitation signal to spectral changes.

The three methods proposed for modelling the excitation signal in case of voiced speech, allow to face to the quasi-stationary nature of speech signal at low bitrates. The way they have to be used depends upon the constraints of the coding scheme:variable or fixed bit rate coding, algorithm computational cost, coding effectiveness and objective performance. The first method has the lowest computational cost and it better fits in a fixed bitrate coding.The second procedure has the highest coding efficiency,when used in a variable bitrate coding scheme.The last algorithm presented gives the best objective and subjective performance while keeping low the bitrate associated.

The main contribution of this work stands in the introduction of the the High Resolution Analysis and of the Analysis-by-Synthesis for low bitrate coding. As a result, the HRAS technique provides both a beneficial introduction of a fine description of the spectral parameters and a high performance signal reconstruction step.

REFERENCES

[1] P. Kroon, B. Atal *"Pitch Predictors with High Temporal Resolution"* Proc. ICASSP, pp.661-664, Albuquerque '90.

[2] J.S. Marques et al. *"Improved Pitch Prediction with Fractional Delays in CELP coding"* Proc. ICASSP, pp.665-668, Albuquerque '90.

[3] M. Fratti, G.A. Mian, G. Riccardi, *"An Approach to Parameter Reoptimization in Multipulse Based Coders "* accepted in IEEE Trans. on Speech Proc.

[4] W. B. Kleijn, W. Granzow *"Method for Waveform Interpolation in Speech Coding"* Digital Signal Processing, vol.1, pp. 215-230, 1991.

[5] S. Ono, K. Ozawa *"2.4 kps Pitch Prediction Multipulse Speech Coding"* Proc. ICASSP, pp. 175-178, 1988, NY.

[6] W. Granzow, B. Atal *"High Quality Digital Speech at 4 kbps"* Proc. GLOBECOM pp. 941-945, 1990.

[7] B. Atal, M. Schroeder *"Predictive Coding of Speech Signals and Subjective Error Criteria"* IEEE Trans. Acoust. Speech Sig. Proc.,vol. ASSP-27,no.3, pp.247-254, June 1979.