An Error – Concealment Technique for Wireless Digital Audio Delivery

Nicolas Tatlas1, Andreas Floros2 Thomas Zarouchas1 and John Mourjopoulos1

1AudioGroup, WCL, Univeristy of Patras
265 00 Rio, Patras, Greece
{ntatlas, thozar, mourjop}@wcl.ee.upatras.gr

Abstract – Real-time digital audio streaming over unreliable wireless IP networks, such as WLANs, encounters playback quality degradation, due to late arrivals or permanent loss of the transmitted packets. In this paper we present a perceptually efficient error concealment technique, which, combined with playback synchronization correction schemes and a high-rate wireless transmission protocol with Quality-Of-Service support, can significantly improve the overall audio playback quality of typical multichannel home-theater wireless applications.

I. INTRODUCTION

During the last decades, the audio technology has met a significant evolution from analogue to high-quality digital applications. The common element in all typical audio systems is that the equipment interconnection is performed through cables that realize physical wired links carrying the audio information. With the recent developments on digital signal representation, storage and processing leading the shift to an all-digital audio reproduction chain, the audio information is mainly distributed in digital formats through synchronous, wired digital interfaces (e.g. SPDIF, FireWire, HDMI etc) and lately asynchronously over packet-based networks and protocols such as the Internet Protocol (IP).

IP-based audio delivery provides the ability to establish numerous audio playback setups that are impossible to realize using conventional, sequential audio systems interconnections, such as music distribution from any digital audio source within a typical multiroom home environment to any networked speaker/playback device. In practice, each IP-enabled audio device can be connected to every other, using the common IP network interface also used for computer communications and Internet access [6].

Nowadays, the IP network interface is frequently wireless. This further extends the supported applications by providing flexible, portable, wire-free and low cost means of packet-based communications. With the recent advances of the Wireless Local Area Networks (WLAN) technology on the offered bandwidth capacity (the recently ratified 802.11g [1] specification offers data rates in the range of 54Mbps, while the forthcoming 802.11n protocol is expected to increase this rate up to 480Mbps), a new set of cable-free, high-quality multimedia applications can be efficiently supported, such as wireless multichannel audio playback for typical home-theater setups.

In order to achieve high-quality digital audio wireless playback, a number of issues related to the wireless interconnection nature must be resolved, such as the minimization of the audible distortions introduced during the wireless transmissions and the absolute clock and packet playout synchronization among the multiple audio wireless devices. Previously published works [2], [3], [6] have introduced novel efficient methods for compensating the effect of hardware clock and packet-based playback synchronization. However, depending on the wireless link quality and in order to maintain the relative receivers’ playback synchronization, a significant number of silence gaps are introduced, due to the permanent loss or excessive delay of the transmitted data packets.

The motivation for this work was the development of a technique for vastly improving the audible effect of temporal playback muting during the wireless delivery of digital audio data. One of the major initial design requirements of the proposed methodology (termed as Fading Pattern Repetition - FPR) was the low implementation and computational complexity, as it should be applied in real-time prior the final audio playback. As it will be explained later in the paper, this requirement was achieved by a perceptually efficient error concealment algorithm that repeats appropriately processed, successfully received audio data packets.

The rest of the paper is organized as follows: Section II presents a brief overview of the technology issues related with the digital audio delivery over WLANs, while Section III analyzes and outlines the implementation of the proposed error concealment technique. In Section IV, the testing procedures and criteria for evaluating the perceptual performance of the audio playback are described, followed by the results obtained during the tests and the conclusions summarized in Section V.

II. OVERVIEW OF AUDIO OVER WLANS

In general, two different types of wireless home audio applications exist, defined in [3] as: a) point-to-point home audio delivery, were a device acts as the “media server” (possibly integrating all available audio sources, including Internet-stored media) that wirelessly streams the same (or different) audio content to a number of wireless audio players and b) point-to-multipoint delivery, consisting of a multichannel digital audio source transmitting audio to multiples receivers (e.g. DVD audio content to 5.1 wirelessly connected loudspeakers) that perform simultaneous and synchronized playback.

In both cases, wireless audio transmission can be theoretically performed using a variety of standardized or proprietary protocols, starting from low-quality analog systems operating in the range of 800-900MHz and up to high-rate digital streaming protocols operating in the
range 2.4 and 5GHz. For example, a previous work [4] has defined the principles for real-time audio data transmissions using the Bluetooth specification [5]. It was found that high-quality audio reproduction can be achieved only through asynchronous links using audio compression, due to the limited Bluetooth bandwidth.

