Dynamically Adding Redundancy for Improved Error Concealment in Packet Voice Coding

Levent Tosun *Peter Kabal*

Department of Electrical and Computer Engineering
McGill University
Montreal, Quebec H3A 2A7

Abstract

This paper presents a method to improve the performance of redundancy-based packet-loss-concealment (PLC) schemes. Many redundancy-based PLC schemes send a fixed amount of extra information about the current packet as part of the subsequent packet, but not every packet is equally important for PLC. We have developed a method to determine the importance of packets and we propose that redundant information should *only* be sent for the important packets. This results in a lower average bit-rate compared to sending a fixed amount of extra information, without sacrificing much from the quality of the concealment. We use a linear prediction (LP) based speech coder (ITU-T G.723.1) as a test platform and we propose that only the excitation parameters should be sent as extra information since LP parameters of a frame can be estimated using the LP parameters of the previous frame.

1 Introduction

Modern speech coders achieve low bit-rates by taking advantage of redundant information found in speech signals. They rely on the assumption that past sections of speech signals provide information about present sections. As long as the bitstream arrives unaltered at the destination, the only concern of a good speech coder is to minimize the bit-rate while keeping the quality sufficiently high. However, with the recent and growing interest in communication over the Internet, the effect of errors (packet loss) occurring in transmission have become a major concern for speech coders.

Data is sent in packets of bits over the Internet. However, packets may not arrive in order or in time for playout. Packet loss is a frequently encountered problem in Voice-over-IP (VoIP) applications. There has been considerable research in this field, proposing several different methods to conceal the effect of lost packets.

This paper is organized as follows. Parametric and hybrid speech coders are discussed in Section 2. In Section 3, packet-loss-concealment (PLC) schemes are explained briefly. The concept of dynamically adding redundancy is introduced in Section 4 where the experimental results are also presented. Finally, conclusions are made in Section 5.

2 Speech Coders

Many parametric speech coders use linear prediction (LP) to model the vocal tract. Speech signals are referred to as quasi-stationary since their characteristics change in time but remain relatively unchanged for a short period of time. Therefore parametric coders operate on a frame basis. They find the parameters to model the vocal tract for each frame and depending on whether the frame is voiced or unvoiced, they use either a periodic signal or white noise to model the excitation signal. Hybrid coders are a combination of waveform and parametric coders — they attempt to find the parameters to model the synthesis of each frame of a speech signal while also providing an excitation signal that minimizes the error in some sense to drive this model. Hybrid coders combine the strengths of waveform and parametric coders, therefore many modern coders are hybrid.

In this research, we used the ITU-T (Telecommunication Standardization Section of International Telecommunication Union) G.723.1 as the test platform [1]. G.723.1 is a dual rate hybrid speech coder designed for multimedia communication. It operates on frames of 30 ms. G.723.1 can use two different methods to generate an excitation signal; algebraic code excited linear prediction (ACELP) and multi-pulse maximum likelihood quantization (MP-MLQ). The former gives a bit-rate of 5.3 kbit/s (158 bits per frame: 24 bits for LP parameters, 134 bits for excitation parameters) whereas the latter gives a bit-rate of 6.3 kbit/s (189 bits per frame: 24 bits for LP parameters, 165 bits for excitation parameters).

3 Packet Loss Concealment Schemes

Speech coders can achieve very low bit-rates by taking advantage of the redundancy in speech signals — they use past information to encode and decode current information. However, speech coding algorithms are not inherently robust to transmission errors [2]. For voice transmission over the Internet, after speech is coded, the bitstream is divided into packets and sent in packets.

Packets experience variable network delays. Real-time voice transmission over the Internet necessitates a limit on the waiting time for the arrival of a packet. A receiver buffer is used to hold packets until their scheduled playout times — the packets which arrive late are considered lost.

The dependence on past frames to decode the current frame introduces the concept of coder state. After the decoding of each packet, some information is saved (state update) to be used in the decoding process of the next packet. This information usually includes past excitation parameters and LP coefficients. In other words, the decoder needs two sources of information to complete its task; information in the current frame and the state information. When a packet loss occurs, due to the dependance of the decoding of a frame to previous frames, the error propagates to subsequent frames [3].

