HIGH PERFORMANCE CELP CODER UTILIZING A NOVEL ADAPTIVE FORWARD-BACKWARD LPC QUANTIZATION

Zijun Yang **Jozsef Vass** **Yunxin Zhao**[†] **Xinhua Zhuang**

Dept. of Computer Engineering
& Computer Science
University of Missouri-Columbia
Columbia, MO 65211

[†] Beckman Institute and Dept. of
Electrical & Computer Engineering
University of Illinois
Urbana, IL 61801

Abstract – A highly efficient algorithm termed *adaptive forward-backward vector quantization* (AFBVQ) is developed for variable bit rate quantization of linear predictive coding (LPC) coefficients and integrated with the FS1016 Federal Standard Code Excited Linear Predictive (CELP) coder. This results in a high performance low bit rate speech coder called as AFBVQ-CELP which brings in two-fold bit rate reduction by backward LPC indexing and by forward LPC VQ.

In AFBVQ, a previously decoded and temporally close speech signal is *re-segmented* into *overlapping* blocks. As the LPC coefficients calculated from one of those synthetic blocks are spectrally close to the current *unquantized* LPC coefficients, the backward LPC indexing is used to encode the current speech block; otherwise, the forward linear prediction is practised with the split vector quantization supported by a very efficient codebook initialization termed Mixture Gaussian Clustering (MGC).

When compared to FS1016 CELP coder, AFBVQ-CELP reduces the LPC bit rate by 18 bit-per-frame (bpf) at the same spectral distortion. It means the overall bit rate is reduced from 4.8 kbps (FS1016 CELP) to 4.2 kbps. Furthermore, the proposed AFBVQ consistently outperforms the traditional forward LPC VQ by 3 bpf with the same spectral distortion. Subjective listening tests show that with AFBVQ-CELP the LPC bit rate can be further reduced to 8.4 bpf, resulting in 3.94 kbps overall bit rate without compromising the decoded speech quality.

INTRODUCTION

Linear prediction plays a center role in various low and intermediate speech coding algorithms [1]. Usually, a new set of linear predictive coding (LPC) coefficients is determined in every 20 to 30 ms and, after quantization, is transmitted to the decoder as side information. To reduce the degradation of the speech quality caused by direct quantization of LPC coefficients, Line Spectral Pairs (LSP) parameters are used for an indirect quantization and interpolation of predictor coefficients. Traditionally, scalar quantization of the

LSP coefficients was used. For example, in the FS1016 Federal Standard Code Excited Linear Predictive (CELP) coder [2], a total of ten LSP coefficients are scalarly quantized to 34 bits-per-frame (bpf). To further reduce the bit rate, the vector quantization schemes for quantization of LSP coefficients were also proposed [3, 4].

In virtually all published CELP algorithms, the predictor coefficients are determined based on the current speech block by using the so-called forward linear prediction. The disadvantages of forward linear prediction are: i) exclusive transmission of predictor coefficients, increasing the required bandwidth; and ii) extensive data buffering, yielding large coding delays. As opposed to forward linear prediction, backward linear prediction requires neither transmission of predictor coefficients nor data buffering. However, the quality of forward linear prediction is usually superior to backward linear prediction.

As is well known, the speech signal is often slowly time-varying and nonstationary. The statistics between the current block and some temporally close previous blocks may often be similar, leading to close sets of predictor coefficients. A method termed Long History Quantization (LHQ) was proposed based on this idea (see Xydeas and So [5]). By allowing previous blocks to be overlapped, the chance for statistical matching between the current block and one of the so-constructed temporally close previous blocks will surely increase. By adaptation of quantizer design to this new strategy, the "global" statistical correlation of speech signals will be more thoroughly exploited and a significant bit rate decrease is expected. In our novel *adaptive forward-backward LPC quantization* (AFBQ) a previously decoded and temporally close speech signal is segmented into overlapping blocks (see Fig. 1). If, and only if, the LPC coefficients calculated from one of those synthetic blocks is sufficiently "close" in some sense to the *unquantized* LPC coefficients calculated from the current speech block, the backward LPC scheme shall be applied, i.e., the LPC coefficients based on the previously decoded optimal speech block are used to encode the current block and only the time delay shall be transmitted.

ADAPTIVE-FORWARD BACKWARD QUANTIZATION

As usual, the input speech x_n is divided into non-overlapping blocks of M samples. For each block, p LPC coefficients $a_1, \ldots, a_p$ are determined by using, e.g., the Levinson-Durbin algorithm.

