LOW-COST SMART TRANSCODING ALGORITHM BETWEEN ITU-T G.729 (8 KBIT/S) AND 3GPP NB-AMR (12.2 KBIT/S)

Mohamed Ghenania, Claude Lamblin

FRANCE TELECOM R&D – DIH/IPS, 2 avenue Pierre Marzin, 22307 Lannion Cedex, FRANCE
email: [mohamed.ghenania, claude.lamblin]@francetelecom.com

ABSTRACT

Networks interconnection causes interoperability problems between different speech coding formats. Today, tandem transcoding solutions (decoding/re-encoding) are generally used in communication chains. However intelligent transcoding solutions which exploit similarities between formats have been recently proposed in order to overcome tandem drawbacks (computational complexity, algorithmic delay, speech degradation). This paper gives an overview of these intelligent transcoding methods for CELP coders. Then, a low-cost intelligent transcoding algorithm between ITU-T G.729 (at 8 kbit/s) and 3GPP NB-AMR (at 12.2 kbit/s) is proposed. It is composed of four parts corresponding to the four CELP parameters conversions: LSP coefficients, fractional pitch lags, fixed codevectors and gains. A novel and computationally efficient method is described here for the fixed codevectors. The ACELP search is strongly focused on privileged positions, what considerably reduces the number of tested combinations. Objective and subjective quality tests show that this smart transcoding algorithm achieves a quality very close to that of the tandem while strongly reducing complexity with a shorter algorithmic delay.

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. Recently, intelligent transcoding solutions have been proposed to overcome these problems: they exploit similarities between CELP standards and are based on parameter conversion.

This paper presents a survey on intelligent transcoding techniques between standard CELP coders (section 2). Then, section 3 proposes an efficient transcoding algorithm between ITU-T G.729 coder at 8 kbit/s [1] and 3GPP NB-AMR at 12.2 kbit/s [2] and describes a novel and fast method for ACELP codevectors transcoding. Section 4 compares the performances of this transcoding solution with those of the tandem according to three criteria: quality with objective and subjective tests, complexity and algorithmic delay. Finally, section 5 concludes this paper.

2. CELP TRANSCODING METHODS OVERVIEW

To translate a frame from a first coding format A into a second coding format B, the most current solution is to cascade a decoder of format A with a coder of format B. However, this conventional tandem transcoding has several drawbacks. First, this operation has a significant cost as a coding operation is computationally expensive. Moreover, the delay is increased by the sum of the delays brought by decoder A and coder B. This can reduce interactivity if several transcodings occur in a communication chain. Finally, a tandem transcoding introduces distortion on the signal since coding B is performed on decoded signal rather than on the original signal.

In order to overcome these problems, several alternatives to tandem transcoding have been proposed. These intelligent transcoding solutions reduce computational complexity and delay while achieving equivalent quality than that of the tandem. They exploit similarities between coding formats. All current CELP coders estimate, quantize and transmit the same type of parameters: LSP coefficients, adaptive codebook (ACB) lags and associated gains, fixed codevectors (FCB) indices and associated gains. The differences between coders can lie in the parameter estimation, the frequency at which they are computed or their quantization method. Consequently, three cases are possible to transcode a parameter.

If a parameter is identically computed and quantized by the two coding formats A and B, a mere copy of the binary field associated with this parameter from bitstream A to bitstream B is performed. Transcoding is done at the binary level. This is the case of the ACB lag in [3, 6, 7] and of the FCB in [6, 7].

If the parameter is quantized differently, it must be re-quantized with the quantization scheme of format B. Moreover, if A and B do not calculate this parameter at the same frequency (for instance if their subframe lengths are not equal), an interpolation (or a decimation) is made on this parameter before quantization. This kind of operation (re-quantization, interpolation, decimation) can be carried out by decoding the parameter without going up until signal level. The conversion is realized at parameter level. LSP coefficients are generally transcoded at this level [3-10]. Parameter level conversion has also been proposed for ACB gain [6, 7] or FCB gain [6, 7, 10].