Modern speech coders have PLC schemes to deal with the problem of packet loss. PLC schemes can be categorized in two groups: receiver-based schemes and sender-receiver-based schemes. Receiver-based schemes try to reproduce the

speech segment that a lost packet corresponds to by using the previous and subsequent segments of the speech or replace it with another waveform.

Sender-receiver-based schemes are those which use the transmitter as well as the receiver for PLC. Sender-receiver-based schemes can further be categorized in three groups: priority-based schemes, redundancy-based schemes and interleaving-based schemes. Priority-based schemes assign priority to the packets according to their importance and assume that the packets will be dropped by a supporting network according to the preassigned priorities. Redundancy-based schemes add redundant information at the transmitter about each packet to either the previous or the next packet, which is then used in the receiver in case of a loss. Interleaving-based schemes distribute the information in a packet into several packets, so that when a packet is lost, only part of the information in that packet is gone and the lost information can be recovered using the part of the information that was distributed to other packets.

With random losses, up to 5% losses can be tolerated when using PLC [4]. However, even a single packet loss at a "critical" frame can be quite audible as we will see later.

There are two key features that a good PLC scheme to be used for parametric coders should have — It should be able to reconstruct a reasonable facsimile of the segment of the speech corresponding to the lost packet and it should be able to update the states for the subsequent packet so as to mitigate the effect of the lost packet on succeeding frames.

The PLC scheme works in two steps — concealment of LP coefficients and concealment of excitation parameters. In G.723.1, LP coefficients are converted to LSFs (line spectral frequencies), and it is the LSFs that are differentially coded for transmission. In case of a loss, a mean LSF vector is used, and for this case, the effect of the previous decoded LSFs on the computation of current decoded LSFs is increased.

G.723.1 allocates memory for past excitation parameters. If there are more than 3 consecutive losses, the memory is cleared. Otherwise one of the two methods is applied according to the frame type (voiced / unvoiced). If it is a voiced frame then a periodic excitation is generated using the period that was previously found. If the frame is declared as unvoiced, each excitation parameter is generated randomly by using a randomly generated number and a gain that was calculated previously. The details of the PLC scheme of G.723.1 can be found in [1].

4 Experimental Results

Many sender-receiver-based PLC schemes that rely on adding redundancy, send extra information regardless of how important each packet is. However, if the data in a lost packet is not crucial in updating the states and if the speech segment that it corresponds to can be adequately regenerated, then we do not need to send extra information about that packet. In other words, the decision as to whether or not to send extra information about a packet should be made depending on how important that packet is for reconstruction. We will describe this as dynamically adding redundancy.

4.1 Importance of Certain Packets

To illustrate that certain packets are much more important than others, the following experiment was carried out. Twenty-two speech files, consisting of 11 different sentence groups each recorded by 11 different female and 11 different male speakers were used. Each speech file consists of 4 sentences and is approximately 10 sec long. After the speech files were decoded, Perceptual Evaluation of Speech Quality (PESQ) was used to evaluate the test results. PESQ is described in ITU-T standard P.862 [5].

In the first part of the experiment, for each file and each mode (6.3 kbit/s, 5.3 kbit/s), the PESQ score is measured under different loss scenarios (each loss scenario is specified by the number and the locations of the lost packets). The standard PLC scheme of G.723.1 is used. The locations of the lost packets are determined randomly; however, consecutive losses are avoided. To find the PESQ score for a speech file under a certain fraction of lost packets, the average of 10 PESQ scores found for 10 different loss scenarios is taken. This procedure is repeated for different fractions of losses and for every file and mode. Since it was observed that the coder had a different performance for male and female speech, separate statistics were kept for male and female speakers. For a specified fraction of losses, an average PESQ score was found by taking the average of the PESQ scores of the speech files recorded by the same gender and coded at the same rate.

