IET Communications

Research Article

Packet loss recovery in audio multimedia streaming by using compressive sensing

ISSN 1751-8628
Received on 18th October 2014
Revised on 12th September 2015
Accepted on 19th December 2015
doi: 10.1049/iet-com.2014.0995
www.ietdl.org

Angelo Ciaramella[1] ✉, *Giulio Giunta*[2]

[1]*Department of Science and Technology, University of Naples' Parthenope', Naples, Italy*
[2]*Isola C4, Centro Direzionale, I-80143, Napoli (NA), Italy*
✉ *E-mail: ciaramella@uniparthenope.it*

Abstract: The aim of this study is to introduce a new scheme, based on a compressive sampling technique, for the reconstruction of lost data in multimedia streaming. The audio streaming data are encapsulated in different packets, at the sender, by using an interleaving technique. The compressive sampling technique is used to recover audio information in case of lost packets, at the receiver. Experimental results are presented for speech and musical audio signals which illustrate the performances and the capabilities of the proposed methodology.

1 Introduction

Streaming technologies and increased bandwidth in access networks have facilitated the transmission of multimedia content on the Internet [1]. This new service gives possible, for example, Internet TV or audio/video services on demand, which in turn creates a great interest in various fields. Users are increasingly turning to this type of services and providers try to offer better quality to meet such needs. The main limitation of this technology is the need for stable transmission conditions to guarantee a certain degree of quality of service. Over the last few years new classes of scalable audio and video streaming applications have been introduced but in most cases the quality of the multimedia content is affected by packet loss, delay and network congestion [1, 2]. Recently, several methodologies for recovering multimedia contents from packet loss have been studied and proposed [2–5]. Voice over internet protocol (VoIP) systems, for example, have become a basic tool on modern Internet phones. However, a high percentage of packet loss can often make speech unintelligible [6–8]. For this reason, VoIP applications regularly incorporate a packet loss recovering or concealment mechanism (packet loss concealment (PLC)).

Over the last few years, several techniques for audio concealment and reconstruction have been introduced. In [9, 10] a loss concealment scheme based on sinusoidal extrapolation and mean-square error criterion is proposed. In [11] the authors introduce an algorithm for audio loss concealment, designed for MPEG-audio streaming, based only on the data available at the receiver. Bahat *et al.* [12] introduced an inpainting-based mechanism to fill the missing data using samples taken from prior recorded audio from the same user. A different study is proposed in [13] where a coded amplitudes scheme suited for quantisation of sinusoidal parameters is used for robust parametric audio coding. Sinusoidal interpolation is also employed in [14]. In [15, 16] different approaches are used based on linear predictive coding and immittance spectral frequency, respectively. In [17, 18] a compressive sampling technique is proposed for audio inpainting and coding, respectively.

The compressive sampling or compressed sensing enables a faithful recovery of signals, images and other data from what appears to be highly sub-Nyquist-rate samples [19]. In real-world applications, most signals are sparse and then compressible with low information loss. Compressible signals can be captured via sampling or sensing protocols that directly condense signals into a small amount of data. In this work, we use an optimisation approach based on the L_1 norm [19, 20] in order to recover signals. There are, however, other algorithmic approaches to compressive sampling based on greedy algorithms such as orthogonal matching pursuit [21, 22], iterative thresholding [23], compressive sampling matching pursuit [24] and many others.

In this paper we propose a new scheme for data loss reconstruction in audio streaming (named packed loss recovery based on compressive sensing, (PLRCS)). In the streaming model, the audio data are encapsulated, at the sender, in different packets using an interleaving technique. At the receiver the information of the lost packets is reconstructed by means of a compressive sampling technique.

The paper is organised as follows. In Section 2 some aspects of the streaming and lost packets are introduced. The compressive sampling methodology is briefly outlined in Section 3. In Sections 4 and 5 we present the proposed methodology and some experimental results, respectively. Finally in Section 6 some conclusions and future remarks are provided.

2 Real-time protocol and interleaving

Multimedia applications require services that differ substantially from the standard ones. These applications are particularly sensitive to the end-to-end delay and they can tolerate only occasional loss of data. Generally real-time applications (e.g. VoIP, real-time events) use the real-time transport protocol [25], which is able to support an IP multicast and data distribution to a group of receivers. Multicast technologies permit the sharing of links between different classes of traffic, generating loss patterns [2, 26]. Routers congestion is one of the main causes for loss packets, as studied in [26].

A multicast channel is typically characterised by high latency, and a large variation in end-to-end delay. The delay variation is a great problem for interactive and loss-tolerant real-time applications (e.g. VoIP, conferences, wireless streaming communications). In fact, packets with a large delay will have to be discarded in order to satisfy the timing requirements of the applications. There are two main classes of methodologies for recovery: active retransmission and passive channel coding. Channel coding techniques can be grouped into the traditional forward error correction and the interleaving-based schemes. Interleaving-based schemes are mainly used for interactive applications (e.g. VoIP) and it is the mechanism applied in our work.

2.1 Interleaving

Interleaving can significantly improve the quality with which we perceive an audio stream [2, 27]. For example, over the last few years, it has been widely adopted for mitigating the bursty losses [6, 8] and in particular in VoIP streaming transmissions [7].

