

Stochastic Packet Reconstruction for Subjectively Improved Audio Delivery over WLANs

Andreas Floros
Dept. of Audiovisual Arts
Ionian University
49 100, Corfu, Greece
+30 26610 23724
floros@ionio.gr

Markos Avlonitis
Dept. of Informatics
Ionian University
49 100, Corfu, Greece
+30 26610 47145
avlon@ionio.gr

Panayiotis Vlamos
Dept. of Informatics
Ionian University
49 100, Corfu, Greece
+30 26610 47145
vlamos@ionio.gr

ABSTRACT

Real time digital audio delivery over Wireless Local Area Networks (WLANs) represents an attractive, flexible and cost effective framework for realizing high-quality, multichannel home audio applications. However, the unreliable nature of WLANs IP link frequently imposes significant playback quality degradation, due to delay or permanent loss of a number of transmitted digital audio packets. In this paper, a novel packet error concealment technique is presented, based on the spectral reconstruction of the statistical equivalent of a previously successfully received audio data packet. It is shown that the proposed data reconstruction scheme outperforms previously published error concealment strategies, in both terms of objective and perceptual criteria.

Keywords

WLAN, Audio streaming, Packet reconstruction

1. INTRODUCTION

The advantages of digital audio technology are nowadays extended to new forms of audio content delivery such as packet-based network and Internet Protocol (IP) distribution. Packet-based audio delivery represents the most attractive approach for establishing multi-point connections that are impossible to realize using legacy, sequential digital interconnections (i.e. S/PDIF). In practice, such setups allow the music delivery from any digital audio source within a typical multi-room home environment to any networked speaker/playback device. Hence, any network-enabled audio device can be connected to any other using the common network interface also used for computer communications and Internet access [1].

The recent ratification of a number of protocols for high-rate guaranteed data transmissions over Wireless Local Area Networks (WLANs) [2]–[3] has launched the development of wireless IP network interfaces. This further extends the range of

the supported audio applications, by providing flexible and wire-free means of digital audio delivery. However, in order to achieve wireless digital audio streaming, a number of issues have to be resolved related to absolute clock [4] and packet synchronization [5] among the multiple audio receivers. Specifically, in order to maintain the relative receivers' playback synchronization a significant number of silence gaps are introduced. Moreover, although out-of-order packets are not presented in a WLAN streaming environment, the permanent loss (or excessive delay) of a number of data packets may also introduce additional silence gaps and clicking sounds imposed by the samples discontinuity between two non-successive packets [5].

The overall audio playback quality over WLANs can be improved by error concealment techniques, which aim to efficiently reconstruct the lost packets or insert non-zero audio data samples instead of silence. Obviously, the optimum reconstruction performance is achieved when the lost packet is substituted by a calculated audio segment that is perceptually indistinguishable, or sufficiently similar to the original one.

In this work, a novel strategy is introduced for reconstructing single missing, high-quality PCM-coded audio packets. The proposed technique (termed here as Stochastic Packet Reconstruction – SPARE) is based on the reconstruction of a missing audio data segment using a stochastic equivalent of the previously successfully received data packet in the frequency domain. As it will be shown in the following paragraphs, the SPARE packet reconstruction scheme significantly improves the audible effect of temporal playback muting during the delivery of audio data over WLANs.

The rest of the paper is organized as following: Section 2 presents a brief overview of error concealment techniques applied to audio and speech applications, while Section 3 analyzes and outlines the proposed SPARE reconstruction scheme. In Section 4, typical objective and subjective results are shown and the effect of the SPARE strategy on the final audio playback quality is assessed and compared to other error concealment methods. Finally, the conclusions are summarized in Section 5.