First, we define the *adaptive forward-backward LPC codebook.* It consists of S code vectors, each having p entries with p being the order of linear prediction. The ith code vector of *adaptive forward-backward LPC codebook* is determined by calculating the LPC coefficients, i.e., $\hat{a}_1^{(i)}, \ldots, \hat{a}_p^{(i)}$, based upon the previously decoded (synthetic) speech block $[y_{n-iK-M}, y_{n-iK-M+1}, \ldots, y_{n-iK-1}]$ that is available at both the encoder and decoder, where M is the length of the LPC block and K is the time delay chosen to be $K = M/4$. Note that only the $M/K = 4$ oldest code vectors of the adaptive LPC codebook need to be updated as illustrated in Fig. 1.

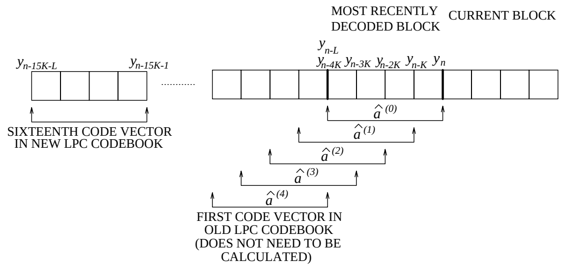

Figure 1. Adaptive forward-backward LPC codebook update scheme.

We then use logarithmic spectral distortion (LSD) [6] to evaluate similarity between the current and previous sets of LPC coefficients. The LSD measure is determined for every candidate code vector in the adaptive forward-backward LPC codebook indexed by $i = 0, \ldots, S-1$. The one that has the smallest spectral distortion, i.e., $\mathrm{LSD}^{(index)}$ with $index = \arg\min_i \mathrm{LSD}^{(i)}$, is selected. If $\mathrm{LSD}^{(index)} > T$, a predefined threshold, then the current LPC coefficients, i.e., $a_1, \ldots, a_p$, are used in speech coding and, after quantization, transmitted to the decoder. If $\mathrm{LSD}^{(index)} \leq T$, then the corresponding LPC coefficients, i.e., $\hat{a}_1^{(index)}, \ldots, \hat{a}_p^{(index)}$, are used in speech coding and only the $index$ to the adaptive LPC codebook needs to be transmitted to the decoder. An additional flag bit is required to notify the decoder whether forward or backward linear prediction is applied at the encoder.

VECTOR QUANTIZATION

Adaptive forward-backward LPC quantization can be implemented as either scalar (AFBSQ) or vector (AFBVQ) quantization. What follows is a description of the design of the vector quantizer. We use split vector quantization [4] of LSP coefficients. After obtaining the initial codebook, the well-known LBG algorithm [3] is used for improvement. In this research, we use a new codebook initialization method termed Mixture Gaussian Clustering (MGC) [7] which represents a merging process. The merging process increases the within-cluster dispersion. Two clusters are merged if, and only if, the increase of within-cluster dispersion is kept at a minimum. Since the total dispersion of training data is equal to the sum of within-cluster dispersion and between-cluster distance, the above merging rule maximizes the decrease of between-cluster distance. Suppose that a pair of clusters j and k containing L_j and L_k training vectors respectively are merged. Then the increase of within-cluster dispersion is derived [7] as

$$\Delta T(j,k) = \frac{L_j L_k}{(L_j + L_k)L} ||\mu_j - \mu_k||^2,$$

where μ_j and μ_k represent the mean vector of clusters j and k respectively, and L is the total number of training vectors in the training set. The equation is used as the distance measure D_{MGC} for selecting the pairs in the merging process. The algorithm starts from a large number of clusters N and successively merges pairs of clusters for which D_{MGC} is the minimum. The algorithm terminates when the desired number of clusters (code vectors) is reached. To make the algorithm computationally more efficient, we apply the well-known pruning algorithm [3] to reduce the number of clusters prior to MGC.

T [dB]	LPC Rate [bpf]	Overall Rate [bps]	segSNR [dB]	LSD [dB]
0	34.0	4800	10.85	1.53
3.0	24.8	4493	10.28	1.81
3.5	21.7	4391	10.16	2.01
4.0	18.8	4290	10.12	2.25
4.5	16.3	4210	10.05	2.52
5.0	14.1	4137	10.01	2.80
5.5	12.3	4076	9.94	3.07
6.0	10.8	4027	9.86	3.33

Table 1. AFBSQ-CELP, $S = 16$ and $M = 240$.

T [dB]	LPC Rate [bpf]	Overall Rate [bps]	segSNR [dB]	LSD [dB]
0	24.0	4467	10.59	1.03
3.0	17.7	4257	10.31	1.45
3.5	15.6	4187	10.27	1.69
4.0	13.6	4120	10.15	1.98
4.5	12.0	4067	10.12	2.25
5.0	10.6	4020	10.04	2.57
5.5	9.4	3980	10.02	2.86
6.0	8.4	3947	9.95	3.14

Table 2. AFBVQ-CELP, $S = 16$, $M = 240$, pruning+MGC initialization.