In the second part of the experiment, in order to find the most important frames for PLC; one frame at a time is deemed to be lost for each file, the standard PLC scheme of G.723.1 is used and a PESQ score is found in each mode (6.3 kbit/s, 5.3 kbit/s). This gives N PESQ scores for a speech file with N frames. For each file and each mode, the PESQ scores are sorted from smallest to largest. The frames that correspond to the lowest PESQ scores are determined to be the most important frames [1]. To measure the performance of the PLC scheme of G.723.1 under a "worst-case-scenario", the location of the losses are selected from the most important frames as opposed to assigning them randomly as it was done in the first part of the experiment. Consecutive losses are again avoided.

The results obtained for female speech files decoded at 5.3 kbit/s (ACELP mode) are given in Fig 1. The top three curves correspond to PESQ scores obtained for random losses when PLC scheme of G.723.1 is used. The three curves in the bottom correspond to PESQ scores obtained under worst-case-scenario losses. For a given loss scenario, the 3 curves correspond to the maximum PESQ score of the 11 sentences, the average of the PESQ scores and the minimum PESQ score. As it can be seen, average, maximum and minimum PESQ scores obtained for worst-case-scenario losses are much lower than those obtained for random losses. The curves for MP-MLQ (6.3 kbit/s) is a slightly upward shifted version of these curves. The curves for male speech files are slightly upward shifted versions of the curves for female speech files. This experiment shows us that certain packets are indeed much more important than others. For example, the average PESQ score under 5% random losses is 2.94 whereas it is 2.05 for the same fraction of worst-case-scenario losses.

The second point in all the curves corresponds to 1 lost packet for all the speech files. As it can be seen, the drop of the PESQ score from the no loss case is significantly higher for a packet loss at a critical frame than for a random packet loss. A subjective listening test conducted with 10 naive

[1] An assessment of the effect of losses on speech packets was also carried out in [6].

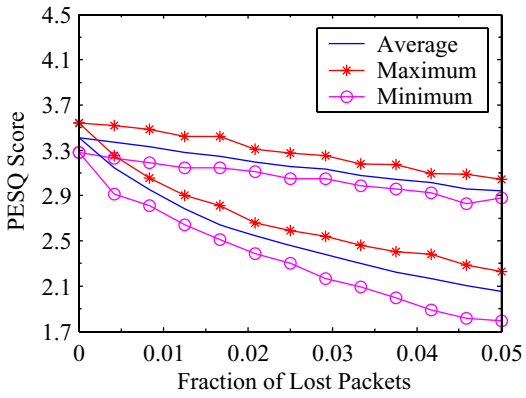

Fig. 1 Illustration of the importance of certain frames. The top 3 curves are for random losses. The bottom 3 curves correspond to "worst-case" losses.

listeners also shows that a single packet loss at a critical frame can be quite audible.

4.2 LP Parameters vs Excitation Parameters

We performed an experiment to figure out if it is the LP parameters or the excitation parameters that the PLC performs poorly to regenerate under worst-case-scenario losses. First LP parameters then the excitation parameters are sent as redundant information to improve the PLC for worst-case-scenario losses. The average PESQ scores are illustrated in Fig 2. The curve at the bottom corresponds to the average PESQ scores obtained for worst-case scenario losses when PLC scheme of G.723.1 is used. The curve in the middle corresponds to sending the LP parameters of the lost packets to improve the PLC. The curve on the top corresponds to sending the excitation parameters. Observing Fig 2, we see that sending the LP parameters as redundant information makes only a small improvement. This is in line with the proposition that LP parameters do not change rapidly from frame to frame and they can be more easily regenerated using past LP parameters. On the other hand, sending the excitation parameters of the most important packets as extra information improves the PLC significantly. Therefore we can conclude that we must consider sending the excitation parameters of the important packets as redundant information.

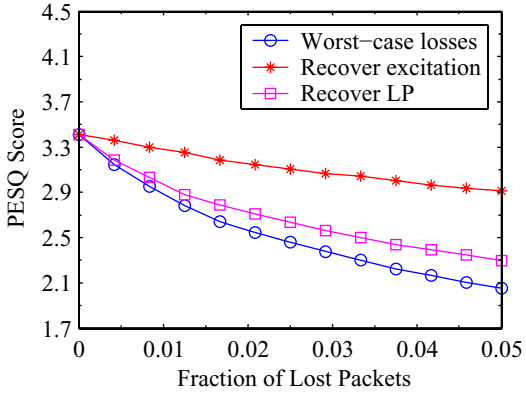

Fig. 2 Illustration of the improvement obtained in PLC by sending excitation parameters as extra information as opposed to LP parameters.

