14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

AN EFFICIENT METHOD FOR MULTI-TAP LONG TERM PREDICTOR (LTP) TRANSCODING: APPLICATION TO ITU-T G.723.1

Claude Lamblin and Mohamed Ghenania

FRANCE TELECOM R&D
2 avenue Pierre Marzin, 22307 Lannion Cedex, FRANCE
phone: + 33 2 96 0513 03, fax: + 33 2 96 05 13 16, email: claude.lamblin@francetelecom.com

ABSTRACT

Network interconnections cause interoperability problems between different speech coding formats. Today, tandem (decoding/re-encoding) is currently used in communication chains. To overcome tandem drawbacks (complexity, quality degradation, delay), intelligent solutions have been proposed to efficiently transcode CELP coders parameters. This paper focuses on CELP Long Term Predictor (LTP) parameters transcoding. First, a survey of LTP transcoding methods is given. Then a novel method is described for multitap LTP transcoding. The multitap gain vector codebook search is restricted to ordered subsets selected from LTP gain parameter of the first coder. This technique is applied to intelligent transcoding towards ITU-T G.723.1. It achieves same quality as tandem while strongly reducing complexity.

1. INTRODUCTION

In recent years the demand for high quality communications has considerably grown with the multiplication of various terminals. To provide a Universal Multimedia Access to users, various networks have been interconnected in which different speech coding standards are adopted. This causes an interoperability problem between these incompatible standards though, at medium bitrates (6-16kbit/s), most of the current speech coding standards are based on the well known code-excited linear prediction (CELP) coding model. The simplest solution to overcome this issue consists in decoding one standard compressed frame and re-encoding the generated signal by a second standard speech coder. This conventional method called tandem transcoding suffers from several problems such as computational complexity, algorithmic delay and speech quality degradation.

To overcome these problems, intelligent transcoding solutions have been proposed that exploit CELP standards similarities and are based on parameter conversion. In [1] a survey of intelligent CELP transcoding techniques has been given highlighting three transcoding levels: binary, parameter, signal. This paper focuses on the Long Term Predictor (LTP) Parameter conversion and is organized as follows. Section 2 reviews LTP parameters transcoding methods. Section 3 describes a novel method for multitap gain vector transcoding based on gain vector codebook ordering according to first coder LTP gains. Section 4 shows how this method can be applied to efficient transcoding from G.729 to G.723.1 and compares its performance with tandem. Section 5 concludes this paper.

2. LTP TRANSCODING METHODS OVERVIEW

Two LTP models are mainly used in CELP coders: the monotap predictor with non integer delay as in ITU-T G.729 [2] or in 3GPP AMR [3] and the multitap predictor with integer lag as in ITU-T G.723.1 [4]. Whatever the model is, LTP analysis is complex: an open loop search is usually performed on the whole range of integer delays followed by a closed-loop search around the open-loop pitch lag. LTP transcoding methods aim at reducing LTP analysis complexity by narrowing the parameter search range.

2.1 Transcoding between two monotap LTP models

This is the simplest (and the most studied) case. When the two coders subframe lengths are equal and the lag codebooks are identical, the lag transcoding can be done at the binary level and the second coder lag is set equal to the first coder lag [5-7]. If lag codebooks differ either in fractional resolution or range, lag transcoding is done at the parameter level with transformation such as truncation or rounding or doubling [6-8]. Moreover, if the subframe lengths are different, lags can be interpolated [9-11]. For instance, the lags of the first coder subframes covering a second coder subframe are interpolated. The interpolated lag may be used only if it is close to the lag of previous subframe; otherwise, a conventional search is done [12]. A straighter method consists in selecting one lag among the first coder lags. Various choices have been proposed: lag of the last first coder subframe [10] or of the subframe with greatest overlapping with the second coder subframe or lag maximising a LTP gain dependent measure [11].

2.2 Transcoding between monotap and multitap LTP models

Unlike the previous case, transcoding between monotap and multitap LTP models is generally done in the signal domain. Intelligent transcoding methods propose to focus the search of the second coder LTP parameters. Most of them aim at open loop search complexity reduction by choosing as second coder open loop lag an interpolation of the first coder lags or one of these lags [13-16]. Yet few methods have addressed the closed loop analysis.

