TOWARDS ILBC SPEECH CODING AT LOWER RATES THROUGH A NEW FORMULATION OF THE START STATE SEARCH

Christopher M. Garrido[†]*, Manohar N. Murthi*[†]*, and Søren Vang Andersen*[‡]

[†] Dept. of Electrical and Computer Engineering
University of Miami, USA
c.garrido@umiami.edu, mmurthi@miami.edu

[‡] Department of Communication Technology
Aalborg University, Denmark
sva@kom.auc.dk

ABSTRACT

The Internet Low Bit-rate Coder (iLBC) has emerged as a candidate for Voice over Internet Protocol (VoIP) applications. By avoiding the inter-frame coding dependencies endemic to many speech coders such as those based on Code Excited Linear Prediction, iLBC is able to achieve superior robustness to packet loss. In addition to robustness to packet loss, a VoIP codec should possess the agility to adjust its source coding rate in order to react to network congestion and to be amenable to joint source channel coding for wireless channels. Towards this end, we develop a new formulation of the iLBC encoding process that allows for a variable rate iLBC. In particular, we demonstrate how the LP excitation signal is constructed from a much shorter vector of 'start state' samples through a non-square synthesis matrix that captures the effects of the Adaptive Codebook operations. With this new framework, the search and quantization of the start state is re-formulated as an Analysis by Synthesis matching problem. We demonstrate how a Multi-Pulse (MP) approach can be utilized to effect a variable rate coding solution for this new framework. A variable rate coder with the MP start state achieves better performance than the Adaptive Multi-Rate (AMR) coder at 12.2 and 10.2 kbps for packet loss rates greater than 4 %.

1. INTRODUCTION

Voice over Internet Protocol (VoIP) has grown in importance in recent years, requiring speech coders that possess the robustness and flexibility to effect high quality communication. Most importantly, voice codecs need to exhibit robustness to packet loss which is endemic in real-time packet-switched communication. Furthermore, coders should have the added flexibility of variable source rates. With the added flexibility of variable rates, a speech coder is amenable to Joint Source-Channel Coding (JSCC)[1] which provides better voice Quality of Service (QoS) over wireless links. In addition, a speech coder featuring variable source coding rates is better able to effect TCP friendly rate/congestion control[2] which is important for the co-existence of heterogeneous applications in the Internet.

To address these performance requirements, differing approaches can be taken. One common trend is to retrofit existing codecs such as Adaptive Multi-Rate (AMR)[3], and G.729A[4] with additional capabilities. Another trend is to provide new codecs specifically designed for VoIP, such as the Internet Low Bit-rate Codec (iLBC)[5]. In contrast to both AMR and G.729A which possess inter-frame coding dependencies inherent in Code-Excited Linear Prediction (CELP) principles, iLBC is an intra-frame codec that produces voice packets that can be decoded independently. iLBC rests upon quantizing a short segment of samples, termed the 'start state', that is used to build-up the rest of the frame's samples through application of the Adaptive Codebook (ACB) both forward and backward in time. By confining the ACB operation to a single frame, iLBC achieves greater robustness to packet loss at a fixed bit-rate.

To provide iLBC with the additional agility to grapple with time-varying wired and wireless networks, we present a new method for quantization of the start state that allows for variable source coding rates. In particular, the effect of the ACB in both directions is first captured in a non-square synthesis matrix that is 'tall and skinny'. The synthesis matrix multiplies the start state (a short vector) to build-up the target residual signal (a long vector). Thus, the search and quantization of the start state can be viewed as an Analysis by Synthesis (AbS) matching problem, and we demonstrate how Multi-Pulse principles[6,7] can be utilized to effect a solution. By varying the number of pulses used to approximate the start state, a variable bit rate is achieved. In combination with entropy coding techniques, the new start-state encoding method achieves the same performance as the original iLBC but allows for flexible rate reduction. In comparison to AMR, the modified iLBC with the MP-based start state provides superior robustness to packet loss for most loss rates.

The remainder of the paper is organized as follows. Section 2 provides an overview of the original iLBC algorithm.

The work of M.N. Murthi was supported in part by the National Science Foundation via CAREER Award CCF-0347229.

Section 3 describes the new AbS formulation for start state quantization. Section 4 describes the quantization of pulses and positions. Section 5 presents some performance results, and Section 6 concludes the paper.