During the interleaving phase, frames of audio signals are sequenced in packets before transmission. In particular, originally adjacent frames are separated in the transmitted stream and returned to their original order at the receiver. This mechanism redistributes the effect of packet losses. If, for example, frames are 5 ms in length and packets 20 ms (i.e. 4 frames/packet), then the first packet will contain units 1, 5, 9, 13; the second units 2, 6, 10, 14; and so on, as illustrated in Fig. 1. We note that the loss of a single packet from an interleaved stream does not cause a single large gap as in a non-interleaved stream, but multiple small gaps in the reconstructed stream. This can be particularly useful for audio tools where transmitted packets are generally similar in length to phonemes in human speech [28]. Interleaving, however, increases latency and its major advantage is that it does not increase the bandwidth requirements of a stream.

3 Compressive sensing

One of the main features of the compressive sensing (CS) or compressed sensing theory is the possibility to recover signals from much fewer measurements with respect to traditional methods [19, 29]. The two principles to satisfy in CS are: *sparsity* of the source signals and *incoherence*, which is related to sensing modality and the representation of the signals. To design efficient sensing or sampling protocols, useful information content, embedded in a sparse signal, must be captured and condensed into a small amount of data. These protocols are non-adaptive and simply require the correlation of the signal with a small number of fixed waveforms.

3.1 The sensing problem

We suppose to use linear functionals to obtain information about a signal $f(t)$

$$y_k = \langle f, \phi_k \rangle \tag{1}$$

with $k = 1, \dots, m$. In fact, we consider a correlation between the signal and the waveforms $\phi_k(t)$. In this work, we focus on discrete signals $\boldsymbol{f} \in \boldsymbol{R}^n$ and a sensing orthogonal matrix $\boldsymbol{\Phi} = [\phi_1, \phi_2, \dots, \phi_n] \in \boldsymbol{R}^{n\times n}$. Now we consider that the number m of available measurements is much smaller than the dimension n of the signal $\boldsymbol{f}$. Letting $\boldsymbol{\Phi}_s$ denote the $m \times n$ sensing matrix, $m \le n$, with the vectors $\boldsymbol{\phi}_1^*, \dots, \boldsymbol{\phi}_m^*$ as rows (a^* is the complex transpose of a), the process of recovering $\boldsymbol{f} \in \boldsymbol{R}^n$ from

$$\boldsymbol{y} = \boldsymbol{\Phi}_s \boldsymbol{f} \in \boldsymbol{R}^m \tag{2}$$

is ill-posed in general when $m < n$, since there are infinitely many candidate signals $\hat{\boldsymbol{f}}$ for which $\boldsymbol{\Phi}_s \hat{\boldsymbol{f}} = \boldsymbol{y}$.

3.2 Sparse representation

Many real signals have condensed representations if expressed in an appropriate basis. Mathematically speaking, a vector $\boldsymbol{f} \in \boldsymbol{R}^n$ is expanded in an orthonormal basis $\boldsymbol{\Psi} = [\boldsymbol{\Psi}_1, \dots, \boldsymbol{\Psi}_n]$ (compressed basis)

$$\boldsymbol{f} = \sum_{i=1}^{n} x_i \boldsymbol{y}_i = \boldsymbol{\Psi x} \tag{3}$$

where $\boldsymbol{x} = [x_1, \dots, x_n]^{\mathrm{T}}$ is the representation of $\boldsymbol{f}$ with respect to the basis $\boldsymbol{\Psi}$. If most of the components of $\boldsymbol{x}$ are zero, then $\boldsymbol{x}$ is referred to as a sparse representation of $\boldsymbol{f}$, and $\boldsymbol{\Psi}$ is a sparsifying basis. Now we consider the pair $(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ of orthobases of $\boldsymbol{R}^n$. The first basis $\boldsymbol{\Phi}$ is used for sensing the object $\boldsymbol{f}$ as in (1) and (2), and the second is used to represent $\boldsymbol{f}$. The coherence $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ measures the largest correlation between any two elements of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$. Considering m measurements, uniformly at random, in the $\boldsymbol{\Phi}$ domain, the smaller the coherence the fewer m samples are needed [19, 29]. In this case, we can measure any set of m coefficients without information loss [19, 29]. For this reason, in CS, one concentrates on low coherence. Since, in our case, $\boldsymbol{\Phi}$ is the identity matrix and $\boldsymbol{\Psi}$ is the discrete cosine transform (DCT) basis, then a maximal incoherence is obtained. Moreover, the m rows of the $\boldsymbol{\Phi}_s$ matrix are randomly selected in the $\boldsymbol{\Phi}$ domain.

3.3 Undersamplig and sparse signal recovery

We suppose to observe a subset of the n coefficients of $\boldsymbol{f}$ and collect the data as in (2).

