AMR TO G.729A SPEECH TRANSCODING WITH FAST CODEBOOK SEARCH

Li-Fang Huang(黃麗芳) and Pao-Chi Chang(張寶基)

Communication Engineering Department, National Central University, Jhongli, Taiwan

E-Mail：pcchang@ce.ncu.edu.tw

摘要

語音轉碼（speech transcoding）是網路語音系統中不可缺少的機制，傳統上最佳的語音轉碼方法是使用完全解碼的方式，在過程上必需進行語音的壓縮及解壓縮處理，造成運算複雜度過高與時間延遲長的缺點。為此，本論文利用脈衝代換之快速碼簿搜尋法，提出一套部份解碼方式的語音轉碼方法，利用語音訊號的特性，以碼框為單位，分析代表各語音所需的語音參數，藉由參數的轉換以達到語音轉碼的效果。該組目標音訊參數亦符合原壓縮方法之壓縮格式。可運用在 AMR 與 G.729A 語音壓縮標準上，並可有效地降低運算複雜度，就每一音框所需的時脈數，約為完全解碼法的 7.2%，且可得到與完全解碼法接近之語音品質。

關鍵詞 :語音轉碼, AMR, G.729A, ACELP

ABSTRACT

Speech transcoding scheme is needed in the voice system over internet. Full decoding technique is an intuitive and traditional speech transcoding method, but it requires high computational complexity and long processing time. In this paper, we propose a partial decoding technique with fast codebook search, which utilizes the pulse replacement method, on ACELP coding architecture. There is no need to redo all the decoding and encoding processes. Partial decoding method can be directly applied to ACELP based speech coding, such as AMR[1] and G.729A[2] speech standards. The proposed method decodes the parameters from the input bit-stream, which includes line-spectral pair (LSP), pitch delay, fixed codevector and codebook gain. It achieves excellent voice quality as the full decoding method does while it only requires 7.2% computation loading calculated byclockticks per frame.

Keyword : speech transcoding, AMR, G.729A, ACELP

1. INTRODUCTION

Mobile telecommunication systems have been evolving towards 3rd generation (3G) rapidly. Voice over IP (VoIP) has become a promising tool for low-cost global telecommunication functionality. As shown in Figure 1. , in applications requiring interoperability between different networks such as wireless and IP network, transcoding is a good choice due to its reduced complexity, delay, and quality degradation [3][4][5].

Intuitively, the simplest solution of transcoding consists of decoding one standard compressed frame and re-encoding the generated signal by a second standard speech coder, as shown in Figure 2. This conventional full decoding method, called tandem transcoding, suffers from several problems such as high computational complexity, long algorithmic delay. Recent years, intelligent transcoding solutions have been proposed to overcome these problems: they exploit similarities between ACELP standards and are operated based on parameter conversion. Since both AMR and G.729A are based on analysis-by-synthesis scheme and their transmitted information is similar [6], transcoding can be applied reasonably well.

In this paper, we propose a partial decoding technique with fast codebook search, which utilizes the pulse replacement method, and describe its working procedure in Section 2. In Section 3, we use C simulation to demonstrate the quality measurement of our proposed approach. Finally, Section 4 concludes the paper.

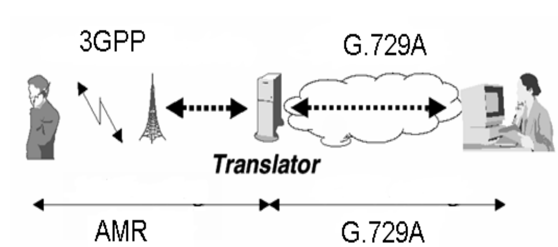

Figure 1. Integration of 3GPP and IP network

2. PROPOSED TRANSCODING SCHEME FROM AMR TO G.729A

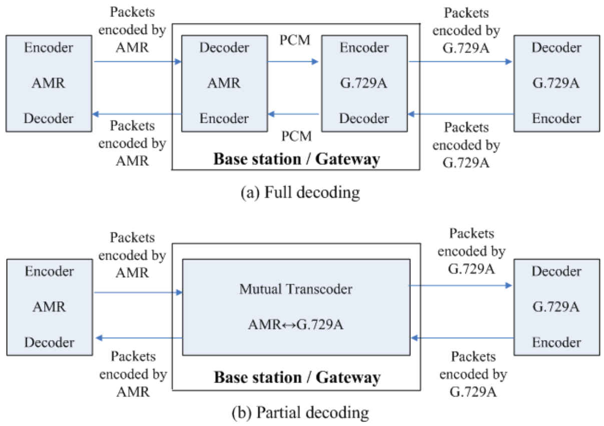

Figure 2. Two speech transcoding schemes (a) Full decoding (b) Partial decoding

To translate one frame from AMR to G.729A, the direct solution is to cascade the decoder of AMR and the encoder of G.729A. However, this conventional method has several problems, including of computation complexity, coding delay, and so on：