Finally, for the same parameter, if formats A and B work in too different ways, it is necessary to decode the signal and to compute the parameter as that is done in the tandem method. This method is often used for FCB index [4, 5, 8, 9], but also for ACB gain [3, 9, 10] and FCB gain [3, 4, 5, 8, 9]. This transcoding can be seen as a partial tandem as only some parameters are computed as in conventional tandem transcoding. In general, it is performed at signal level. Nevertheless, in-between alternatives exist: parameter can be computed in excitation level (e.g. gains in [7]). In addition, some solutions have been developed to reduce complexity in signal level by taking into account some of first format information. For example for ACB lag estimation, the range of open-loop pitch search can be constrained around the lag value of A [4, 5, 8, 9, 10]. Similarly, the exploration of the G.723.1 multitap adaptive gain codebook can be restricted according to the first format monotap gain [4, 5, 8]. Some solutions have also been proposed to limit the exploration of the FCB codebook [3, 10].

3. PROPOSED TRANSCODING ALGORITHM FOR ITU-T G.729 AND 3GPP NB-AMR

These three conversion levels (binary, parameter or signal) have been used to develop an efficient transcoding algorithm between ITU-T G.729 coder at 8 kbit/s and 3GPP NB-AMR at 12.2 kbit/s. Based on the algebraic code-excited linear prediction (ACELP) technology, these coders are multirate standards: ITU-T G.729 speech coder has three modes (main mode at 8 kbit/s, and two other modes at 6.4 and 11.8 kbit/s) and 3GPP NB-AMR has eight modes from 12.2 down to 4.75 kbit/s. In this paper, only ITU-T G.729 main mode and the 3GPP NB-AMR highest bit rate mode 12.2 kbit/s, also called Enhanced Full Rate (EFR), are considered.

	G.729	EFR
subframes	2×5ms	4×5ms
LSP	1 VQ per 10ms	1 SMQ per 20ms
ACB lag	Monotap 1/3	Monotap 1/6
FCB pulses	4 pulses	10 pulses
ACB gain	VQ on 7bits	SQ (4bits)
FCB gain		SQ (5bits)

Table 1: Similarities and differences of G.729 and EFR

The two coders have similarities and differences summarized in table 1. G.729 operates on 10 ms frames (80 samples) whereas EFR operates on 20 ms frames, yet their subframe lengths are equal (5 ms/ 40 samples). For each parameter of the four CELP parameters (LSP, adaptive lags and fixed codevectors, associated gains), intelligent transcoding solutions have been proposed that exploit these similarities and overcome these differences. In particular, we introduce a novel and fast method to transcode the ACELP fixed-codevectors.

3.1 LSP conversion

Both coders perform a 10th order LP analysis every 10ms (i.e. once per G.729 frame and twice per EFR frame) and use LPC to LSP conversion before MA predictive quantization. Nevertheless, G.729 quantizes only one set of LSP coefficients with a switched 4th order predictive vector quantization on 18 bits. Note also that G.729 LP analysis introduces a 5 ms look-ahead. In EFR, LP analysis is performed twice per frame. The two sets of LP coefficients per frame are jointly quantized on 38 bits with split matrix quantization (SMQ) of 1st order MA prediction LSF residuals.

Here, a conversion at parameter level is used to transcode: LSP from format A are decoded and re-quantized using format B quantization scheme. Before quantization, the LSP mapping between G.729 and EFR is:

EFR→ G.729:

$$LSP_{2n}^{G729} = LSP_{n\_1}^{EFR} \text{ and } LSP_{2n+1}^{G729} = LSP_{n\_3}^{EFR}$$

G.729→EFR:

$$LSP_{n\_1}^{EFR} = LSP_{2n}^{G729} \text{ and } LSP_{n\_3}^{EFR} = LSP_{2n+1}^{G729}$$

where LSP_{n}^{G729} is the set of LSP for G.729 n-th frame and $LSP_{n\_k}^{EFR}$ is the set of LSP for the k-th subframe (k=0, 1, 2 or 3) of EFR n-th frame.

The computational complexity is strongly reduced because LP analysis and LPC into LSP conversion are omitted. Furthermore, the algorithmic delay is reduced by 5ms because the LP analysis look-ahead is removed.