2. STANDARD ILBC SPEECH CODEC

iLBC is a narrowband speech codec operating at 15.2kbps with 20 ms frames, and 13.3kbps with 30ms frames. iLBC is essentially a linear predictive coder utilizing block based coding of the linear prediction (LP) residual signal through a combination of scalar quantization and adaptive codebooks. After the computation of the LP residual signal which we denote by the $M \times 1$ vector $\mathbf{t}_{res}$ (with $M = 160$ for 20ms frames and $M = 240$ for 30ms frames), the start state is identified through a constrained search of $\mathbf{t}_{res}$. We denote the start state as $\mathbf{v}_{ss}$ which is an $N \times 1$ vector (with $N = 57$ for 20ms and 58 for 30ms) of contiguous high energy samples within the residual signal vector $\mathbf{t}_{res}$. Then each term of the start state vector $\mathbf{v}_{ss}$ is scalar quantized at 3 bits/sample. It is vital to underscore the importance of $\mathbf{v}_{ss}$ and its ability to capture a good representation of periodicity in voiced speech or high energy noise in unvoiced speech which is used to exploit long-term redundancies in the residual signal.

The remainder of the residual signal $\mathbf{t}_{res}$ is quantized through ACB operations. First, the initial ACB memory is populated with the quantized start state samples $\mathbf{v}_{ss}$. Then the ACB is run both forwards and backwards in time, with 3 ACB stages per 40 sample subframe. In this manner, the short start state vector $\mathbf{v}_{ss}$ builds up the rest of the larger residual signal vector $\mathbf{t}_{res}$ through the ACB operations. At the decoder, the location and quantization information for $\mathbf{v}_{ss}$ and the ACB gains and lags are utilized to reconstruct an approximation of $\mathbf{t}_{res}$ which is then used as an excitation to LP synthesis filters to produce speech.

By avoiding the use of previous frames' samples in the initialization of the ACB memory, iLBC is able to achieve frame independence which leads to robustness to packet loss at the cost of start state quantization. In terms of extending the capabilities of iLBC, one could relax the frame independence assumption and allow for occasional use of inter-frame coding, leading to a system comparable to the one proposed in [8] for CELP-based G.729A. In this paper, we take a different approach, concentrating on the quantization of the start state.

3. START STATE SEARCH AND QUANTIZATION AS ANALYSIS BY SYNTHESIS MATCHING

In the current iLBC coder, the quantization of $\mathbf{v}_{ss}$ results in a total of $3N$ bits. This represents a large percentage of the total encoded payload. We now demonstrate how we can represent $\mathbf{v}_{ss}$ in a more efficient manner that allows for variable rate source coding without many changes to the existing iLBC decoders.

Our approach is based on the observation that the short start state vector $\mathbf{v}_{ss}$ is used to build up the much larger residual target vector $\mathbf{t}_{res}$. First, we can analyze how the ACB encoding operations, initially based solely on $\mathbf{v}_{ss}$, eventually approximate $\mathbf{t}_{res}$, capturing this relationship in a synthesis matrix. Thus, one can form a non-square $M \times N$ synthesis matrix $\mathbf{H}$ which describes a linear mapping between the samples in $\mathbf{v}_{ss}$ and the residual target signal $\mathbf{t}_{res}$, resulting in the system of equations

$$\mathbf{t}_{res} \approx \mathbf{H}\mathbf{v}_{ss} \tag{1}$$

Thus, the 'tall' synthesis matrix $\mathbf{H}$ captures the ACB operations that allow the short $N \times 1$ vector $\mathbf{v}_{ss}$ to build up the much larger $M \times 1$ vector $\mathbf{t}_{res}$.

To form $\mathbf{H}$, the iLBC encoder is run as before, with the start state and all the ACB lags and gains identified. To compute the k^{th} column of $\mathbf{H}$, one first creates an artificial start state vector consisting entirely of zeros except for a 1 in the k^{th} element of the vector, essentially a unit impulse vector. With an initial codebook consisting of the unit impulse vector, the ACB decoding operations are run with the existing gains and lags, resulting in an $M \times 1$ vector which is the k^{th} column of $\mathbf{H}$. Thus the k^{th} column of $\mathbf{H}$ determines how the k^{th} sample of $\mathbf{v}_{ss}$ contributes to the build up of $\mathbf{t}_{res}$.

