Enhanced recovery technique for improving voice quality degraded by packet loss in data networks

S.A. Napoleon, S.A. Khamis and M.E. Nasr
Electronics and Communication Dept. Faulty of Eng., Tanta University, Tanta, Egypt

On transferring voice over a packet switched network, packet loss is unavoidable. Hence it is necessary to recover lost voice packets. Many recovery techniques were developed to recover in real-time as voice is a real-time traffic. These recovery techniques repair only during loss periods neglecting the effect of loss on the subsequent speech segments. This paper has developed an enhanced technique called Parallel Recovery Technique (PRT) to alleviate this shortcoming in other recovery techniques. PRT is found to perform better than other methods with a close to zero added complexity. The results show improvement in perceived voice quality of about 10% at 20% loss rate.

خلال نقل الحزم الصوتية عبر شبكات تبديل الحزم، لايمكن تجنب فقد الحزم. لذا لابد من وسيلة لاسترجاع الحزم المفقودة. وبما أن الصوت هو تطبيق من تطبيقات الزمن الحقيقى، فقد تم تطوير عدة تقنيات لاسترحاع الحزم آنيا. هذه تقنيات تعالج منطقة الفقد فقط مهملة تأثير الفقد على مابعد منطقة الفقد. فى هذه الورقة قد تم تطوير تقنية محسنة تدعى تقنية الاسترجاع المتوازى لتعالج المشكلة السابق ذكرها. و بالفعل أدت الى تحسن فى جودة الصوت باضافة قدر بسيط من التعقيد. النتائج اظهرت تحسين فى جودة الصوت بنسبة ١٠% عند معدل فقد قدره ٢٠%.

Keywords: Quality of service, Packet loss recovery, Voice over IP, Data networks

1. Introduction

On transmitting packetized voice over data networks, packet loss is more likely to happen which greatly degrades perceived voice quality. For Voice over Packet (VoP) and other real-time applications, loss recovery must be done as soon as the loss detected making it impossible to use recovery mechanisms used with data packets. A lot of recovery techniques, which are called receiver-based recovery techniques, are developed to recover lost voice packets. These techniques repair the decoded waveform during loss periods only. Most codecs used for data networks are LPC-based or differential-based codecs, which generate low bit-rate traffic. These codecs build the current waveform using current received packet and the decoder history. It is noted by [1] that the loss affects the after-loss waveform also which makes excess quality degradation even when using a sophisticated recovery technique. In this paper a recovery technique called the Parallel Recovery Technique (PRT) is developed to enhance the quality of the speech via repair in the loss location while minimizing the effect of loss on the after-loss stream as well. Further, three voice compression methods are used, to test the PRT, the GSM standard full rate coder, GSM Half Rate coder and the FS1016 CELP coder.

The paper is arranged as follows: section 2 is dedicated to take a look on the three used coders; while an overview for conventional recovery techniques are presented in section 3. The PRT is then explored in section 4. A study for codecs performance with different recovery techniques and the performance and results of the PRT are then shown in section 5. Finally the paper concludes in section 6.

2. Voice compression algorithms

The three coders that will be overviewed in this section are all based on Analysis-by-Synthesis (AbS) techniques in which, the speech is segmented into segments of 10 to 30 ms, then coefficients of adaptive filters and an excitation signal are calculated for each speech segment and transmitted to be used in speech synthesis at the. An excitation signal is derived from the input speech signal in such a manner that the difference between the input and the synthesized speech is quite small.

© Faculty of Engineering Alexandria University, Egypt.

2.1 GSM full rate coder (RPE-LTP)

SM [2, 3] (GSM Full Rate) (FR) standard coder, is a version of the RPE-LTP which stands for Regular Pulse Excited (REP) coder with Long Term Predictor (LTP). Fig. 1 shows the encoder and decoder for this codec. This coder has an Short Term Predictor (STP) of order 10 and one tap LTP. Its excitation signal is a set of regularly spaced ten pulses per each sub frame that has 40 samples. The frame size of this coder is 160 samples, i.e. each frame has four sub frames. The excitation pulses can take one of four positions. The coder output frame contains the 10 STP filter coefficients represented by Line Spectral Frequencies (LSFs) and the LTP gain and delay and the ten pulses amplitudes and the position of them. The bit rate for this codec is 13 Kbps.

2.2. GSM Half Rate coder (GSM HR)

Also called or Vector Sum Excited Linear Prediction (VSELP) coder [3]. Its bit rate is 5.6 Kbps. Its excitation is two stochastic codebooks with short length to reduce codebook index search. LTP with adaptive codebook approach exists. STP order is 10. The overall complexity is more than that of the GSM full rate coder.

2.3. The Federal Standard CELP (FS1016)

CELP stands for CODE Excited Linear Prediction (CELP) [3]. The excitation signal is chosen form a stochastic codebook. The STP order is 10 and the LTP uses the adaptive codebook approach. The bit rate for this codec is 4.8 Kbps.

