*International Journal of Innovative Computing, Information and Control*
Volume **7**, Number **12**, December **2011**

ICIC International ©2011 ISSN 1349-4198
pp. 6773–6783

# AN ADAPTIVE PACKET LOSS RECOVERY METHOD BASED ON REAL-TIME SPEECH QUALITY ASSESSMENT AND REDUNDANT SPEECH TRANSMISSION

Jin Ah Kang and Hong Kook Kim

School of Information and Communications
Gwangju Institute of Science and Technology (GIST)
1 Oryong-dong, Buk-gu, Gwangju 500-712, Korea
{ jinari; hongkook }@gist.ac.kr

Received August 2010; revised February 2011

Abstract. *In this paper, an adaptive packet loss recovery (APLR) method that improves the speech quality of a real-time speech streaming (RSS) system over IP networks is proposed. The proposed APLR method estimates the packet loss rate (PLR) of network via a real-time speech quality assessment (RSQA) at the receiver side of the RSS system, and then requests the opposite RSS system to transmit redundant speech frame data (RSD). Thus, it assists the speech decoder employed in the RSS system to reconstruct lost speech signals when the estimated PLR is high. In particular, according to the estimated PLR, the proposed APLR method then controls the bitrate of speech coding for the RSS system. In other words, a speech packet combines the bitstreams of the current speech frame data (CSD) and the RSD for a high PLR. Otherwise, for a low PLR, a speech packet consists of the CSD bitstreams alone. The effectiveness of the proposed APLR method is finally demonstrated by using an adaptive multirate-narrowband (AMR-NB) speech codec and ITU-T Recommendation P.563 as the scalable speech codec and RSQA, respectively. It is shown from experiments that an RSS system employing the proposed APLR method significantly improves the speech quality under packet loss conditions.*
**Keywords:** Packet loss recovery, Real-time speech quality assessment, Redundant speech transmission, Adaptive multi-rate speech coding

1. **Introduction.** Due to the rapid development of Internet protocol (IP) networks over the past few decades, audio and video streaming services are increasingly available via the Internet. Moreover, as these services are extended to wireless networks, the quality of service (QoS) of audio and video streaming is becoming ever more critical. To this end, there have been a number of previous studies reported that have attempted to guarantee the quality of speech that could be impaired due to environmental factors such as network conditions [1, 2] or noise [3]. Specifically, real-time speech streaming (RSS) systems require a minimum level of speech communication quality. For example, the guarantee of speech quality is important for e-learning applications, where such speech quality is largely related to packet losses and end-to-end packet delays [4]. When speech streaming is performed via user datagram protocol/IP (UDP/IP) networks, however, packets may be lost or arrive too late for playback due to inevitable delays. In this case, a typical RSS system can only tolerate a few packet losses for real-time services, though these packet losses frequently occur in wireless networks due to bandwidth fluctuations [5].

Several packet loss recovery methods, implemented via the Internet and wireless networks, have been proposed for RSS systems. For instance, the techniques proposed in [6, 7] were sender-based packet loss recovery methods using forward error correction (FEC). In regards to wireless networks, the techniques proposed in [8, 9] were based on unequal error protection (UEP) methods. In addition, the modified discrete cosine transform (MDCT) coefficients of audio signals were used as redundant data in order to assist an MP3 audio decoder to reconstruct lost audio signals [10]. However, these methods did not take into account time-varying network conditions, i.e., packet loss rate (PLR). In other words, in order to recover the lost packets based on conventional FEC methods, the redundant data are designed to be constantly transmitted even if the network conditions are declared as no packet losses. Therefore, a new packet loss recovery method is needed to efficiently recover the speech quality by considering the time-varying network conditions.

Towards this goal, this paper focuses on the use of a real-time speech quality assessment (RSQA) as well as redundant speech frame data (RSD) transmission, where the RSQA is used as the PLR estimation under time-varying network conditions. In addition, a real-time transport protocol (RTP) payload format is newly suggested as a means of supporting the proposed adaptive packet loss recovery (APLR) method. In other words, a speech packet can combine the bitstreams of the current speech frame data (CSD) and the RSD when the PLR is estimated to be high. Thus, even if a speech packet is lost, the speech decoder employed at the opposite side of the RSS system can reconstruct speech signals corresponding to lost packets having a minimum speech quality by using the RSD of the previous packet. On the other hand, when the PLR is estimated to be low, a speech packet is organized by using the CSD bitstreams alone. The effectiveness of the proposed APLR method is finally demonstrated by using the adaptive multirate-narrowband (AMR-NB) speech codec [11] and ITU-T Recommendation P.563 [12] as the scalable speech codec and RSQA, respectively.

