

A Low Complexity Packet Loss Recovery Method for Audio Transmission

Xuming Lu, Huijian He, Hongzhou Tan
 School of information science and technology
 Sun Yat-Sen University
 Guangzhou, China
 E-mail: luxuming@189.cn

Abstract—A packet loss recovery technique based on interleaving is proposed for uncompressed high-fidelity audio transmission. The audio samples are interleaved for transmission to avoid burst errors. The lost samples in the receiver are estimated by using quadratic fit interpolation (QFI). Finally a low pass filter (LPF) is applied to the estimated data for suppressing high-frequency aliasing caused by data loss. Objective evaluation of the performance is given by the peak signal-to-noise ratio (PSNR). Simulation results show that the proposed method reconstructs the lost data effectively, and has good performance with respect to the PSNR.

Keywords—packet loss recovery; interpolation; interleaving; low pass filter

I. INTRODUCTION

With the development of communication technology, the data rates of communication systems keep increasing, which gives rise to the demand of high-quality multimedia services [1], such as high-definition video, high-fidelity music, etc. However, absolutely reliable communication does not exist, especially in a wireless communication environment. Due to channel fading, noise interference, and transmission collisions, packet loss is a very common problem. In the transmission of high-fidelity audio, occasional packet loss will result in quality degradation [2]. Therefore, an appropriate packet loss recovery algorithm is required in these audio transmission systems.

Many techniques have been researched and applied for packet loss in audio transmission. Recovery techniques may be divided into two classes: sender-based repair techniques, including retransmission, forward error correction (FEC) and interleaving; and receiver-based error concealment techniques, including insertion, interpolation and regeneration [3]. Error concealment schemes rely on producing a replacement for the lost packet. This is possible since audio signals exhibit large amounts of short-term self-similarity [4]. In the absence of extra information from the sender, receiver-based error concealment techniques just try to conceal the error rather than correct it. Therefore, sender-based techniques outperform receiver-based techniques, generally. Among sender-based techniques, FEC is the most effective one. It consumes lower bandwidth compared with retransmission, and provides an exact recovery whereas interleaving just provides an approximate recovery. FEC has been adopted in many applications, especially in the architectures that have codecs in the transmission [5].

However, adding an extra codec for FEC in the transmission of high-quality audio will greatly increase the complexity. For hardware implementation, the high complexity limits the application of FEC.

In this paper, a low complexity packet loss recovery scheme is proposed for uncompressed audio transmission. Pulse code modulation (PCM) audio data are interleaved in the sender. Then the lost audio data are recovered by quadratic fit interpolation (QFI) in the receiver. The estimated data are processed by a low pass filter (LPF) to reduce the effect of high-frequency aliasing caused by data loss. To evaluate the performance, the proposed algorithm is simulated to analyze the peak signal-to-noise ratio (PSNR) under different packet loss rates.

The rest of the paper is organized as follows. The model of audio transmission systems is introduced in Section II. In Section III a packet loss recovery scheme of low complexity is described in detail. Section IV gives the simulation results. The conclusion is presented in Section V.

II. SYSTEM MODEL

Fig. 1 is the model of an audio transmission system. Audio data are first processed by pre-process module. Then media access control (MAC) module controls the data transmission over the physical layer (PHY). PHY in the receiver demodulates the signals from channel, and uploads the data to MAC. Finally the received data are processed in post-process module to reconstruct the transmitted audio.

The pre-process module processes source data for transmission, such as source coding for compression, forward error correction coding, and data encapsulation for upper layer protocols. MAC controls data transmission. In the sender, it implements channel state detection, conflict judgment, retransmission, etc. In the receiver, it checks the correctness of the received data, and determines the destination of the received packet, etc. PHY performs modulation and demodulation according to the physical channel. The post-process module is basically the inverse process of the pre-process module.