In [13], a monotap fractional delay model is derived from a multitap integer delay model. The second coder fractional lag λ' is directly derived from the first coder integer lag λ and gain vector (β_j):

$$\lambda' = \lambda - \frac{\sum_{j=-2}^{2} j\beta_j^2}{\sum_{j=-2}^{2} \beta_j^2}$$

Transcoding from monotap model to multitap model is addressed in [14-17]. The gain vector closed-loop search is done in a codebook subset determined by the first coder monotap gain. The subsets are built with an analytical or statistical approach [14-16]. In [14-15], a global gain of each vector gain is computed then subsets of the multitap vector gain codebooks are determined from these global gains. In [16], the subsets are designed by training. First the dynamic range of the first coder monotap gain is split in subsections. Then a study on the monotap LTP first coder to multitap LTP second coder tandem determines the most probable gain vector indices for each subsection. For example, the LTP transcoding algorithm from 3GPP AMR to ITU-T G.723.1 splits the AMR pitch gain range in 8 subsections and for each subsection, the most probable 40 (respectively 85) gain vectors among 85 (respectively 170) codebook vectors are searched.

3. MULTITAP LTP TRANSCODING

Most smart LTP transcoding methods address lag. There are few transcoding methods from or to multitap gain vectors but transcoding between two multitap LTP models has not yet been investigated. This configuration occurs in ITU-T G.723.1 transrating from its high rate mode to its low rate mode (i.e. from 6.3 to 5.3 kbit/s). A fast transrating algorithm is proposed here.

3.1 ITU-T G.723.1 transrating algorithm

G.723.1 operates on 30 ms frames (240 samples) split in four subframes (7.5 ms/60 samples) and uses a 5^{th} order LTP predictor. Per frame two open-loop lags are found. Then for each subframe a pitch value and a gain vector are jointly searched in closed loop. The lag search is limited around the open-loop lag for even subframes and around the previous lag for odd subframes. Both modes use the same LTP lag codebook: 128 integer lags from 18 to 145. The high rate uses two LTP gain vector codebooks: a small codebook (85 vectors) if the pitch value is less than 58; a larger (170 vectors) otherwise. The low rate only uses the large codebook. The LTP analysis is complex: it represents 46%of the low rate mode complexity and 26% of the high rate.

Since the two modes have the same subframes and lag codebook, lag transcoding is done at the binary level as it is the case for monotap LTP model transcoding with identical subframe lengths and lag codebooks (see §2.2). The closed-loop lag of the low rate mode is simply set equal to the closed-loop pitch lag of the high rate mode. So the open-loop analysis and the closed-loop focused search are bypassed.

For gain vector transcoding, even if both rates use the large codebook, binary mapping cannot be used without degrading the quality; nor small codebook to large codebook parameter conversion. Thus gain vector transcoding must be performed at signal level to determine low rate LTP gain vector by closed-loop search. However, it can be noted that thanks to the closed-loop lag efficient transcoding the gain vector large codebook is only searched once instead of three or four times as in the conventional tandem search. Furthermore, a novel method has been designed to limit this search.

3.2 LTP gain vector transcoding

This new method also exploits the high rate gain *a priori* information to focus the low rate gain vector codebook search. However, instead of using fixed depth search as the solutions cited in §2.2, the search depth may vary to offer a flexible complexity-quality tradeoff. The method is based on codebook orders: at each subframe, an order is selected according the high rate gain vector, and the codebook search is limited to the first ordered elements.

3.2.1. LTP gain vector codebook ranking

The orders are determined off-line by finding preference rankings of the low rate gain codebook according to the high rate gain vectors. A ranking is based on a similarity measure of the low rate filters with a high rate filter (the first filters being the most similar). For each filter derived from the two high rate gain codebooks, a ranking of the low rate gain vectors is defined. Although the similarity measure may vary with the pitch, the ranking is hardly affected. Three measures have been tested: Euclidean distance, Residual Energy Minimization and CELP criterion. The first two measures are analytical whereas the last one is based on a statistical study (done on 100 clean and noisy speech sentences (10 French talkers)).

