LOW-COMPLEXITY PERCEPTUAL PACKET MARKING FOR SPEECH TRANSMISSION OVER TINY MOTE DEVICE

M. Petracca, J.C. De Martin

Dipartimento di Automatica e Informatica
Politecnico di Torino,
Corso Duca degli Abruzzi 24, Torino, Italy
Email: [matteo.petracca|demartin]@polito.it

G. Litovsky, M. Tacca, A. Fumagalli

Department of Electrical Engineering
The University of Texas at Dallas,
800 West Campbell Road, Richardson, TX, USA
Email: [ggl052000|mtacca|andreaf]@utdallas.edu

ABSTRACT

Multimedia applications in Wireless Sensor Networks (WSNs) require different approaches with respect to traditional networks due to the adopted devices' limitations in energy consumption and computational capabilities. This paper proposes a low-complexity algorithm for selecting speech packets according to their perceptual importance for voice transmission over WSNs. The proposed algorithm allows wireless sensor nodes to select which packets to protect during transmission in order to increase speech quality while at the same time minimizing the necessary energy. Experimental results based on cooperative transmission among devices show that the proposed algorithm achieves a good speech quality level while reducing the need to protect packets by 40% when compared to random selection.

Index Terms— Wireless Sensor Networks, Perceptual Speech Transmission

1. INTRODUCTION

Wireless Sensor Networks (WSNs) have been experiencing a rapid growth in the last several years as they replace old wired and wireless systems that are more expensive and harder to setup. Some scenarios in which WSNs have been successfully used are environmental monitoring, human tracking, biomedical research, military surveillance and, more recently, multimedia transmission [1].

Multimedia content diffusion over WSNs is a very promising and challenging research area which has only recently received attention by the research community. The possibility of sending voice and video using tiny and low power consumption devices enables the development of new and interesting applications, such as emergency calls or jointly environmental- and video-surveillance.

The transmission of multimedia contents over packet switched networks has been extensively studied over the years and a large number of algorithms, both for compression and transmission, have been developed. In the case of multimedia transmission over WSNs, the whole network is composed of tiny devices, which have constrained computational capabilities and require low-power consumption (a device which is installed in a location has to survive as long as possible with a small battery). With these constraints, enabling speech or video transmission requires the use of low-complexity algorithms, which in some cases have to be developed for a particular hardware.

Speech in WSNs has been explored to a certain degree in [2]. In their work, Mangharam et al. conducted speech transmission experiments in a coal mine by sending voice packets coded with the ITU-T G.726 standard [3], otherwise known as Adaptive Differential Pulse Code Modulation (ADPCM), which is capable of bit rates from 16 kb/s to 40 kb/s. The authors measured the received stream speech quality, according to the Mean Opinion Score (MOS) [4] scale, showing speech quality values ranking between poor (annoying distortion), and fair (slightly annoying distortion), in a single hop transmission with low values of packet loss rate. The authors obtained small improvements by transmitting redundant speech frames at the cost of increased bandwidth and a reduction in the benefit of the selected speech coding standard.

In this work we address the problem of enabling quality speech communications over WSNs. More in particular, we propose and develop a low-complexity perceptual packet marking algorithm suitable for implementation in tiny mote devices due to its simplicity. The aim of the proposed algorithm is to evaluate the perceptual importance of a speech packet before its compression, so as to enable selective protection of packets by the network. This results in a reduction of transmission bandwidth and power consumption. Experiments conducted using the UTMOST platform [5] developed at the University of Texas at Dallas in 2007 and a network-based protection mechanism of speech packets show the benefits of the proposed algorithm.

The rest of the paper is organized as follows: In Section 2 we address the perceptual packet selection problem by presenting a low-complexity perceptual evaluation measure and the marking algorithm. The network-based protection mechanism used for experiments is described in Section 3. Experimental results and conclusions follow in Section 4 and 5.

2. PERCEPTUAL PACKET MARKING

2.1. Packet-based voice transmission overview

The transmission of speech flows over packet-based networks involves three main entities: the sender, the network itself and the receiver. At the *sender* the voice signal is acquired, compressed according to a speech standard [3, 6] to reduce the bit rate, and packetized before transmission through the network. The *device network* receives the packet and delivers it to the destination by applying any number of internal mechanisms (e.g. buffer queues, link layer retransmission, etc.) to overcome packet losses which can still occur, particularly in a wireless scenario. At the *receiver* two main operations can be realized before sending the voice samples to the audio device: speech decoding and packet loss concealment. If a packet is received correctly it can be decoded according to the selected transmission standard, but if a loss occurs it has to be concealed so as to increase the perceived speech quality.

