Low Complexity Multi-track Search Scheme for G.723.1 MP-MLQ

R. S. Lin[a], J. Y. Wang*[a], H. B. Jhang[a] and J. S. Pan[b]

Department of Computer Science and Information Engineering Southern Taiwan University[a]

Department of Electronic Engineering National Kaohsiung University of Applied Sciences[b]

Abstract **—The Multi-pulse Maximum Likelihood Quantization (MP-MLQ) algorithm with high computational complexity and high quality has been used in the G.723.1 standard codec. To reduce the computational complexity of MP-MLQ, in this paper, we propose a fast search algorithm by using a designed energy function and the multi-track positions structure of the stochastic excitation signals to predict the candidate pulses for each subframe. Simulation results show that the degradation of PESQ for the proposed approach is about 0.036 relative to the original search procedure. However, the proposed approach reduced the computational complexity about 98.54% with perceptually negligible degradation.**

I. INTRODUCTION

Speech Communication is the most dominant and common service in telecommunication networks and multimedia world. It has remained the most desirable medium of communication between groups of people. The Skype and VOIP which uses Internet and speech compression technology primarily, is one of the most popular issues in recent years. However, with rapidly increasing of population growth rate of Browsing the Internet, there are intermittent communication caused by packets interruptions to degrade the speech quality of high bit-rate voice streams in a limited network bandwidth resources. To achieve "continuity", speech codec with high compression rate has been used to generate a low-rate data stream, and then is to maintain the continuous voice in the congestion of the network environment, however the high compression rate codec required very high computational complexity, therefore reducing the bit rate and improving speech quality of codecs are the common objectives of many scholars' studies, however using of low-rate and low computation complexity codec which has to take the voice quality and real-time voice communications into account. In low-rate codec, Telecommunication Standardization Sector of International Telecommunication Union (ITU-T) recommends two speech codecs, G.723.1 [1] and G.729 [2], which have high speech quality and low bit-rate. These two speech coders are the major audio compression standards for H.323 Internet phone systems and H.324 digital videophone service in public switching telephone network (PSTN) systems. G.723.1 has been used on Internet extensively, such as the built-in applied software "NetMeeting" in Microsoft's windows operation system. In NetMeeting, G.723.1 is used as one of the speech compression standards in its Internet phone. The G.723.1 codec is dual rate 6.3, 5.3kbit/s and the G.729 is 8 kbit/s bit-rate.

The both coders adopt the vocal tract model, which can be characterized by Code-Excited Linear Predictive Coding (CELP), and the parameter coding of the vocal tract model is performed in various formats. A general speech coder is composed of three main parts including short-term prediction (STP), analysis of human vocal tract model and long-term prediction (LTP), the analysis of the speech periodic signals, finally the Stochastic Codebook Excitation (S.C.E) search. The coding structure is based on the Analysis-by-Synthesis (A.b.S) search mechanism [3] - [7]. It can achieve a good voice quality and low-rate, but requires high computational complexity [8]. In this paper, we propose a fast search algorithm to improve the search computational complexity of MP-MLQ algorithm in the G.723.1 speech coder. In this paper, we utilized energy function $b[n]$ and design a multi-track positions structure of the stochastic excitation signals, which arrange even and odd pulses positions in the subframe. We adopt to predict the predetermined candidate pulse positions for every track by $b[n]$ energy signal function before the MP-MLQ search stochastic codebook procedure. Restrict the search combination number of pulses to achieve reduction of search computation required in MP-MLQ algorithm.

In section II, we will briefly review the standard MP-MLQ search algorithm of the G.723.1 speech coder. Then, we introduce a low complexity multi-track fast search algorithm for MP-MLQ in Section III, and In Section IV, experiments and the estimation of voice quality. Finally, Section V gives a conclusion.