1. Computation complexity ：The conventional full decoding method needs to perform decode and encoder at least once. The computation load is tremendous in some compression procedure.
2. Coding delay：Some processing delay is generated by frame buffering and windowing look-ahead from LPC analysis. Therefore, this conventional transcodng method increases the processing delay.

To solve these problems, we make use of the similarities between both codecs. Both AMR and G.729A are based on ACELP. ACELP transmits four kinds of parameters that correspond with the speech characteristics：LPC coefficients, pitch delay, index of fixed codebook, and gains of the adaptive codebook, and fixed codebook. AMR and G.729A are different from each other in frame size and method of quantization, as shown in Table I and Table II .

Table I. Specification of AMR and G.729A coding standards

	AMR	G.729A
Algorithm	ACELP	CS-ACELP
Bit-rate	4.75, 5.15, 5.9, 6.7, 7.4, 7.95, 10.2, 12.2 kbits/s	8 kbits/s
Frame size	20ms	10ms
Subfrmae size	5ms	5ms

Table II. Difference of techniques between AMR and G.729A

	AMR	G.729A
LSP	SMQ	VQ
Adaptive codebook lag	12.2k mode： 1/6	1/3
	Other modes： 1/3	
Fixed codebook pulses	2~10 pulses	4 pulses
Gains	VQ	SQ

AMR operates on 20ms (160 samples) per frame whereas G.729A operates on 10ms (80 samples) per frame. However, their subframe sizes are equal to 5ms (40 samples). Based on the same subframe size, it is easy to map AMR to G.729A on parameter layer. In particular, we propose a speech transcoding with fast codebook search from AMR eight modes to G.729A in this work.

2.1. LSP conversion

Both AMR and G.729A perform a 10th order LPC analysis and use LPC to LSP conversion before MA predictive quantization. However, the quantization methods of the two codecs are different. Here, a conversion at parameter layer is used. LSP from AMR are decoded and re-quantized using G.729A quantization scheme. Before quantization, the formula we use to mapping the LSP parameter between both codecs can be divided into two types：

1. AMR 12.2 kbit/s mode, as shown in Figure 3

AMR performs the LP analysis once per 10ms (i.e. twice per frame), so we can get two sets of LSP parameters per frame.

We can map the LSP parameters by using the formula：

$$p_2^{(2n-1)} = \left(\hat{q}_4^{(n-1)} + \hat{q}_2^{(n)}\right)/2 \ ; \ p_2^{(2n)} = \left(\hat{q}_2^{(n)} + \hat{q}_4^{(n)}\right)/2 \qquad (1)$$

where $p_2^{(n)}$ denotes the 2nd subframe in frame n of G.729A and $\hat{q}_2^{(n)}$ denote the 2nd subframe in frame n of AMR

2. AMR other seven modes, as shown in Figure 4

AMR performs the LP analysis once per 20ms (i.e. once per frame), so we can get just one set of LSP parameters per frame. And G.729A performs the LP analysis per 10ms.

We can map the LSP parameters by using the formula：

$$p_2^{(2n)} = \hat{q}_4^{(n)} \ ; \ p_2^{(2n-1)} = \left(\hat{q}_4^{(n-1)} + \hat{q}_4^{(n)}\right)/2 \qquad (2)$$

where $p_2^{(2n)}$ denotes the 2nd subframe in frame $2n$ of G.729A and $\hat{q}_4^{(n)}$ denote the 4th subframe in frame n of AMR.