Following this introduction, Section 2 discusses an RSS system that can accommodate the proposed APLR method, and Section 3 describes the detailed procedure of the proposed APLR method. After that, Section 4 evaluates the performance of the proposed APLR method by measuring the mean opinion score (MOS) according to different packet loss rates. Finally, we summarize our findings in Section 5.

## 2. An RSS System Based on the Proposed APLR Method.

### 2.1. Overview.

An RSS system extends the traditional speech communication function over a public switched telephone network (PSTN) to IP networks in order to provide various applications. It transmits a continuous speech signal as segmented data packets via an IP protocol. For this transmission, the RSS system transforms a continuous input speech signal to digital speech frame data by using an analog-to-digital (A/D) converter, and then encodes each speech frame data to a bitstream using a speech compression algorithm. Next, the bitstreams are transmitted via a real-time streaming protocol in the form of a predefined packet. When the transmitted packets arrive at the opposite RSS system over the IP network, the bitstreams are extracted after unpacketizing. Subsequently, the extracted bitstreams are decoded into speech frame data, and the decoded speech frame data are finally transformed into a continuous output speech signal using a digital-to-analog (D/A) converter.

Figure 1 shows the packet flow for the RSS system implemented in this paper, where *Subsystems A* and *B* represent the parties of the speech stream communication, which both employ the proposed APLR method. First, the sender side of *Subsystem A* performs scalable speech encoding for the input CSD. Next, the sender side generates a packet according to a real-time transport protocol (RTP) payload format, where the packet includes the CSD bitstreams with the RSQA score that has been sent back from *Subsystem B*. Note here that the RSD bitstreams should be incorporated in this payload when the estimated PLR is high. After that, the formatted RTP packet is transmitted. Meanwhile, as the RTP packet arrives at the receiver side of *Subsystem B*, the receiver side analyzes the received packet according to the RTP payload format, and then extracts the CSD bitstreams and the RSQA score. In the case that the RTP payload format includes the RSD bitstreams, the RSD bitstreams are used to recover a lost packet in the future. Next, the extracted CSD bitstreams are decoded using a scalable speech decoder and stored in a speech buffer in order to obtain an RSQA score. Finally, the RSQA score is transmitted back to *Subsystem A*.