3. Conventional recovery methods

Packet Loss recovery methods can be classified, according to [2], into either receiver-based recovery, or sender-based recovery. In the receiver based-recovery techniques, the receiver is responsible for the reconstruction of missing voice segments by using whatever available voice information. Common receiver only techniques are silence / waveform substitution, packet repetition, sample interpolation and pitch waveform replication. In the sender-based recovery techniques, both source and destination are responsible for recovering the missing voice segments. Common sender-based recovery methods are, Forward Error Check (FEC), Congestion control, Interleaving and Retransmission. These recovery methods usually need more processing delay and look ahead delay, and cannot recover all kinds of the packet loss. In the following, a general overview for distinctive features of receiver-based missing cell recovery methods is presented.

3.1. Silence/Waveform Substitution (S/WS)

In the silence method, it is not able to maintain an acceptable quality of playback voice in the event of high-lost rate and large packet size. In this technique, lost cells are simply replaced with silence. Hence, applications using this technique do not incur much additional processing power and is therefore suitable for an environment where the probability of losing cells is low and the computers do not have much processing power [2]. Whereas in the waveform method, when a segment of voice fails to arrive at the destination on time, the previous segment of voice is used to replace the missing segment of voice. The assumption of this technique is that the speech characteristics have not changed much from a preceding speech segment and it is logical to use the previous segment of speech to reconstruct the missing portion. This method does not work for large packet size as the voice characteristics are most likely to change noticeably from one previous cell to the next. Moreover, it also does not guard against the continuous loss of multiple cells where voice characteristics do not remain the same over the duration of cells loss. As with silence substitution, it does not demand lots of processing power. Hence, it is used in some of the interactive voice communication applications [2].

3.2. Sample Interpolation (SI)

This technique is similar to waveform substitution; however, it does not directly replace all missing voice segments with the previously received segments. It modifies the previous audio packets before substituting the missing voice segments with it. The method assumes that the voice characteristics change slightly over a short period of time. In order to use previously received samples to replace the missing voice segments and at the same time accommodating the slight change in voice attribute, the missing samples are estimated based on the previous samples' characteristics. A simple form of sample modification is linear interpolation of voice [4]. In comparison, it requires more processing power than the previous methods, but it offers a better contingency solution. As with waveform substitution, it is not usable in a prolonged duration of cells loss as it is likely that the voice characteristics will change significantly.

3.3. Pitch Waveform Replication (PWR)

A refinement on waveform substitution is by using a pitch detection algorithm on either side of the loss. Losses during unvoiced speech segments are repaired using packet repetition and voiced losses repeat a waveform of appropriate pitch length. The technique, known as pitch waveform replication, was found to work marginally better than waveform substitution.

3.4. Double Sided Pitch Waveform Replication (DSPWR)

This method proposed by [4] and found to be tolerant to higher loss rates than other mentioned methods but with higher complexity and delay. Double Sided Pitch Waveform Replication (DSPWR) treated two shortcomings in PWR method. First, PWR only copes with the continuity for the boundaries between the reconstructed packets and their previous packets when the speech is voiced, whereas the continuity for the boundaries between the reconstructed packets and the subsequent ones is not properly dealt with. This is referred to as the discontinuity problem of PWR. The second shortcoming of the PWR method is that it uses the repetition method to recover the lost packet when the speech is unvoiced or when the pitch detection fails. However, the reconstructed speech will not have acceptable quality when there is a transition from voiced to unvoiced at the substitution packets. This is referred to as the repetition problem of PWR. It is noted that there are some reconstruction techniques, such as the phase-matching reconstruction method [4] and the double-sided periodic substitution method, proposed to solve the first problem of PWR on discontinuity. However, the phase-matching method tends to introduce extra frequency discontinuity, and as a result there is still audible noise. The double-sided periodic substitution method does not take the phase discontinuity into consideration. Neither the phase-matching method nor the double-sided periodic substitution method addressed the second problem of PWR on repetition mentioned above. DSPWR remedies the above mentioned problems of the PWR by searching for pitch in both preceding and following packets surrounding loss. This adds the ability to detect if during loss a transition to a voiced segment occurred, in this case a careful repair is required since the loss during such a transition causes sever impact on speech quality. DSPWR can tolerate loss rates up to 30%. According to the status of the two pitch detectors (for preceding and following packets); there are four repair procedures described in [4] Also DSPWR has amplitude and pitch adjustment capabilities.

