Comparison Between Silence Substitution and Packet Repetition for Real-Time Speech Communications

M. Sarwar Ehsan and *G. Kubin*
Signal Processing and Speech Communication Lab.
Graz University of Technology, Austria
email: `sarwar@sbox.tugraz.at`
`g.kubin@ieee.org`

Abstract

Real-time speech communication over packet switched networks requires low delay packet loss concealment (PLC) methods. There are several PLC methods used in IP telephony to cope with packet losses. Two commonly used methods are silence substitution and packet (waveform) repetition. We compare these methods according to the rate distortion criterion by introducing a penalty for packet (waveform) repetition. This analysis allows us for fair comparison between the two methods. We also compare the results with that of ITU-T's E-model and find them to be in agreement.

Keywords
Rate Distortion Function, PLC, E-model

1 Introduction

Packet switched networks are primarily designed for non real-time data communication, therefore, the methods used for error recovery such as Automatic Repeat reQuest (ARQ) are not suitable for real-time communication like internet telephony. For real-time communication we require those methods which offer low delay. Packet loss concealment (PLC) methods are used in real-time communication over packet switched networks. These methods can be divided into two categories *a*) Sender Based Techniques and *b*) Receiver Based Techniques. Sender based techniques include FEC (Forward Error Correction), MDC (Multiple Description Coding), interleaving and multiplexing [1], [2], [3]. Some of these techniques such as FEC require extra information to be added at the sender so that the receiver can utilize it to recover from errors. In MDC, the source information is divided into multiple sub-sources and sent to the receiver. Each sub-source is good enough for a minimum acceptable quality [4]. Whereas interleaving does not require additional bandwidth but the receiver must wait for all the interleaved information to arrive [5]. Receiver based techniques use substitution of silence, random noise, repetition of the previously received packet, waveform substitution, pitch replication or interpolation [1] [2]. Although some of these techniques really perform well, but except silence substitution and packet repetition most of the techniques are computationally expensive and require higher computational power and delay.

In [6] Kim and Kleijn compare the distortion rate bound for FEC and MDC. They concluded that MDC performs better than FEC if there is some feedback about the network condition. We compare the distortion rate bound for silence substitution and packet repetition. Here by packet repetition we mean, not simply repeating the previously received packet but to perform some signal processing and synchronization so that we can approximate lost speech signal. This means that the repetition is performed at waveform level rather than bit level. In the remainder of the paper we use waveform repetition and packet repetition interchangeably. We have evaluated this bound for packet repetition by introducing a scaling factor which models the degradation due to waveform repetition. The scaling factor also model how well the reconstructed waveform is similar to the original waveform. The rest of the paper is organized as follows: in section 2 we will calculate distortion bounds for silence substitution and waveform repetition. We will compare these bounds in section 3. In the last section we will present some conclusions.

2 Calculation of Distortion-Rate Bounds

The Rate Distortion Function specifies the lowest average rate possible for a given average distortion or, equivalently, the Distortion Rate Function specifies the lowest average distortion possible for a given average rate. Mathematically, for a source X with reconstruction $\hat{X}$ and with

M. S. Ehsan acknowledges the financial support given by HEC, Govt. of Pakistan.

distortion measure $d(x, \hat{x})$, the rate distortion function is defined as [7]

$$R(D) = \min_{p(\hat{x}|x):\sum_{(x,\hat{x})} p(x)p(\hat{x}|x)d(x,\hat{x})\leq D} I(X;\hat{X}), \quad (1)$$

where $I(X;\hat{X})$ is the mutual information and minimization is over the all conditional distributions $p(\hat{x} \mid x)$ for which the joint distribution $p(x,\hat{x}) = p(x)p(\hat{x} \mid x)$ satisfies the expected distortion constraint.

Analytical expressions for the rate distortion functions are not available for most of the sources. For a Gaussian *iid* source with unit variance and a squared error criterion the distortion rate (DR) function is given by [7]

$$D(R) = 2^{-2R}. \quad (2)$$

Although real sources such as speech are not Gaussian, this model is still useful, because it is well known from information theory that the channel requirements for a Gaussian source are greater than or equal to the channel requirements of any other source [7]. Therefore, this result gives a lower bound on the distortion for a given rate even for sources such as speech signals. This model is also used frequently for autoregressive (AR) Gaussian sources [8]. In this case it is assumed that the side information is known perfectly by the receiver.