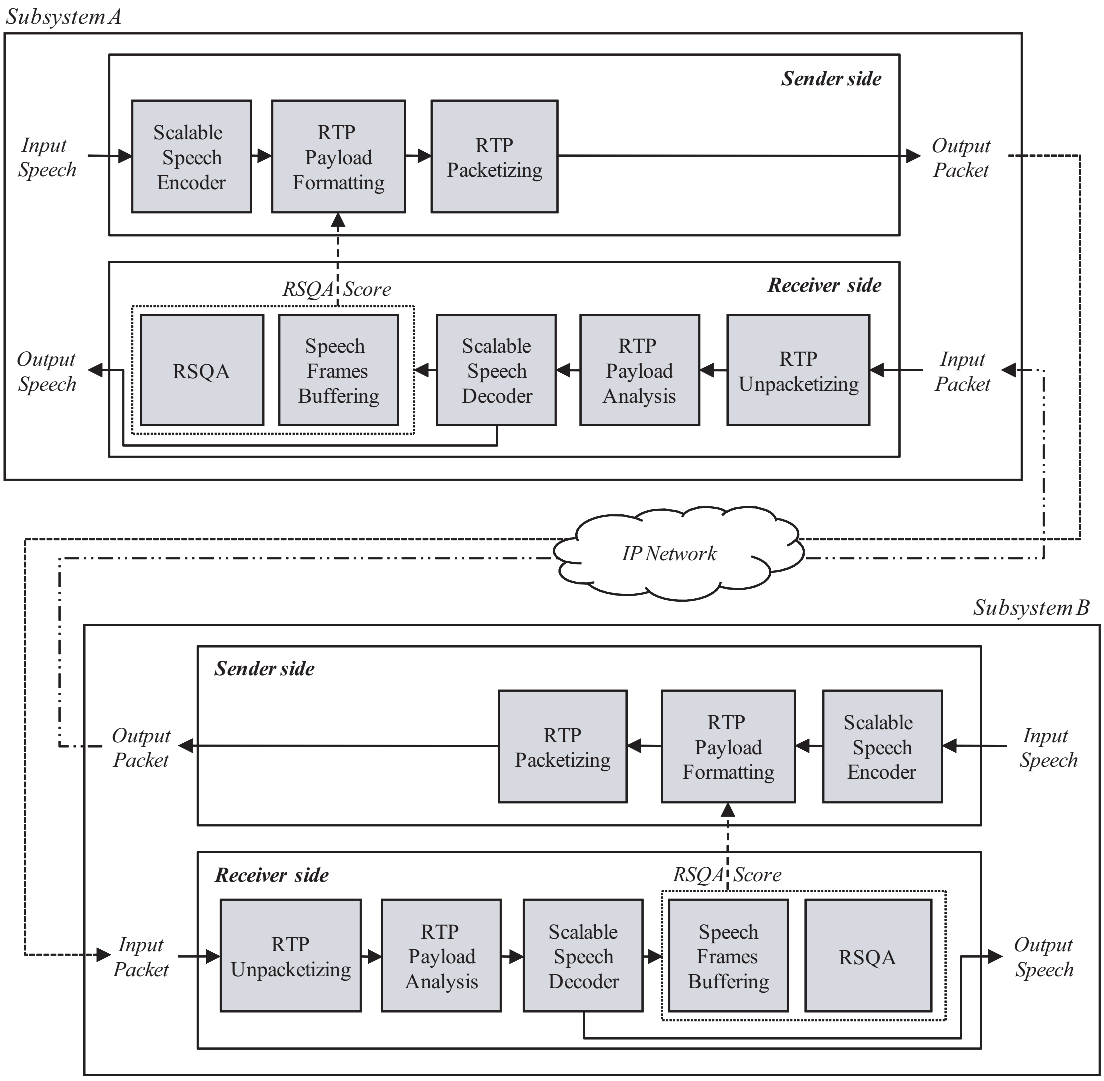


FIGURE 1. Packet flow for an RSS system employing the proposed APLR method, where *Subsystems A* and *B* represent the two communication parties

2.2. **Payload format.** As mentioned in Section 2.1, an RSS system employing the proposed APLR method can have an indicator for a scalable bitrate of speech coding. Moreover, in order to deliver RSQA scores from *Subsystem A* to *Subsystem B*, and vice versa, there should be fields reserved in the format to accommodate the transmission of RSD bitstreams and RSQA scores. Thus, we select the RTP payload format defined in IETF RFC 3267 for the AMR-NB speech codec [13], as shown in Figure 2.

In the payload format, an 'F|FT|Q' sequence of control fields is used to describe each speech frame data. Note here that a codec mode request (CMR) field is applied to all

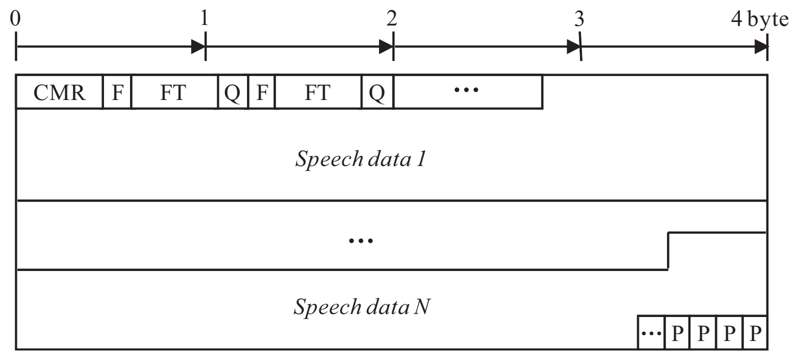


FIGURE 2. Example of RTP payload format for the AMR-NB speech codec defined in RFC 3267

speech frame data. In other words, a one-bit F field indicates whether this frame is to be followed by another speech frame data (F = 1) or if it is the final speech frame data (F = 0). In addition, an FT field, comprised of 4 bits, then indicates if this frame is actually coded by a speech encoder or if it is just comfort noise. That is, bits in this field are assigned from 0 to 7, corresponding to encoding bitrates of 4.75, 5.15, 5.90, 6.70, 7.40, 7.95, 10.2 and 12.2 kbit/s, respectively. However, if comfort noise is encoded, the assigned number only ranges from 8 to 11. Note also that the number 15 indicates the condition in which there is no data to be transmitted, and that the numbers 12 to 14 are reserved for future use. Next, a Q field, a one-bit indicator of the speech quality, is set at 0 when the speech frame data are severely damaged. Otherwise, it is set at 1. Finally, the CMR field, comprised of four-bits, is used to deliver a mode change signal to the speech encoder. For example, it is set to one of eight encoding modes, corresponding to different bitrates of the AMR-NB speech codec. At the end of the payload, P fields are used to ensure octet alignment.

In order to realize the proposed APLR method in this payload format, two new frame indices for the RSD bitstream and RSQA score are incorporated into the FT field, which are denoted using the numbers 12 and 13, respectively.

The use of the RTP payload format described above has several advantages. First, the control ability for a speech encoder, such as the CMR field, is retained for the employed speech codec in the implemented RSS system. Next, the overhead of the control fields for each RSD bitstream is required to be as small as six-bits in 'F|FT|Q'. Finally, no additional transport protocol for the speech quality feedback is needed since this feedback is conducted using RTP packets that are currently being used to deliver the speech bitstreams. Therefore, the transmission overhead for the speech quality feedback is significantly reduced compared with existing transport protocols designed for feedback such as the RTP control protocol (RTCP) [14].