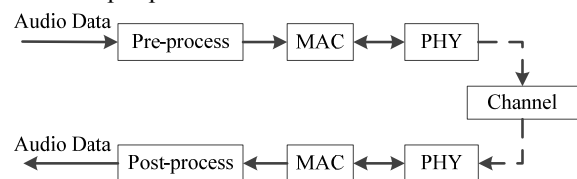


Figure 1. Model of audio transmission systems

In practice, there is no guarantee that data packets are all received correctly, although both PHY and MAC have taken efforts to decrease the packet loss rate [6]. For general TCP/IP applications, reception of the packets can be more reliably with upper layer protocols. However, it increases the complexity and time delay, and limits its use in real-time applications.

For the transmission of high-fidelity audio, a new transmitting scheme is proposed. The PCM data are interleaved in the pre-process stage in Fig. 1, whereas the data are recovered by de-interleaving and interpolation of lost packets in the post-process stage in the receiver.

III. PACKET LOSS RECOVERY SCHEME

In frame-based data transmission, a packet loss in the transmission means a burst of errors. Interleaving is an effective method to avoid burst errors. It disperses burst errors into data stream, and makes it possible to estimate the lost data with the received data.

The interleaving is based on data blocks. Each block consists of $M \cdot N$ samples. Here, M represents the number of packets for a block, and N represents the number of samples in a packet. Let k , i and j be the index of a sample at the original sequence, index at the interleaved stream, and index at the de-interleaved stream, respectively. Then the permutations of interleaving and de-interleaving can be formulated as follows:

$$i = \text{mod}(k, M) \cdot N + \text{floor}(k/M), \quad (1)$$

$$j = \text{mod}(i, N) \cdot M + \text{floor}(i/N), \quad (2)$$

where $\text{mod}(x, y)$ gets modulus after x divided by y , and $\text{floor}(x)$ keeps the integer part of the real number x . $k, i, j = 0, 1, \dots, M \cdot N - 1$.

Fig. 2 shows the interleaving and de-interleaving procedure with $M=4$ and $N=4$. It is shown that the loss of a single packet from the interleaved stream results in multiple small gaps in the reconstructed stream, as opposed to the single large gap in a non-interleaved stream. For the error concealment in the receiver, a single lost sample can be well estimated by using its surrounding samples.

Considering the packet loss rate in practice, the number of packets for a block is set to 3, i.e., $M=3$. The number of samples N in a packet depends on the physical layer. Assume that only a single packet is lost in the 3-packet block. Fig. 3 shows the flow of the proposed packet loss recovery scheme, where packet 2 is lost, and packet 1 and 3 are used to restore the lost packet. In Fig. 3, b_k denotes the estimation of the lost sample a_k , and c_k is the output of the LPF corresponding to b_k .

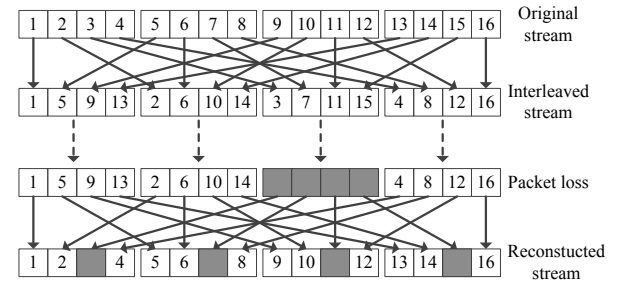


Figure 2. Interleaving and de-interleaving of multiple packets

There are several methods to find b_k : silent substitution, repetition, and average interpolation. Silent substitution fills the gaps with zeros directly. Repetition replaces a lost sample with the copy of its previous one. Average interpolation generates the estimation by averaging the two samples on both sides of the lost sample.