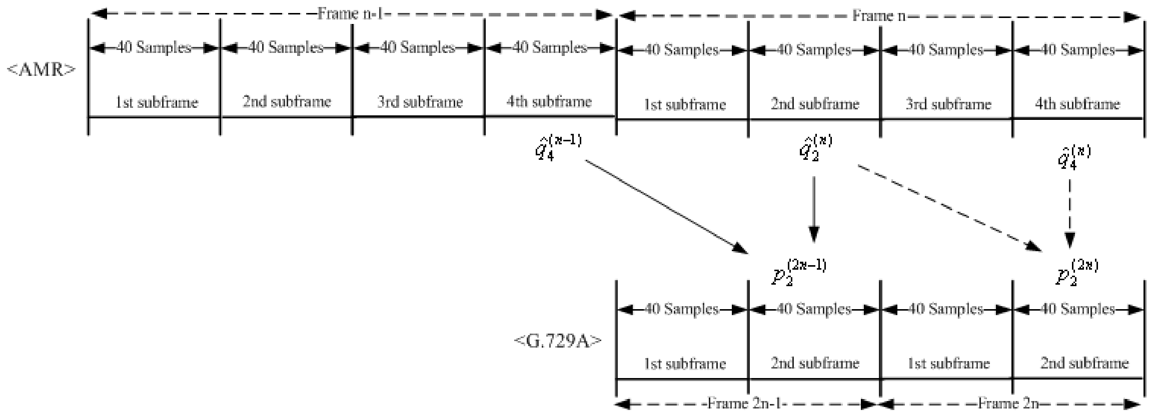

Figure 3. The LSP mapping method from AMR to G.729A (12.2 kbit/s)

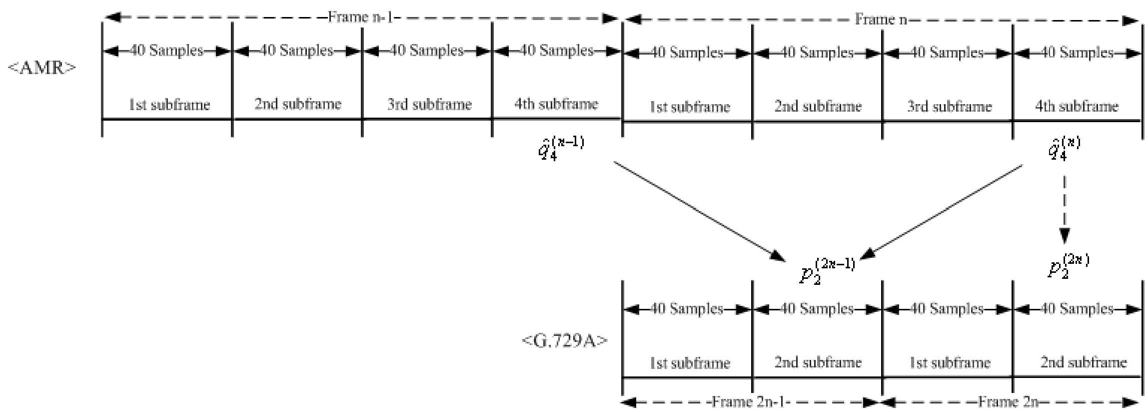

Figure 4. The LSP mapping method from AMR to G.729A (other 7 modes)

2.2. Pitch delay conversion

For both AMR and G.729A, the adaptive codebook is searched by combining open-loop with closed-loop searches. Both codecs share the following features. For even subframes, the pitch lag is absolutely 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 previous subframe with a fractional resolution. The pitch search range and fractional resolution between AMR and G.729A are slightly different, as shown in Table III .

We can find the strong similarities between AMR and G.729A, and therefore the complexity can be reduced. We discuss the procedure in two groups, AMR 12.2 kbit/s mode and other seven modes. First, we choose the integer part of pitch every two subframes of AMR to be the open-loop result of G.729A. Second, we search the fractional part of pitch by using the closed-loop search method of G.729A. The block diagram is shown in Figure 5.

Table III. The pitch search range of AMR and G.729A

	Odd subframe			Even subframe	
Precision	Fractional		Integer	Fractional	
	Frac.	Range	Range	Frac.	Range
G.729A	1/3	$19_{1/3}$ $84_{2/3}$	85 143	1/3	$-5_{2/3}$ $4_{2/3}$
AMR (5.9、6.7k)	1/3	$19_{1/3}$ $84_{2/3}$	85 143	1/3	$-1_{2/3}$ $0_{2/3}$
AMR (7.4、10.2k)	1/3	$19_{1/3}$ $84_{2/3}$	85 143	1/3	$-5_{2/3}$ $4_{2/3}$
AMR(7.95k)	1/3	$19_{1/3}$ $84_{2/3}$	85 143	1/3	$-10_{2/3}$ $9_{2/3}$
AMR(12.2k)	1/6	$17_{3/6}$ $94_{3/6}$	94 143	1/6	$-5_{3/6}$ $4_{3/6}$

.