As modern IP WLAN technologies overcome the above bandwidth restrictions, audio compression is not necessary, leading to feasible high-quality wireless streaming. However, the IP wireless transmission used imposes a number of problems, including the lack of in-time delivery guarantees and the conversion of the asynchronous packet-based data delivery to synchronous. As described in [6], such a conversion can be efficiently realized using appropriate buffering and time-stamping of audio packets on the IP playback device side. However, the variable interference level presented in the wireless medium frequently raises unacceptable packet delivery delays, especially under network traffic congestion.

Considering a simple example of a CD player sending audio data to 2 wireless loudspeakers (see Fig. 1), excessive packet-delay can introduce: a) inter-packet gaps during playback, due to the lack of the appropriate packet and/or b) relative channel phase shifting (one audio channel leading or lagging), which causes loss of the sound spatial information. In both cases, the distortion effect is audible.

As it shown in [3], the audible effect of the variable delay wireless point-to-multipoint audio delivery can be partially decreased by employing a) Quality of Service (QoS) networking mechanisms, such as the IEEE802.11e [7] polling-based access, which represents the most efficient QoS mechanism for real-time traffic delivery [8] and b) a playback synchronization correction scheme (such as CoDeS introduced in [3]) that minimizes the variable channel phase shift and ensures that the reproduction is kept synchronized among the wireless loudspeakers by disregarding received packets when necessary and/or by adding an appropriate number of zero samples in the playout buffer. Hence, although the above approach significantly improves the multichannel playback quality in terms or relative channel synchronization, silence playback gaps may be introduced. Under heavy channel degradation, due to the instantaneous packet delay increment (which is practically equivalent to permanent packet losses), the length of these gaps is significantly high, rendering the playback quality unacceptable.

From the above analysis, it is clear that an error concealment mechanism combined with the playback synchronization correction scheme would minimize the length of the silence gaps (and consequently the perceptual degradation they induce). Towards this aim, the Fading Pattern Repetition (FPR) strategy was developed during this work and is presented in the next Section.

III. IMPLEMENTATION

The FPR strategy proposed here is based on the well-known Pattern Repetition (PR) algorithm, used in speech, audio and audiovisual applications, where in case a packet is permanently lost, it is substituted by directly repeating a previously correctly received data segment. PR presents an attractive choice, because of its extremely low complexity and computational cost; however, the possible amplitude and phase mismatch between the audio stream and the segment to be repeated may cause audible clicking sounds [9], leading in many cases to further deterioration of the overall audio quality, as will be also shown in Section IV.

In order to avoid the audible effect of such discontinuities between the substitute packet limits and the correctly received packets, the FPR strategy employs time-domain window functions, as illustrated in Fig. 2. The playout buffer for each reproduction device is monitored to detect a forthcoming gap in reproduction, while a number of successfully received packets are kept in a different buffer, to be used as possible data-frame substitutes.

![Fig. 2. Error concealment using the FPR strategy: (a) original sequence (b) lost packet substituted with silence (c) PR strategy (d) FPR strategy](image)

According to the FPR scheme, in case the receiver queue is empty while a predetermined number of samples remain to be reproduced, a linear descending gain function – similar to a fade-out process –is imposed. The substitute segment is added to the playout buffer in packet-by-packet basis, while an ascending gain function – similar to a fade-in process –is imposed to the first packet. When a new packet is received, then the reverse process is followed: The linear descending function is imposed to the last samples of the substitute packet being reproduced, while the ascending function is used for the first samples of the received packet.

For the network packet sizes considered in this work, a typical length of 50 samples was selected for the gain functions, while the buffer containing packets for possible substitution can hold 5 packets. As already mentioned, one of the major initial design requirements of the proposed methodology was the low complexity. For the above parameters, the physical memory requirement for
the FPR scheme is less than 5Kbytes, while the processing power required for applying the windowing function is very low.

IV. TEST METHODOLOGY AND RESULTS

For the purposes of this work, a CD-quality wireless audio playback application was considered, that is a typical CD-DA wireless source transmitting 16-bit, stereo PCM data (sampled at 44.1kHz) concurrently to 2 wireless loudspeakers, performing the left and right channel playback respectively. However, it should be noted that the above application scenario can be easily extended to more channels (e.g. 6 audio channels as used in the DVD format). Taking into account the derived data bit-rates for stereo linear PCM (nearly 1.4 Mbps for CD-audio), the legacy, low-cost IEEE802.11b standard was selected with a transmission rate equal to 11Mbps.