With this information, the source signal is recovered by solving an L_1-norm constrained minimisation problem. The proposed reconstruction $\bar{\boldsymbol{f}}$ is given by $\bar{\boldsymbol{f}} = \boldsymbol{\Psi}\bar{\boldsymbol{x}}$, where $\bar{\boldsymbol{x}}$ is the solution to

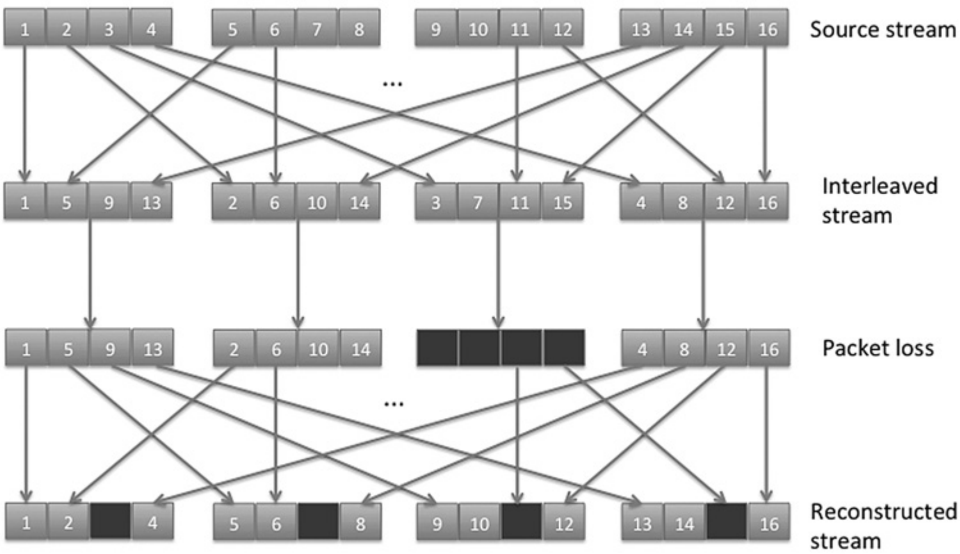

Fig. 1 *Interleaving units across multiple packets*

IET Commun., 2016, Vol. 10, Iss. 4, pp. 387–392

the convex optimisation program ($\|\boldsymbol{x}\|_{L_1} = \sum_i |x_i|$):

$$\min_{x \in R^n} \quad \|\boldsymbol{x}\|_{L_1} \quad \text{s.t.} \quad y_k = \langle \phi_k, \boldsymbol{\Psi x} \rangle \quad \forall k \in M \tag{4}$$

That is, among all vectors $\boldsymbol{x}$ consistent with the data, we pick the one with minimal L_1-norm.

The keys to CS are sparsity and the L_1 norm. If the expansion of the original signal as linear combination of the selected basis functions has many zero coefficients, then it is often possible to reconstruct the signal exactly (see [19, 29] for more details and proofs). In principle, computing this reconstruction should minimise the L_0 pseudo-norm of $\boldsymbol{x}$, i.e., the number of its non-zero components. The latter is a combinatorial problem whose computational complexity is NP-hard. Fortunately, in [19, 29] it has been shown that, in most cases, L_0 can be replaced by L_1.

4 Signal reconstruction

In order to explain the proposed methodology, we consider a multimedia streaming scheme as shown in Fig. 2, where a client receives packets from a server. A signal $f(t)$ is sampled, on the server, by means of a PCM encoding technique (e.g. 64 kbit/s). We suppose, for example, that the server collects data every 20 ms, thus obtaining four packets composed of 160 bytes (or 160 samples). A raw signal can be regarded as a vector $\boldsymbol{f}$ that can be represented as a linear combination of certain basis functions, as in (3)

$$\boldsymbol{f} = \boldsymbol{\Psi x}. \tag{5}$$

The basis functions must be suitable for a particular application (e.g. wavelet, gammatone etc.) and in our experiments, $\boldsymbol{\Psi}$ is the DCT. We remark that in order to use DCT as sparsifying basis, we have to rely on a moderately low number of samples (640 samples, 20 ms). Before applying the interleaving approach, the components of $\boldsymbol{f}$ are randomly permuted to ensure a random distribution of the missing information. If we consider a random permutation matrix $\boldsymbol{P}_\pi$, then the resulting sequence is

$$\boldsymbol{f}_\pi = \boldsymbol{P}_\pi \boldsymbol{f}. \tag{6}$$

Applying the interleaving mechanism described in Fig. 1 (permutation matrix $\boldsymbol{I}_\pi$) to $\boldsymbol{f}_\pi$, then from the sequence $\boldsymbol{f}_\pi$ a new sequence $\boldsymbol{f}_\text{I}$ of four blocks $\boldsymbol{f}_\text{I}^{(i)}$, with $i = 1, 2, 3, 4$ is obtained:

$$\boldsymbol{f}_\text{I} = \boldsymbol{I}_\pi \boldsymbol{f}_\pi = [\boldsymbol{f}_\text{I}^{(1)}\, \boldsymbol{f}_\text{I}^{(2)}\, \boldsymbol{f}_\text{I}^{(3)}\, \boldsymbol{f}_\text{I}^{(4)}]^\text{T}. \tag{7}$$

Now we could consider that in a streaming communication process some packets may be lost (for example, two lost packets in Fig. 2). In this case, the client receives only two packets

$$\tilde{\boldsymbol{f}}_\text{I} = [\boldsymbol{f}_\text{I}^{(1)} \text{ Null } \boldsymbol{f}_\text{I}^{(3)} \text{ Null}]^\text{T}. \tag{8}$$