2. ERROR CONCEALMENT OVERVIEW

As mentioned previously, during the real-time wireless transmission of audio data, some packets may be lost or delayed for a time interval greater than their intended playout time. In order to minimize the above audible effect of missing frames, an

error concealment technique must be employed. According to [6], the error concealment techniques for audio and video communications can be categorized in two major classes depending on their interactions with source coders: source-code dependent and source-code independent methods. For source-code independent algorithms, no prior awareness of the coding algorithms is necessary. On the contrary, source-code dependent techniques achieve the concealment by utilizing specific coding features. Additionally, source-code independent methods can be further separated into sender-based, receiver-based and sender receiver-based, according to where the error concealment strategy is applied.

Sender receiver-based concealment algorithms are more effective when priority-based schemes are applied that assign different priorities to different streaming packets, or subclass several characteristics of compressed domain audio bitstreams within secondary bitstream transmissions [7]. However, as the bandwidth available through a typical WLAN allows the transmission of uncompressed PCM-coded audio, such algorithms are not optimized for audio / music applications. Instead, receiver-based methods are preferred for audio IP-based transmissions. Typical receiver-based error concealment methods for audio/music applications include white noise generation or zeros padding [8], pattern matching using small segments of samples immediately before or after lost packets [9] – [10] or repeating the last successfully received packet [11]. Moreover, a number of previously published packet concealment techniques [12] – [13] employ time-based windowing functions for improving the audibility of the packet inserted, while other additionally use interleaving techniques for enhancing the concealment performance under the presence of burst packet losses [14].

In this work the novel SPARE receiver-based error concealment method is introduced, which reconstructs a missing audio data packet by calculating a statistical equivalent replica of the spectrum of the last successfully received audio packet. As mentioned previously, the SPARE reconstruction scheme is designed for real-time audio distribution over WLANs, where no out-of-order packets are presented. Hence, only lost or delayed packets are considered in this work.

3. STOCHASTIC RECONSTRUCTION

The main idea of the SPARE reconstruction algorithm is to reconstruct a missing packet by using as a source signal the last successfully received audio packet. The procedure lies in the assumption that in most cases there is a relative continuity between successive packets regarding their statistical properties in time. To this end we propose the reconstruction of the missing packet in the sense that its statistical properties coincides with those of the last successfully received one. This can be done by means of a novel stochastic generator machine reported recently [15], which is briefly described in the next paragraph.

3.1 Stochastic generator analysis

Assume an arbitrary random signal. A statistical treatment by means of a Taylor expansion around the mean in terms of an arbitrary white noise process is performed. As a result, an analytical decomposition of the initial signal can be derived, where the appropriate signal information is included into the corresponding expansion coefficients [15].

Indeed, let S be the original audio signal, the statistical properties of which must be estimated. Quite general, we may assume that $S=S(W)$, where W is a white noise stochastic process with the standard statistical properties, i.e.

$$\langle W \rangle = 0 \quad (1)$$

and

$$\langle W_i W_j \rangle = \sigma^2 \delta_{ij} \quad (2)$$

with σ^2 being the noise amplitude. In a first approximation we may approximate $S(W)$ with a parabola performing a Taylor expansion around the mean $\langle W \rangle$ as described in [16] using the equation:

$$S(W) = S(\langle W \rangle) + (S'|_{\langle W \rangle})W + \frac{1}{2}S''|_{\langle W \rangle}W^2 \quad (3)$$

where with prime we denote differentiation. From the above equation, following reference [16] it can be shown that

$$S''|_{\langle W \rangle} = \pm \sqrt{\sigma_s^2 / \sigma^2} \quad (4)$$

where σ_s^2 is the variance of the original input signal.

Taking the time averages of both sides of eq. (3) it is important to note that, as usually assumed in physical systems, we can distinguish two distinct time scales: a) time averages that are calculated for time periods long enough for the rapid fluctuating noise W and b) time averages calculated over sort enough time intervals in order the expansion above has meaning in the vicinity of $S(\langle W \rangle)$, that is

$$\langle S(W) \rangle = S(\langle W \rangle) + \frac{1}{2}S''|_{\langle W \rangle}\sigma^2 \quad (5)$$