4. Parallel Recovery Technique (PRT)

In this section a modified Recovery technique named PRT for improving the voice is described. Most codecs used for data networks are LPC-based or differential-based codecs, which generate low bit-rate traffic. These codecs build the current waveform using current received packet and the decoder's history. It is noted by [1] that the loss affects the after-loss waveform also which makes excess quality degradation even when using a sophisticated recovery technique because, after the loss period, the decoder starts to reconstruct the waveform using the parameters enclosed in the received packets and the decoder's history which is greatly far from the right value at the beginning of decoding after-loss packets, that leads to errors in decoding. The decoder's history deviation is a consequence of the fact that, the mentioned recovery techniques in section 3 recover the lost waveform using the decoded waveforms during the no-loss periods neglecting the status of the decoder itself. In order to minimize the error on decoded waveform after loss period, the history of the decoder must set as close as possible to the decoder history if no loss is encountered. This may be done by re-encoding the recovered waveform using one of the recovery techniques, in section 3, to generate parameters which may resemble that enclosed in the lost packets. However this extremely increases the recovery technique complexity. Another method is to use Packet Repetition (PR) to copy the parameters in the last received packet before loss so it is decoded instead of decoding null packets. In this way, decoder history is maintained as close to the correct history as possible with minimal increase in the recovery technique complexity. Using packet repetition and any other known receiver-based recovery technique is called PRT. The PRT is implemented for test as in the block diagram in fig. 2. The depacketizer depackes the incoming voice packets, detect missing packets and tell the controller. The palyout buffer used to reduce the effect of delay jitter. When the depacketizer detects a missing sequence number, it triggers the controller to send the control signals a, b to turn the switches SW1, SW2 to positions 2 making the decoder decodes packets repeated by the packet repetition algorithm. At the same time one of the mentioned recovery techniques used to build a waveform instead of the lost part using last decoded waveform. After loss period ends, the controller turns SW1 and SW2 to positions 1 again to decode the incoming voice packets with the decoder's history not deviated so much from the right value.

5. Performance evaluation and results

The performance of voice compression techniques under packet loss conditions which are given in section 2 are studied using different recovery techniques in section 3. On testing the codecs, a packet loss generator and

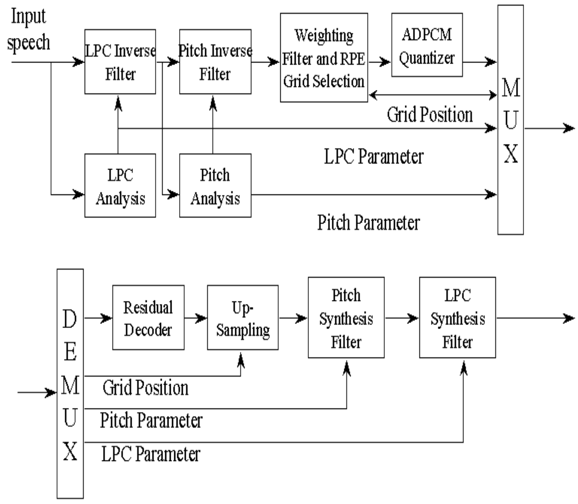

Fig. 1. Coder and Decoder for the GSM FR [2].

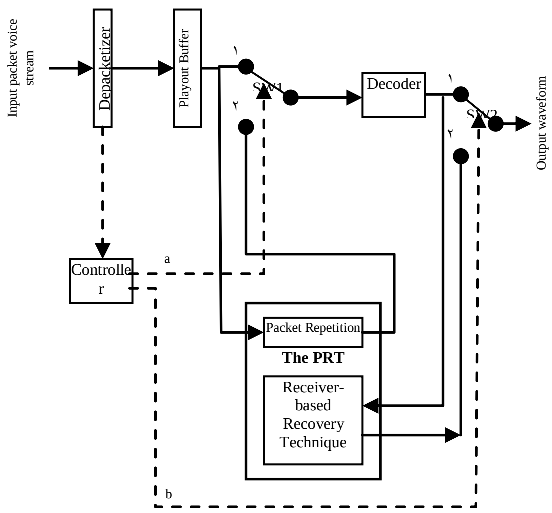

Fig. 2. Block diagram for the PRT.

a quality measure have to be used. The simulation system used to test recovery techniques is shown in fig. 3. The network packet loss simulator is a 2-State Markov model loss generator. It is found by [5], to bethe best model that describes loss distributions in a network. The quality is measured using the Perceptual Evaluation of Speech Quality (PESQ) the ITU-T standard P.862 [6], which is a software implementation for the human ear, and the coder and decoder blocks are for the coder under study. In addition, a Packetizer and Depacketizer used to pack the generated parameters by the coder