Moreover, such transcoded LSP match the original signal better than the tandem ones. Figure 1 shows for a voiced subframe that LPC spectrum obtained by transcoding is closer to the original spectrum which is the G.729 LPC spectrum in this case (transcoding G.729 to EFR).

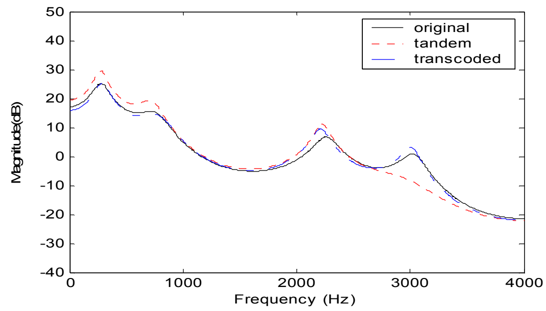

Figure 1: Comparison of LPC spectrum

SD (dB)	G.729 to EFR			EFR to G.729		
	Mean	Outliers (2-4dB)	Outliers (>4dB)	Mean	Outliers (2-4dB)	Outliers (>4dB)
Tandem	2.71	52.38%	5.40%	2.29	32.59%	1.01%
Transcoder	1.08	0.52%	0%	1.14	1.18%	0%

Table 2: Spectral distortion on LSP of tandem and transcoder

Table 2 shows that transcoded LSP are closer to "original" (i.e. quantized only once) LSP than those obtained by the tandem. Indeed, the mean spectral distortion (between tandem or transcoded LSP and original quantized LSP) calculated on 18 French sentences (6 for male, 6 for female, 6 for children) is smaller for transcoder case. The tandem quantizes and transmits the LSP calculated on a synthesis signal, different from the original signal because of the distortion brought by the first coding/decoding whereas the transcoded LSP are obtained by a mere re-quantization of original LSP already quantized/dequantized by codec A.

3.2 Pitch delay conversion

For both coders, the adaptive codebook is searched with a combined open-loop/closed-loop search. For even subframes, the pitch lag is absolute coded with a fractional resolution for lags below a lag bound and integer resolution only for greater lags. For odd subframes, the lag is delta coded relative to the lag of the previous subframe with a fractional resolution. Besides these common features, pitch models used by G.729 and EFR are slightly different: EFR fractional resolution is finer (1/6) than G.729 one (1/3), EFR absolute lags range is wider [17 3/6; 143] than G.729 one ([19 1/3; 143]), the lag bound between fractional and integer resolutions is 95 for EFR and 85 for G.729. These differences are summarized in table 3.

Pitch conversion can be made either at binary level (i.e. without decoding the pitch index chosen by coder A) or at parameter level. The conversion uses tables exploiting the inclusion of G.729 absolute lags codebook in the EFR absolute lags codebook and the rounding of finer fractional resolutions.

Precision	Even subframes			Odd subframes	
	Fractional		Integer	Fractional	
	frac	Range	Range	frac	Range
G.729	1/3	$19_{1/3}$ $84_{2/3}$	85 143	1/3	$-5_{2/3}$ $4_{2/3}$
EFR	1/6	$17_{3/6}$ $94_{3/6}$	95 143	1/6	$-5_{3/6}$ $4_{3/6}$

Table 3: Pitch models of G.729 and EFR

Here, the strong similarities between the two pitch models allow to greatly reduce the computational complexity of the pitch lag conversion as none of the different analysis (open-loop analysis, integer closed-loop analysis-by-synthesis and fractional exploration) need to be performed.

3.3 Fixed codevector conversion

In EFR and G.729, algebraic code-excited linear prediction (ACELP) technology is used for fixed codevector model. EFR uses a 35 bits ACELP codebook of 10 pulses with amplitudes +1 or -1 while G.729 uses an ACELP fixed-codebook of 4 non-zero pulses with amplitudes +1 or -1. In this paper, we present an original conversion method between a 4 pulses ACELP fixed codebook (G.729) and a 10 pulses ACELP one (EFR).