where we have used that $\langle W \rangle = 0$ and $\langle W^2 \rangle = \sigma^2$. To estimate $S''|_{\langle W \rangle}$ we apply eq. (5) for the function $S^3(\langle W \rangle)$. Since its second derivative at $\langle W \rangle$ equals to $6S(W)S''|_{\langle W \rangle}^2 + 3S(W)^2 S''|_{\langle W \rangle}$, after some elementary calculations and neglecting higher orders we conclude that,

$$S''|_{\langle W \rangle} = \left(\frac{1}{\sigma^2} - 1 \right) S - \frac{2}{3} \frac{\mu_3}{\sigma_s^2} \quad (6)$$

where the abbreviation $S = S(\langle W \rangle)$ was used and μ_3 is the third central moment of the experimentally observed signal. In order to estimate S we return to eq. (5) and inverting,

$$S = \langle S(W) \rangle - \frac{1}{2}S''|_{\langle W \rangle}\sigma^2 \quad (7)$$

For low noise amplitude, using eq. (6) and substituting finally in eq. (3) we end up with

$$S_{est}(W) = \frac{2}{3}\langle S(W) \rangle + \frac{2}{9}\mu_3 \frac{\sigma^2}{\sigma_s^2} + \left(\pm \sqrt{\frac{\sigma_s^2}{\sigma^2}} \right) W +$$

$$+ \frac{1}{2} \left[\frac{1}{\sigma^2} \left(\frac{2}{3} \langle S(W) \rangle + \frac{2}{9} \mu_3 \frac{\sigma^2}{\sigma_S^2} \right) - \frac{2}{3} \frac{\mu_3}{\sigma_S^2} \right] W^2 \quad (8)$$

From the above analysis it is clear that the original signal $S(W)$ may be written as a polynomial of an arbitrary white noise process where the corresponding coefficients are explicitly expressed as functions of the statistical moments (of any order depending on the desired accuracy) of the initial input signal. Furthermore one can note that the values of the coefficients depend also on the noise amplitude of the arbitrary white noise used.

3.2 Stochastic reconstruction description

Let $s_i(n)$ denote the samples of the i -th transmitted audio data packet with $(i-1)L_p \leq n \leq iL_p - 1$. L_p is the length of the transmitted data packet in samples. Moreover, let us assume that the i -th wirelessly transmitted audio packet was lost, while the $(i-1)$ th data packet transmission was successfully performed. The SPARE reconstruction process produces a data packet $s'_i(n)$ with spectral characteristics equivalent to the successfully received packet $s_{i-1}(n)$. For this reason, the Fast Fourier Transform (FFT) of the $s_{i-1}(n)$ audio signal is calculated, expressed as a complex sequence of numbers $S_{i-1}(k)$, that is

$$S_{i-1}(k) = FFT\{s_{i-1}(n)\} = a_k + jb_k \quad (9)$$

To proceed further the stochastic generator presented in the previous paragraph is used in order to construct two statistically equivalent signals for the real and imaginary parts of eq. (9) respectively, namely a'_k and b'_k . Next, the reconstructed signal in the frequency domain is formed as

$$S'_{i-1}(k) = a'_k + jb'_k \quad (10)$$

and the reconstructed signal $s'_i(n)$ in the time domain is finally produced by simply calculating the inverse FFT transform of $S'_{i-1}(k)$.

Following the mathematical analysis presented in Section 3.1, the reconstructed data packet has the same spectral statistical properties with the last successfully received one. Note that the reconstructed signal depends on two parameters of the stochastic generator: the white noise amplitude used and the time window over which averages are calculated. In practice the values of these parameters are chosen in order optimum characteristics of the reconstructed signal are derived. While the choice of the noise amplitude is directly associated with the amplitude of the fluctuations of the previous audio packet, the size of the average window depends on its scale of fluctuations: for small (or large) scale of fluctuations a relatively small (or large) average window length is appropriate. Specifically, in the present application of audio signals used, it was observed that small average window lengths are appropriate in order to achieve equivalent signal energy between the last successfully received and the reconstructed audio packets.