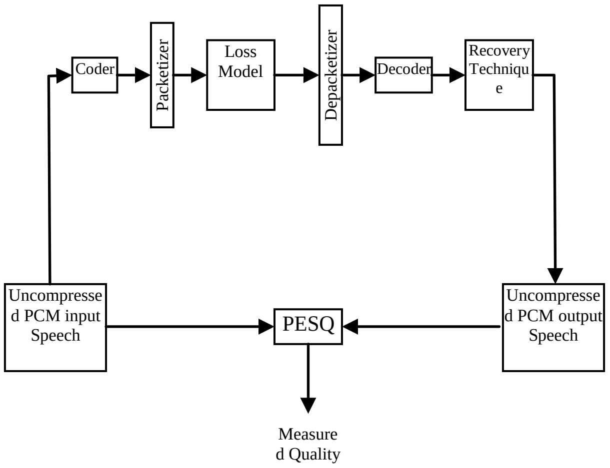

Fig. 3. Simulation system used for testing recovery techniques and PRT.

and depack them at the receiver end.
The results collected are the average quality from using two speech sentences Arabic and English:
"يعتبر برنامج ميكروسوفت أكسس من أهم برامج قواعد البيانات"
"An apple a day keeps the doctor away" and for every given loss rate the simulation is done for ten times to measure the average quality.The obtained results for this simulation are given in fig. 4, fig. 5, and fig. 6 for the three coders in section 2. These results show that quality for a stream repaired using DSPWR is better than repair using PWR and PWR is better than SI and WS and PR. SS is the worst recovery technique for all codecs. For complete comparison, the complexity, measured in terms of execution time, for each recovery technique normalized to the SS method can be found in table 1 [4].

The enhanced recovery using PRT will be called PRT-“Recovery Technique” for example the enhanced WS will be called PRT-WS.
The results in figs. 7, through 18, for the three focused coders show a noticeable improvement in quality of the repaired stream. It is remarkable that as the loss rate increases, the enhancement becomes more noticeable. This can be explained as follows; for low loss rates the loss tends to be for individual packets, so the decoder history will not deviate so much than the coder. Hence, the improvement using the packetrepetition in the PRT will be small. However for large loss rates, the loss tends to occur in multiple packet bursts as well as single packet bursts, that forces the decoder’s history away from the encoder. The use for packet repetition in the PRT will reduce the

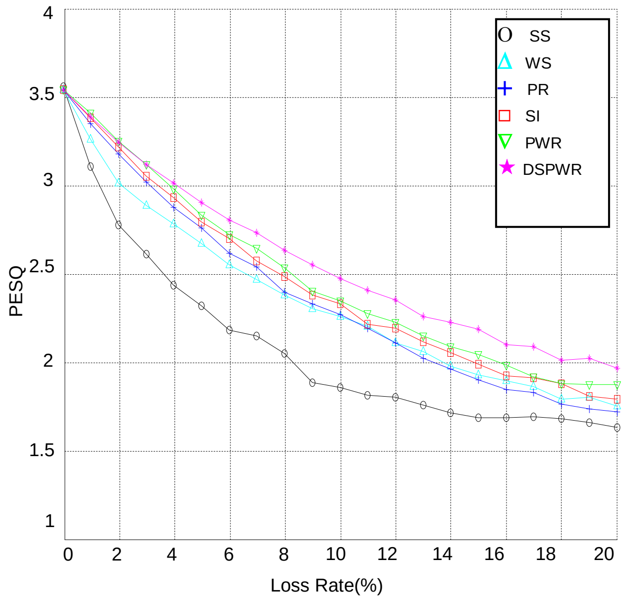

Fig. 4. PESQ vs. loss for the GSM full rate coder (RPE-LTP).

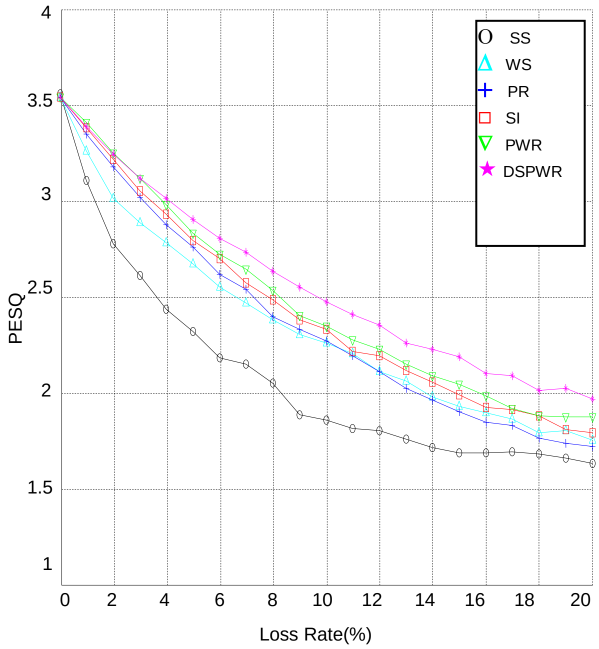

Fig. 5. PESQ vs. loss for the GSM half rate coder (VSELP).