Regarding the squared error distortion measure, we note that perceptual distortion measures [9] are considered more appropriate for speech at low rate compression. For packet switched networks the overhead of packet headers etc., normally reduces the compression requirements for the source codec, and in the high rate regime, the squared error distortion measure is not so different from a perceptual measure [10].

We use the Gilbert-Elliott model to model packet losses in the Internet. It is a two state model with state '0' and '1' representing the good and the bad state as shown in Fig.1. Stationary probabilities of theses two states are:

$$\Pi_0 = \frac{q}{p+q} \quad (3)$$

and

$$\Pi_1 = \frac{p}{p+q} \quad (4)$$

where p and q are the transition probabilities from the good state to the bad state and vice versa.

The Burst ratio is used to characterize the busrtiness of losses and always constant and defined as [11]:

$$\text{Burst Ratio} = \frac{L_B}{L_R} \quad (5)$$

where L_B is the average length of observed bursts in an arrival sequence and L_R is the average length of bursts expected for the network under random loss.

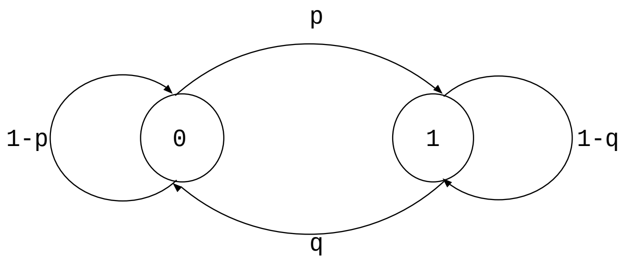

Figure 1. Gilbert-Elliott Channel Model

L_B is defined as [12]

$$L_B = \frac{1}{q}. \quad (6)$$

For random losses $q = 1 - p$, therefore, L_R is

$$L_R = \frac{1}{1-p}. \quad (7)$$

For random losses burst ratio is equal to 1.

In the following subsection we will calculate distortion bound silence substitution, perfect repetition and imperfect repetition.

2.1 DR Bound for Silence Substitution

To calculate the DR bound for silence substitution, we consider the case where two packets are sent by the sender. The distortion bound for each combination of received packets is given in Table 1. Therefore, the overall distortion is given by

$$D(R) = \frac{q(1-p)}{p+q}2^{-2R} + \frac{pq}{p+q} + \frac{pq}{p+q}2^{-2R} + \frac{p(1-q)}{p+q}.$$

$$D(R) = \frac{q}{p+q}2^{-2R} + \frac{p}{p+q}. \quad (8)$$

We interpret the above equation as a combination of information theoretic concepts such as rate, R and distortion, D and networking concepts such as packet loss probability, Π_1 and average burst length L_B. In terms of these parameters, Eq.(8) can be rewritten as:

$$D(R, \Pi_1, L_B) = (1 - \Pi_1)2^{-2R} + \Pi_1. \quad (9)$$

Which shows that in this case the distortion is independent of the burstiness of the packet losses.

2.2 DR Bound for Perfect Waveform Repetition

If the previously received packet is used to replace the lost packet, then the distortion bound for each case given

Table 1. Probabilities of Receiving Packets and Distortion Bound for Silence Substitution

Previous	Current	Probability	Distortion
Good	Good	$\frac{q(1-p)}{p+q}$	2^{-2R}
Good	Bad	$\frac{pq}{p+q}$	1
Bad	Good	$\frac{pq}{p+q}$	2^{-2R}
Bad	Bad	$\frac{p(1-q)}{p+q}$	1