## 3. **Proposed APLR Method.**

### 3.1. **Packet loss recovery and RSQA at the receiver side.**

Figure 3 shows the procedure for packet loss recovery and RSQA at the receiver side for the proposed APLR method. First, the packet loss occurrence is verified through RTP packet analysis. Then, the received CSD bitstreams are decoded if it is decided that there are no packet losses. On the other hand, if it is decided that there are packet losses, the lost speech signals are recovered by using the RSD bitstreams or by using the packet loss concealment (PLC) algorithm of the speech codec, depending on the availability of the RSD bitstreams. Finally, the speech decoder reconstructs the speech frame data from these CSD bitstreams, and the RSQA is conducted using the speech frame data once the amount of speech frame data is sufficient for computing an RSQA score.

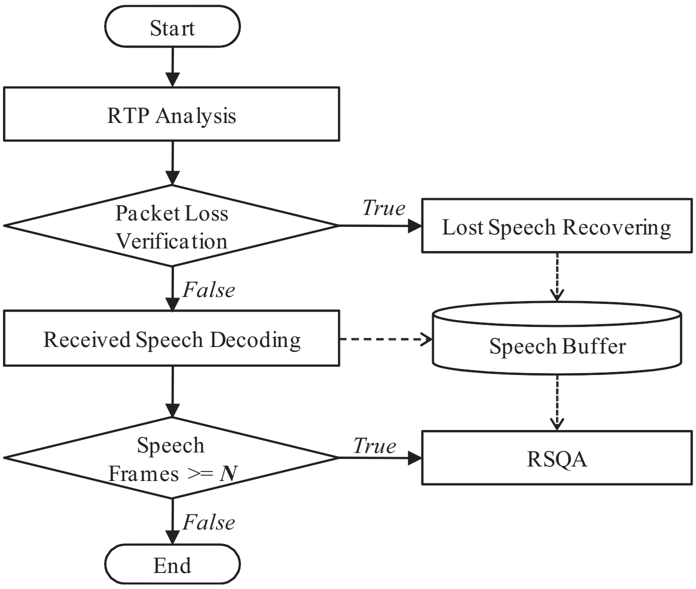


FIGURE 3. Procedure for packet loss recovery and RSQA at the receiver side

For the RSQA computation, the speech frame data in the speech buffer are used by overlapping, as shown in Figure 4. In the figure, $\hat{s}(m)$ is the $m$-th speech frame data input to the speech buffer, $N$ is the total number of frames to be used for each RSQA, and $P$ is the number of frames to be overlapped for the next RSQA. In other words, the RSQA is conducted when every $(N{-}P)$ frames are newly received from the opposite RSS system.

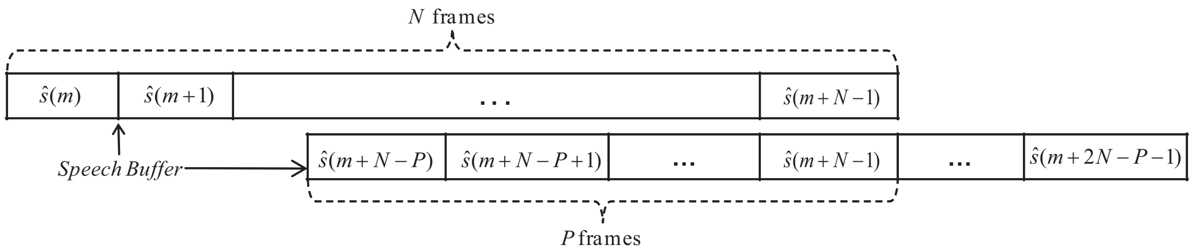


FIGURE 4. Overlap of speech frame data for the RSQA at the receiver side

3.2. **Scalable speech coding and adaptive RSD transmission at the sender side.** Figure 5 presents a flowchart illustrating how to transmit scalable speech coding bitstreams and the RSD bitstreams at the sender side for the proposed APLR method. First, the sender side estimates the PLR by comparing the received RSQA score with a predefined threshold, $\boldsymbol{MOS_T}$, and changes the bitrate of scalable speech coding according to the estimated PLR. In other words, when the received RSQA score $\mathrm{MOS}_S$ is higher than $\boldsymbol{MOS_T}$, the CSD bitstreams are encoded with the highest bitrate $Bitrate_F$ with no RSD transmission as shown in Figure 6. However, when $\mathrm{MOS}_S$ is lower than $\boldsymbol{MOS_T}$, the CSD bitstreams are encoded with a smaller bitrate $Bitrate_H$ than $Bitrate_F$, which enables the remaining bitrate $Bitrate_R$ to be assigned for the RSD transmission. Finally, the RTP payload format described in Section 2.2 is configured according to this adaptive RSD transmission.