The client applies the inverse of the interleaving process, obtaining a subset of coefficients of $\boldsymbol{f}_\pi$

$$\tilde{\boldsymbol{f}}_\pi = \boldsymbol{I}_\pi^\text{T} \tilde{\boldsymbol{f}}_\text{I}. \tag{9}$$

Moreover, applying the inverse of the permutation process, the following subset of samples of $\boldsymbol{f}$ are obtained:

$$\tilde{\boldsymbol{f}} = \boldsymbol{P}_\pi^\text{T} \tilde{\boldsymbol{f}}_\pi. \tag{10}$$

We note that, in this way, the signal received by the client is a vector containing few random samples of $\boldsymbol{f}$ (not *null* elements of $\tilde{\boldsymbol{f}}$). Mathematically, we can consider a linear operator $\boldsymbol{\Phi}_\text{s}$ [as in (2)] such that

$$\tilde{\boldsymbol{f}} = \boldsymbol{\Phi}_\text{s} \boldsymbol{f}. \tag{11}$$

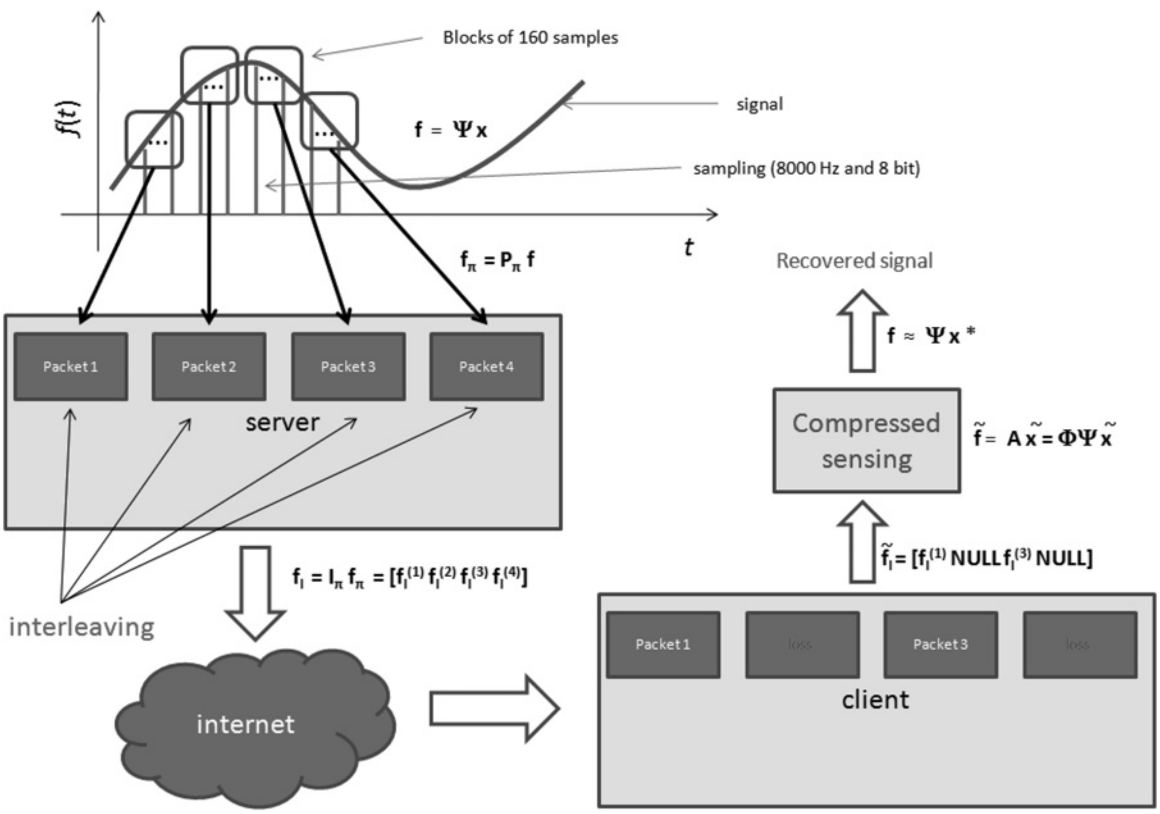

Fig. 2 *Multimedia streaming process*

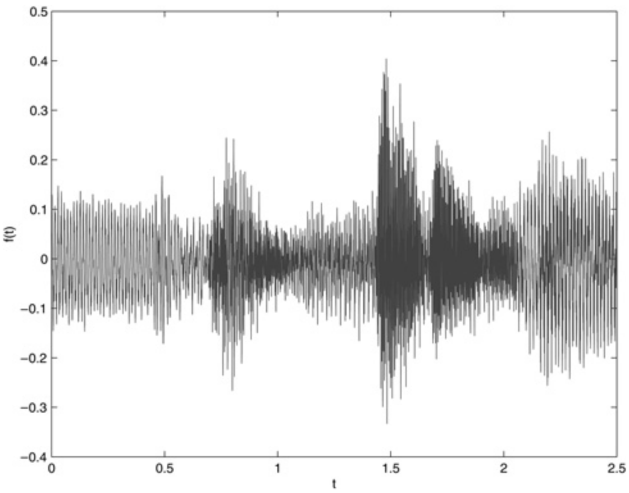

Fig. 3 *Audio signal of a female speaker*

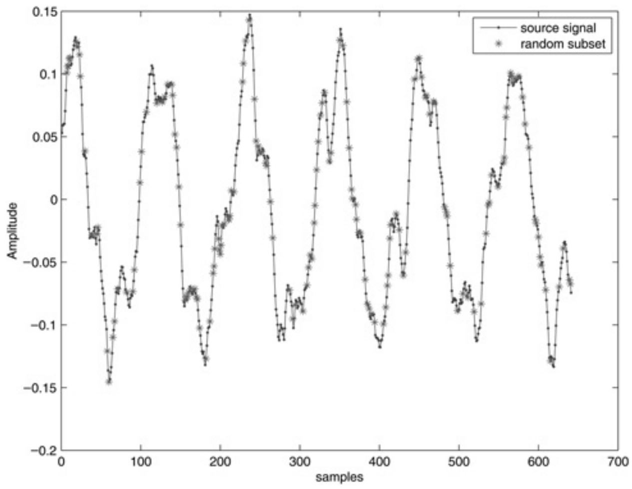

Fig. 4 *Frame information: three packets loss*

In our case, $\mathbf{\Phi}_s$ is a subset of the rows of the identity operator, with row indexes corresponding to the not *null* elements of $\tilde{\boldsymbol{f}}$. To reconstruct the signal, the client recovers the sparse representation coefficients by solving the undetermined linear system

$$\boldsymbol{A}\boldsymbol{x} = \tilde{\boldsymbol{f}} \tag{12}$$

where $\boldsymbol{A} = \mathbf{\Phi}\mathbf{\Psi}$ is the CS matrix, i.e., by computing the solution $\bar{\boldsymbol{x}}$ to the convex optimisation problem in (4). Then one can recover the original signal $\boldsymbol{f}$ by means of the approximated reconstructed signal

$$\bar{\boldsymbol{f}} = \mathbf{\Psi}\bar{\boldsymbol{x}}. \tag{13}$$

In the next section, we also compare the solution obtained by using the L_1 norm with the one obtained by using an optimisation approach based on the L_2 norm ($\|\boldsymbol{x}\|_{L_2} = \sum_i x_i^2$), i.e., the least-squares solution of (12).

5 Experimental results