The real-time wireless audio transmission and playback process was performed using a computer-based simulation platform presented in [10]. The platform consists of (a) a pre-processing stage, which converts the digital audio data into inputs to the wireless network simulator (b) the wireless network simulator, conforming to the latest 802.11e draft specification [7] and (c) a post-processing stage which produces an audio file containing a “reproduced” PCM version of the original digital audio data. Both the synchronization algorithm (CoDeS) and the proposed FPR error concealment mechanism are realized in the post-processing module to derive the “corrected” version of the received data.

As it is shown in Fig. 3, the audio data exchange between the above modules is realized using trace files, which contain information for the frame generation time and the corresponding packet length. A trace file maps the transmitted PCM data packets to specific segments of stereo wave files containing the original digital audio. This mapping includes parameters such as the desired pure audio data packet length (in bytes) plus the packet header.

![Fig. 3. Block diagram of the wireless audio playback simulation methodology](image)

The wireless simulator produces one output trace file per serviced Traffic Stream (TS), containing information such as whether the corresponding packets were transmitted or not, and the packet delay induced. These files are mapped to the stereo audio samples using an application realizing a First-In First-Out (FIFO) reception (Rx) queue, which is filled with the successfully received data packets. Upon playback, samples are read from the Rx queue at a rate equal to the audio sampling frequency \( f_s \) (Hz). The Rx queue length and the initial pre-buffering interval are defined by the user in multiples of 100ms (the nominal beacon transmission period defined by the 802.11 protocol). For all test cases considered here, the Tx and Rx queue lengths were equal to 5000 and 10000 bytes (or 5000 samples) respectively, while the initial playback latency was set to 1 beacon interval.

From the networking point of view, the transmitted packet length value represents a critical parameter that must be carefully selected. During this work, this length was set to 294 and 882bytes, a selection implied by the requirement of deriving a wireless transmission scheme using the basic traffic scheduling scheme defined in the current 802.11e specification [7]. 48 header bytes are added to the above lengths, where 8 bytes represent the User Datagram Protocol (UDP) overhead and 40 additional bytes are reserved for future control purposes.

The audible effect of the wireless network distortions was assessed using the well-established Noise-to-Mask Ratio (NMR) criterion [11] which utilizes the masking functions of the human ear and determines the distance of the distortions imposed by any audio system and the audibility 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 for perceptually insignificant distortions, the values of NMR must be as low as possible.

![Fig. 4. Measured NMR vs packet length for PCM audio](image)

Fig. 4 shows the measured NMR values for the packet length values considered in this work and for both cases of PR, and FPR error concealment mechanism employment. In the same Figure, the perceptually measured performance of the CoDeS synchronization algorithm is also shown, where lost or excessively delayed packets are substituted with silence (zero samples), allowing the direct assessment of the PR and FPR scheme effect.

From these measurements it is clear that in the case of combining the CoDeS+FPR algorithms, the network-induced distortions perceptual effect (silence gaps) significantly decreases. Namely, for the case of the 294byte packet length, a reduction of 5dB in NMR values is achieved, while for the larger packet size of 882bytes, this reduction is 3dBs. For the case of the larger packet lengths, this reduction is lower due to the larger silence...
gaps that are introduced by the loss of a single (or multiple) data frames. However, for large packet lengths, the overall NMR is significantly lower for all examined cases; hence, the playback quality is always higher.

V. CONCLUSIONS

Wireless IP networks represent a very promising and flexible mean for audio delivery, especially after the recent advantages on QoS support over WLANs. However, during wireless audio streaming, the variable link quality degradations can lead to excessive packet delivery delays and equivalent data losses. In typical multichannel audio playback setups (e.g. stereo or 5.1 home theater environments) in order to avoid the relative audio channels phase shifts caused by packet losses, muting is performed, controlled by a playback synchronization scheme.

In this paper, the FPR error concealment technique is proposed, which aims to minimize the perceptual effect of muting introduced by the playback synchronization correction scheme, using a low complexity and easy-to-implement algorithm that replaces the silence gaps with appropriately processed data samples derived from packets that are already successfully received. Using a well-established perceptual criterion it was shown that the FPR technique achieves a perceptually significant reduction of the audibility of the silence gaps, allowing high-quality, uncompressed digital audio playback over WLANs.

Future plans and extensions of this work include the enhancement of the FPR technique using a) networking mechanisms, such as optimized 802.11e transmission schedulers and concurrent (multicast) audio channel transmissions that will potentially reduce the possibility of packet losses and excessive delays at the expense of higher network traffic and b) perceptually optimized shaping techniques of the audio samples that replace the silence gaps for further improving the overall transmission and playback system performance.

REFERENCES