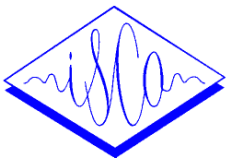

4th European Conference on
Speech Communication and Technology
EUROSPEECH '95
Madrid, Spain, September 18-21, 1995

A DIFFERENTIAL ENCODING METHOD FOR THE LTP DELAY IN CELP CODERS

Andrei Popescu and Nicolas Moreau

ENST, 46 rue Barrault,
75013 PARIS FRANCE

Claude Lamblin

CNET/Centre Lannion A - 2 av. P. Marzin
22307 LANNION FRANCE

ABSTRACT

We propose a differential encoding method for the long term predictor (LTP) delay in CELP coders. This method permits saving bits while preserving the coded speech quality. Fast delay changes between successive LTP updates are allowed and no buffering of long signal frames is necessary, which makes our method suited for use in low- and medium-delay coders.

1. INTRODUCTION

The LTP [5] is an important component of CELP coders. It models the periodicity of the speech signal.

The LTP parameters are periodically updated; they are usually maintained constant between two successive updates, for a duration that we call here an "LTP block", or simply "a block".

In principle, the LTP predicts the coder excitation for the current block as a scaled version of a past excitation block. Typically, the LTP parameters are:

- The LTP delay - the time-lag between the current LTP block and the past excitation block used to predict it.
- The LTP gain - the scaling factor multiplying the past excitation bloc to produce the predicted excitation block.

The LTP parameters are either estimated from the input signal ("open-loop" calculation), or computed by an analysis-by-synthesis procedure , which can be interpreted [2] as an "adaptive codebook search" ("closed-loop" calculation).

In this work, we suppose that the LTP parameters are calculated in closed-loop [6]. The encoder searches a set of LTP delays $\mathcal{S}$ for that LTP delay which results in the least perceptually weighted distortion of the synthesized signal, during the current LTP block. The set $\mathcal{S}$ contains M LTP delays, covering the typical range of the speech signal period (pitch). Transmitting the LTP delay requires $\log_2 M$ bits. This is what we call "full-search" LTP (FS-LTP).

This work was supported by France Telecom/CNET.

The LTP delay computed in closed-loop is not necessarily close to the actual pitch, although most of the time it is. A better match of the current signal block is sometimes obtained using an LTP delay which is a multiple of the signal period.

Because the pitch of the speech signal during voiced segments varies relatively slowly at the scale of an LTP block, there is a lot of redundancy in the series of LTP delays sent to the decoder. Several approaches have been proposed in order to exploit this redundancy [6] [3] [1].

In the frame-oriented CELP coder from [6], for several LTP blocks belonging to a same 20 ms frame, only small differences of the LTP delay around a value T_0 are allowed. The LTP delays of the frame are then differentially encoded with respect to T_0. A refinement of this method can be found in [3], where a differential LTP delay encoding is only used for the speech frames classified "voiced".

The authors of [1] use a Huffman encoding of the difference $\Delta = \delta_i - \delta_{i-1}$, between the LTP delay δ_{i-1} of the previous block, and the LTP delay δ_i of the current block. During voiced segments, $|\Delta|$ is most of the time close to zero, and these values of Δ are assigned shorter codewords. The saved bits are used to enhance the stochastic excitation, resulting in a dynamic bit allocation scheme.

In [2], the adaptive codebook search is constrained to delays corresponding to the normalized-correlation peaks produced by a previous open-loop analysis. This is viewed as a means to reduce the complexity of the adaptive codebook search, while preserving an efficient long-term prediction.

The method we propose performs differential encoding of the pitch delay using a fixed number of bits per transmitted LTP delay. In our method, in function of the LTP delay δ_{i-1}, we determine a subset $\mathcal{S}_i \subset \mathcal{S}$ of LTP delays which are likely to be chosen in the closed-loop search, for block i. The set $\mathcal{S}_i$ contains N delays ($N < M$).

In contrast with previous work, the originality of

our method consists in the fact that $\mathcal{S}_i$ is not a simple neighborhood of δ_{i-1}. We try to make the operation of our differentially encoded LTP (D-LTP) as close as possible to that of the full-search LTP (FS-LTP). Based on the conditional probability $p(\delta_i|\delta_{i-1})$, we choose $\mathcal{S}_i$ in function of δ_{i-1} in an attempt to maximize the probability of $\mathcal{S}_i$ to include the best delay for the encoding of block i. Here $\mathcal{S}_i$ also includes neighborhoods of the multiples and submultiples of δ_{i-1} and other delays, as shown below.