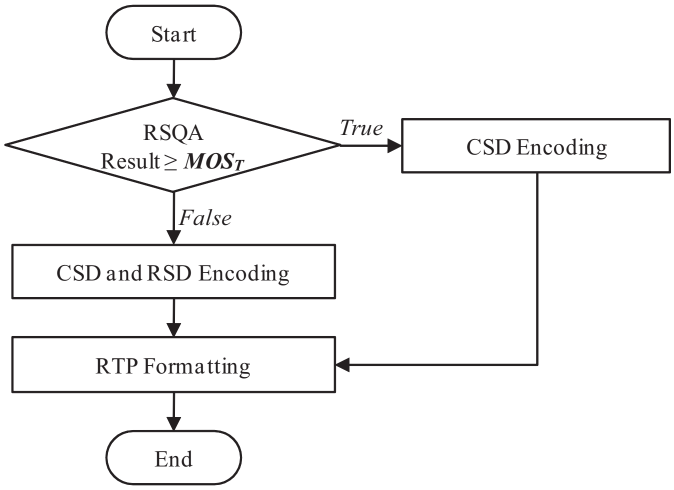


Figure 5. Flowchart of scalable speech coding and adaptive RSD transmission at the sender side

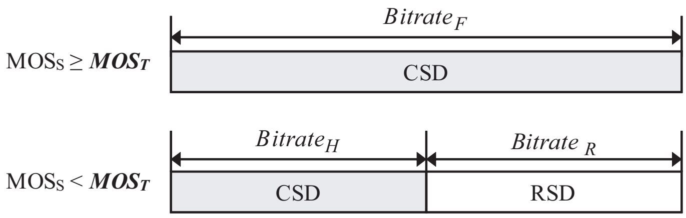


Figure 6. Bitrate assignment according to the estimated PLR using the received RSQA score

As described above, at the sender side of the RSS system, the PLR is estimated from the RSQA scores that are periodically reported from the opposite RSS system. This PLR estimation is effective for the following reasons. First, the speech quality of the received speech data is strongly correlated with the PLR [15]. Second, the speech quality measured as mean opinion scores can be considered as a clearer indicator of the speech quality than any other parameter in the RSS system [16].

4. **Performance Evaluation.** In order to demonstrate the effectiveness of the proposed APLR method, the AMR-NB speech codec [11] was used as a scalable speech codec. The AMR-NB speech codec was operated at eight bitrates ranging from 4.75 to 12.2 kbit/s, and the RSQA was performed using ITU-T Recommendation P.563 [12]. In this experiment, the input speech signals were sampled at 8 kHz and encoded using the AMR-NB speech codec at a bitrate of 10.2 kbit/s. In order to apply the ITU-T Recommendation P.563, $N$ was set to 200 frames, corresponding to 4 seconds, which satisfied the requirements of ITU-T Recommendation P.563 [12]. In addition, $\boldsymbol{MOS_T}$ was set to 3.6, regarded as the guidance value of speech communication availability [17]. Finally, the RSD bitrate $Bitrate_R$ was set to 4.75 kbit/s, which was almost half the bitrate of 10.2 kbit/s. For the CSD bitstreams, when the estimated PLR is assume to be high the bitrate $Bitrate_H$ was also set to 4.75 kbit/s.

Next, we simulated two different packet loss conditions, including random and burst packet losses. During these simulations, PLRs of 3, 5, 7, 10 and 13% were generated by the Gilbert-Elliot model defined in ITU-T Recommendation G.191 [18]. Under the burst packet loss condition, the burstiness of the packet losses was set to 0.5, and the mean and maximum consecutive packet losses were measured at 1.4 and 4.6 frames, respectively.

4.1. **Performance evaluation of RSQA according to the number of overlapped frames.** As an important factor for the RSQA, we considered an appropriate number of frames to be overlapped in the received speech frame data (Figure 4), denoted as $P$. In addition, $N$ was set to 200 frames, as described above, and $P$ was increased from 0 to 180 frames in a step of 20 frames in order to investigate the effect of the number of overlapped frames on the overall speech quality.

In order to perform the test, we prepared speech data consisting of 48 sentences spoken by 8 speakers (4 males and 4 females) from the NTT-AT speech database [19], where each sentence was about 4 seconds long and sampled at a rate of 16 kHz. These speech data were first filtered using a modified intermediate reference system (IRS) filter followed by an automatic level adjustment [18]. Then, the speech data were down-sampled from 16 to 8 kHz. Each speech sentence was processed by the proposed APLR method, and subsequently evaluated using the perceptual evaluation of speech quality (PESQ) defined by ITU-T Recommendation P.862 [20].