4. RESULTS

The real-time wireless audio transmission process was performed using computer-based simulations. The transmitted content was 16-bit PCM audio sampled at 44.1kHz. Four different types of audio content were tested, namely Orchestra, Rock, Violin and Vocals. From the networking point of view and following the results obtained in [17], the transmitted packet length value was selected equal to $L_p=882$ samples, a value that represents the optimum choice for wireless audio transmissions over typical WLAN protocols with QoS support. The packet loss probability was equal to 8% in all test cases considered herein.

The noise amplitude parameter of the SPARE algorithm was set to 0.98, while the average calculation window length value was 10 samples. After a sequence of tests it was found that these values produce optimal results in terms of the objective and subjective criteria that are measured below. In order to relatively assess the performance of the proposed SPARE technique, two additional error concealment algorithms were tested, namely the Zeros Reconstruction (ZR) algorithm which performs zeros padding (or muting during playback at the absence of an audio data packet) and the Previous Packet Replication (PPR) concealment scheme which simply repeats the previously successfully received packet as mentioned in Section 2.

Figures 1 and 2 illustrate a typical example of the output of the SPARE algorithm in frequency and time domains respectively, compared with the spectrum and the waveform of the previous successfully received packet $s_{i-1}(n)$ that is used for reconstructing the i -th lost packet. Clearly, the reconstructed signal represents a close (but not identical) statistical match of the previous packet signal.

The effect of the above reconstruction scheme on the audio playback quality was systematically assessed using both objective and subjective (i.e. perceptual) criteria. More specifically, in order to estimate the degree of approximation of the original and the reconstructed audio signal, the reconstruction error in the time domain was calculated as

$$e_i(n) = s'_i(n) - s_i(n) \quad (11)$$

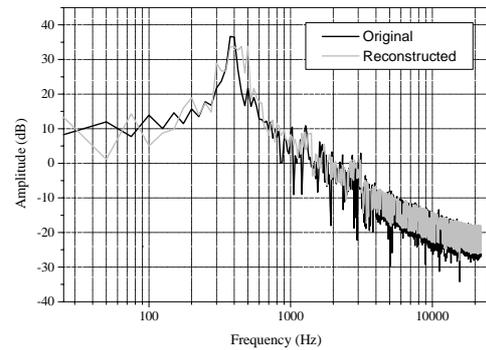


Figure 1. Original vs SPARE reconstructed spectrum

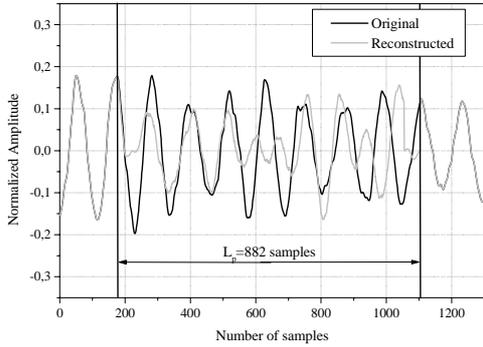


Figure 2. Original vs SPARE reconstructed signal

Additionally, the (%) Reconstruction Energy Error Ratio (REER) was measured, defined as the (%) ratio of the energy of the above error for all lost data packets to the total energy of the original audio signal. The aim of introducing the REER value here is to collectively take into account the error for all lost packets. Moreover, it is obvious that the REER value is directly related to the similarity of the original audio signal with the reconstructed one, as low REER values correspond to low reconstruction error.

Figure 3 shows the measured (%) REER values obtained during the test cases considered in this work. From this Figure it is obvious that the SPARE packet reconstruction algorithm error is minimum for all four kinds of audio material tested. It should be also noted that the ZR algorithm also imposes low reconstruction error (but always higher than the SPARE technique). This is due to the zero padding process which, according to eq. (11) renders the per packet reconstruction error energy equal to the energy of the last successfully received audio packet.