Furthermore, our differentially-encoded LTP leads to no additional coding delay compared to the full-search LTP, so it can be used in medium- and low-delay coders. The number of LTP delays evaluated in the closed-loop search is reduced M/N times.

2. LTP DELAY DIFFERENTIAL ENCODING

We introduce the function ϕ:

$$\begin{aligned} \mathcal{S} &\xrightarrow{\phi} \mathcal{S}^N \\ \delta_{i-1} &\longrightarrow \mathcal{S}_i = (\delta_i^1, \delta_i^2, ..., \delta_i^N) \end{aligned}$$

which specifies the ordered set of candidate LTP delays for the encoding of block i, if the coder used the delay δ_{i-1} for the encoding of the previous block.

The function ϕ is the same for the encoder and the decoder, so the remote decoder can determine $\mathcal{S}_i$ knowing δ_{i-1}. Then, only one index among N has to be transmitted to the decoder to determine $\delta_i \in \mathcal{S}_i$.

In order to limit the effect of errors, the LTP delay can be reset to a given value during silence periods, or the set $\mathcal{S}_i$ can be periodically forced to contain only fixed delays (independent of δ_{i-1}).

In order for the operation of the differentially encoded LTP (D-LTP) to be as close as possible to that of the full-search LTP (FS-LTP), the probability of $\mathcal{S}_i$ to include the best delay (from $\mathcal{S}$) for the encoding of vector i should be maximized. This observation can be used in a design of the function ϕ by training. A straightforward design of ϕ consists of estimating the probabilities $p(\delta_i|\delta_{i-1} = k)$ for all $k \in \mathcal{S}$ using a large enough training set. Then $\phi(k)$ is the ordered set of the N delays δ_i $(i = 1..N)$ maximizing $p(\delta_i|k)$. To avoid excessive memory requirements, in this work we use an analytical expression for ϕ, approximating the one resulting from training.

In what follows we denote V_α a set of delays from $\mathcal{S}$, whose values are located in the neighborhood of $\alpha\delta_{i-1}$. We denote N_α the size of V_α.

Experimenting with the full-search LTP, we established the following:

1. During voiced sounds, the adaptive codebook search chooses most of the time delays δ_i close to the pitch period. The situation where the previous delay and the current delay are both close to the pitch period is very likely and we take this into account by including in $\mathcal{S}_i$ a neighborhood V_1 of δ_{i-1}.
2. Also during voiced sounds, sometimes a multiple of the pitch period is chosen. For the situation where δ_{i-1} was close to the pitch period and δ_i is a multiple, we include the neighborhoods V_2 and V_3 of the multiples of δ_{i-1} into $\mathcal{S}_i$. For the case where δ_{i-1} was a multiple of the pitch period and δ_i is close to the pitch period, we include the neighborhoods $V_{1/2}$ and $V_{1/3}$ of the submultiples of δ_{i-1} into $\mathcal{S}_i$.
3. During unvoiced sounds there is no correlation between δ_i and δ_{i-1}. To deal with this and to allow a fast adaptation of the LTP delay at the beginning of voiced sounds, we also include in $\mathcal{S}_i$ a number $N - \sum N_\alpha$ of equally spaced delays from[1] $\mathcal{S} \setminus \{\cup V_\alpha\}$.

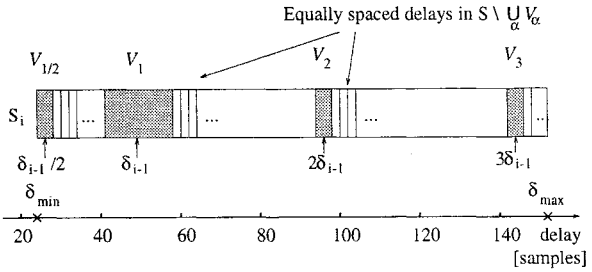

Figure 1: The part of the adaptive codebook searched for the encoding of vector i, if the previous LTP delay was δ_{i-1}.

In conclusion, the subset $\mathcal{S}_i$ contains delays from the neighborhood of the previous LTP delay δ_{i-1} (denoted V_1) as well as neighborhoods of submultiples and multiples of δ_{i-1} (denoted $V_{1/2}$, $V_{1/3}$, V_2 and V_3). To deal with unvoiced sounds and with rapid pitch changes, we also include into $\mathcal{S}_i$ uniformly spaced delays from $\mathcal{S} \setminus \{\cup V_\alpha\}$. Figure 1 shows how $\mathcal{S}_i$ is built, in function of δ_{i-1}.