	1st subframe			2nd~4th subframe		
Precision	Fractional		Integer	Fractional		
	Frac.	Range	Range		Frac.	Range
AMR (4.75、5.15k)	1/3	$19_{1/3}$ $84_{2/3}$	85 143	step1	1	-5 4
				step2	1/3	$-1_{2/3}$ $0_{2/3}$

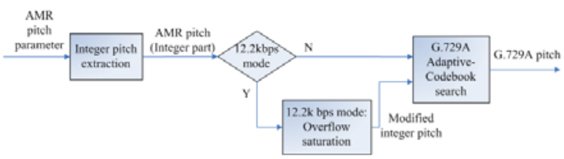

Figure 5. Two speech pitch delay conversion schemes (a) Full decoding (b) Partial decoding

2.3. Proposed fixed codevector conversion

An algebraic codebook structure is adopted in both AMR and G.729A as a fixed codebook. A given number of non-zero pulses are specified for the position and amplitude, either +1 or -1. Because the codebooks of AMR are different from G.729A except 7.4 and 7.95 kbit/s modes, the codebook can not be shared in general. The procedure of the fixed codevector conversion is shown in Figure 6.

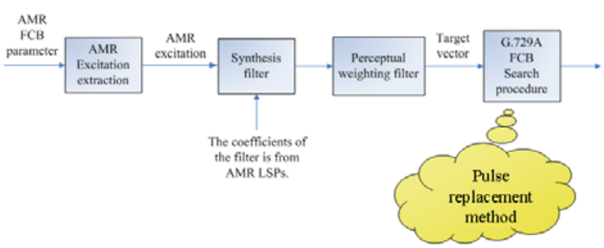

Figure 6. The procedur of fixed codebook search

We propose a partial decoding method using a fast codebook search algorithm. The pulse replacement method was used in ACELP fast fixed codebook search. We utilize this method in speech transcoding. In pulse replacement procedure, it is necessary to measure the contribution of each pulse and replace the least important pulse with a new one. The search block diagram is shown in Figure 7. The contribution is to compute the similarity between synthesized signal and the target signal.

$$S(d, c_k) = \frac{\left(\sum_{n=0}^{39} d(n) c_k(n)\right)}{c_k^t \Phi c_k} \qquad (3)$$

where $d(n)$ denotes the correlation of the target signal and the impulse response, c_k denotes the kth fixed codebook vector, and Φ denotes the autocorrelations of the impulse response of the weighted synthesis filter.

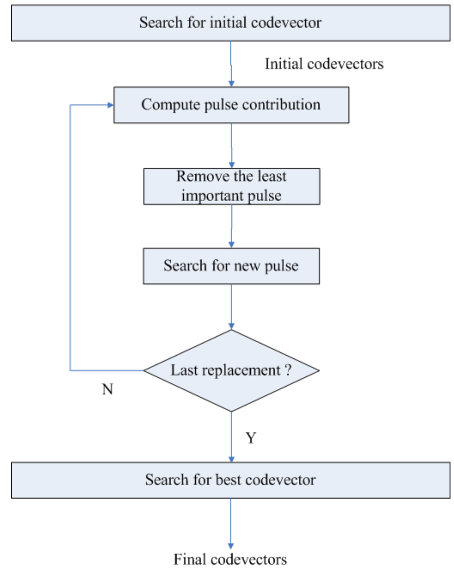

Figure 7. The pulse replacement procedure

In searching the initial codevector, because the pulse numbers of AMR eight modes are different, which is shown in Table IV , they are classified into three classes：

1. AMR 4.75, 5.15, 5.9 and 6.7 kbit/s modes ：Randomly generating one or two extra pulses.
2. AMR 7.4 and 7.95 kbit/s modes ：Mapping the AMR codevector to G.729A directly.
3. AMR 10.2 and 12.2 kbit/s modes：Calculating the contributions of all the 8(10) pulses and reserving the most important four.

Table IV. The pulse number of AMR eight modes

Bit-rate (kbit/s)	4.75	5.15	5.9	6.7	7.4	7.95	10.2	12.2
Number of pulses	2	2	2	3	4	4	8	10

After the initial codevector is determined, the pulse replacement procedure is applied to the initial codevector to enhance the performance. The complexity of this method is very low because the search is run on a limited number of pulses only.

3. SIMULATION RESULTS

To evaluate the performance of our proposed system, we use C code to simulate all digital audio signal process and MATLAB for signal analysis on Pentium-4 3.2G PC. The input digital speech test sequence is sampled at 8kHz as 16-bit PCM, including

Chinese and English with 9 females and 10 males. The length of each sentence is at least longer than 9 seconds.