II. MP-MLQ SEARCH ALGORITHM

After the short-term analysis and long-term prediction, the weighted residual signal, $r[n]$ is obtained as a new target vector for the stochastic excitation codebook processing block. The stochastic excitation search procedure, which performs estimation and quantization for the target vector, involves the determination of pulses position and pulses amplitude. The target vector $r[n]$ will be approximated as

$$r'[n] = \sum_{j=0}^{n} h[j] \cdot v[n-j] \ ,0 \le n \le L-1 \qquad (1)$$

where L is subframe length and

$$v[n] = G\sum_{k=0}^{P-1} \alpha_k \delta[n-m_k] \ ,0 \le n \le L-1 \qquad (2)$$

is the coded excitation vector of the vocal tract filter, in which $\delta[n]$ is a Dirac function, G is the gain factor, and $\{\alpha_k\}, k = 0 \ldots P-1$ are the signs (± 1). If the number of pulses is P, the coded positions of pulses in the MP-MLQ stage will be $\{m_k\}, k = 0 \ldots P-1$. In G.723.1, P is 6 for even subframe and is 5 for odd subframe. There is a restriction on pulse positions in the G.723.1 coder, the positions can be either all odd or all even, which is indicated by a grid bit. Therefore, the problem of MP-MLQ is to find the unknown parameters, G, $\{\alpha_k\}, k = 0 \ldots P-1$, and $\{m_k\}, k = 0 \ldots P-1$, such that the mean square error

$$err[n] = r[n] - r'[n] = r[n] - G\sum_{k=0}^{P-1}\alpha_k h[n-m_k] \tag{3}$$

is minimized.

According to the property of maximum likelihood, the cross-correlation function $d[j]$ between the impulse response, $h[n]$ and the new target vector, $r[n]$ is first computed as

$$d[j] = \sum_{n=j}^{L-1} r[n]\cdot h[n-j] \ ,0 \le j \le L-1 \tag{4}$$

and, the parameter $G_{\max}$ is estimated by

$$G_{\max} = \frac{\max\{|d[j]|\}_{j=0..L-1}}{\sum_{n=0}^{L-1} h[n]\cdot h[n]} \tag{5}$$

For each of these gain values the signs and locations of the pulses are sequentially optimized. This procedure is repeated for both the odd and even indices. Finally the combination of the quantized parameters that yields the minimum mean square of $err[n]$ is selected.

The speech coder G.723.1 is based on the analysis-by-synthesis technology, such codec structure can be achieved high speech quality and low-bit rate. But the shortcoming of this technology is encoder required much computational complexity to search stochastic excitation pulses and the estimated coding gain calculation. As Fig. 1 shown, the fact is that the MP-MLQ codebook search procedure takes up above 55% [9]-[10] computation required in the encoding process of the G.723.1 coder.

It is necessary to reduce computation of MP-MLQ search algorithm in ITU-T G.723.1 to improve the efficiency [11]-[13] of the speech coder to be real-time coding speech signal.

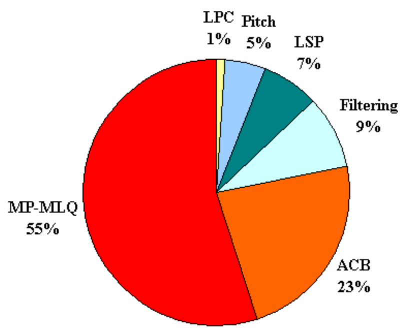

Fig. 1 The distribution of computational load in the encoding process for high-rate G.723.1

III. MULTI-TRACK SEARCH PREDICTION SCHEME

The MP-MLQ search algorithm calculates the minimize mean square error (MSE) of odd pulse positions and even pulse positions for subframe, respectively, and then the least MSE is selected. In G.723.1, the optimal combination of stochastic excitation pulse positions is $2\times C_m^{30}, m = 6,5$,and in (3)-(5), it is noted that the MP-MLQ search algorithm requires much computational complexity.

In this paper, we use the energy function $b[n]$ [14] of the excitation signal, and design a multi-track excitation pulses positions structure, these arrange to reduce the search complexity of MP-MLQ and to achieve real-time implementation, the flow chart of proposed approach is shown in Fig. 2. First, residual signal $r[n]$ was filtered by $A(z)$ filter to generate the excitation signal $res_{LTP}[n]$ for each subframe, in which $A(z)$ filter is defined as

$$A(z) = 1 - \sum_{i=1}^{10} a[i]z^{-i} \tag{6}$$

and, the excitation signal $res_{LTP}[n]$ is defined as

$$res_{LTP}[n] = r[n] - \sum_{i=1}^{10} a[i]*r[n-i] \ ,0 \le n \le 59 \tag{7}$$

By using $res_{LTP}[n]$ and $d[n]$, pulse-position likelihood-estimate energy function vector is defined as

$$b[n] = \frac{res_{LTP}[n]}{\sqrt{\sum_{i=0}^{L-1} res_{LTP}[i]\cdot res_{LTP}[i]}} + \frac{d[n]}{\sqrt{\sum_{i=0}^{L-1} d[i]\cdot d[i]}} \ ,0 \le n \le L-1 \tag{8}$$

We adopt $b[n]$ as energy function to predict candidate excitation pulses for every track, so we use $|res_{LTP}[n]|$ and $|d[n]|$ instead of $res_{LTP}[n]$ and $d[n]$ in (8).