Depending on δ_{i-1}, some of the delays from V_α may find themselves outside the allowed LTP delay range. The indices of these delays are also assigned to equally spaced delays from $S \setminus \{\cup V_\alpha\}$.

Table 1: The size of the neighborhoods V_α of the multiples and submultiples of δ_{i-1}.

Size of V_α (number of delays)				
$N_{1/3}$	$N_{1/2}$	N_1	N_2	N_3
7	7	21	1	1

In our simulations the neighborhoods V_α had a fixed number of delays. The values N_α were optimized in subjective quality tests; the resulting values are given in Table 1.

[1]Here $\mathcal{S} \setminus \{\cup V_\alpha\}$ is the part of $\mathcal{S}$ remaining after removing all the neighborhoods V_α.

Table 2: Comparison of the FS-LTP and the D-LTP for the Encoding of Nine Utterances.

		Female				Male				Child	Average
		no. 1	no. 2	no. 3	no. 4	no.1	no. 2	no. 3	no. 4	no. 1	
Segmental SNR [dB]	FS-LTP	10.84	8.20	10.66	9.78	7.55	9.02	8.48	9.03	13.33	9.65
	D-LTP	10.60	7.79	10.15	9.26	6.97	8.55	7.99	8.54	13.06	9.21
Pair comparison	FS-LTP	1.0	3.5	2.5	1.5	2.5	2.0	3.0	1.0	1.5	53.6%
	D-LTP	3.0	2.0	1.0	1.0	1.5	2.0	0.5	2.0	3.0	46.4%

We present these values in order to give an idea about the relative sizes of the neighborhoods V_α. The values of N_α should be optimized again for use in other coders, as the conditional pmf $p(\delta_i|\delta_{i-1})$ would probably differ.

3. PERFORMANCE EVALUATION

We used a noninteger-delay LTP with 24-sample LTP blocks in an 8 kbit/s CELP coder (see [4]). In Table 2, the FS-LTP ($M = 256$ delays) and the corresponding D-LTP ($N = 64$ delays) are compared in terms of segmental SNR.

We also give the results of an informal subjective pair comparison test involving five trained listeners. The listeners are presented pairs of encodings of test signals (utterances), A and B. In each pair one encoding uses the FS-LTP and the other our D-LTP. The duration of each utterance is approximately 6 seconds. The whole set of pairs (9 test utterances were used) is presented to the listeners twice. Within each pair the order of the encodings (FS-LTP first or D-LTP first) is random. After listening to a pair of encodings, each listener can give one of 5 possible answers: prefer rather A, prefer definitely A, A and B are equivalent, prefer rather B, prefer definitely B. If one coding scheme is "rather preferred" by a listener for the encoding of a given utterance, then it gets a score of 0.5, if it is "definitely preferred" it gets 1 point. The cumulated subjective scores for each utterance are given in Table 2.

Although there is a SEGSNR loss of 0.4 dB on the average, the subjective scores obtained by the full-search LTP and the differential LTP are roughly equivalent. We save two bits every 24 samples, i.e. 0.67 kb/s. Besides reducing the coding rate, this method also reduces the complexity of the adaptive codebook search.

4. CONCLUSION

We propose a differential encoding of the LTP delay for use in CELP coders. Our method does not require the processing of long signal frames in the encoder, so it can be used in medium/low-delay coders. It permits saving bits and computation while preserving the coded speech quality.

5. REFERENCES

[1] T. Eriksson and J. Sjőberg. Dynamic bit allocation in CELP excitation coding. In *Proceedings ICASSP*, pages II–171–II–174, 1993.

[2] I.A. Gerson and M.A. Jasiuk. Techniques for improving the performance of CELP type speech coders. In *Proceedings ICASSP*, pages 205–208, 1991.

[3] K. Ozawa, M. Serizawa, T. Miyano, and T. Nomura. M-LCELP speech coding at bit rates below 4 kbps. In *Proceedings EUROSPEECH*, pages 51–54, 1993.

[4] A. Popescu, N. Moreau, and C. Lamblin. CELP coding using trellis-coded vector quantization of the excitation. In *Proceedings ICASSP*, volume 1, pages 13–16, 1995.

[5] R.P Ramachandran and P. Kabal. Pitch prediction filters in speech coding. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(4):467–478, Avril 1989.

[6] M. Yong and A. Gersho. Efficient encoding of the long-term predictor in vector excitation coders. In *Advances in Speech Coding*, pages 329–337. Kluwer Academic Publishers, 1991.