Table 1
Complexity vs. recovery techniques

	SS	PR	WS	SI	PWR	DSPWR
Complexity in terms of execution time normalized to SS complexity	1	1.5 to 2	4 to 6	10 to 15	25 to 50	450 to 600

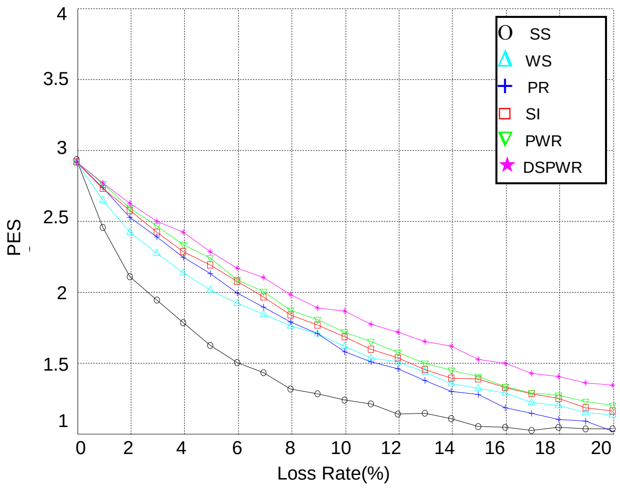

Fig. 6. PESQ vs. loss rate for federal standard coder (FS1016 CELP).

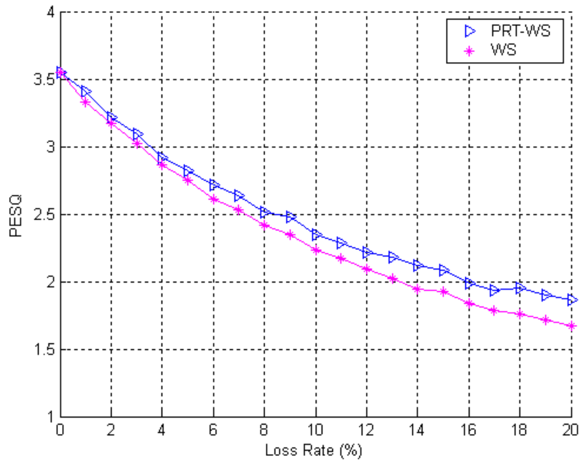

Fig. 7. Enhancement for the WS technique by the PRT-WS technique for GSM Full Rate coder (RPE-LTP.)

decoder's history deviation, which will greatly improve the quality.

Comparing the PRT to the Switched Recovery Technique (SRT) developed in [2]; the average complexity in PRT will not be reduced but slightly increased by the amount added by using the PR technique at the same time the

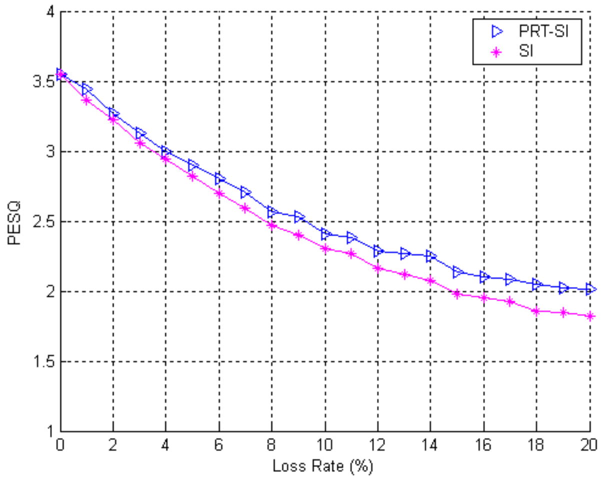

Fig. 8. Enhancement for the SI technique by the PRT-SI technique for GSM Full Rate coder (RPE-LTP).

other recovery techniques used. Another difference is the use of the desired recovery technique for all loss rates while in SRT the recovery technique in use depends on the instantaneous loss rate which is the inverse of the time distance between every two consecutive losses.

6. Conclusions

On transferring voice over packet switched networks, packet loss is unavoidable. Ordinary receiver-based recovery techniques do not take the decoder history into account that causes quality degradation due to miss-tracking of the encoder. The PRT cured this problem by employing Packet Repetition in parallel to the recovery technique used. This improved the Quality of the recovered voice stream during and after loss without adding too much complexity. It is noted that the PRT's complexity always greater than the complexity of the ordinary recovery method used in it by approximately three times the complexity of the SS recovery method.