The predicted candidate excitation pulse positions has larger value $b[n]$ in every track, and these candidate pulse positions with odd and even indices will be the multi-track structure as shown in the table I and II. The MP-MLQ algorithm searches the predicted candidate excitation pulse positions, which involve odd and even indices. These candidate positions are dented as $\Phi_p(b[n_j]|Str),\ p = 5,6.\ j = 0,1\ ,\ 1 \le Str \le 5$. The suffix j indicates even or odd pulse position indices in every track. However the [12] predict candidate excitation pulse positions only even or odd pulse positions alternatively, therefore it is maybe lose the best excitation pulse vector. Furthermore, only Str positions with larger value $b[n]$ for every track will be searched, the number of candidate pulse positions will be reduced from L=60 to ($2\times Str\times p$). Therefore, the computational complexity of MP-MLQ stage can be reduced by using the proposed approach. The quality of the coded speech will be evaluated in the following.

TABLE I
MULTI-TRACK STRUCTURE OF EVEN SUBFRAME

Pulses positions
0(1), 12(13), 24(25), 36(37), 48(49)
2(3), 14(15), 26(27), 38(39), 50(51)
4(5), 16(17), 28(29), 40(41), 52(53)
6(7), 18(19), 30(31), 42(43), 54(55)
8(9), 20(21), 32(33), 44(45), 56(57)
10(11), 22(23), 34(35), 46(47), 58(59)

TABLE II
MULTI-TRACK STRUCTURE OF ODD SUBFRAME

Pulses positions
0(1), 10(11), 20(21), 30(31), 40(41), 50(51)
2(3), 12(13), 22(23), 32(33), 42(43), 52(53)
4(5), 14(15), 24(25), 34(35), 44(45), 54(55)
6(7), 16(17), 26(27), 36(37), 46(47), 56(57)
8(9), 18(19), 28(29), 38(39), 48(49), 58(59)

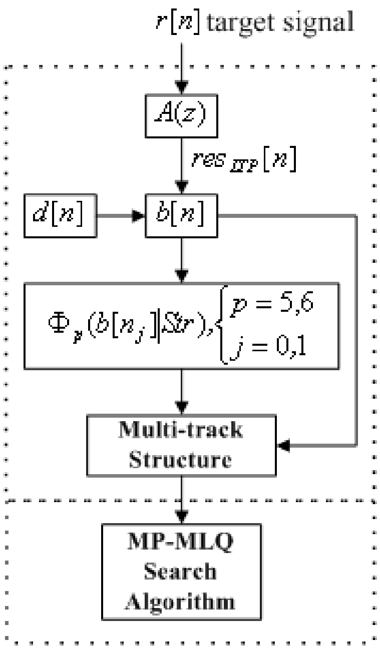

Fig. 2 The flow chart of proposed approach.

IV. EXPERIMENT RESULTS

In this paper, we proposed multi-track fast search method for the ITU-T G.723.1 standard speech codec (STD) MP-MLQ search algorithm. In order to evaluate the speech quality for the proposed approach, 20 speeches are performed. These speech files are produced by 10 males and 10 females. To show the perceptual speech quality, the objective MOS, called PESQ, is performed [15]. In the following experiments, we compared the proposed approach and the [12] to the STD codec. First, we use multi-track table I structure. Fig. 3 and Fig. 4 show the computational complexity and the average PESQ degradation with different Str numbers in every track. Observing Fig. 3, a gap is presented between Str =1 and Str =2. By setting Str =2, Fig. 5 shows the PESQ degradation for these 20 test speeches. It can be found that average PESQ degrades 0.055 is achieved.