The conversion is made at the signal level, but the B fixed codebook is limited to some combinations of privileged pulse positions derived from the pulse positions chosen by coder A. These privileged positions are selected in the close neighborhood (right and left) of coder A pulses positions and are used to define a subset of coder B fixed codebook. Then, encoder B focused search is performed in a restricted sub-codebook instead of the full codebook as in conventional tandem. The neighborhood size is chosen according to quality/complexity trade-off.

Track	Pulse	Sign	Positions
T_0	i0 , i5	± 1	0, 5, 10, 15, 20, 25, 30, 35
T_1	i1 , i6	± 1	1, 6, 11, 16, 21, 26, 31, 36
T_2	i2 , i7	± 1	2, 7, 12, 17, 22, 27, 32, 37
T_3	i3 , i8	± 1	3, 8, 13, 18, 23, 28, 33, 38
T_4	i4 , i9	± 1	4, 9, 14, 19, 24, 29, 34, 39

Table 4: EFR 10 pulses ACELP codebook

Track	Pulse	Sign	Positions
T_0	i0	± 1	0, 5, 10, 15, 20, 25, 30, 35
T_1	i1	± 1	1, 6, 11, 16, 21, 26, 31, 36
T_2	i2	± 1	2, 7, 12, 17, 22, 27, 32, 37
T_3 T_4	i3	± 1	3, 8, 13, 18, 23, 28, 33, 38 4, 9, 14, 19, 24, 29, 34, 39

Table 5: Positions of the G.729 pulses (8 kbit/s)

By exploiting the common Interleaved Single-Pulse Permutation structure of ACELP codebooks used in G.729 and EFR, the sub-codebook positions can be very quickly extracted from the positions chosen by coder A. Both EFR and G.729 ACELP codebooks use the same 40-samples length subframe structure into five interleaved tracks of length 8. Table 4 shows that EFR has two pulses per track. G.729 (table 5) has one pulse per track in the first three tracks and a last one can be placed either on the fourth or fifth track.

Each pulse position p is represented by the couple (t, r), where t is the pulse track and r its rank in its track t:

$$p = 5r + t \text{ with } 0 \leq t \leq 4 \text{ and } 0 \leq r \leq 7,$$

In the proposed conversion method, from each pulse position of format A, right and left neighborhoods are extracted according to the following method:

Let d the neighborhood size, the neighbors $p_k^{(i)}$ of the i[th] pulse position $p^{(i)} = \left(t_j, r\right)$ are:

$$p_k^{(i)} = p^{(i)} + k = \left(t_k^{(i)}, r_k^{(i)}\right) \text{ with k in [-d, d],}$$

$$\text{where } t_k^{(i)} = t^{(i)}_{(j+k)\equiv 5}\text{, and } r_k^{(i)} = r + \left\lfloor \frac{j+k}{5} \right\rfloor$$

Note that $0 \leq r \leq 7$. Note also that $p_0^{(i)} = p^{(i)}$.

Then, the allowed positions for each track T_j ($0 \leq j \leq 4$) are restricted to be in the union of the N neighborhoods of the N coder A pulse positions: $S = \bigcup_{i=1}^{N} \left\{ \bigcup_{k=-d}^{d} p_k^{(i)} \right\}$

The track size reduction leads to a restricted subcodebook. For instance, if neighborhood size d is set to 1, in EFR to

G.729 transcoder, the maximum number of allowed positions in each track is 6 instead of 8. With neighborhood size d equal to 2, in G.729 to EFR way, 5 rank positions are allowed instead of 8. Table 6 shows for the two transcoding ways that in most cases, CELP Criteria (CC_{foc}) obtained by a restricted focused exploration is close to the standard one CC_{std} obtained with the conventional focused codebook exploration. Informal experts listening pair comparison test indicates that this proposed procedure does not introduce any audible quality degradation while strongly reducing the number of tested combinations.

	d	$CC_{foc} = CC_{std}$	$CC_{foc} > 0.8\ CC_{std}$
G.729 to EFR	2	12.4%	75.2%
EFR to G.729	1	60.1%	87.4%

Table 6: CELP criteria comparison

3.4 Gains conversion

In G.729 adaptive and fixed codebooks are jointly vector quantized on 7 bits whereas EFR adaptive and fixed codebooks gains are separately scalar quantized with 4 and 5 bits respectively. For these parameters, gains are recomputed on synthesized signals and then re-quantized like in conventional tandem.