The first measure computes the Euclidean distance between two filters spectra. The second measure (REM) computes the energy of the residual obtained by filtering a first filter impulse response by a second filter.

The third measure is derived from the CELP criterion computed during the adaptive codebook search. It uses the Signal to Noise Ratio between the target vector and the synthesized vector. First, a statistical study is done in transcoding configuration on a speech corpus to store for each subframe, the gain vector selected by the high rate and the 170 SNRs for the 170 low rate gain vectors. Then, for each high rate gain vector, noted $B^{(h)}$, a preference ranking of the 170 low rate gain vectors is built by an iterative procedure that sequentially selects in the large codebook the subset of p gain vectors which maximizes the SNR:

<u>step 0:</u> selection of the best gain vector index k_0

$$k_0 = Arg\left\{ \max_{k\in[0\cdots169]} \left(\sum_{i=1}^{N^h} SNR_k(i) \right) \right\}$$

where $SNR_k(i)$ is the SNR associated to the low rate gain vector k for the i^{th} subframe having $B^{(h)}$ as high rate gain vector and N^h the number of such subframes.

step p+1: selection of gain vector index k_{p+1}

$$k_{p+1} = Arg\left\{ \max_{k\in[0\cdots169]-K^h(p)} \left(\sum_{i=1}^{N^h} \max_{l\in\{k_0,\ldots,k_p,k\}} (SNR_l(i)) \right) \right\}$$

where $K^h(p)$ denotes the subset of the first p sorted gain vectors.

This step is repeated until all the 170 vectors are ranked.

The rankings are gathered in two look-up tables corresponding to the two high rate gain vector codebooks. The huge memory required (43350=170*(170+85)) can be reduced by storing only the first M (M<<170) indices. The high rate gain vectors could also be grouped *a priori* by classification techniques. Similar rankings can also be merged: a same ranking is mapped with different high rate gain vectors and the ranking procedure iterated.

3.2.2. LTP gain vector codebook flexible search

The intelligent transcoder exploits the low rate codebook ordering according to the high rate gain vectors: only the first ordered gain vectors are tested. The SNR ratio *D(p)* evaluates the degradation induced by a search limited to *p* first gain vectors compared with an exhaustive search:

$$D(p) = \frac{\sum_i \max_{k\in K^h(p)} (SNR_k(i))}{\sum_i \max_{k\in[0\cdots169]} (SNR_k(i))}$$

Figure 1 shows for the three ranking methods (*1-D(p)*) in function of the search depth *p*. The statistical approach trounces the analytical methods as it has been confirmed by informal listening. Thanks to the fast decrease of (*1-D(p)*), the search can be strongly focussed. For instance, *p* can be reduced without any sensible degradation from 170 to 40. For the statistical ranking, Table 1 gives *D(p)* for some *p*.

Thanks to the codebook orders, *p* may vary according to available computational power. It also depends on the high rate index. To allow a better quality control, a *fixed-distortion* search can replace a *fixed-depth* search. Indeed the degradation curve varies from one high rate gain vector to another. Figure 2 shows for some gain vectors a fast increase whereas for others this increase is slower. The rankings allow a flexible choice of the complexity/quality trade-off.

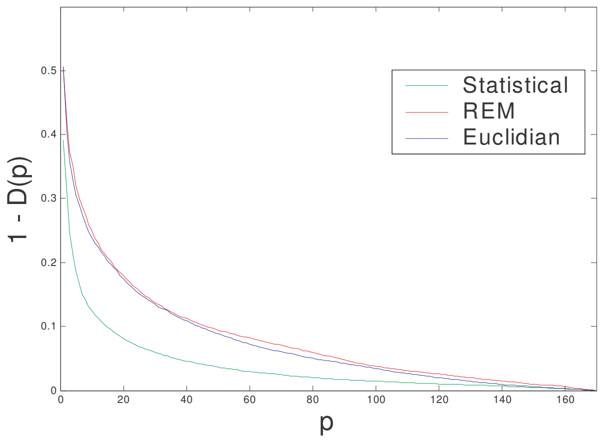

Figure 1 - SNR loss for the three similarity measures

h	small codebook			large codebook		
p	20	40	60	20	40	80
D(p)	0.93	0.96	0.98	0.92	0.95	0.98

