

# ILBC - A LINEAR PREDICTIVE CODER WITH ROBUSTNESS TO PACKET LOSSES

*S. V. Andersen<sup>†</sup> W. B. Kleijn<sup>††</sup> R. Hagen<sup>††</sup> J. Linden<sup>††</sup> M. N. Murthi<sup>††</sup> J. Skoglund<sup>††</sup>*

<sup>†</sup>Department of Communication Technology  
Aalborg University, Denmark

<sup>††</sup>Global IP Sound  
Stockholm, Sweden and San Francisco, USA

## ABSTRACT

In this paper, we discuss the internet low bit rate codec (iLBC) with an emphasis on the frame-independent long-term prediction. The frame-independent long-term prediction is a method to exploit pitch-lag correlations in the encoding of speech without suffering multiple-frame speech degradation in connection with transmission loss. We present mean opinion scores for the iLBC codec and show by means of signal examples how the nature of degradation in a predictive codec based on frame-independent long-term prediction differs from that of traditional CELP codecs.

## 1. INTRODUCTION

This paper describes an algorithm for the coding of narrow-band speech at 13.867 kbit/s. The algorithm exploits long-term linear predictive coding without introducing data dependency across frame boundaries. This results in a codec suitable for transmission of speech over lossy packet networks. The described algorithm is an integral part of the internet low bit rate codec (iLBC) developed by Global IP Sound. This codec has recently been handed to the internet engineering task force (IETF) where it is housed in the audio/video transport (AVT) working group [1].

The described algorithm results in a speech coding system with a controlled response to packet losses similar to what is known from pulse code modulation (PCM) with packet loss concealment (PLC), such as the ITU-T G.711 standard, which operates at a fixed bit rate of 64 kbit/s. At the same time, the described algorithm enables fixed bit rate coding with a quality-versus-bit-rate tradeoff close to what is known from code-excited linear prediction (CELP) [2].

The remainder of this paper is organized as follows. Section 2 describes the general idea of a frame-independent predictive coding that exploits long-term correlation. In Section 3 the important encoding of the start state for the frame-independent predictive coding is discussed. Sections 4 and 5 gives specifics for the internet low bit rate codec with which the experimental results were obtained. Finally, results are given in Section 6 and conclusions are drawn in Section 7.

## 2. FRAME-INDEPENDENT LONG-TERM PREDICTIVE CODING

The essence of the codec is linear predictive coding (LPC) and block based coding of the LPC residual signal using an adaptive codebook. The common practice in CELP coders

is to populate the adaptive codebook with excitation signal prior in time. That approach holds a number of problems in it. The following states a few of these problems.

- If the past signal is lost or error contaminated during the transmission, the adaptive codebook in the decoder will differ from the one in the encoder. This leads to a poor decoded signal quality.
- At the onset of a voiced speech segment, the adaptive codebook is insufficient to properly describe the pitch cycle. This leads to a slow voicing onset in the decoded signal.
- The desired fast build up of high periodicity of voiced regions, and the undesired feedback of coding noise to the adaptive codebook are conflicting performance goals in CELP [3, 4]. In practice, this means that CELP encoding results in a decoded signal with a noisy character in voiced regions. This noisy character is perceived especially for high-pitched voices.

In contrast to CELP, our codec applies the adaptive codebook both forward and backward in time, starting from a segment inside the speech frame, which we denote as the start state vector. The start state can be determined as a segment of samples with the highest residual energy. This segment will typically capture at least one dominating pitch pulse when the segment has voiced speech in it. We consider this signal as the start state for the long-term predictive coding. The location and waveform of the start state is encoded and transmitted for each frame. The adaptive codebook is first populated with segments of the decoded start state. Then the adaptive codebook is used for long-term predictive coding in one time direction, e.g., forward in time for the remaining signal frame from the end of the start state to the end of the signal frame. During this encoding the adaptive codebook is continuously adapted with the most recent decoded signal. Subsequently, the adaptive codebook is populated with segments of the decoded start state and the first encoded signal segment. After this, the adaptive codebook is used for long-term predictive coding in the other time direction, e.g., backward in time for the remaining signal frame from the beginning of the start state to the beginning of the signal frame. In this way, a data packet can be made to contain start state information, and lag and gain information sufficient for the correct decoding of a complete signal frame.