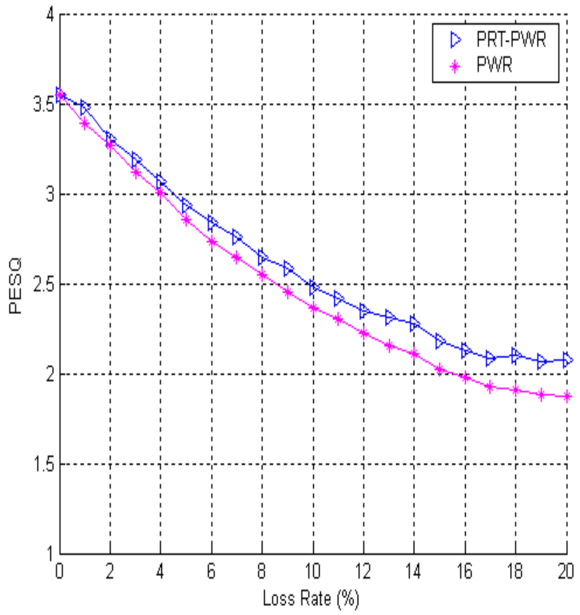

Fig. 9. Enhancement for the PWR technique by the PRT-PWR technique for GSM Full Rate coder (RPE-LTP).

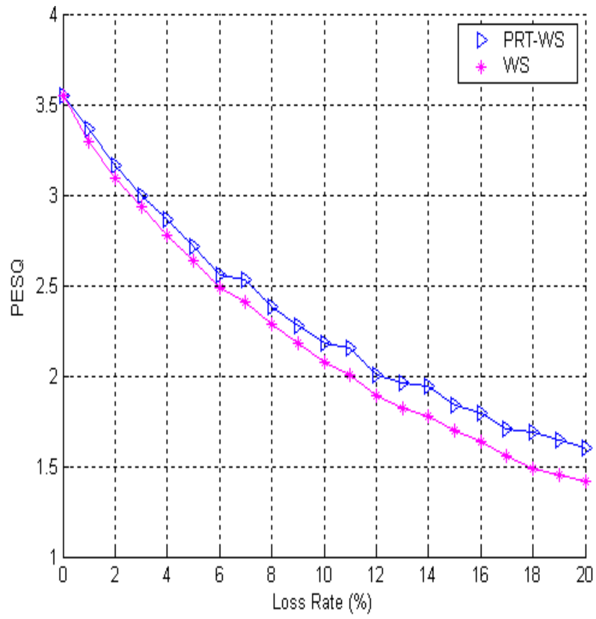

Fig. 11. Enhancement for the WS technique by the PRT-WS technique for HR-GSM coder

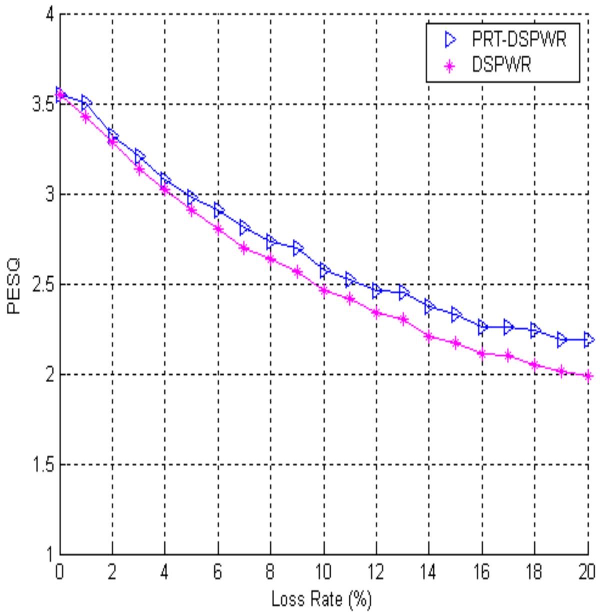

Fig. 10. Enhancement for the DSPWR technique by the PRT-DSPWR technique for GSM Full Rate coder (RPE-LTP).

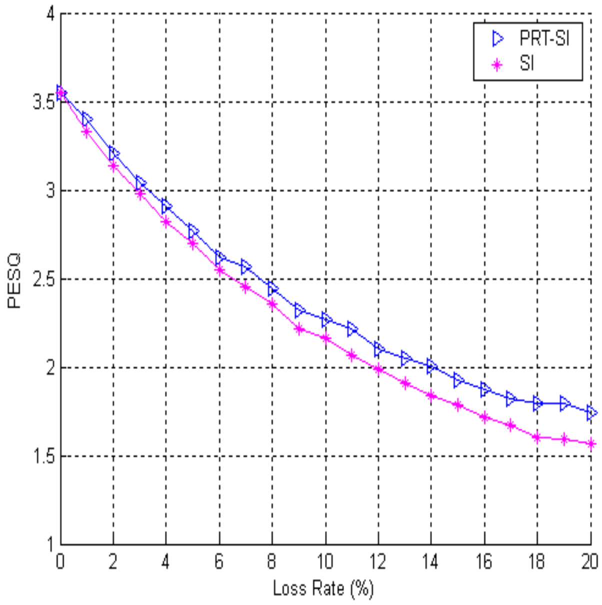

Fig. 12. Enhancement for the SI technique by the PRT-SI technique for HR-GSM coder.