Table 1- Fixed-depth search performance

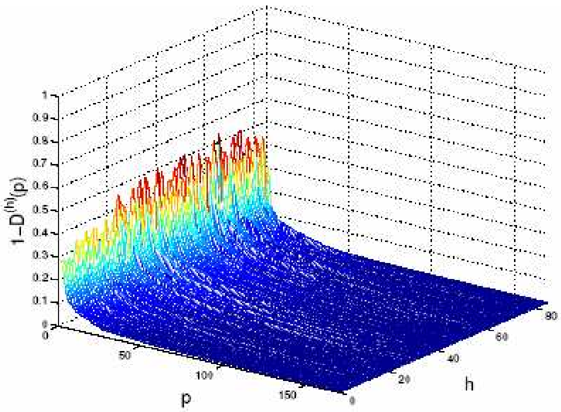

Figure 2 - Distortion decrease in function the depth search p and high rate gain vector index h (small codebook)

3.3 ITU-T G.723.1 LTP transrating performance

Informal pair comparison tests indicate that this method does not bring any audible quality degradation while strongly reducing the complexity. The complexity of G.723.1 low rate LTP analysis that stands for 46% of the total computational load is greatly reduced by the pitch lag analysis removal and the limited gain vector search. A conventional closed-loop search tests 510 (respectively 680) filters for even (resp. odd) subframe whereas with a fixed-depth search only 40 filters per subframe are tested.

4. ITU-T G.729 ↔ ITU-T G.723.1 TRANSCODING

The new method can also be applied to transcoding from G.729 LTP gain to G.723.1 LTP gain vector [17].

4.1. G.723.1 LTP gain vector codebook rankings

To overcome the dissimilarities (subframe length, …) of the two coders, a hybrid configuration shown in Figure 3 has been built to insert the G.729 LTP model in the G.723.1 encoder and so to collect the statistics needed to rank the G.723.1 LTP gain vectors according G.729 LTP gains.

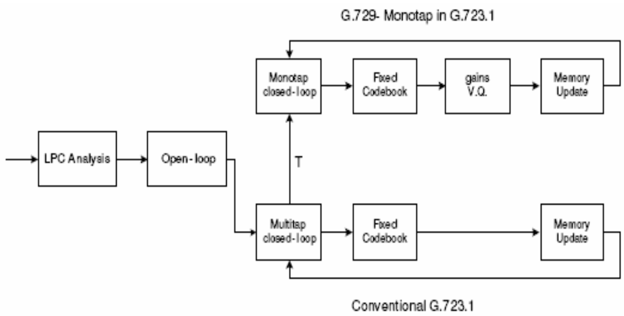

Figure 3 – Hybrid configuration: G.723.1 encoder with G.729 LTP model inside

The orders are obtained by the iterative method given in §3.2.1. As *D(p)* (average value on the 128 G.729 LTP gains) quickly increases to 1, a *fixed-distortion* search performs as well as exhaustive search with an average p of 49 (instead of 85) and 58 (instead of 170). A strongly focused *fixed-depth* search also works well (D=0.99): p=30 (resp. 40) instead of 85 (resp. 170).

4.2. LTP Parameters transcoding

Adaptation is needed to use the hybrid coder outcome in a transcoder. Two G.729 gains (and so two orders) can be associated with one G.723.1 subframe as it overlaps two G.729 subframes. Various options were tried [17]. The order linked to the gain of G.729 subframe fully covered by G.723.1 subframe was found to perform well.

For lag transcoding, to further reduce the complexity, the lag jitter was neglected in the approximation of a multitap LTP model by a monotap LTP model, closed-loop lags are found by direct mapping for even subframes and narrowed range [-1,1] for odd subframes.

4.3. Performance

The comparison of the processing times of tandem and LTP transcoding gives a rough estimation of the complexity. The computational load of G.723.1 high rate LTP analysis was found reduced by a factor of 68%.

Perceptual evaluation of speech quality (PESQ) on 54 French sentences (9 sentences of 3 male and 3 female talkers) has been used to compare the quality of the proposed transcoding algorithm and of the tandem. The scores are very close (less than 0.05 difference).