This method implies that the adaptive codebook in the decoder will be independent of the reception or loss of a previous packet. Furthermore, if a voicing onset occurs

within the frame then at least one significant pitch pulse will be contained in the adaptive codebook as a starting point and thereby the first pitch cycle in the frame can be accurately encoded. This scheme no longer has an inherent compromise between fast built-up of high periodicity and the feedback of coding noise to the adaptive codebook. This is because the start state typically will contain a fully built-up pitch pulse. In our experience, once the codebook built-up problem is removed, the adaptive codebook can advantageously be used in all stages of a multi-stage coding configuration, rather than using the adaptive codebook in a single stage refined with a stage that uses a large fixed codebook, which is common practice in CELP.

While the problem of fast built-up of high periodicity, known from other applications of an adaptive codebook, seems resolved in this scheme, we may expect it to be replaced by other voicing related problems. For example, the encoding of voiced regions now gets a high sensitivity to the coding noise in the encoding of the start state. Also, whereas the coding noise introduced by a CELP coder in the encoding of a highly periodic speech segment will itself evolve in small steps from one pitch cycle to the next and thereby show a high degree of desired periodicity, the coding noise vector in our proposed scheme may shift abruptly at a frame boundary. In the iLBC codec these problems are resolved by extensive long-term postfiltering [5].

### 3. IDENTIFICATION AND ENCODING OF THE START STATE

For the frame-independent long-term predictive coding the efficient encoding of the start state is essential. As a basis, we apply a simple scale normalization over the entire start state followed by a scalar quantizer with noise feedback to perceptually shape the quantization noise.

For voiced speech, the start state will contain at least one dominating pitch pulse. It is important that this pulse is represented well in the decoded start state. A scalar quantization of an impulse-like signal leads to a poor trade-off between overload and granular noise for the quantizer. Therefore, we preprocess the start state with an all-pass filter to evenly distribute the signal energy between samples in the preprocessed start state. While this all pass filter can possibly be optimized for the given signal, we have instead chosen a simple empiric design that avoids the transmission of side information to specify this all-pass filter to the decoder. We realize the all-pass filter as a circular convolution with the decoded LPC synthesis filter forward in time cascaded with the decoded LPC analysis filter backward in time. The resulting preprocessed residual becomes, except for the circular effect in the convolution, equal to the prediction residual that results if the LPC analysis filter is applied backward in time. Therefore, rather than resulting in a concentrated peak at the beginning of a pitch cycle, this tends to give a wider distribution of the prediction error over the pitch cycle. After decoding of the start state, an inverse all-pass filtering is made to compensate the effect of this preprocessing.

### 4. ENCODER SPECIFICS

Our experiments with the frame-independent long-term predictive coding and the encoding of start states are based on the internet low bit rate codec (iLBC) [1], specifics of which are given in the following.

The codec operates on a 240 sample frame, on which two LPC analyses are obtained: one for a window centered toward the beginning of the frame and one for a window centered toward the end of the frame. The two LPC filters are jointly encoded using line spectral frequencies (LSF). After analysis filtering with a smoothly interpolated LPC analysis filter, the residual frame is divided into 6 subframes, each of length 40. Two consecutive subframes having the highest residual energy are identified. Subsequently the 57 trailing or tailing samples of these two subframes is chosen as the target for start state encoding. The choice between trailing and tailing segments is again made so as to maximize energy. The start state is then encoded using 6 bits for a scale normalization and 3 bits per sample.

Based on the decoded start state, the adaptive codebook is initialized for the encoding of the remaining 23 samples of the two maximum energy subframes. For the remaining subframes forward in time, all available 80 decoded LPC excitation samples are used as initialization. Similarly, all available decoded samples, or maximally 127, are used for the initialization of the adaptive codebook backward in time. In all three situations, the adaptive codebook contains segments of the decoded LPC excitation that are shifted in time by one sample. Additionally the adaptive codebook is populated with vectors obtained as linear combinations of the time shifted LPC excitation vectors. The linear combination coefficients are time-invariant and were trained so as to minimize the summed-squared error when the adaptive codebook is applied. The adaptive codebook is applied in a multi-stage gain-shape configuration: the representation of each subframe is refined in 3 stages.

Since codebook encoding with squared-error matching is known to produce a coded signal of less power than the scalar DPCM coded signal, a gain correction factor is calculated by comparing the power loss in the codebook encoding to the power loss in the scalar DPCM coding. The gain correction factor is quantized with 4 bits and is used to scale down the start state to produce a signal with a smooth power contour over the frame.

The encoding results in a bit-allocation per frame as specified in Table 1.

### 5. DECODER SPECIFICS

The decoder first decodes the start state, then subframes forward in time, and finally subframes backward in time. Before synthesis filtering, the decoder applies a pitch post-filtering in the residual domain [5]. When a packet is not received in time for playback, a packet loss concealment is applied. This concealment also works in LPC excitation domain. Details on the post-filtering and packet loss concealment can be found in [1].