With Eq.(1), one can perform more judicious quantization of the start state vector $\mathbf{v}_{ss}$, possibly abandoning the current scalar quantization method of allocating an equal number of bits to all elements of $\mathbf{v}_{ss}$. One can take this synthesis matrix formulation a step further by placing it in the perceptually weighted domain. By using a standard LP-based perceptual weighting filter, one can define $\tilde{\mathbf{H}}$ to be the synthesis matrix that includes the effect of the perceptual weighting, and $\tilde{\mathbf{t}}_{pw}$ to be the perceptually weighted target. Then one approach to representing the start state is to find an approximation vector $\hat{\mathbf{v}}$ that minimizes

$$\|\tilde{\mathbf{t}}_{pw} - \tilde{\mathbf{H}}\hat{\mathbf{v}}\|^2. \tag{2}$$

This is an Analysis-by-Synthesis (AbS) matching problem for determining a short vector $\hat{\mathbf{v}}$ that best synthesizes $\mathbf{t}_{res}$ in a perceptually weighted sense.

Although many differing solution methods are possible within this framework, in this paper we focus on utilizing the well-known principles of Multi-Pulse excitations[6,7] to determine an approximation vector $\hat{\mathbf{v}}$ to the start state. That is, $\hat{\mathbf{v}}$ consists of P pulse locations with P corresponding gains, and zeros elsewhere. By utilizing Multi-Pulse principles, the AbS matching problem is of a reasonable complexity as the P ($1 \leq P \leq N$) pulse locations must be chosen from only N possible locations (with $N = 57$ or 58) to match a target of dimension $M = 160$ or 240. While the

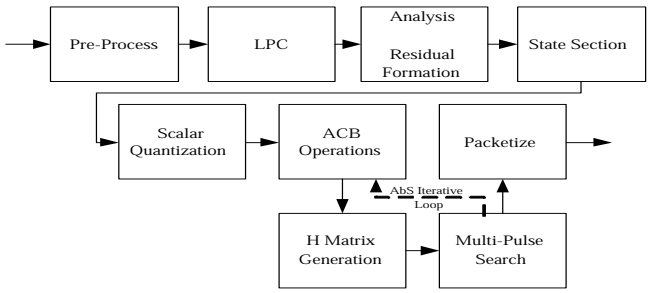

Fig. 1. Modified iLBC encoder.

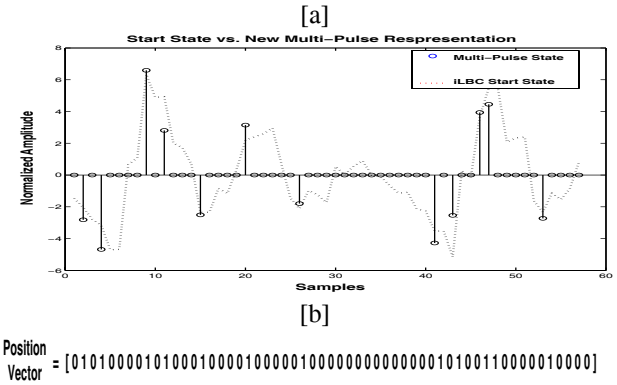

Fig. 2. (a) Original and Multi-Pulse($P = 12$) start states, (b) Pulse Position Vector

matching is principally in dimension M, the search operations can be simplified by using standard procedures such as backfiltered targets resulting in search computations on vectors of dimension N only. The number of pulses P can be varied to ultimately effect a variable rate coding scheme.

Thus, the P pulses are found sequentially with the existing gains re-optimized after each pulse is found as this provides better performance when P is large. Once all P pulses of $\hat{\mathbf{v}}$ are determined, the vector $\hat{\mathbf{v}}$ could be used as a new representation of the start state $\mathbf{v}_{ss}$.

However, the original ACB lags and gains may no longer be accurate as they describe the build up of the residual $\mathbf{t}_{res}$ with respect to the original start state $\mathbf{v}_{ss}$. Therefore, it is quite advantageous to repeat the procedure. That is, new ACB lags and gains are found based on the initial Multi-Pulse start state $\hat{\mathbf{v}}$, a new synthesis matrix is found, and the Multi-Pulse search is repeated to find a new vector of pulses that forms the final representation of the start state $\hat{\mathbf{v}}_{mp}$. In principle, this procedure can be repeated many more times, but we have found that the additional performance gains are negligible. Once this final multi-pulse vector $\hat{\mathbf{v}}_{mp}$ is determined, its parameters must be quantized for transmission.