In this section, we show some experimental results obtained by applying PLRCS for reconstructing streaming audio signals (e.g. VoIP and musical audio streaming). We consider audio signals codified by a PCM encoding scheme (sampling frequency of 8000 Hz and 8-bit quantisation).

The first results are presented by comparing the source signals with those obtained from the optimisation approaches based on the L_1 and L_2 norms (named PLRCS-L_1 and PLRCS-L_2, respectively). The software, the source and the reconstructed wav files are available on request.

The first audio recording corresponds to a female voice reading the news in English. The audio recording has duration of 6.25 s. Streaming data are collected each 20 ms, obtaining a stream of four packets (or windows) composed by 160 samples ($4 \times 160 = 640$ samples totally) as in the schemes of Figs. 1 and 2.

We stress that in our approach the overall recording duration is irrelevant since the reconstruction is made at the receiver on the effectively received stream without considering any other temporal information. Fig. 3 shows a section of this audio signal. In Fig. 4 a stream of 640 samples [$\tilde{\boldsymbol{f}}$ in (11)] after the interleaving phase and the loss of three packets is shown. In Fig. 5 we compare a source frame of this audio signal with those recovered by using both PLRCS-L_1 and PLRCS-L_2, when the loss of three packets is considered. In this case the parameters n and m of the CS scheme (see Section 3.2) are 640 and 160, respectively. In Fig. 6 the residua between the entire source signal and the reconstructed ones are visualised, when the random loss of zero, one, two or three

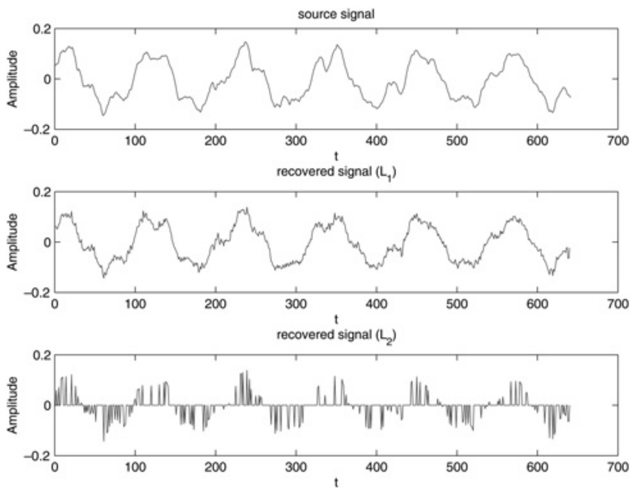

Fig. 5 *Comparison between a frame of the source signal and those of the reconstructed signals by using L_1 and L_2 norms*

packets is simulated. Finally, we compare the cross-correlation coefficients between the entire source signal and the recovered ones simulating the random loss of zero, one, two or three packets