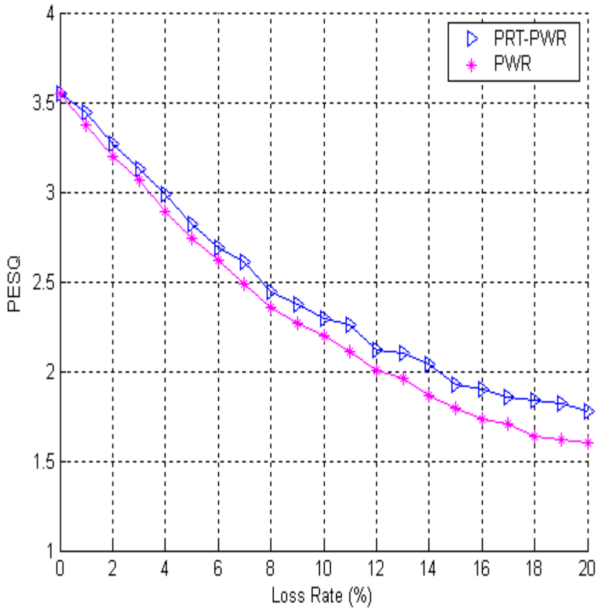

Fig. 13. Enhancement for the PWR technique by the PRT-PWR technique for HR-GSM coder.

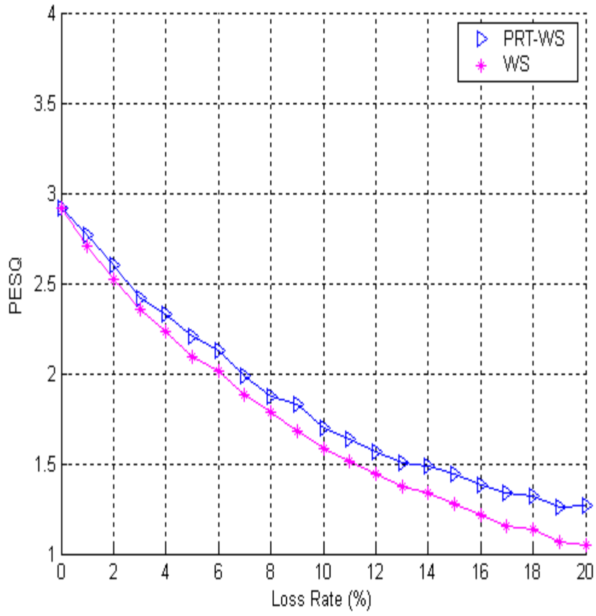

Fig. 15. Enhancement for the WS technique by the PRT-WS technique for FS1016 CELP.

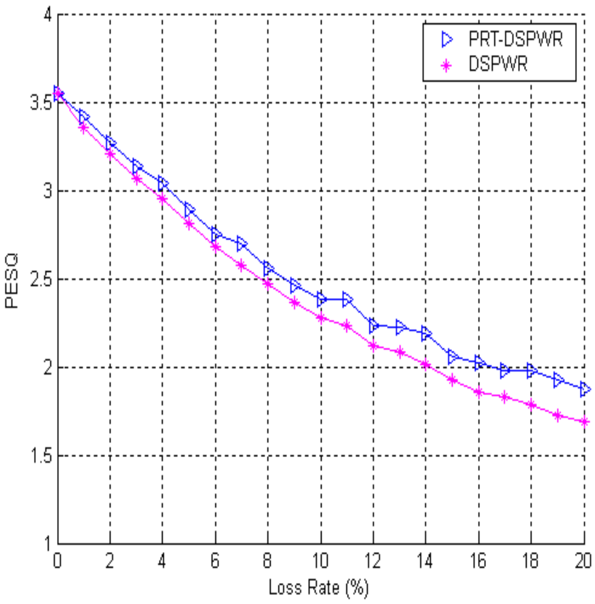

Fig. 14. Enhancement for the DSPWR technique by the PRT-DSPWR technique for HR-GSM coder.

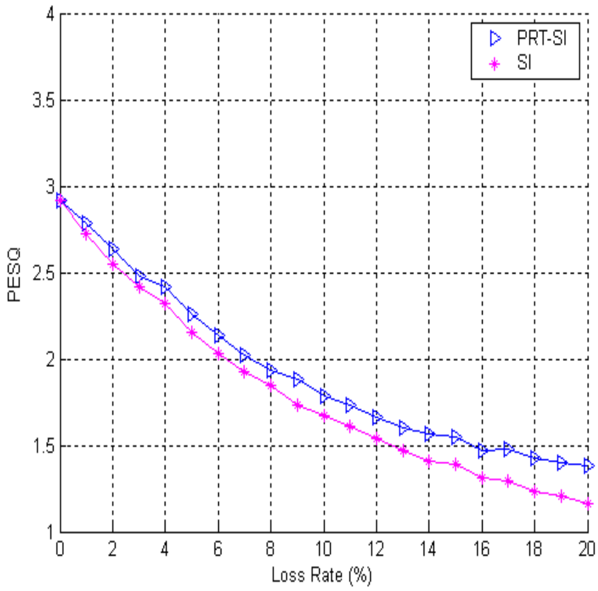

Fig. 16. Enhancement for the SI technique by the PRT-SI technique for FS1016 CELP.