Tables 1 and 2 show the speech qualities measured in MOS using PESQ for the speech data processed by the proposed APLR method for different $P$ values under random and burst PLR conditions, respectively. It was shown from the tables that the proposed APLR method could provide better speech quality with a longer $P$. Moreover, when $P$ was set to 180 frames, the speech quality was the best under both random and burst PLRs. Based on these results, the number of overlap frames, $P$, for the proposed APLR method was set to 180 frames for the experiments in the following subsection.

4.2. **Performance evaluation of the proposed APLR method.** In this subsection, a speech quality comparison was performed on an RSS system employing the AMR-NB speech codec with/without the proposed APLR method. Note here that the PLC algorithm embedded in the AMR-NB speech decoder was always applied without regards to the proposed APLR method. In the test, 48 speech sentences from the NTT-AT speech database were used, and they were prepared by the same procedure described in Subsection 4.1 even if these sentences were different from those used in Subsection 4.1.

As the experimental conditions, the RSS system was first implemented by using the AMR-NB and the ITU-T Recommendation P.563 as a scalable speech codec and the RSQA, respectively. Here, the input speech signals were sampled at 8 kHz and encoded using the AMR-NB speech codec at a bitrate of 10.2 kbit/s. Using the RSQA, the speech data were transmitted for up to 4 seconds (at a period of 0.4 seconds), by overlapping the speech data by as long as 3.6 seconds. Second, the threshold for the estimated speech quality required to transmit the RSD was set to 3.6 MOS [17], and the bitrate was set to 4.75 kbit/s for each RSD and CSD bitstream when the PLR was assumed to be high. Third, the packet loss conditions were generated under PLRs of 3, 5, 7, 10 and 13% with random and burst patterns by using the Gilbert-Elliot channel model [18]. Here, the burstiness of the packet losses was set to 0.5, and the mean and maximum consecutive packet losses were measured at 1.4 and 4.6 frames, respectively, in the burst patterns.

Table 3 compares the speech quality measured in MOS using PESQ for the processed speech data with/without the proposed APLR method under random PLR conditions, where the PLR ranged from 3 to 13%. As expected, if the PLR was 0, i.e., there were no packet losses, the speech quality was not affected by the proposed APLR method, as shown in the first column of the table. In addition, it was shown from the table that as the PLR increased, the speech quality obtained by applying the proposed APLR method was relatively improved compared with that without the proposed APLR method. As a

Table 1. MOS scores measured by PESQ for the speech data recovered with the proposed APLR method using different $P$ values under random PLRs ranged from 3 to 13%

| $P$ (frames) | PLR (%) | | | | | | Average |
|---|---|---|---|---|---|---|---|
| | 0 | 3 | 5 | 7 | 10 | 13 | |
| 0 | 3.64 | 3.18 | 2.99 | 2.80 | 2.76 | 2.52 | 2.85 |
| 20 | 3.64 | 3.21 | 3.13 | 2.97 | 2.85 | 2.78 | 2.99 |
| 40 | 3.64 | 3.20 | 3.13 | 2.92 | 2.86 | 2.77 | 2.98 |
| 60 | 3.64 | 3.22 | 3.13 | 2.95 | 2.86 | 2.78 | 2.99 |
| 80 | 3.64 | 3.15 | 3.03 | 2.81 | 2.81 | 2.63 | 2.88 |
| 100 | 3.64 | 3.19 | 3.10 | 2.94 | 2.86 | 2.77 | 2.97 |
| 120 | 3.64 | 3.18 | 3.10 | 2.99 | 2.89 | 2.83 | 3.00 |
| 140 | 3.64 | 3.16 | 3.12 | 3.03 | 2.90 | 2.82 | 3.01 |
| 160 | 3.64 | 3.18 | 3.11 | 2.99 | 2.91 | 2.84 | 3.00 |
| 180 | 3.64 | 3.18 | 3.14 | 3.05 | 2.90 | 2.85 | 3.02 |

Table 2. MOS scores measured by PESQ for the speech data recovered with the proposed APLR method using different $P$ values under burst PLRs ranged from 3 to 13%

| $P$ (frames) | PLR (%) | | | | | | Average |
|---|---|---|---|---|---|---|---|
| | 0 | 3 | 5 | 7 | 10 | 13 | |
| 0 | 3.64 | 3.28 | 3.10 | 2.87 | 2.79 | 2.57 | 2.92 |
| 20 | 3.64 | 3.15 | 3.03 | 2.81 | 2.81 | 2.63 | 2.88 |
| 40 | 3.64 | 3.17 | 3.04 | 2.87 | 2.77 | 2.61 | 2.89 |
| 60 | 3.64 | 3.17 | 2.99 | 2.86 | 2.80 | 2.57 | 2.88 |
| 80 | 3.64 | 3.25 | 3.11 | 2.89 | 2.84 | 2.62 | 2.94 |
| 100 | 3.64 | 3.19 | 3.02 | 2.91 | 2.81 | 2.67 | 2.92 |
| 120 | 3.64 | 3.19 | 3.02 | 2.94 | 2.83 | 2.66 | 2.93 |
| 140 | 3.64 | 3.18 | 3.02 | 3.91 | 2.82 | 2.65 | 2.92 |
| 160 | 3.64 | 3.17 | 2.99 | 2.92 | 2.81 | 2.73 | 2.92 |
| 180 | 3.64 | 3.17 | 3.03 | 2.88 | 2.83 | 2.72 | 2.93 |

result, the proposed APLR method improved the average speech quality from 2.79 to 3.00 MOS under random PLR conditions.