4.3 Using Excitation Parameters in PLC

The excitation parameters that are sent as extra information can either be used only in updating the states, as it was also done in [7], or in the reconstruction of the excitation parameters of the lost frame (the states of the subsequent frame are updated consequently). The latter obviously gives a better performance whereas the former avoids an additional delay.

We performed an experiment to observe the improvements that these two methods provide in the PLC. The results are illustrated in Fig 3. The curve at the bottom corresponds to PESQ scores obtained for worst-case scenario losses when PLC scheme of G.723.1 is used. The curve in the middle corresponds to using the excitation parameters only to update the states to improve PLC and the curve on the top corresponds to using them to regenerate the lost excitation parameters. Figure 3 shows that using the excitation parameters that are sent as redundant information only in updating the states makes only a small improvement in the PLC — as it was also noted in [7]. Therefore we can conclude that when an important packet is lost, its excitation parameters should be reconstructed using the extra information, in which case the states of the subsequent frame are updated consequently.

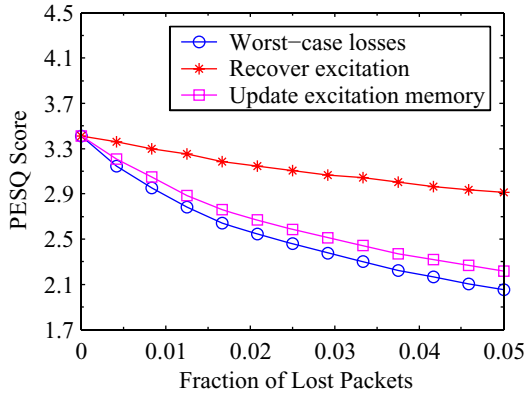

Fig. 3 Comparison of using excitation parameters in the PLC to using them only to update the states

The 7^{th} and 13^{th} points in Figures 1, 2 and 3 correspond to 2.5% and 5% losses. For these conditions, subjective listening tests were conducted using 10 naive listeners. The relative ordering predicted by PESQ scores was confirmed. The conclusions we made based on PESQ scores were verified with these subjective tests.

4.4 Determining Packet Importance

To figure out why the excitation parameters corresponding to certain frames are more important than others, we compared the excitation signals of the important frames to the excitation signals of the frames right before them and excitation signals generated by the PLC scheme of G.723.1 when they are lost. An example is given in Fig 4. The excitation signal of an important frame corresponds to a voiced frame. The previous frame's excitation signal, on the other hand, indicates that it is an unvoiced frame. We can also observe that the excitation signal generated by the PLC scheme of G.723.1 resembles the excitation signal of the previous frame. The PLC scheme of G.723.1 uses past excitation parameters to regenerate the excitation parameters of a lost frame. Therefore, when a voiced frame following an unvoiced one is

lost, the PLC scheme of G.723.1 performs poorly to regenerate the lost excitation parameters. This experiment shows that voiced frames following an unvoiced frame are important in the sense that their excitation parameters cannot be adequately reconstructed in case of a loss. It also reinforces the conclusion that excitation parameters of the important frames should be sent as extra information and be used in the regeneration of the lost excitation parameters.

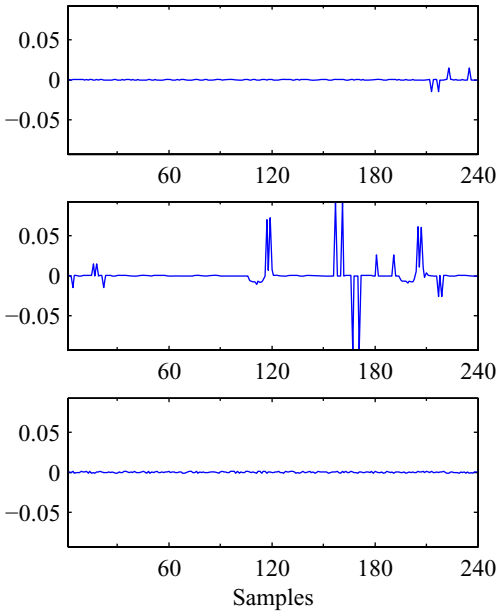

Fig. 4 Comparison of excitation signal of an important frame (the graph in the middle) with the excitation signal of the previous frame (the graph on top) and the excitation signal generated by the PLC scheme of G.723.1 when it is the only lost packet (graph at the bottom)