Table 2. Distortion Bounds for Perfect and Imperfect Waveform Repetition

Previous	Current	Distortion (P.R.)	Distortion (I.R.)
Good	Good	2^{-2R}	2^{-2R}
Good	Bad	2^{-2R}	$(\alpha+2^{-2R})$
Bad	Good	2^{-2R}	2^{-2R}
Bad	Bad	1	1

in Table 1 changes to the bounds given in Table 2. While calculating these bounds, we have assumed that the previous packet can be treated as a perfect substitute of the lost packet. Although this assumption is not true, we use it for baseline comparison purposes, we will present a more realistic model of the effect of packet repetition on the quality in the next sub-section. This assumption holds true if duration of the loss is small. The effect of reconstructed signal is inaudible for the losses of small duration. We also assume that if more than one packets are lost in a row, then only the first can be recovered by packet repetition and all subsequent lost packets are replaced by zero substitution. Under these assumptions the overall distortion is now given by the following equation

$$D(R) = \frac{q(1-p)}{p+q}2^{-2R} + \frac{pq}{p+q}2^{-2R} + \frac{pq}{p+q}2^{-2R} + \frac{p(1-q)}{p+q}$$

$$D(R) = \frac{q(1+p)}{p+q}2^{-2R} + \frac{p(1-q)}{p+q}. \quad (10)$$

2.3 DR Bound for Imperfect Waveform Repetition

To incorporate the degradation in quality due to repeating the previously received packet in place of the lost packet, we introduce a factor α. We call this factor as packet change ratio (PCR). α can be think of as a ratio of the change between the two successive speech frames (packets). Introduction of PCR allow us to have fair comparison between the two methods. $\alpha = 0$ means that the lost packet is perfectly reconstructed form the previously received packet. Large values of α represent greater difference between the actual and reconstructed waveform. We also assume that if number of packets lost in a row is more than the buffer length then we use silence substitution instead of packet repetition.

To calculate distortion rate bound for imperfect waveform repetition we use Table 2. Therefore, DR bound is given by the following equation

$$D(R) = \frac{q(1-p)}{p+q}2^{-2R} + \frac{pq}{p+q}(\alpha+2^{-2R}) + \frac{pq}{p+q}2^{-2R} + \frac{p(1-q)}{p+q}$$

$$D(R) = \frac{q}{p+q}2^{-2R} + \frac{pq}{p+q}(\alpha+2^{-2R}) + \frac{p(1-q)}{p+q}. \quad (11)$$

We can express this bound in terms of Π_1 and L_B as

$$D(R,\Pi_1,L_B) = (1-\Pi_1)2^{-2R} + \frac{\Pi_1}{L_B}(\alpha+2^{-2R}) + \frac{\Pi_1(L_B-1)}{L_B}. \quad (12)$$

For the random losses the above equation reduces to

$$D(R,\Pi_1,L_B) = (1-\Pi_1)2^{-2R} + \Pi_1(1-\Pi_1)(\alpha+2^{-2R}) + \Pi_1^2. \quad (13)$$

3 Experimental Results

In this section we compare the theoretical bounds calculated in the previous section under different packet loss conditions. We will also compare our results with ITU-T recommendations G.113.

3.1 Comparison Between Different PLC Methods

In Fig. 2 represents results for random packet losses. Both in figure 2(a) and 2(b), we consider $p = 0.1$ and $q = 0.9$. It is clear from the figure that waveform repetition performs much better than silence substitution. The difference of performance is 8dB and 7dB for α is 0.05 and 0.1 respectively. $p = 0.2$ and $q = 0.8$ are used in Fig. 2(c) and 2(d).

Fig. 3 represents results for bursty losses. Both in figure 3(a) and 3(b) values of p and q is 0.1 and 0.8 respectively but with $\alpha = 0.1$ and $\alpha = 0.4$. In this case waveform repetition method performs 5dB better than the silence substitution case. We can see from Fig.2 and 3 that PLC methods perform better in case of random losses than bursty losses.