Parameter	Bits
LSF	52
Position of Start State	4
Scale Factor for Start State	6
Scalar quantization of start state	171
Shapes, Short frame	24
Shapes, Subframe 1	24
Shapes, Subframe 2	27
Shapes, Subframe 3	27
Shapes, Subframe 4	27
Gains	50
Gain correction factor	4
Sum	416

Table 1: Bit allocation for codec parameters.

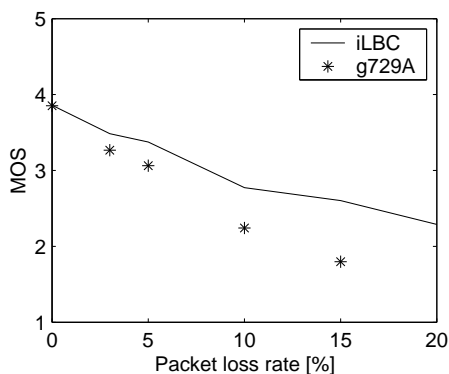


Figure 1: Mean opinion score (MOS) results for the iLBC codec and ITU-T G.729A at packet loss rates between 0% and 20%.

## 6. RESULTS

We compared this method with ITU-T standard G.729A when applied in an IP network with packet loss. Packet loss rates from 0 to 20% were simulated with a basis in loss statistics observed in a real IP network. On the simulated data, a formal mean opinion score (MOS) test was obtained. The formal test was conducted by an independent test lab (Dynastat, Inc.). The resulting MOS scores are plotted in Figure 1. We observe that the iLBC codec results in a speech quality that is not significantly higher than that of ITU-T G.729A when no packets are lost. However, with a significant rate of lost packets, the iLBC codec maintains a significantly higher quality than ITU-T G.729A.

When inspecting the data, we observe the reason for this improved quality. Often in connection with a packet loss within a voiced speech segment, the decoded signal from the iLBC codec suffers minor distortion while the decoded signal from the CELP codec is severely distorted for several frames after the packet was lost. One such example is given in Figure 2. This advantage of the iLBC codec is a direct consequence of the frame-independent long-term prediction.

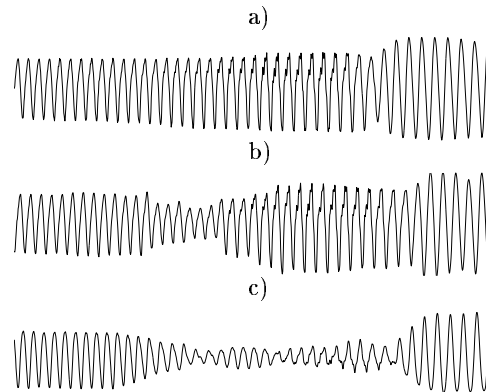


Figure 2: Robustness of the Frame-Independent Long-Term Prediction compared to a standard CELP method, ITU-T standard G.729A: a) original signal segment; b) decoded signal resulting from iLBC after a single IP packet was lost in the transmission; c) same as b) but for G.729A.

## 7. CONCLUSION

The frame-independent long-term prediction supplies a method by which prediction based speech coders can be made more robust against packet losses. This method leads to significantly improved MOS scores. The cost related to this method is that long-term redundancy across frame boundaries is no longer exploited. Instead the encoding of the start state occupy a significant portion of the total bit rate.

In our experiments we thus compared a codec operating at 13.867 kbit/s with the ITU-T G.729A codec which operates at 8 kbit/s. For application in IP networks we find this comparison worthwhile: because of overhead from IP packet headers, and because of packet saturation of IP routers, this difference in bit rate is not guaranteed to result in practical difference in network capacity.

## 8. REFERENCES

- [1] S. V. Andersen *et. al.*, “Internet low bit rate codec,” *IETF internet-draft* <http://search.ietf.org/internet-drafts/draft-andersen-ilbc-00.txt>, Feb. 2002.
- [2] M. Schroeder and B. Atal, “Code-excited linear prediction (CELP): high quality speech at very low bit rates,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (Tampa), pp. 937–940, 1985.
- [3] Y. Shoham, “Constrained-excitation coding of speech,” in *Abstracts IEEE Workshop on Speech Coding for Telecomm.*, (Vancouver), p. 65, 1989.
- [4] W. B. Kleijn, “On the periodicity of speech coded with linear-prediction based analysis-by-synthesis coders,” *IEEE Trans. Speech and Audio Process.*, vol. 2, no. 4, pp. 539–542, 1994.
- [5] W. B. Kleijn, “Enhancement of coded speech by constrained optimization,” *proc. IEEE Speech Coding Workshop*, 2002, submitted for publication.