It is easy to see that the main difference between the excitation signal of an important frame and that of the previous one is the lack of peaks in the latter. Figure 4 shows the excitation signal of a frame that was determined to be the most important one in the first experiment. The difference between the excitation signal of an important frame and that of the previous frame is not always as obvious as in this example. However, it is observed that the energy of the peaks of excitation signals of important frames is significantly larger than the energy of the peaks of excitation signals of the previous frames.

We have developed a method to determine the importance of packets. We calculate the ratios of the average peak magnitude and the rms of the excitation signal of a frame to those of the previous one. The frame is determined to be important if either one of the ratios is greater than 5. Using this method, on average, 11% of the packets are determined to be important. This method for selecting important frames chooses frames, which to a large extent coincide with those determined to be the most important based on PESQ scores.

4.5 New Redundancy-Based PLC Scheme

If the excitation parameters of a frame are determined to be important for PLC, they are sent with the subsequent packet. Sending excitation parameters twice for the important packets results in an *average* bit-rate of 6.9 kbit/s for MP-MLQ and 5.8 kbit/s for ACELP. On the other hand, sending excitation parameters twice for each and every packet results in a fixed bit-rate of 11.8 kbit/s for MP-MLQ and 9.7 kbit/s for ACELP. Hence, the method provides an improved PLC at a modest increase in the average bit-rate. Table 1 summarizes these results.

Table 1 The comparison of sending extra information for important packets to sending them for every packet

Coder Mode	Standard PLC (G.723.1)	EXC par. (imp. frames)	EXC par. (each frame)
ACELP	5.3 kbit/s	5.8 kbit/s	9.7 kbit/s
MP-MLQ	6.3 kbit/s	6.9 kbit/s	11.8 kbit/s

With the improvement that this method provides on the PLC, the effect of the worst-case-scenario losses is reduced to that of random losses. In other words, 5% worst-case-scenario losses can be tolerated.

5 Conclusion

In this paper we showed that excitation parameters should be sent as extra information for redundancy-based PLC schemes and that they should be used both in the regeneration of the excitation parameters of the lost packet and in updating the states. We further showed that it is not necessary to send the excitation parameters of each and every packet as extra information because not all the packets have the same importance for PLC. We have developed a method to determine the important packets. We showed that duplicating the excitation parameters of the packets that are determined to be important according to our method results in an 11% increase in the bit-rate on average, which is much lower compared to the increase that would be obtained by duplicating the excitation parameters of all the frames.

References

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

[2] B. Wah, X. Su, and D. Lin. A Survey of Error-Concealmant Schemes for Real-Time Audio and Video Transmissions over the Internet. *Proc. IEEE Int. Symp. Multimedia Software Engineering*, pp. 17–24, Dec. 2000.

[3] A. Shah, S. Atungsiri, A. Kondoz and B. Evans. Lossy multiplexing of low bit rate speech in thin route telephony. *IEE Electronic Letters*, vol. 32, pp. 95–97, Jan. 1996.

[4] N. Jayant. Effects of packet losses on waveform coded speech. *Proc. Int. Conf. Computer Commun.*, pp. 275–280, Oct. 1980.

[5] ITU-T, *Recommendation P.862, Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, Feb. 2001.

[6] C. Hoene, B. Rathke and A. Wolisz. On the importance of a VoIP packet. *Proc. ISCA Workshop Auditory Quality of Systems*, pp. 55–62, Apr. 2003.

[7] R. Lefebvre, P. Gournay and R. Salami. A study of design compromises for speech coders in packet networks. *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing.*, vol. 1, pp. 265–268, May. 2004.