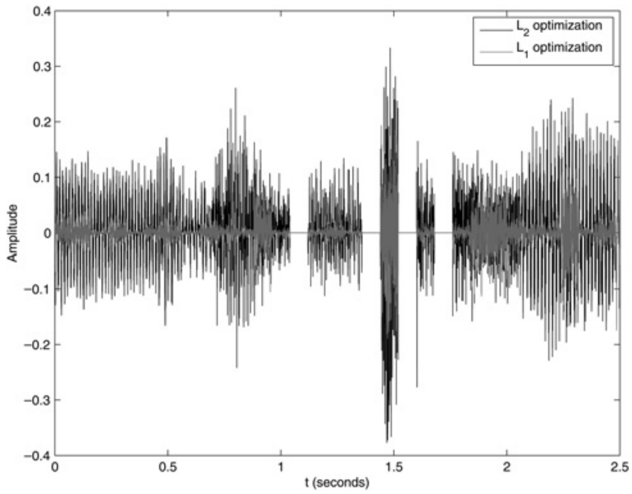

Fig. 6 *Residua between the source signal and the reconstructed signals by using L_1 and L_2 norms*

IET Commun., 2016, Vol. 10, Iss. 4, pp. 387–392

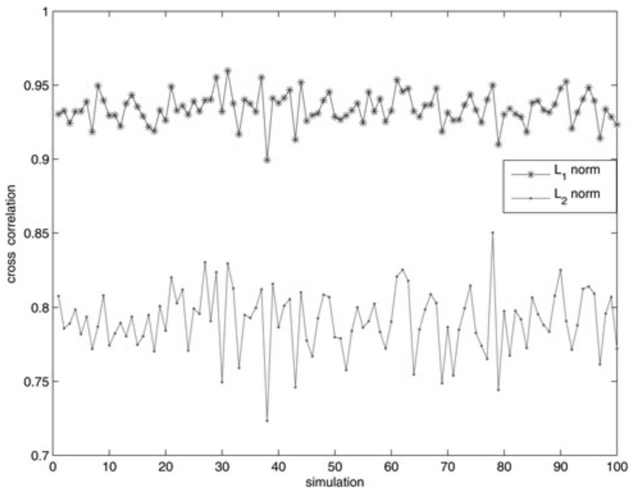

Fig. 7 *Cross-correlation coefficients after 100 simulations: audio female speaker*

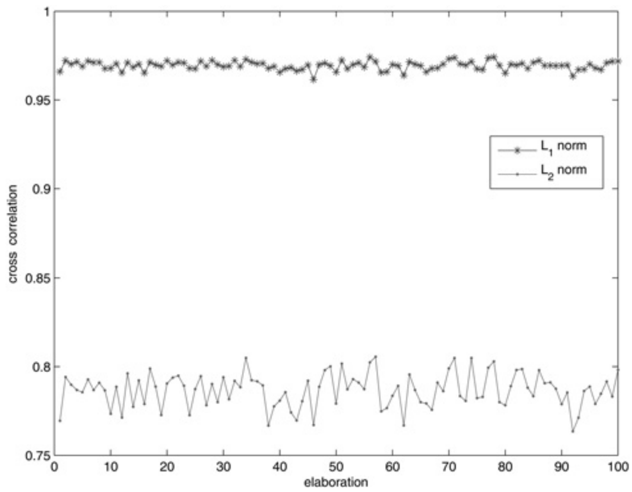

Fig. 9 *Cross-correlation coefficients after 100 simulations: audio song*

for each interleaving block. The results of 100 simulations are visualised in Fig. 7.

In the second experiment the audio recording is a male voice reading the news in English (the recording duration is 6.25 s). The cross-correlation coefficients obtained after 100 simulations are shown in Fig. 8.

In a further experiment we consider an example of musical audio signal. In particular, we consider a piece of the Jazz song played by Chet Baker, titled 'Blue Room' (the recording duration is 50 s). The results of the cross-correlation coefficients are presented in Fig. 9. We can observe that in all the cases PLRCS-L_1 provides the best results.

Finally, we compare PLRCS (both L_1 and L_2 based optimisation approaches) and ITU-G.723.1 [12, 30], a standard designed for VoIP applications [31], which contains a PLC mechanism. The performances are compared considering the cross-correlation coefficients and a speech quality assessment approach. For the latter, we observe that the goal of the PLRCS and G.723.1 methodologies is to produce a perceptually plausible audio signal. Degradation in perceived audio quality can be measured by speech quality assessment approaches (e.g. mean opinion score [32]). A notable disadvantage of some of these methods is that they are very time consuming. We have used an automatic perceptual evaluation of speech quality (PESQ) measure [32, 33], in order to assess the quality of the speech enhancement algorithms. This method gives the best results in the sense of the highest correlation with subjective measures [34] and the quality is estimated on a fixed interval ([0, 4.5]). In our experiments the number of samples for each packet is 240.