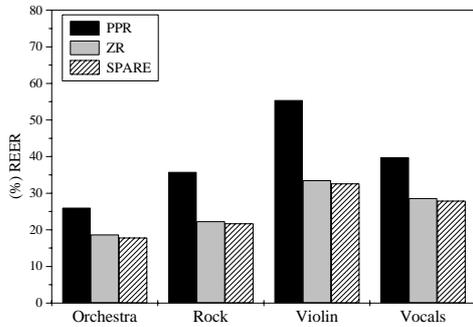


Figure 3. Measured (%) Error Energy Ratio

Although the (%) REER value represents an accurate, objective approximation indicator of the original and the reconstructed audio signals, it provides no information on the audibility of the distortions imposed due to the insertion of the reconstructed packet $s'_i(n)$ in the audio playback queue. Apart from the different frequency content of the reconstructed and the original – lost audio data, the waveform discontinuity that may appear at the first and the last samples of the reconstructed data can potentially impose additional distortions perceived as “clicking” sounds.

In order to assess the audibility of the wireless network distortions, the well-established Noise-to-Mask Ratio (NMR) criterion [18] was employed, which utilizes the masking functions of the human ear and determines the distance of the distortions imposed by any audio system and the masking threshold. For NMR estimation, the original PCM audio signal (prior to wireless transmission) was used as reference and a single NMR value was calculated based on the averaged (segmental) NMR value. It should be noted that NMR values above 0dB indicate the presence of audible distortions, while negative NMR values indicate an audio signal with less audible distortions. The objective metric utilized for the quality assessment of audio streams was based on the averaged (segmental) NMR values.

Figure 4 shows the measured NMR values for all four audio content types considered in this work and for all cases of SPARE, ZR and PPR error concealment mechanisms. From these measurements it is clear that in the case of the SPARE algorithm, the packet-loss perceptual effect is significantly decreased as a reduction of nearly 3dB for the NMR value is observed in all test cases.

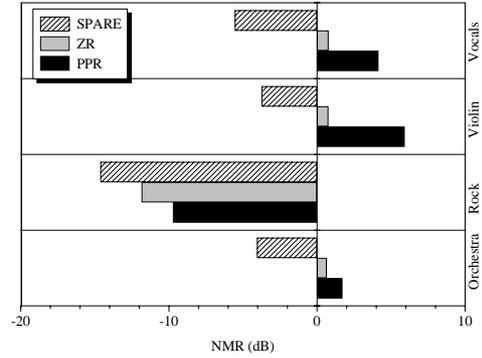


Figure 4. Measured Noise-to-Mask Ratio (dB)

5. CONCLUSIONS

WLANs represent a very flexible and attractive mean for audio delivery, especially following the recent advantages on supported data rates and Quality-of-Service (QoS) support. However, the frequent wireless link quality degradations can lead to excessive packet delivery delays and equivalent audio data losses. In typical multichannel audio playback setups (e.g. stereo or 5.1 home theater environments) in order to overcome the playback quality degradation imposed by the above packet losses, error concealment strategies are applied.

In this paper, the SPARE error concealment technique is proposed, which aims to improve the overall audio playback quality, using a novel stochastic algorithm that replaces the silence gaps with appropriately reconstructed data samples derived as spectral statistical equivalents of packets that are already successfully received. Using well-established objective and subjective measurements it was shown that compared to relative error concealment techniques the SPARE technique achieves a perceptually significant reduction of the audibility of the above packet-loss induced distortions, allowing high-quality, uncompressed digital audio playback over WLANs.

Future plans and extensions of this work will include the assessment of the SPARE algorithm performance under variable transmitted packet lengths (L_p) as well as the enhancement of the SPARE technique using optimum time-domain windowing functions for minimizing the discontinuities among successive packet samples and variable/ adaptive noise amplitude and window length of the stochastic generator unit according to the input signal dynamic characteristics.

6. ACKNOWLEDGMENTS

The authors wish to thank Mr. Thomas Zarouchas (Audio Technology Group, WCL, University of Patras) for his contribution to this work.