Figure 1 shows a block diagram of the new proposed encoder. Figure 2a depicts an example of $\mathbf{v}_{mp}$ (P=12). This vector is simply substituted in place of $\mathbf{v}_{ss}$ in the initialization of the ACB in the decoder, thereby requiring no major changes to the iLBC decoder.

4. PULSE QUANTIZATION

The gains, positions, and number of pulses of $\hat{\mathbf{v}}_{mp}$ must be transmitted to a decoder. For gain quantization, we trained 4-bit scalar quantizers based on a training set extracted from the TIMIT database. To describe the positions of the P pulses, we take different approaches. For very low values of P, ($1 \leq P \leq 9$) the pulse positions are specified using $\lceil \log_2(N) \rceil = 6$ bits for each position. For larger values of P, the pulse positions are described by a dimension N pulse position vector with a '1' signifying the location of the pulse, and a '0' indicating no pulse, resulting in N bits for the locations which is subsequently entropy coded using basic arithmetic coding.

	Number of Pulses (P)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Source Rate (kbps)	7.15	7.65	8.15	8.65	9.15	9.65	10.02	10.35	10.67	10.89	11.23	11.46	11.74	12.02	20 ms
	7.87	8.20	8.53	8.87	9.20	9.53	9.78	10.00	10.22	10.36	10.59	10.74	10.92	11.11	30 ms
	15	16	17	18	19	20	21	22	23	24	25	26	27	28	
Source Rate (kbps)	12.31	12.70	12.90	13.10	13.30	13.50	13.70	13.90	14.10	14.30	14.50	14.70	14.90	15.10	20 ms
	11.31	11.57	11.73	11.87	12.00	12.13	12.27	12.40	12.53	12.67	12.80	12.93	13.07	13.20	30 ms

Table 1. Number of pulses and corresponding rates of modified iLBC

These methods result in the total bit rates illustrated in Table 1. The rates range from 7.15kbps for 20 ms frames and 7.87 kbps for 30 ms frames both only using one pulse to a total of 15.1 kbps for 20ms and 13.2 kbps for 30 ms using the full 28 pulses. Thus by varying the number of pulses used to represent the start state, iLBC is able to achieve a variety of bit rates. At the decoder, the multi-pulse vector is formed and used as the start state which is then processed by the existing iLBC decoding process.

5. PERFORMANCE EVALUATION

As a benchmark for testing, we utilized over one hour of speech, including sentences from 100 male and female speakers, from the TIMIT database. The modified ($P = 28$) iLBC coder operating at the original iLBC coding rate provided the same performance as the standard iLBC coder as measured by the PESQ[9] algorithm. Figure 3 shows a PESQ comparison of the modified iLBC and AMR coders over their respective source rates without any packet loss. Here we see that under ideal conditions AMR scores higher at all rates. AMR has an advantage in that it takes advantage of interframe coding in contrast to the modified iLBC

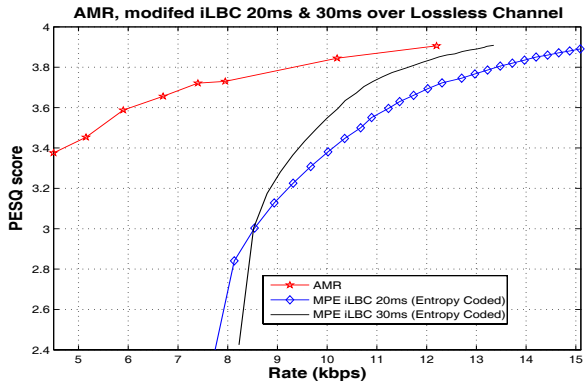

Fig. 3. Comparison of AMR and modified iLBC for differing rates and no packet loss

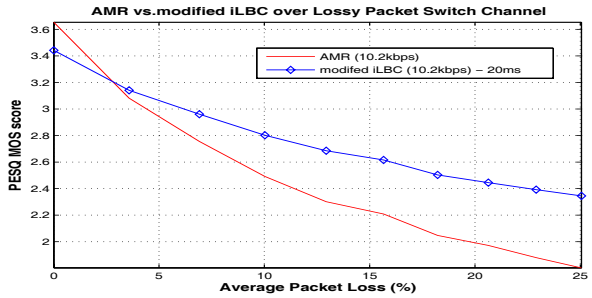

Fig. 4. PESQ Comparison of modified iLBC at 10.35 kbps and AMR at 10.2kbps transmitted over a packet loss channel. Modified iLBC is better for average loss above 4 % .