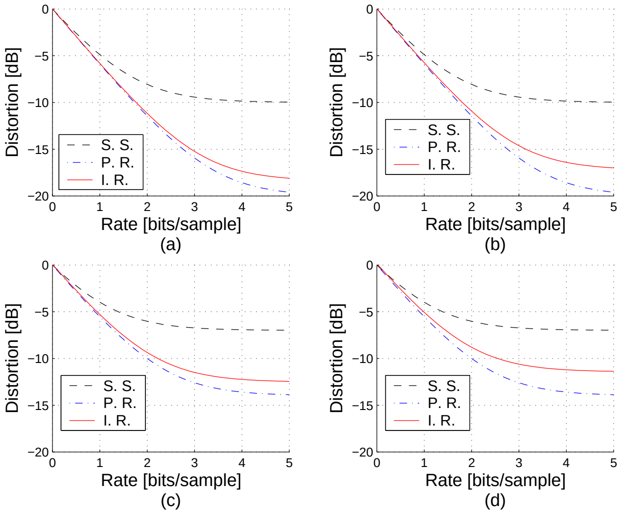

Figure 2. DR Comparison (for Random Loss) between Silence Substitution (S. S.), Perfect Repetition (P. R.) and Imperfect Repetition (I. R.) For $p = 0.1$, $q = 0.9$ (a) $\alpha = 0.05$ (b) $\alpha = 0.1$, and for $p = 0.2$, $q = 0.8$ (c)$\alpha = 0.1$ (d) $\alpha = 0.2$

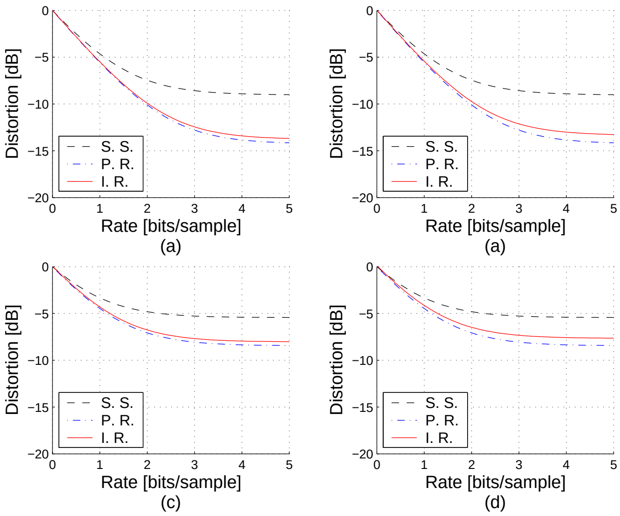

Figure 3. DR Comparison (for Bursty Loss) between Silence Substitution (S. S.), Perfect Repetition (P. R.) and Imperfect Repetition (I. R.). For $p = 0.1$, $q = 0.7$ (a) $\alpha = 0.05$, (b) $\alpha = 0.1$, and for $p = 0.2$, $q = 0.5$ (c) $\alpha = 0.1$ (d) $\alpha = 0.2$

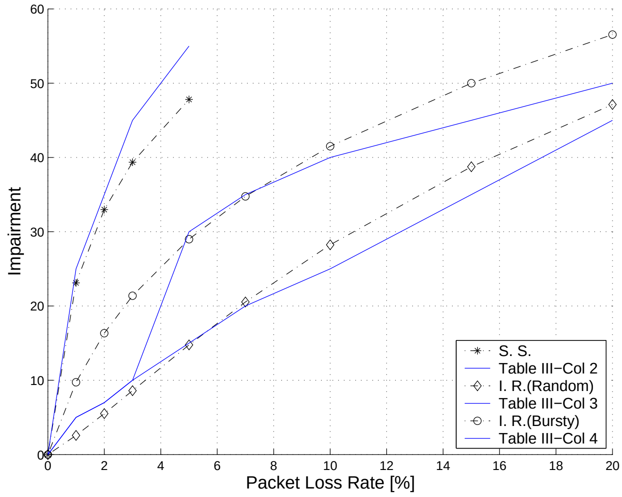

Figure 4. Comparison between Silence Substitution (S. S.), Perfect Repetition (P. R.), Imperfect Repetition (I. R.) and ITU-T G.113:

3.2 Comparison with G.113

ITU-T recommendation G.113 deals with the effect of speech processing on the transmission quality [11]. G.113 includes effects of low bit rate codecs and packet losses on the overall speech quality. This effect is known as Equipment Impairment Factor, I_e. Table 3 gives values of I_e for the G.711 codec for different packet loss rates and PLC methods. Values of the equipment impairment factor, I_e are used to calculate the transmission rating factor of the E-model [13]. To compare our results with ITU-T values of Table 3, we will convert distortions given by (9), (12) and (13) to impairment by using the following equation

$$\text{Impairment} = A \cdot \log_{10}(D(R, \Pi_1, L_B)) \qquad (14)$$

where A is a scaling constant. The constant A is selected such that the results of (14) match with Table 1. Here we consider $R = 4$, $A = 43$. Column 1 of Table 1 is taken as Π_1.

For the bursty losses we use quantity burst ratio as defined in 5. Fig. 4 shows the comparison results between Table 1 and Eq. (9) and (12). We select burst ratio equal to 1.3. By examining Fig. 4, we see that our results for silence substitution and random losses are in agreement with ITU-T recommendation. For bursty case, there is a mismatch, which can be attributed to insufficient buffer length.

4 Conclusion

We compare the DR performance of the silence substitution and packet repetition in case of a Gilbert-Elliott chan-

Table 3. ITU-T G.113 Equipment Impairment Factor for G.711 codec

Packet Loss Rate %	G.711 without PLC	G.711 with PLC	
		Random Loss	Bursty Loss
0	0	0	0
1	25	5	5
2	35	7	7
3	45	10	10
5	55	15	30
7	NA	20	35
10	NA	25	40
15	NA	35	45
20	NA	45	50

nel. These methods are commonly used in speech communication over packet switched networks. We have found that if we repeat waveform constructed from the previously received packet the DR performance is better than that of silence substitution. Although the performance of both methods is the same at low rates, a significant difference in the performance between the two methods is observed if the packet loss probability is increased.

We also compare our results with the ITU-T recommendation G.113 and found our results in agreement with the recommendation.

We have introduced packet change ratio (PCR) as a measure to model the degradation of quality due to packet repetition. We also consider the effect of playout buffer length on distortion bound.

References

[1] B. W. Wah, X. Su, and D. Lin, "A Survey of Error-Concealment Schemes for Real-Time Audio and Video Transmission over the Internet," in *Proc. ICASSP'00*, (Istanbul, Turkey), pp. 17–24, 2000.

[2] O. Hodson, C. Perkins, and V. Hardman, "A survey of packet loss recovery techniques for streaming audio," *IEEE Network*, pp. 40–48, Sep/Oct 1998.

[3] S. V. Anderson and G. Kubin, "Multiplexed Predictive Coding of Speech," in *Proc. ICASSP'01*, (Utah, USA), pp. 741–744, 2001.

[4] A. A. El-Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Transactions on Information Theory*, vol. 28, pp. 851–857, 1982.

[5] N. S. Jayant and S. W. Christensen, "Effect of packet loss in waveform coded speech and improvements due to an odd-even sample-interpolation procedure," *IEEE Transactions on Communications*, pp. 101–109, Feb 1981.

[6] M. Y. Kim and W. B. Kleijn, "Rate-distortion comparisons between FEC and MDC based on Gilbert channel model," *IEEE International Conference on Networks*, pp. 495–500, Sep 2003.

[7] T. M. Cover and J. A. Thomson, *Elements of Information Theory*. Englewood Cliffs NJ: Prentice Hall, 1988.

[8] R. M. Gray, "Information rates of autoregressive processes," *IEEE Transactions on Information Theory*, vol. 16, pp. 412–421, Jul 1970.

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

[10] H. Pobloth, "Perceptual and squared error aspects in speech and audio coding," 2004.

[11] "Transmission impairment due to speech processing," *ITU-T recommendation G.113*, 2001.

[12] H. Sanneck and G. Carle, "A framework model for packet loss metrices based on loss runlengths," *In Proceedings of SPIE/ACM SIGMM*, pp. 177–187, Jan 2000.

[13] "E-model, a computational model for use in transmission planning," *ITU-T recommendation G.107*, 2002.