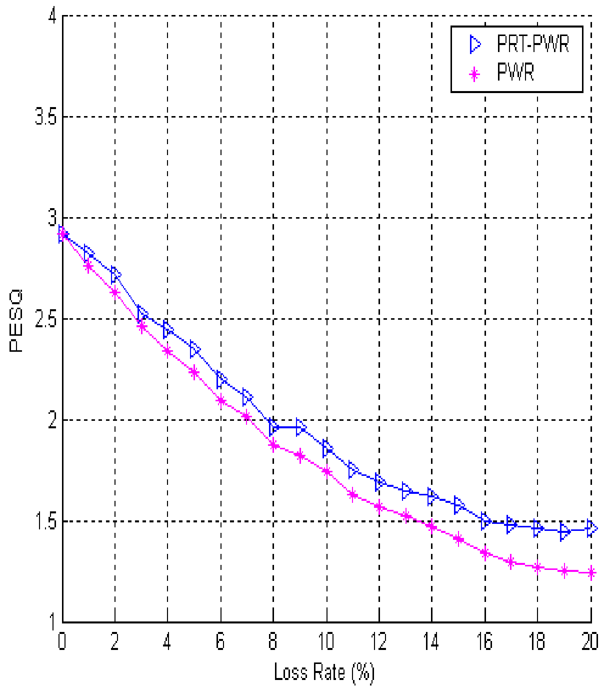

Fig. 17. Enhancement for the PWR technique by the PRT-PWR technique for FS1016 CELP.

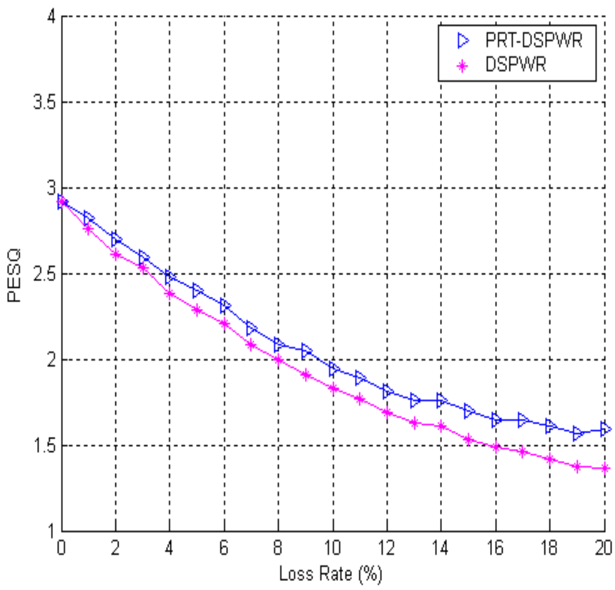

Fig. 18. Enhancement for the DSPWR technique by the PRT-DSPWR technique for FS1016 CELP.

References

[1] L.F. Sun, G. Wade, B.M. Lines and E..C.I Feachor, "Impact of Packet Loss Location on Perceived Speech Quality", Proceedings of 2nd IP-Telephony Workshop (IPTEL '01), Columbia University, New York, pp. 114-122. (2001).

[2] Mohamed E. Nasr and A. Sameh Napoleon, "On Improving Voice Quality Degraded by Packet Loss in Data Networks", The 22nd National Radio Science Conference (NRSC' 2005), pp. 15-17 (2005).

[3] C.Wai Cau, "Speech Coding Algorithms. Foundation and Evolution of Standardized Coders", John Wiley and Sons, Inc. (2003).

[4] Wen-Tsai Liao; Jeng-Chun Chen; Ming-Syan Chen, "Adaptive Recovery Techniques for Real-Time Audio Streams", INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, Vol. 2, pp. 815 – 823, Vol. 2, pp. 22-26 (2001).

[5] L. Carvalho, J. Angeja and A. Navarro, "A New Packet Loss Model of the IEEE 802.11g Wireless Network for Multimedia Communications", IEEE Transactions on Consumer Electronics, Vol. 51, Issue: 3 pp. 809 – 814 (2005).

[6] ITU-T Recommendation "Perceptual Evaluation of Speech Quality (PESQ)", p. 862 (2001).

Received July 1, 2006
Accepted December 21, 2006