coder. Nevertheless, the modified iLBC coder provides only a 0.1 PESQ loss for a savings of 2 kbps as compared to the regular iLBC. In Figures 4 and 5, modified iLBC and AMR are compared in terms of robustness to packet loss (simulated using a Gilbert model) for source rates of 10.2 and 12.2kbps respectively. As one can see, iLBC performs better than AMR for most packet loss conditions.

6. CONCLUSIONS

Through re-formulating the iLBC start state search and quantization process as an Analysis by Synthesis matching problem in which a synthesis matrix captures the build-up of the LP excitation signal from a much shorter vector of start state samples, we allow for more judicious quantization of the start state. By utilizing a Multi-Pulse approach within this AbS matching framework, we demonstrate how the iLBC coder can be provided with additional rate flexibility. In comparisons with the AMR coder, this variable rate iLBC provides superior performance for most packet loss rates, though AMR provides superior performance for zero packet loss. Thus, the variable rate iLBC provides greater performance flexibility than the standard iLBC coder without re-

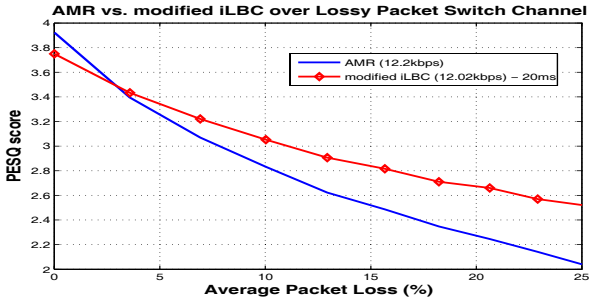

Fig. 5. Comparison of modified iLBC at 12.02 kbps and AMR at 12.2 kbps transmitted over a packet loss channel. Modified iLBC is better for average loss above 4 % .

quiring much change to the decoding architecture. However, additional gains in performance can be achieved within this AbS matching framework by further modifications to the existing iLBC decoder. For example, the number of ACB stages per subframe can be reduced to provide more bits to the start state quantization (which allows for more pulses within the Multi-Pulse solution). Furthermore, the length of the start state vector can be reduced when the number of pulses is limited, allowing the ACB a better ability to build out the rest of the frame. Therefore, other solution methods and refinements within this AbS matching framework are possible and could lead to better performance for frame independent predictive coding at lower source coding rates.

7. REFERENCES

[1] J. Hagenauer, and T. Stockhammer, "Channel coding and transmission aspects for wireless multimedia," In *Proceedings of the IEEE*, vol. 87, no. 10, Oct. 1999.

[2] S. Floyd, M. Handley, and J. Padhye, "Equation-Based Congestion Control for Unicast Applications: the Extended Version" ICSI Report Number TR-00-003, March 2000

[3] Ekudden, E., R. Hagen, I. Johansson, and J. Svedberg, "The adaptive multi-rate speech coder" In *1999 IEEE Speech Coding Workshop Proceedings*, pp. 117-119.

[4] R. Salami, et al, "Design and Description of CS-ACELP: A Toll Quality 8 kb/s Speech Coder" In *IEEE Transactions on Speech and Audio Processing*, March 1998.

[5] Andersen, S.V., W.B. Kleijn, R. Hagen, J. Linden, M.N. Murthi, and J. Skoglund, "iLBC-A Linear Predictive Coder with Robustness to Packet Losses," In *2002 IEEE Speech Coding Workshop Proceedings*, pp.23-25.

[6] B.S. Atal "High-quality speech at low bit rates: multi-pulse and stochastically excited linear predictive coders" In *Proceedings of ICASSP 1986*

[7] A.M. Kondoz, *Digital Speech: Coding for Low Bit Rate Communication Systems*, Wiley

[8] R. Lefebvre, P. Gournay, and R. Salami, "A Study of Design Compromises for Speech Coders in Packet Networks," In *Proceedings of ICASSP 2004*

[9] ITU-T P.862 "Perceptual Evaluation of Speech Quality (PESQ)."