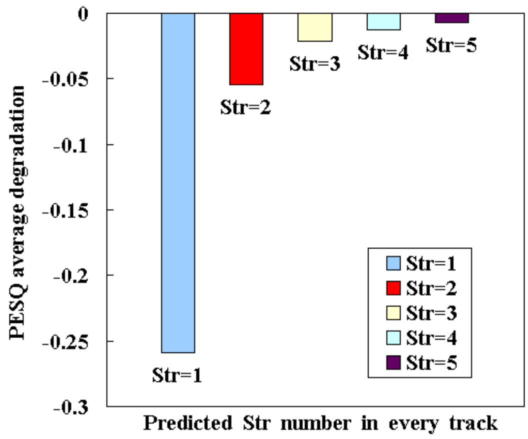

Fig. 3 PESQ average degradation for different Str.

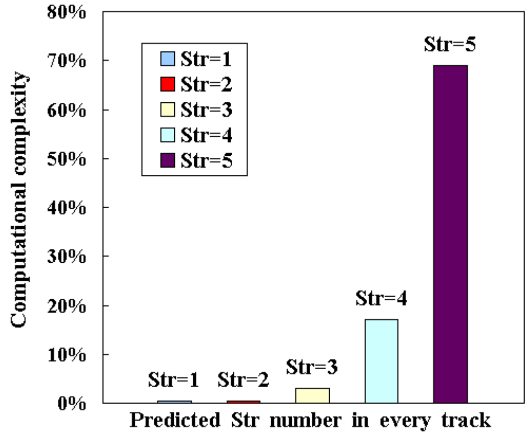

Fig. 4 Computational complexity for different Str.

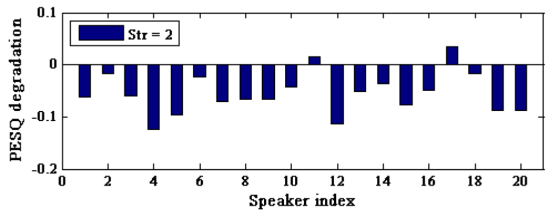

Fig. 5 PESQ degradation with Str=2, using table I

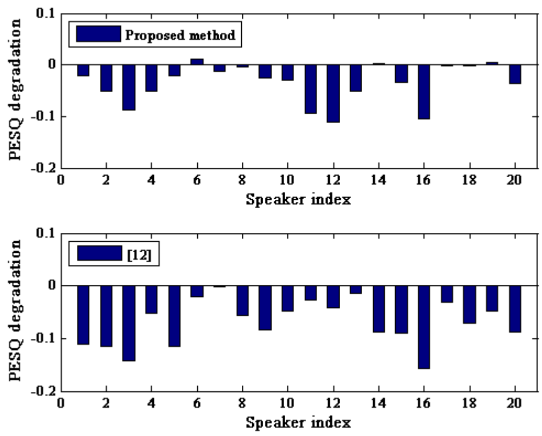

Fig. 6 PESQ degradation of compared the Proposed and [12] to STD

To improve speech quality, we propose the table I structure and every track predict candidate positions ($Str=2$) for even subframe and. As for, odd subframe use table II structure and every track predict candidate positions ($Str=3$). Fig. 6 and Fig. 7 show the PESQ degradation for these 20 test speeches. It can be found that average PESQ degrades only 0.036. However, [12] degrades 0.069. Table III shows computational complexity of the proposed approach compared to STD and [12]. In the table III, $P_0[n]$ and $P_1[n]$ is combination numbers of candidate pulse positions for even and odd subframe respectively.

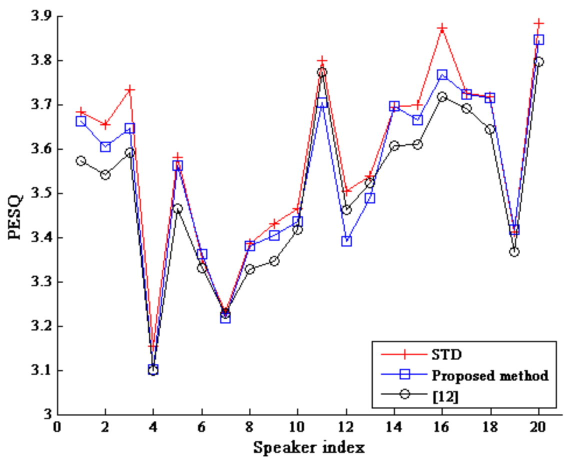

Fig. 7 PESQ for different method.