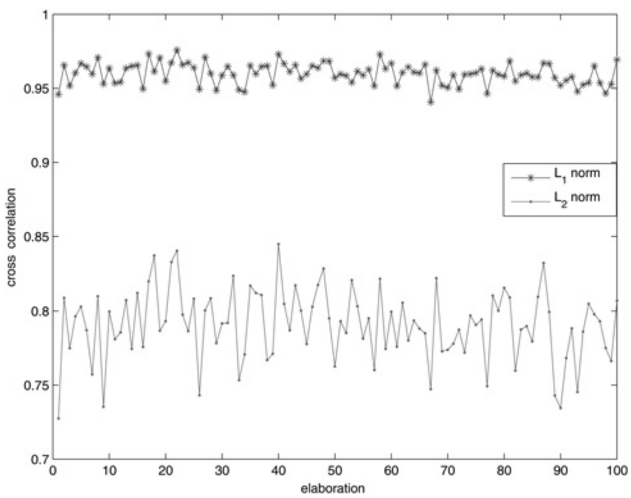

Fig. 8 *Cross-correlation coefficients after 100 simulations: audio male speaker*

In Table 1 we report the PESQ values obtained on the female audio recording in the first experiment. We simulated the percentage of loss packets, on the overall stream, from 5 to 80%. Table 2 shows the comparisons also considering the cross-correlation coefficients.

We highlight that the proposed methodology has been also compared with the techniques proposed in [35] and iLBC [36] (one of the codecs used by several streaming products). These two methodologies give low reconstruction quality and in particular when the packet loss rates are higher than 10%, this low quality is clearly perceived. We also remark that similar experiments and comparisons have been carried out on several other voice and musical audio recordings. In all cases we observed that PLRCS-L_1 provides the best results even for a high percentage of loss packets.

Table 1 PESQ quality evaluations: G.7.231 and PLRCS with L_1 and L_2 norms

Loss packets, %	G.723.1	PLRCS-L_2	PLRCS-L_1
5	3.21	2.68	3.87
10	2.65	2.09	3.59
15	2.39	1.90	3.03
20	2.07	1.71	2.84
30	1.80	1.49	2.43
50	0.95	1.01	1.72
70	0.34	0.22	0.59
80	0.07	0.00	0.33

Table 2 Cross-correlation coefficients: G.7.231 and PLRCS with L_1 and L_2 norms

Percentage	G.723.1	PLRCS-L_2	PLRCS-L_1
5	90.33	97.35	99.48
10	88.86	91.95	99.04
15	79.18	92.11	98.41
20	68.25	88.39	97.67
30	58.39	81.89	93.63
50	46.40	72.52	90.55
70	24.72	58.42	80.07
80	18.57	46.45	71.08

6 Conclusions

In this paper a new scheme for data loss reconstruction, based on a CS technique, in multimedia streaming has been introduced. Audio streaming data are encapsulated, at the sender, in different packets by using an interleaving technique. Information contained in the loss packets is recovered, at the receiver, by using a compressive sampling technique based on the L_1 norm. The experimental results highlighted that L_1 norm in the optimisation scheme performs better than L_2 norm. In particular, the methodology based on the L_1 norm, gives a better performance comparing to other known methodologies even when considering a high percentage of loss packets. In the future, the authors will focus on the use of different optimisation approaches and on the application of the proposed scheme for real-life situations (e.g. VoIP, conferences, wireless streaming communications), also in the case of dedicated hardware.

7 Acknowledgment

The authors would like to thank Mary Frances Donegan for the precious linguistic support. This work was funded by the University of Naples "Parthenope" (project "Sostegno alla ricerca individuale per il triennio 2015–2017").