Table 4 compares the speech quality measured in MOS using PESQ for the decoded speech data with/without the proposed APLR method under burst PLR conditions, where the PLR ranged from 3% to 13%. Similarly to the results shown in Table 3, the RSS system employing the proposed APLR method provided relatively higher MOS scores for higher PLRs, compared with that without the proposed APLR method. Consequently, it was shown from this experiment that the proposed APLR method could improve the average speech quality by as much as 0.21 and 0.15 MOS under random and burst packet loss conditions, respectively, compared with that of an RSS system without the proposed APLR method in which only the AMR-NB PLC algorithm was used. Thus, it could be concluded here that the proposed APLR method improved the speech quality under both random and burst packet loss conditions, compared with the AMR-NB PLC algorithm. Moreover, it did not require any additional transmission overhead.

Table 3. MOS scores measured by PESQ for the speech data recovered with/without the proposed APLR method under random PLRs ranging from 3 to 13%

| Method | PLR (%) | | | | | | Average |
|---|---|---|---|---|---|---|---|
| | 0 | 3 | 5 | 7 | 10 | 13 | |
| Without the proposed APLR method | 3.66 | 3.15 | 2.94 | 2.79 | 2.60 | 2.46 | 2.79 |
| With the proposed APLR method | 3.66 | 3.16 | 3.11 | 3.05 | 2.90 | 2.81 | 3.00 |
| Improvement | 0.00 | 0.01 | 0.17 | 0.26 | 0.30 | 0.35 | 0.21 |

Table 4. MOS scores measured by PESQ for the speech data recovered with/without the proposed APLR method under burst PLRs ranging from 3 to 13%

| Method | PLR (%) | | | | | | Average |
|---|---|---|---|---|---|---|---|
| | 0 | 3 | 5 | 7 | 10 | 13 | |
| Without the proposed APLR method | 3.66 | 3.23 | 2.99 | 2.88 | 2.70 | 2.45 | 2.85 |
| With the proposed APLR method | 3.66 | 3.27 | 3.08 | 3.01 | 2.90 | 2.72 | 3.00 |
| Improvement | 0.00 | 0.04 | 0.09 | 0.13 | 0.20 | 0.27 | 0.15 |

By using one frame of RSD in this experiment, the improvement was marginal for burst packet losses; however, the improvement was significant for higher PLRs under random and burst packet loss conditions. For example, we achieved 0.35 and 0.27 MOS for 13% random and burst PLR conditions, respectively. This improvement was due to the fact that the proposed APLR method could assist the speech decoder to recover lost speech signals by using the transmitted RSD bitstream, when the PLR was estimated to be high. Hence, it could be also concluded that the proposed APLR method was useful for both random and burst packet loss patterns, especially for high PLRs, by adaptively transmitting the RSD bitstreams according to the estimated network conditions.

In order to realize the proposed APLR method on an RSS system, we needed to optimize the computational complexity of the RSQA module included in the proposed APLR method. In the proposed method, the computations for the speech coding were actually reduced because the speech coding bitrate was lowered according to the estimated PLR. However, the proposed APLR method needed additional computations for the PLR estimation. Thus, a computational optimization for the single-ended speech quality assessment should be performed in order to implement the proposed APLR method in real-time on resource-limited devices such as portable multimedia devices.

5. **Conclusions.** In this paper, we proposed an adaptive speech packet loss recovery (APLR) method to improve the speech quality of a real-time speech streaming (RSS) system. The aim of this work was to design an RSS system such that a minimum speech quality was guaranteed for reliable speech communications by considering time-varying network conditions, i.e., the packet loss rate (PLR). Here, the proposed APLR method was optimized to accommodate packet loss recovery capabilities for a given transmission bandwidth and a desired speech quality, and thus it could provide an adaptive redundant

speech frame data (RSD) transmission according to the estimated PLR. For the PLR estimation, the proposed APLR method performed the speech quality assessment for the received speech data. Based on the feedback from the opposite RSS system, the proposed APLR method could then control the RSD transmission and optimize the speech coding bitrate.