4. PERFORMANCES EVALUATION

4.1 Computational Complexity

As a rough estimation of the proposed algorithm computational complexity, its processing time is compared with the tandem one for each parameter.

parameter	G.729 → EFR	EFR → G.729
LPC	50%	40%
ACB	99%	99%
FCB	38%	43%
Gains	0%	0%

Table 7: Complexity reduction in parameters computation

4.2 Speech quality

Perceptual evaluation of speech quality (PESQ) measure on 12 French sentences (6 for male and 6 for female) has been used to compare the quality of the proposed transcoding algorithm and of the conventional tandem. Table 8 shows that the proposed algorithm and qualified are very close.

	G.729 → EFR		EFR → G.729	
	Male	Female	Male	Female
Tandem	3.650	3.346	3.716	3.448
Transcoded	3.619	3.225	3.632	3.256

Table 8: Comparison of PESQ scores

Pair comparison tests have been performed involving 16 subjects, listening on handsets, on 16 samples (3 male, 3 female and 2 children voices). Tandem was found equivalent to the proposed algorithm in more than half of the cases. Yet tandem was preferred to our algorithm more often than the opposite, especially in EFR to G729 case. Degradations produced by the two methods were perceived as different: tandem introduces spectrum distortion (slightly masked by the handset) whereas transcoding introduces local hitches.

	Tandem	No preference	Transcoding
G.729 → EFR	28.5%	53.5%	18%
EFR → G.729	37.5%	50.8%	11.7%

Table 9: Pair comparison tests results

4.3 Delay

Here, only algorithmic and processing delays are considered. Processing delay is reduced because of the important reduction of computational complexity. Furthermore, in EFR to G.729 transcoder, LP analysis look-ahead is removed. The algorithmic delay is then reduced by 5 ms.

5. CONCLUSION

In this paper, a smart transcoding algorithm between ITU-T G.729 and 3GPP EFR speech coders was proposed. An original approach for fixed-codevectors transcoding which favors neighborhoods of the first coder pulse positions has been proposed. This method has also been successfully applied to other transcoding between CELP coders using constrained multipulse fixed-codebooks such as ITU-T G.723.1 and 3GPP NB-AMR. Performance evaluation shows that the proposed method achieves a quality very close to the conventional tandem one while reducing complexity and delay.

REFERENCES

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

[2] 3GPP, "AMR speech codec: Transcoding functions", December 1999, 3G TS 26.090.

[3] Shu-Min Tsai, Jar-Ferr Yang, "GSM to G.729 speech transcoder", in *Proc. ICECS 2001*, pp. 485-488.

[4] Sung Wan Yoon and al., "An efficient Transcoding Algorithm for G723.1 and G.729A speech coders", in *Proc. Eurospeech 2001*, pp. 2499-2502.

[5] Kyung Tae Kim and al., "An efficient transcoding algorithm for G723.1 and EVRC speech coders", in *Proc. IEEE VTS 2001*, pp.1561-1564.

[6] Ysuji Ota and al., "Speech Coding Translation for IP and 3G Mobile Integrated Network", in *Proc. ICC 2002*, pp.114-118.

[7] H.G. Kang and al., "Improving transcoding capability of speech coders in clean and frame erasured channel environments", *Proc. IEEE Workshop on Speech Coding 2000*, pp78-80.

[8] Sung-Wan Yoon and al., "Transcoding algorithm for G723.1 and AMR speech Coders: For interoperability between VoIP and Mobile Networks", *Proc. Eurospeech 2003,* pp. 1101-1104.

[9] Seongho Seo and al., "A novel Transcoding Algorithm for SMV and G723.1 speech coders via direct parameter Transformation", in *Proc. Eurospeech 2003*, pp.2861-2864.

[10] Seongho Seo and al., "A novel Transcoding Algorithm for AMR and EVRC speech codecs via direct parameter Transformation", in *Proc. ICASSP 2003*, pp. 177-180, vol. II.