3.1. Objective Measurements

We adopt Perceptual Evaluation of Speech Quality (PESQ) [7] as objective measurement. PESQ attempts to incorporate more than just speech codecs but also end-to-end network measurement. As shown in Figure 8, it takes the human psycho-acoustic model into account. From [7], the grade of PESQ MOS is very close to human subjective MOS and the correlation between them is as high as 0.95. Therefore, PESQ can not only present the objective measurement but reflect the subjective quality.

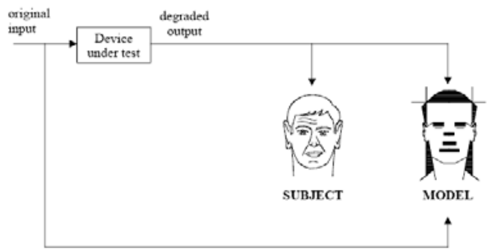

Figure 8. The evaluation model of PESQ

The MOS grade ranges 5~1 which denotes from excellent to unacceptable respectively. And generally, the 3 point means “fair” denoting human can accept that voice quality.

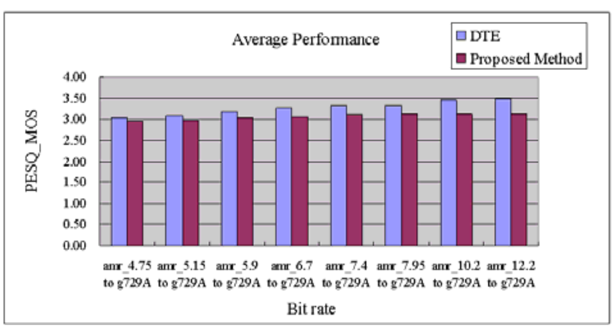

Figure 9. The average performance of conventional and proposed method

Our simulation is shown in Figure 9. We can find that in our partial transcoding method, the PESQ MOS grade is closed to conventional DTE method.

As a rough transcoding of proposed method computation complexity, its clockticks is compared with conventional full decoding method. The result is shown in Table V The computation complexity of the conventional method is as high as 13.89 times of ours.

Table V. The complexity between conventional and proposed methods

Test files	Conventional method	Proposed method
TSTSEQ1.pcm	249,369,138	17,986,692
female7.pcm	248,984,368	17,894,348
Average	249,176,753	17,940,520

4. CONCLUSIONS

We have proposed a partial decoding method for speech transcoding. The method can be applied to all eight modes of AMR to G.729A. The speech quality is close to the conventional method with only 7.2% computation complexity.

5. REFERENCES

[1] ETSI, “Digital cellular Telecommunications System （Phase 2+）；Adaptive Muliti-Rate（AMR）speech transcoding, ”" EN 301 704, Apr. 2000.

[2] ITU-T Recommendation G.729A, "Coding of speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)," Mar. 1996.

[3] Y. Ota, M. Suzuki and Y. Tsuchinaga, "Speech coding translation for IP and 3G mobile integrated network," Proc. of ICC, pp. 114-118, Apr. 2002.

[4] S. Lee, S. Seo and D. Jang, "A novel transcoding algorithm for AMR and EVRC speech codecs via direct parameter transformation," Proc. of ICASSP, pp. 177-180, vol. 2, Apr. 2003.

[5] J. Choi, C. Lee, H. Kang, Y. Park and D. Youn, "Improvement issues on transcoding algorithms for the flexible usage to the various pairs of speech codec," Proc. of ICASSP, pp. I-269 ~ I-272, May. 2004.

[6] A. S. Spanias, "Speech Coding: A Tutorial Review," Proc. of the IEEE, vol. 82, no. 10, pp. 1541-82, Oct. 1994.

[7] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone network and speech codecs," Feb. 2001.

[8] R. V. Cox, “Three New Speech Codecs from the ITU Cover a Range of Application,” IEEE Comm. Magazine, Sept. 1997.

[9] M. R. Scheroeder, B. S. Atal, "Code-excited linear prediction (CELP): high quality speech at very low bit rate," Proc. of ICASSP, pp. 937-940, Mar. 1985.

[10] P.C. Chang and H.M. Yu, "Dither-like Data Hiding in Multistage Vector Quantization of MELP and G.729 Speech Coding," in Thirty-Sixth Asilomar Conference on Signals, Systems, and Computers, vol. 2, pp.1199-1203, Monterey, CA, Nov. 2002.

[11] C. Hoene, H. Karl, and A. Wolisz, "A perceptual quality model for adaptive VoIP Applications," Proc. of SPECTS, San Jose, CA, July 2004.