To simplify the description, let a_0 be the lost sample at sampling time t_0 . Its estimated value is denoted by b_0 . The four samples around it are represented as (t_{-2}, a_{-2}) , (t_{-1}, a_{-1}) , (t_1, a_1) , (t_2, a_2) . Therefore, silent substitution, repetition, and average interpolation can be expressed, respectively, as follows:

$$b_0 = 0, \quad (3)$$

$$b_0 = a_{-1}, \quad (4)$$

$$b_0 = \frac{1}{2}(a_{-1} + a_1). \quad (5)$$

In order to achieve more smooth interpolation result with the adjacent samples, quadratic fit is used to estimate the lost sample. Suppose the fitting curve is

$$a = k_2 t^2 + k_1 t + k_0, \quad (6)$$

where k_2 , k_1 , and k_0 are real coefficients. Substituting the four samples (t_{-2}, a_{-2}) , (t_{-1}, a_{-1}) , (t_1, a_1) , (t_2, a_2) into (6), it then results in

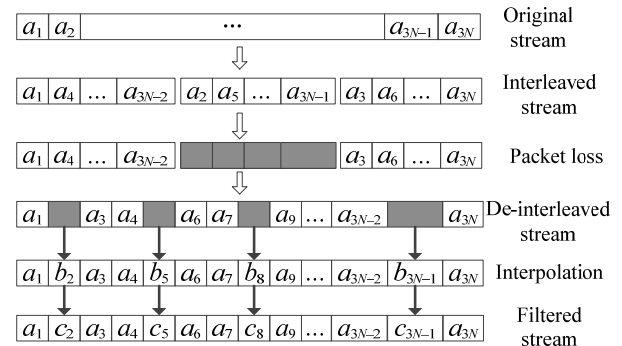


Figure 3. The proposed packet loss recovery scheme

$$\mathbf{a} = \begin{bmatrix} t_{-2}^2 & t_{-2} & 1 \\ t_{-1}^2 & t_{-1} & 1 \\ t_1^2 & t_1 & 1 \\ t_2^2 & t_2 & 1 \end{bmatrix} \mathbf{k}, \quad (7)$$

where $\mathbf{a}=[a_{-2}, a_{-1}, a_1, a_2]^T$, $\mathbf{k}=[k_2, k_1, k_0]^T$. Let \mathbf{T} denote the matrix in (7), then the least squares solution of \mathbf{k} is given by

$$\mathbf{k} = \arg \min_{\mathbf{k}} \{(\mathbf{a} - \mathbf{T}\mathbf{k})^T (\mathbf{a} - \mathbf{T}\mathbf{k})\}. \quad (8)$$

By taking the derivation on the block in the braces of (8) and setting the result to zero, the solution becomes

$$\mathbf{k} = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{a}. \quad (9)$$

Without loss of generality, let $t_{-2}=2$, $t_{-1}=1$, $t_0=0$, $t_1=1$, $t_2=2$. Then equation (9) becomes

$$\mathbf{k} = \begin{bmatrix} 1/6 & -1/6 & -1/6 & 1/6 \\ -1/5 & -1/10 & 1/10 & 1/5 \\ -1/6 & 2/3 & 2/3 & -1/6 \end{bmatrix} \mathbf{a}. \quad (10)$$

Since the estimation of the lost sample is $b_0 = f(0) = k_0$. With the expression in (10), the estimation can be obtained by

$$b_0 = \frac{2}{3}(a_{-1} + a_1) - \frac{1}{6}(a_{-2} + a_2). \quad (11)$$

As shown in (11), because the coefficients are constant, the multiply operation can be realized by shift-and-add circuits. Since no multiplier is required, the complexity is low for hardware implementation.

As shown in Fig. 3, the de-interleaved audio sequence with a packet lost is essentially a down-sampled audio sequence. According to the theory of multirate digital signal processing, the down-sampled signal will be corrupted by high-frequency aliasing [7]. In order to suppress the aliasing artifact after interpolation, an appropriate low pass filter can be applied to the interpolation results.

The transfer function of the ideal low pass filter and its impulse response can be expressed as

$$H_d(e^{j\omega}) = \begin{cases} 1 & |\omega| \leq \omega_c \\ 0 & \omega_c < |\omega| \leq \pi \end{cases}, \quad (12)$$

$$h_d(n) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{j\omega n} d\omega = \frac{\sin \omega_c n}{\pi n}. \quad (13)$$

Here, ω_c is the cutoff frequency of the filter. Since the length of $h_d(n)$ is infinite, it can not be used in actual

practice. In order to get finite impulse response, a rectangular window $w(n)$ can be applied to $h_d(n)$.