7. REFERENCES

- [1] Blank, T., Atkinson, B., Isard, M., Johnston J. and Olynyk, K. An Internet Protocol (IP) Sound System. Presented at the *117th AES Convention*, San Francisco, October 28-31 2004 (preprint 6211).
- [2] IEEE802.11 WG, IEEE802.11g, *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 Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band*. June 2003.
- [3] IEEE802.11 WG, IEEE802.11e/D13.0, *IEEE Standard for Information Technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements - Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment: Medium Access Control (MAC) Quality of Service Enhancements*. January 2005.
- [4] Blum P., and Thiele L. Trace-Based Evaluation of Clock Synchronization Algorithms for Wireless Loudspeakers. In *Proceedings of the 2nd Workshop on Embedded Systems for Real-Time Multimedia (ESTIMedia'04)*, Stockholm, September 6-7, 2004.
- [5] Tatlas, N. A., Floros, A., Zarouchas T. and Mourjopoulos, J. WLAN Technologies for Audio Delivery under review at the *EURASIP Advances on Multimedia*.
- [6] Wah, B.W., Xiao S. and Dong L. A survey of error-concealment schemes for real-time audio and video transmissions over the Internet. In *Proceedings of International Symposium on Multimedia Software Engineering*, December 2000, pp. 17-24.
- [7] Wang, Y., Ahmaniemi, A., Isherwood D. and Huang, W. Content-based UEP: A New Scheme for Packet Loss Recovery in Music Streaming. In *Proceedings of 2003 ACM Multimedia Conference*, November 02-08 2003, Berkeley, CA, USA.
- [8] Suzuki J. and Taka, M. Missing packet recovery techniques for low-bit-rate coded speech. *IEEE Journal on Selected Areas in Communications*, 7, 5 (June 1989), 707-717.
- [9] Tang, J. Evaluation of double sided periodic substitution (DSPTS) method for recovering missing speech in packet voice communications. In *Proceedings of the Tenth Annual Int'l Phoenix Conf. on Computers and Communications*, March 1991, 454 - 458.
- [10] Wasem, O. J., Goodman, D. J., Dvordak, C. A. and Page, H. G. The effect of waveform substitution on the quality of PCM packet communications. *IEEE Transactions. on Acoustics, Speech, and Signal Processing*, 36, 3 (March 1988), 342-348.
- [11] Tucker R. C. E. and Flood J. E. Optimizing the performance of packet-switch speech. In *Proceedings of the IEEE Conference on Digital Processing of Signals in Communications*, April 1985, 227-234.
- [12] Tatlas, N. A., Floros, A., Zarouchas T. and Mourjopoulos, J. An Error - Concealment Technique for Wireless Digital Audio Delivery. In *Proceedings of the 5th International Conference on Communication Systems, Networks and Digital Signal Processing (CSNDSP'06)*, July 2006, Patras, Greece, 181-184.
- [13] Stenger, A., Younes, K. B., Reng, R. and Girod, B. A new Error Concealment Technique for Audio Transmission with Packet Loss. In *Proceedings of the European Signal Processing Conference (EUSIPCO 96)*, Trieste, Italy, September 1996.
- [14] Hellerud, E., Voldhaug, J. E. and Svensson, U. P. Perceptually Controlled Error Protection for Audio Streaming over IP Networks. In *Proceedings of the International Conference on Digital Telecommunications (ICDT 2006)*, Cap Esterel, Crte d'Azur, France, Aug. 2006.
- [15] Avlonitis M. and Vlamos, P. On an analytical signal reconstruction in the presence of additive white noise. Under review at the *Statistics and Probability Letters*.
- [16] Papoulis, A. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 2nd edition, 1984, ISBN: 0070484686.
- [17] Tatlas, N. A., Floros A. and Mourjopoulos, J. Wireless Digital Audio Delivery Analysis and Evaluation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2006*, Toulouse, France, May 2006, V-201 - V-204.
- [18] Brandenburg K. and Sporer, T. NMR and Masking Flag: Evaluation of Quality Using Perceptual Criteria. presented at the *Audio Eng. Soc. 11th International Test & Measurement Conference*, April 1992.