In the above discussion, it is noted that the proposed approach reduced the computational complexity about 98.54%, and the [12] approach reduce 98.35%. However, the preprocessing for deciding the candidate positions need extra computational load as shown in Table IV.

TABLE III
COMPUTATIONAL COMPLEXITY FOR DIFFERENT METHOD

Method	Combination
STD	$P_0[n]=2\times C_6^{30}=1187550$
	$P_1[n]=2\times C_5^{30}=285012$
[12]	$P_0[n]=C_6^{6\times3}=18564$
	$P_1[n]=C_5^{5\times3}=3003$
Proposed	$P_0[n]=2\times C_6^{6\times2}=1848$
	$P_1[n]=2\times C_5^{5\times3}=6006$

TABLE IV
EXTRA COMPUTATIONAL LOAD FOR THE PROPOSED APPROACH

Procedure	Computational load (L=60)	Total
$res_{LTP}[n]$	$(L-10)\times10+55$, (×) $(L-10)\times10+55$, (+)	797, (×) 733, (+)
$b[n]$	$4\times L+2$, (×) $3\times L-2$, (+)	

V. CONCLUSION

In this paper, we proposed a fast MP-MLQ search algorithm for stochastic codebook of ITU-T G.723.1 codec. By using pulse-position likelihood- estimate energy function vector and pulses multi-track structure, the proposed approach can reduce the computational complexity about 98.54% and the PESQ estimation degrades about 0.036. Simulation results show that the coded speech quality evaluated by using the standard subjective and objective quality measurements is with perceptually negligible degradation.

REFERENCES

[1] ITU-T Recommendation G.723.1, *Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s*, March 1996.

[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] P. Kroon and B. S. Atal, "On the use of pitch predictors with high temporal resolution," *IEEE Trans. Signal Processing*, vol. 39 Issue 3, pp.733-735, March. 1991.

[4] W. B. Kleijn, R. P. Ramachandran and P. Kroon , "Interpolation of the pitch-predictor parameters in analysis-by-synthesis speech coders ," *IEEE Trans. on Speech and Audio Processing*, vol. 2 Issue 1 part 2, pp. 42-54, January 1994.

[5] P. Kroon and B. S. Atal, "Pitch predictors with high temporal resolution," *IEEE Proc.*, pp. 661-664, 1990.

[6] C. Zhang, K. Tang, H. Cui, W. Du, and J. Li, "Efficient pitch predictor algorithm in ITU-T G.723.1," *IEEE Conf. Communication Technology Proceedings*, Vol. 02, pp. 1727-1729, April 2003.

[7] S. K. Jung, K. T. Kim, Y. C. Park, and H. G. Kang, "A fast adaptive-codebook search algorithm for G.723.1 speech coder," *IEEE Signal processing letters*, vol. 12, pp. 75-78, January 2005.

[8] M. Olausson, L. Dake, "Instruction and hardware acceleration for MP-MLQ in G.723.1," *Signal Processing Systems*, pp. 235-239, October 16-18, 2002.

[9] F. K. Chen and J .F. Yang, "Candidate Scheme for MP-MLQ Search in G.723.1," *Third JEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, Taoyuan, Taiwan, pp. 20-23, March 2001.

[10] S. M. Lee, S. Park and Y. Jang, "Cost-effective implementation of ITU-T G.723.1 on a DSP chip," *Proceedings of 1997 IEEE International Symposium on Consumer Electronics*, pp. 31-34, December 1997.

[11] C. Negrescu," Optimization algorithm for the MP-MLQ excitation in G723.1 encoder," *IEEE Conf. Electronics, Circuits and Systems*, pp. 1003-1006, June 2006.

[12] R. S. Lin, Y. C. Chen and F. K. Chen, "Lower Complexity Search Method for G.723.1 MP-MLQ Algorithm," *The eighth International Conference on Intelligent Systems Design and Applications (ISDA-2008)*, pp. 158-161, November 26-28, 2008.

[13] R. S. Lin, J. Y. Wang, "Low Complexity Fast Search Method by residual signal for G.723.1 Codec," *Information Technologies, Applications and Management Conference (ITAMC-2009),* June 2009.

[14] ITU-T Recommendation G.729.1, *An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729*, May 2006.

[15] ITU-T Rec. P.862, *Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Qualit Assessment of Narrowband Telephone Networks and Speech Coders*, February 2001.