GSM TO G.729 SPEECH TRANSCODER

Shu-Min Tsai and Jar-Ferr Yang

Department of Electrical Engineering,
National Cheng Kung University
1 University Road, Taiwan, ROC
Tel: 06-2757575 ext. 62362 Fax: 06-2345482
Email:jfyang@ee.ncku.edu.tw

ABSTRACT[1]

With increasing demand of wireless and Internet accesses, the interoperability crossing these two networks becomes increasingly important for modern communications. A transcoding system that translates the coding parameters directly between the GSM and G.729 speech coders is discussed in the present paper. The GSM coder is used for mobile communications while the G.729 is the most favorable speech coder in Internet communications. For Internet/wireless gateway servers, it is evident that the proposed transcoding system requires much less computation than the conventional decode-then-encode (DTE) approach.

1. INTRODUCTION

For wireless communications, the Global System for Mobile communication (GSM) suggests an efficient speech coding system called the GSM codec and gains a lot of popularity internationally [1]. In the same time, the G.729 speech coder standardized by International Telecommunication Union (ITU) is introduced for Internet applications [2]. Internet telephony, the G.729 figured with the reduced rate G.729D and the enhanced rate G.729E becomes the most favorable speech coding system, which can provide variable rate in Internet application and the possible digital speech in the future Public Switch Telephone Network (PSTN). It is obvious that transcoding between the GSM and the G.729 speech compression standards will become more and more important when we need to integrate these two speech phone standards to establish wireless and Internet telephony connections.

A straightforward way for the transcoding is through a so-called Decode-Then-Encode (DTE) approach. It first starts from the GSM decoding process to reconstruct the compressed speech then perform the G.729 encoding to complete the transcoding. In this study, therefore, we propose a speech transcoder capable of direct coding parameter translation that not only reduce the most of computation, but also keep the speech quality as the DTE approach. Our transcoding technique successfully reduces the cost of interoperability and promotes the feasibility in links of wireless mobile phone and Internet phone.

This paper is organized as follows. From reduction of computation, the transcoding methods are described in Section 2. To verify effect of the proposed transcoding techniques, some experimental results will be depicted in Section 3. Conclusions follow in Section 4.

[1] This research was supported by National Science Council under Contract #NSC-89-2213-E006-094, Taiwan, ROC.

0-7803-7057-0/01/$10.00 ©2001 IEEE.

2. TRANSCODING METHOD

In this paper, we present a new transcoder, which can effectively and directly transferring all the coding parameters from the GSM to the G.729 coded speech. The first problem being faced for the coder transformation is different frame size from individual system. Frame synchrony transformation is extremely important since different frame size will cause data loss or overlap. Figure 1 shows the transformation of speech frame length. To transform each other directly, we adopt two G.729 speech frame and one GSM speech frame with 160 samples each.

Table 1 listed the main function for each frame. The CPU time requirement is estimated from the C program of the G.729 encoder with Intel Pentium II CPU. In Table 1, the main three parts are short-term synthesis filter, long-term synthesis filter, and the excitation. These three parts are required 63.1% CPU time. So the coder transformation is based on these three parts, will reduce the computation. Each of these transformations is addressed below.

2.1. LPC Parameters Transformation

The LPC parameters of G.729 and GSM are both calculated from Levinson-Durbin algorithm, the vocal tract model is

$$H(z) = \frac{1}{1+\sum_{i=1}^{P} \alpha_i z^{-i}}, \tag{1}$$

where α_i, for all i, are the LP filter coefficients with P=10.

The G.729 uses line spectrum pair (LSP) to represent the LPC parameters. The LSP are quantized by the two-stage VQ. As for GSM, a three-segment vector quantizer of the reflection coefficients is applied. The concept of two-stage VQ is also employed here.

The frame size of these two systems must be considered. To transform one GSM frame and two G.729 frames simultaneously. The LSP of subframe from GSM are obtained named G_1, G_2, G_3 and G_4. And the LSP of the G.729 for the second subframe are named A_2 and B_2. When LSP be obtained from GSM, we use the parameters in G.729 coder with different methods. The methods are listed as follows:

Method 1: LSP of A_2= LSP of G_2
LSP of B_2= LSP of G_4
Method 2: LSP of A_2= LSP of G_1 (2)
LSP of B_2= LSP of G_3
Method 3: LSP of A_2=
1/4*LSP of G_1+3/4* LSP of G_2
LSP of B_2=
1/4* LSP of G_3+3/4* LSP of G_4

We then use logarithmic spectral distortion (LSD) to measure the similarity between the DTE and methods of LPC coefficients defined above. For a detailed comparison, plot of the spectral distortion in dB with three different methods is depicted in Figure 2. We learn that Method 3 (solid line) gives the most preferred result due to the existence of the correlation between adjacent subframes.