$$w(n) = \begin{cases} 1 & -N \leq n \leq N \\ 0 & \text{otherwise} \end{cases}, \quad (14)$$

where $2N+1$ is the length of $w(n)$. Therefore, the windowed filter coefficients are

$$h(n) = c \cdot h_d(n) \cdot w(n), \quad (15)$$

where c is a constant for energy normalization.

$$c = \left[\sum_{k=-N}^{k=N} h_d(k) \right]^{-1}. \quad (16)$$

In the case of one packet lost in a 3-packet block, it is reasonable to set the cutoff frequency $\omega_c = 2\pi/3$ and $N=5$. With (13) and (15), the corresponding tap coefficients from -5 to 5 are $[0.0568, 0.0711, 0, 0.1421, 0.2842, 0.6874, 0.2842, 0.1421, 0, 0.0711, 0.0568]$. It is interesting to note that $h(-3)$ and $h(3)$ are 0, and the samples at these corresponding positions are obtained by interpolation. Therefore, only the received samples and b_0 contribute to the LPF. The amplitude-frequency response of the LPF with parameters above is plotted in Fig. 4.

IV. EXPERIMENTS RESULTS

This section simulates the proposed packet loss recovery algorithm. The performances of different algorithms, including silent substitution, repetition, and average interpolation, are compared. The audio sources are PCM data which are sampled at frequency 44.1 kHz and 16 bits per sample.

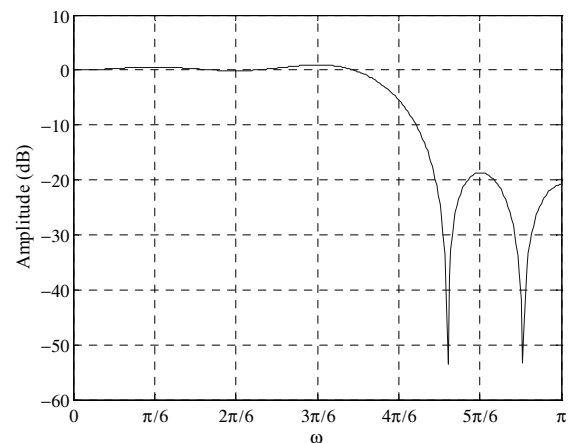


Figure 4. Amplitude-frequency response of the LPF

The influence of different algorithms on the signal spectrum is studied first. A 3-packet block with the second packet lost is chosen for simulation. The lost data are recovered by different algorithms. Then the spectrums of recovered results are calculated, as illustrated in Fig. 5. Fig. 5(a) is the spectrum of the original signal. Fig. 5(b) is the part of the spectrum labeled by the circle in Fig. 5(a). The method of silent substitution introduces a lot of high-frequency noise, as shown in Fig. 5(c). The results of repetition method, average interpolation, and the proposed QFI are drawn in Figs. 5(d), 5(e), and 5(f), respectively. It is shown that QFI has the best result from these subfigures. As shown in Fig. 5(g), high-frequency aliasing caused by data loss is further suppressed by filtering the QFI results with LPF.

In order to evaluate the performance objectively, the simulation results are given by PSNR, which is defined as

$$\text{PSNR} = 10 \log_{10} \left[\frac{(2^D - 1)^2}{\text{MSE}} \right], \quad (17)$$

where D is the number of bits for a sample, i.e., 16 in this simulation. MSE is the mean square error between the original data and the restored data.

In the simulation, the length of a packet is set to 256 audio samples. 12000 packets are processed to get the results. Fig. 6 shows the results of PSNR versus the packet loss rate.

As shown in Fig. 6, with the packet loss rate increasing, the performances of all methods decrease. Given a packet loss rate, the silent substitution has the worst performance, whereas QFI performs better than repetition and average interpolation. It is also shown that the LPF improves the performance of QFI effectively. This is because the original audio contains few high-frequency components as shown in Fig. 5. If the original audio contains more high-frequency components, the improvement from the LPF will be less. However, considering the aliasing caused by packet loss, a LPF can suppress the high-frequency noise and improve the quality of the restored audio, even for the audio with rich high-frequency components.