8 References

1 Pozueco, L., Paneda, X.G., Garcia, R., *et al.*: 'Adaptable system based on scalable video coding for high-quality video service', *Comput. Electr. Eng.*, 2013, **39**, (3), pp. 775–789
2 Perkins, C., Hodson, O., Hardman, V.: 'A survey of packet loss recovery techniques for streaming audio', *IEEE Netw.*, 1998, **12**, (5), pp. 40–48
3 Banu, J.F., Ramachandran, V.: 'Study of QoS management techniques for voiceapplications', *Int. J. Comput. Sci. Electron. Eng. (IJCSEE)*, 2013, **1**, (1), pp. 80–84
4 Feamster, N., Balakrishnan, H.: 'Packet loss recovery for streaming video'. Twelfth Int. Packet Video Workshop, 2002
5 Lu, X., He, H., Tan, H.: 'A low complexity packet loss recovery method for audio transmission'. Proc. 2nd Int. Conf. on Computer Science and Electronics Engineering (ICCSEE 2013), 2013, pp. 1526–1529
6 Botta, A., Pescapé, A.: 'IP packet interleaving for UDP Bursty losses', *J. Syst. Softw.*, 2015, **109**, pp. 177–191
7 Kheddar, H., Boudraa, B.: 'Implementation of interleaving methods on MELP 2.4 coder to reduce packet loss in the Voice over IP (VoIP) transmission', *Int. J. Eng. Res. Appl.*, 2015, **5**, (3), pp. 1–4
8 Liang, Y.J., Apostolopoulos, J.G., Girod, B.: 'Model-based delay-distortion optimization for video streaming using packet interleaving'. Conf. Record of the 36th Asilomar Conf. on Signals, Systems and Computers, 2002, vol. 2, pp. 1315–1319
9 Lindblom, J., Hedelin, P.: 'Packet loss concealment based on sinusoidal modeling'. IEEE Workshop Proc. on Speech Coding, 2002
10 Rodbro, C.A., Christensen, M.G., Andersen, S.V., *et al.*: 'Compressed domain packet loss concealment of sinusoidally coded speech'. Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, 2003
11 Ofir, H., Malah, D.: 'Packet loss concealment for audio streaming based on the GAPES and MAPES algorithms'. Proc. IEEE 24th Convention of Electrical and Electronics Engineers in Israel, 2006
12 Bahat, Y., Schechner, Y.Y., Elad, M.: 'Self-content audio in painting', *Signal Process..*, 2015, **111**, pp. 61–72
13 Jensen, J.R., Christensen, M.G., Jensens, M.H., *et al.*: 'Robust parametric audio coding using multiple description coding', *IEEE Signal Process. Lett.*, 2009, **16**, (12), pp. 1083–1086
14 Kleijn, W.B., Shabestary, T.Z., Skoglund, J.: 'Sinusoidal interpolation across missing data'. Proc. 14th Int. Workshop on Acoustic Signal Enhancement, 2014, pp. 70–74
15 Toyoshima, M., Shimamura, T.: 'Packet loss concealment for VoIP based on pitch waveform replication and linear predictive coding'. Proc. 2014 IEEE Asia Pacific Conf. on Circuits and Systems (APCCAS), 2014, pp. 89–92
16 Xiang, K., Hu, R.: 'An improved packet loss concealment method form mobile audio coding', *Open Autom. Control Syst. J.*, 2014, **6**, pp. 188–193
17 Adler, A., Emiya, V., Jafari, M.G., *et al.*: 'Audio inpainting', *IEEE Trans. Audio Speech Language Process.*, 2012, **20**, (3), pp. 922–932
18 Griffin, A., Hirvonen, T., Tzagkarakis, C., *et al.*: 'Single-channel and multi-channel sinusoidal audio coding using compressed sensing', *IEEE Trans. Audio Speech Lang. Process.*, 2011, **19**, (5), pp. 1382–1395
19 Candès, E.J., Wakin, M.B.: 'An introduction to compressive sampling', *IEEE Signal Process. Mag.*, 2008, **25**, (2), pp. 21–30
20 Romberg, J.: 'l_1-Magic'. www.acm.caltech.edu/l1magic
21 Mallat, S., Zhang, Z.: 'Matching pursuits with time-frequency dictionaries', *IEEE Trans. Signal Process.*, 1993, **41**, (12), pp. 3397–3415
22 Tropp, J.A., Gilbert, A.C.: 'Signal recovery from random measurements via orthogonal matching pursuit', *IEEE Trans. Inf. Theory*, 2007, **53**, (12), pp. 4655–4666
23 Fornasier, M., Rauhut, H.: 'Iterative thresholding algorithms', *Appl. Comput. Harmon. Anal*, 2008, **25**, (2), pp. 187–208
24 Needell, D., Tropp, J.A.: 'CoSaMP: iterative signal recovery from noisy samples', *Appl. Comput. Harmon. Anal.*, 2008, **26**, (3), pp. 301–321
25 Schulzrinne, H., Canser, S., Frederick, R., *et al.*: 'RTP: a transport protocol for real-time applications'. IETF Audio/Video Transport WG, RFC 1889, January 1996
26 Handley, M.: 'An examination of Mbone performance'. USC/ISI res. rep. ISI/RR-97-450, 1997
27 Ramsey, J.L.: 'Realization of optimum interleavers', *IEEE Trans. Inf. Theory*, 1970, **16**, pp. 338–345
28 Miller, G.A., Licklider, J.C.R.: 'The intellegibility of interrupted speech', *J. Acoust. Soc. Amer.*, 1950, **22**, (2), pp. 167–173
29 Donoho, D.L.: 'Compressed sensing', *IEEE Trans. Inf. Theory*, 2006, **52**, (4), pp. 1289–1306
30 I. Telecommunication Union: 'ITU-TG.723.1'. http://www.itu.int/rec/T-REC-G.723.1/_page.print
31 Kabal, P.: 'ITU-TG.723.1 SpeechCoder: a Matlab implementation'. Technical Report, MMSPLab Technical Report, Department of Electrical and Computer Engineering, McGill University, August 2011
32 ITU: 'Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs'. ITU-T, Recommendation P.862, http://www.itu.int/rec/T-REC-P.862/
33 Loizou, P.: 'Speech enhancement: theory and practice' (CRC Press, Boca Raton, FL, 2007)
34 Hovorka, J.: 'Methods for evaluation of speech enhancement algorithms', *Adv. Military Technol.*, 2009, **4**, (2), pp. 27–45
35 Suzuki, J., Taka, M.: 'Missing packet recovery techniques for low-bit-rate coded speech', *IEEE J. Sel. Areas Commun*, 1989, **7**, (5), pp. 707–717
36 Andersen, S., Duric Telio, A., Astrom, H., *et al.*: 'Low bit rate codec (iLBC), RFC 3951'. Copyright (C) The Internet Society, 2004