2.2. Pitch Transformation

Pitch predictor is used to predict the quasi-periodic signal of speech. So pitch of long-term prediction is an important parameter for speech encoding. The fractional pitch resolution of G.729 is *1/3* while that of GSM is *1/6*, *1/3* or *1/2*.

Pitch search is separated into two parts, open loop search and closed loop search. This part requires heavily calculation. How to utilize the received pitch information from the other coder to reduce the search complexity is the research concern.

Since subframe size of both GSM and G.729 are 40 samples, the transformation can be done one-by-one. Speech signal is classified as voiced and unvoiced in the GSM system. In order to reduce the loop search in the G.729 encoder, the G.729 directly uses the pitch lag found from GSM.

2.3. VSELP and ACELP Transformation

Both VSELP and ACELP methods determine the random excitation from residual speech after the pitch estimation [3, 4]. The calculation of random excitation is also time-consuming. There would be a great contribution on transcoding system if this part could be successfully transformed.

The proposed method in G.729 coder is shown as Figure 3. The LP coefficients and pitch lags are obtained from the GSM coder. The random excited signal of the G.729 is expressed by

$$e'_{random}[n] \cong e_{pitch}[n] + e_{random}[n] - e'_{pitch}[n], \quad (3)$$

where $e_{pitch}[n]$ and $e_{random}[n]$ respectively denote the adaptive and random excitations, which could be directly obtain from the GSM coding parameters, and $e'_{pitch}[n]$ is the pitch excitation of the G.729 obtained in Subsection 2.2. Now, we will find pulses $\tilde{x}$ from $e'_{random}[n]$, than let the algorithm

$$|Y - H\tilde{x}| = 0, \quad (4)$$

where Y is the target signal from G.729 coder, H is the synthesis filter. It requires a lot of computation to obtain the optimum solution. Thus, a suboptimal solution could be developed by minimizing Equation (4).

In order to speed up the search procedure, it is necessary to reduce the number of the pulse position combination. Thus, we calculate the energy of target signal per track as

$$E_i = \sum_{j=0}^{7} [y(i+5j)]^2, \qquad i = 0,\dots,4 \quad (5)$$

$$\Delta = (E_{\max} - E_{\min}) / 4 \quad (6)$$

$$m_i = (E_i - ave(E_i)) / \Delta, \quad i = 0,\dots,4 \quad (7)$$

$$n_i = m_1 + m_i, \qquad i = 0,\dots,4 \quad (8)$$

where n_i is the pulse number per track, m_1 is the initial value. We use the excitation signal from the GSM to find the search range. We will pick up n_i leading pulses from GSM excited signals and put them into G.729 system to search the suboptimal pulses. The pulse location is more important when the energy is higher. So, we will pick more pulses in the track.

3. SIMULATION RESULTS

The simulation results in objective performance are tabulated in the Table 2. The SNR of conventional method is 10.574 and the SegSNR is 8.174. With the proposed method, the SNRs are slightly decreased. Table 3 shows the computational saving of each block proposed method of GSM/G.729 transcoding. Estimating the pitch requires 23% CPU time in G.729 encoder. We can save 98.5% CPU time with our proposed transcoding method. Just in this part we have saved a lot of computation. So combining three parts, a total computational saving of more than 39.6% CPU time is achieved and the speech quality is still good. Moreover, The speech signal through the GSM/G.729 transcoding still performs well in subjective measure. We provided the listening test in http://netcity.hinet.net/m85611a.

4. CONCLUSIONS

In this paper, the concept of transcoding between two speech coding systems is proposed. We use the parameters from the GSM encoded speech to reduce the most of search loops in the G.729 encoder. In G.729 speech coder, the complexity of the encoder is about 11.86 times than that of the decoder. A substantial computation reduction is achieved over the direct DTE transcoding. Thus, the proposed algorithm promotes the possibility of the wireless mobile phone roaming in the worldwide Internet networks.