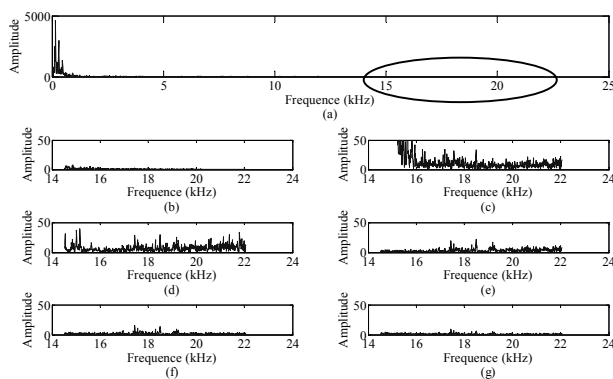


Figure 5. (a) Spectrum of original audio; (b) high-frequency components of original audio; (c) result of silent substitution; (d) result of repetition; (e) result of average interpolation; (f) result of QFI; (g) result of QFI-LPF

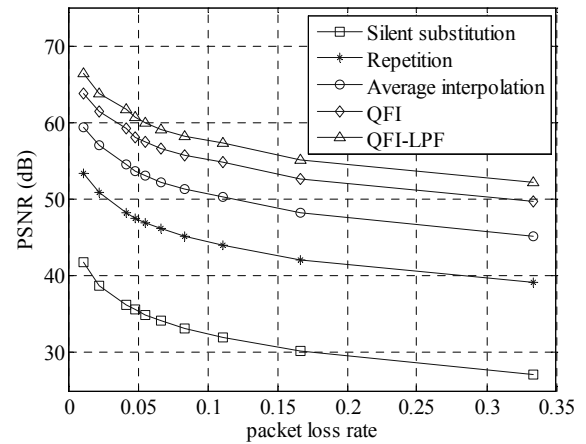


Figure 6. PSNR versus packet loss rate

V. CONCLUSION

For the transmission of uncompressed high-fidelity audio, a packet loss recovery scheme has been proposed. The audio samples are interleaved to avoid burst errors caused by packet loss. The quadratic fit interpolation with a LPF is applied for recovering the lost samples. It has been shown that the proposed method restores the lost samples effectively and has good performance with respect to the PSNR. The advantage of the proposed scheme is the low complexity, because it can be realized by shift-and-add circuits. Therefore, it is suitable for hardware implementation in high-fidelity audio transmission.

REFERENCES

- [1] A. Khalifeh and H. Yousefi'zadeh, "Optimal audio transmission over error-prone wireless links," *IEEE Trans. Multimedia*, vol. 12, Apr 2010, pp. 204-214.
- [2] R. R. Pastrana-Vidal and C. Colomes, "Perceived Quality of an Audio Signal Impaired by Signal Loss: Psychoacoustic Tests and Prediction Model," *International Conference on Acoustics, Speech and Signal Processing (ICASSP 07)*, 2007, pp. 277-280.
- [3] C. Perkins, O. Hodson, and V. Hardman, "A Survey of Packet Loss Recovery Techniques for Streaming Audio," *IEEE Network*, vol. 12, October 1988, pp. 40-47.
- [4] B. Rahnama, and A. Elci, "A novel no-latency simple-to-implement sender-based packet-loss recovery technique for multimedia streams," *International Symposium on Computer Networks*, IEEE Press, 2006, pp. 192-196.
- [5] J. Liu, S. Zhao, J. Wang, and J. Kuang, "FEC-based packet loss recovery for AVS-M Audio codec," *International Conference on Multimedia Technology (ICMT)*, 2011, pp. 3069-3072.
- [6] IEEE STANDARD 802.11-2012 - IEEE Standard for Information technology--Telecommunications and information exchange between systems Local and metropolitan area networks--Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Available: <http://standards.ieee.org/about/get/802/802.11.html>.
- [7] F. J. Harris, *Multirate Signal Processing for Communication Systems*. Upper Saddle River, New Jersey: Prentice Hall PTR, May 2004.