4.4. ITU-T G.723.1 → ITU-T G.729 Transcoding

This method has also been applied to G.729 pitch lag transcoding from G.723.1 LTP lag and gain vector [17]. The closed-loop search of the integer lag around the open-loop lag was focused using the G.723.1 selected multitap filter to restrict further the limited range in the neighbourhood of the open-loop lag to an ordered subset of integers lags. However, although the study gave interesting results, the complexity gain was overcome by the computation of the fractional part.

5. CONCLUSION

A smart transcoding for multitap LTP model coders has been presented It proposes a new method for gain vector transcoding that determines gain codebook rankings according to the first coder LTP gain(s) to focus the search. It offers flexible conversion from Multi (or mono)-tap to multitap LTP models. *Fixed-depth* or *fixed-distortion* searches allow control of either complexity or quality.

This method has been extended to other cases. The main idea is to define preference rankings in a second coder parameter codebook according to a first coder parameter. Then these first coder parameter dependent orders are used to focus the second coder parameter search.

6. REFERENCES

[1] M. Ghenania, C. Lamblin "Low-Cost Smart Transcoding Algorithm between ITU-T G.729 (8 kbit/s) and 3GPP NB-AMR (12.2 kbit/s)", in Proc. Eusipco, Vienna, September 2004.

[2] ITU-T Rec. G.729, "Coding of speech at 8 kbit/s using conjugate structure algebraic-Code-Excited Linear Prediction (CS-ACELP)", March 1996.

[3] 3GPP, "AMR speech codec : Transcoding functions", 3G TS 26.090, December 1999.

[4] ITU-T Rec. G.723.1, "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s", March 1996.

[5] HG Kang, HK Kim, R.V. Cox, "Improving transcoding capability of speech coders in clean and frame erasured channel environments" in Proc. IEEE Workshop on Speech Coding, 2000. Page(s): 78 –80.

[6] Y. Ota, M. Suzuki, Y. Tsuchinaga, M. Tanaka, S. Sasaki, "Speech coding translation for IP and 3G mobile integrated network", in Proc. ICC 2002, Volume 1, pp. 114 –118.

[7] S. Masano, O. Yasuji, T. Yoshiteru, "Transcoder for prevention of tandem coding of speech", Patent EP 1 202 251 A2, 2002-05-02,

[8] M. Atsushi, "Method of converting codes between speech coding and decoding systems, and device and program therefore", Patent Number: EP1267328, 18 December 2002.

[9] S. Masahiro, "Voice Code Sequence Converting Device and Method", Patent EP1363274, 19 November 2003.

[10] O. Yasuji, S. Masanao, T. Masakiyo, T. Yoshiteru, "Voice code conversion method and apparatus", Patent US2003142699, 31 July 2003.

[11] J. Marwan, W.J. Wei, G. Sameh, I. Michael, "Method for adaptative pitch-lag computation in audio transcoder", Patent WO 03079330, 12 March 2003.

[12] S. Lee; S. Seo; D. Jang; C.D. Yoo; "A novel Transcoding Algorithm for AMR and EVRC speech codecs via direct parameter Transformation", in Proc. ICASSP 2003, pp. 177-180, vol. II.

[13] J. Marwan, "A transcoding scheme between CELP-based speech codes", Patent WO03058407, 8 January 2003.

[14] SW Yoon, SK Jung, YC Park, DH Youn, "An Efficient Transcoding Algorithm For G.723.1 And G.729A Speech Coders," in Proc. Eurospeech 2001, pp.2499-2502.

[15] KT Kim, SK Jung, YC Park, YS Choi, DH Youn, "An efficient transcoding algorithm for G723.1 and EVRC speech coders", in Proc. IEEE VTS 2001, 54th Vehicular Technology Conference (VTC 2001), vol.3, pp.1561-1564, Oct.7-10, 2001.

[16] S.W. Yoon, J.K. Choi, H.G. Kang, D.H Youn, "Transcoding algorithm for G723.1 and AMR speech Coders: for interoperability between VoIP and Mobile Networks", in Proc. Eurospeech 2003, pp. 1101-1104.

[17] M. Ghenania, "Transcoding between standardized CELP coders" Ph. D. thesis, Rennes 1 University, France, June 2005.