REFERENCES

[1] European Telecommunication Standard, “Digital Cellular Telecommunications System (Phase 2+); Half Rate Speech; GSM 06.20 Version 5.1.1”, May 1998.

[2] ITU-T Recommendation G.729, “Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP)”, March 1996.

[3] I. M. Trancoso and B. S. Atal, "Efficient Search Procedures for Selecting the Optimum Innovation in Stochastic Coders", *IEEE Trans. on ASSP*, vol. 38, no. 3, March 1990.

[4] Lan Juan, Lin Biqin, and Fu Qiuliang, "An 8-kb/s Conjugate-Structure Algebraic CELP (CS-ACELP) Speech Coding", *Proceedings of ICSP*, 1998.

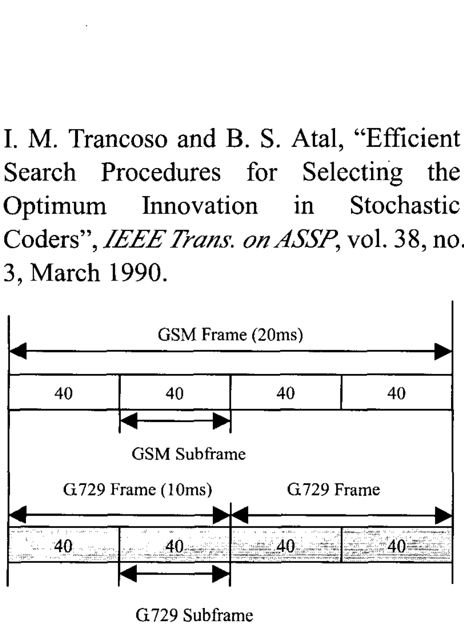

Fig. 1. GSM/G.729 frame transform.

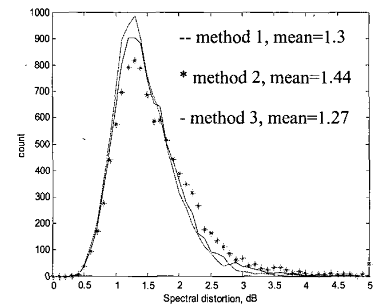

Fig. 2. Histogram of spectral distortion in

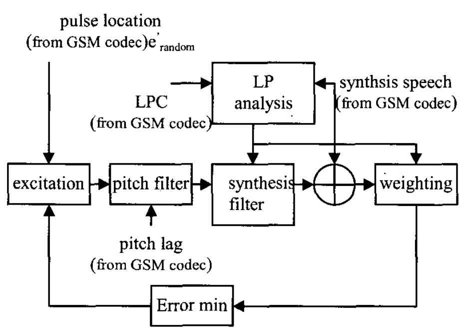

Fig. 3. Proposed method in G.729 coder.

Table 1. The CPU time requirement of G.729 encoder

Function description	CPU time requirement
LPC, quantised and interpolated LSP	20.68%
Open-loop & close-loop analysis	23%
Fixed codebook search	19.42%
The others	36.9%

Table 2. Performance evaluation of the transcoder

Performance	Traditonal Method	LPC+Pitch Excitation$_1$	LPC+Pitch Excitation$_2$	LPC+Pitch Excitation$_3$	LPC+Pitch Excitation$_4$
SNR	10.574	9.893	10.036	10.191	10.242
SegSNR	8.174	7.186	7.473	7.644	7.723

(Excitation$_1$ with m_1=2, Excitation$_2$: m_1=3, Excitation$_3$: m_1=4, Excitation$_4$: m_1=5)

Table 3. Computational saving in main parts of the transcoder

Main Parts	LPC+Pitch Excitation$_1$	LPC+Pitch Excitation$_2$	LPC+Pitch Excitation$_3$	LPC+Pitch Excitation$_4$
Short-term	15.7%	15.7%	15.7%	15.7%
Long-term	98.5%	98.5%	98.5%	98.5%
Excitation	83.2%	81.9%	78.3%	70.8%

(Excitation$_1$ with m_1=2, Excitation$_2$: m_1=3, Excitation$_3$: m_1=4, Excitation$_4$: m_1=5)