PERFORMANCE EVALUATION

The proposed AFBQ with both scalar and vector quantization is integrated with the FS1016 Federal Standard CELP coder [2] resulting in AFBSQ-CELP and AFBVQ-CELP, respectively. Both the segmental signal-to-noise ratio (segSNR) and logarithmical spectral distortion (LSD) are used to evaluate the performance of the new coder. The test speech signal contains 600 seconds of speech spoken by both male and female speakers. Based on experiments in [8] the size of the forward backward codebook $S = 16$. Tables 1 and 2 show the performance of AFBSQ-CELP and AFBVQ-CELP, respectively, when the threshold T changes from $T = 3$ dB to $T = 6$ dB in

0.5 dB increments. Fig 2. compares the performance of AFBSQ, AFBVQ and split-VQ reported in [4]. When compared to AFBSQ, at a given bit rate AFBVQ decreases LSD by 0.72–0.85 dB. As also shown in Fig. 2, AFBVQ gives 0.2–0.3 dB smaller LSD than forward split VQ. This is equivalent to reducing the LPC bit rate by 3–4 bpf having the same spectral distortion. Fig. 3 compares the performance of different vector quantization codebook initializations. Pruning+MGC initialization outperforms random initialization by 2 bpf. Finally, computer experiments show [8] that the proposed AFBQ outperforms LHQ by 1bpf. The overall bit rate of FS1016 CELP is reduced to 4.2 kbps by AFBVQ. Subjective listening test show that the bit rate can be further reduced to 8.4 bpf (3.94 kbps overall bit rate) without compromising the decoded speech quality.

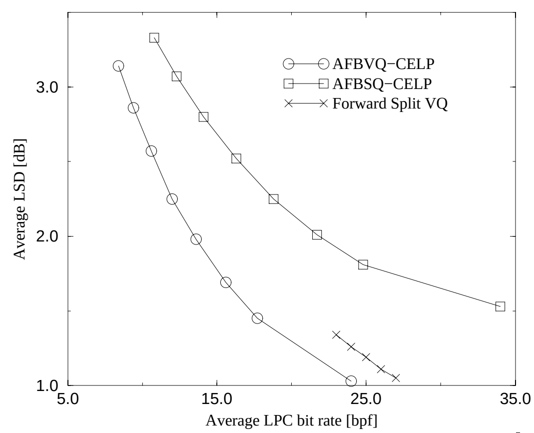

Figure 2. Comparison of AFBSQ, AFBVQ and split VQ [4] in terms of LSD.

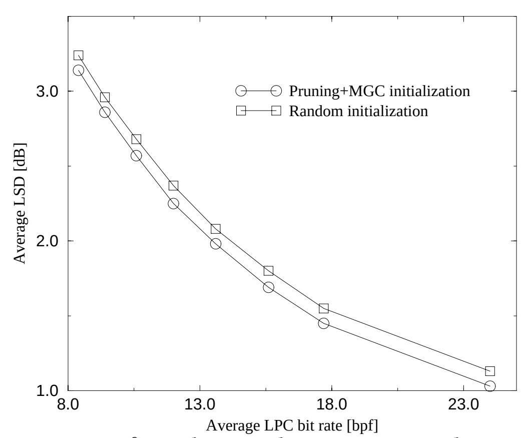

Figure 3. Comparison of random and pruning initialization in terms of LSD.

CONCLUSIONS

In this paper a new variable rate LPC quantization scheme was proposed and integrated with the FS1016 CELP coder. AFBQ scheme can also be applied to other speech coding algorithms for which exclusively forward linear prediction is used. As mentioned above, the bit rate in AFBQ can be easily controlled by deciding the between-block similarity in terms of the threshold T. This would further provide a valuable feature with most cellular mobile applications.

References

[1] A. Gersho, "Advances in speech and audio compression," *Proceedings of IEEE*, vol. 82, pp. 900–918, 1994.

[2] J.P. Campbell, V.C. Welch and T.E. Tremain, "An expandable error-protected 4800 BPS CELP coder (U.S. Federal Standard 4800 BPS voice coder)," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1989, pp. 735–738.

[3] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Norwell, MA, 1992.

[4] K.K. Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 1, pp. 3–14, Jan. 1993.

[5] C.S. Xydeas and K.K.M. So, "A long history quantization approach to scalar and vector quantization of LSP coefficients," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1993, vol. 2, pp. 1–4.

[6] A.H. Gray and J.D. Markel, "Distance measures for speech processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, pp. 380–391, Oct. 1976.

[7] Y. Zhao, "A speaker-independent continuous speech recognition system using continuous mixture density HMM of phoneme-sized units," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 3, pp. 345–361, 1993.

[8] J. Vass, Y. Zhao, and X. Zhuang, "Adaptive forward-backward quantizer for low bit rate high quality speech coding," *to appear in IEEE Transactions on Speech and Audio Processing*.