To show the effectiveness of the proposed method, we compared the speech quality recovered with/without the proposed APLR method under packet loss conditions. Here, the speech quality without the proposed APLR method was measured when the packet losses were recovered by the packet loss concealment (PLC) algorithm embedded in the AMR-NB speech decoder at the highest bitrate of 10.2 kbit/s. For the proposed APLR method, the bitrate of the AMR-NB was set to 4.75 kbit/s in order to encode each RSD and current speech frame data (CSD) for a comparison with the equivalent conditions of the transmission bandwidth. Consequently, the experiments showed that the speech quality of the RSS system employing the proposed APLR method was improved from 2.79 to 3.00 MOS and from 2.85 to 3.00 MOS under random and burst PLRs, respectively, compared with that of the RSS system without the proposed APLR method. Thus, it could be concluded that the proposed APLR method could improve the speech quality under both random and burst packet loss conditions, compared with the AMR-NB PLC algorithm. Moreover, it did not require any additional transmission overhead.

**Acknowledgment.** This work was supported in part by "Fusion-Tech Developments for THz Information & Communications" Program of the Gwangju Institute of Science and Technology (GIST) in 2011, by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2010-0000135), and by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2010-C1090-1021-0007).

## REFERENCES

[1] K.-T. Yang, W. K. Lai and C.-S. Shieh, A cross-layer approach to packet size adaptation for improved utilization in mobile wireless networks, *International Journal of Innovative Computing, Information and Control*, vol.5, no.11(B), pp.4335-4346, 2009.

[2] D. Gao, Y. Xu and Y. Gu, A novel algorithm of wavelet packet speech enhancement using time-frequency threshold, *ICIC Express Letters*, vol.1, no.2, pp.131-136, 2007.

[3] T. Guan and Q. Gong, A study on the effects of spectral information encoding in mandarin speech recognition in white noise, *ICIC Express Letters*, vol.3, no.3(A), pp.415-419, 2009.

[4] C.-F. Wu, C.-L. Lee and W.-W. Chang, Perceptual-based playout mechanisms for multi-stream voice over IP networks, *Proc. of Interspeech, Antwerp*, Belgium, pp.1673-1676, 2007.

[5] Q. Zhang, G. Wang, Z. Xiong, J. Zhou and W. Zhu, Error robust scalable audio streaming over wireless IP networks, *IEEE Transactions on Multimedia*, vol.6, no.6, pp.897-909, 2004.

[6] J.-C. Bolot, S. Fosse-Parisis and D. Towsley, Adaptive FEC-based error control for Internet telephony, *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, New York, NY, vol.3, pp.1453-1460, 1999.

[7] W. Jiang and H. Schulzrinne, Comparison and optimization of packet loss repair methods on VoIP perceived quality under bursty loss, *Proc. of the 12th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, Miami, FL, pp.73-81, 2002.

[8] C. Yung, H. Fu, C. Tsui, R. S. Cheng and D. George, Unequal error protection for wireless transmission of MPEG audio, *Proc. of IEEE International Symposium on Circuits and Systems*, Orlando, FL, vol.6, pp.342-345, 1999.

[9] J. Hagenauer and T. Stockhammer, Channel coding and transmission aspects for wireless multimedia, *Proceedings of the IEEE*, vol.87, no.10, pp.1764-1777, 1999.

[10] A. Ito, T. Sakai, K. Konno and S. Makino, Packet loss concealment for MDCT-based audio codec using correlation-based side information, *International Journal of Innovative Computing, Information and Control*, vol.6, no.3(B), pp.1347-1361, 2010.
[11] 3GPP, TS 26.090 V6.0.0, *Adaptive Multi-Rate (AMR) Speech Codec: Transcoding Functions*, 2004.
[12] ITU-T Recommendation P.563, *Single-Ended Method for Objective Speech Quality Assessment in Narrow-Band Telephony Applications*, 2004.
[13] IETF RFC 3267, *Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-rate Wideband (AMR-WB) Audio Codecs*, 2002.
[14] IETF RFC 1890, *RTP Profile for Audio and Video Conferences with Minimal Control*, 1996.
[15] L. F. Sun, G. Wade, B. M. Lines and E. C. Ifeachor, Impact of packet loss location on perceived speech quality, *Proc. of the 2nd IP-Telephony Workshop*, New York, NY, pp.114-122, 2001.
[16] L. F. Sun and E. C. Ifeachor, Voice quality prediction models and their application in VoIP networks, *IEEE Transactions on Multimedia*, vol.8, no.4, pp.809-820, 2006.
[17] D. P. Hole and F. A. Tobagi, Capacity of an IEEE 802.11b wireless LAN supporting VoIP, *Proc. of IEEE International Conference on Communications*, Paris, France, vol.1, pp.196-201, 2004.
[18] ITU-T recommendation G.191, *Software Tools for Speech and Audio Coding Standardization*, 1996.
[19] NTT-AT, *Multi-Lingual Speech Database for Telephonometry*, 1994.
[20] 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*, 2001.