2.2. Perceptual Evaluation

The perceptual importance of a speech packet can be expressed as the distortion that would be introduced by its loss. The bigger the distortion value is, the more perceptually important is the packet. A common measure used to evaluate the distortion introduced by a speech packet loss is the Spectral Distortion (SD), which measures the power spectra distance between the original and the concealed packets. The SD equation is shown in (1), where S_X and $\widehat{S}_X$ are respectively the power spectra of the original and concealed speech frame X. It is important to underline that we are referring to a speech packet as a set of PCM samples acquired by means of analog to digital conversion to which a compression algorithm can be applied.

$$SD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[10log_{10}(S_X(i)) - 10log_{10}(\widehat{S}_X(i))]^2} \quad (1)$$

The SD computation requires simulating the lost packet at the sender, after which the concealment is applied and the reconstructed packet is evaluated. A Discrete Fourier Transform (DFT) is then performed on both the original and the reconstructed packets. This technique is not suitable for tiny mote devices with limited computational capabilities, thus a low-complexity distortion measure is needed, which will necessarily cause some drop in performance.

A first simplification of equation (1) can be reached by choosing a simple packet loss concealment based on silent insertion of lost frames. Even if this choice does not guarantee higher performance than interpolation or predictive algorithms, it can be easily implemented with minimal computational requirements. With the selected concealment, the log power spectrum $10log_{10}(\widehat{S}_X)$ can be considered equal to zero, all PCM samples of the reconstructed packets are equal to zero, thus the SD for a single packet depends only on the power spectrum of the original signal. According to Parseval's equality, shown in equation (2), the power spectrum of a signal is equal to the square of the magnitude of its samples in the time domain.

$$\sum_{i=0}^{N-1}|X(i)|^2 = \frac{1}{N}\sum_{i=0}^{N-1}|S_X(i)|^2 \quad (2)$$

Equation (1) with silent insertion concealment shares similarities with Parseval's equality, thus a low-complexity perceptual measure can be performed using only the PCM samples which compose a speech packet. The proposed measure is reported in equation (3), and it's based on the sum of the absolute values of the acquired PCM samples, $X(i)$.

$$P(X) = \sum_{i=1}^{N}|X(i)| \quad (3)$$

Fig. 1 shows the proposed measure as a function of the spectral distortion for a large subset of the NTT Multi-lingual Speech Database and packet length of 20 ms. Although the compared measures are not equal, they are strictly correlated, and bigger values of spectral distortion coincide with bigger values of (3).

2.3. Marking Algorithm

In packet-based network scenarios in which it is possible to provide Quality of Service (QoS) just for a small number of packets, a protection percentage is usually imposed at the source node. According to the imposed protection percentage, the source node chooses to protect packets which can maximize a particular parameter. The adoption of a perceptual selection of voice packets maximizes the received stream speech quality.

The selection of the most important speech packets is done according to two main parameters: the perceptual importance and the desired percentage of protected packets. Once the perceptual importance of a packet X is evaluated,

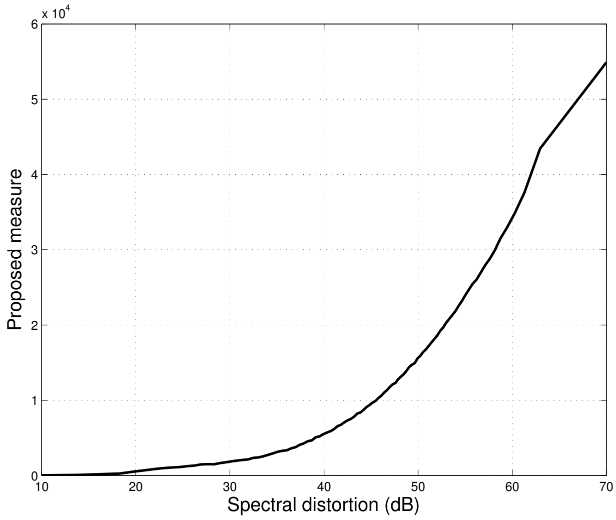

Fig. 1. Proposed measure versus Spectral Distortion.

it is classified as perceptually important if $P(X) > T(p)$, where $T(p)$ is a threshold which is a function of the protection percentage p. The various thresholds have been evaluated and tabulated starting from the cumulative distribution function (cdf) of $P(X)$, which is depicted in Fig. 2 for the previously adopted database. Also shown in the picture is the choice of the threshold which guarantees a protection percentage equal to 20%. This imposed percentage means that 80% of the packets will not be protected, thus packets with $P(X)$ less than the threshold with a cumulative probability equal to 0.8 will not be selected as perceptually important.

Even if the threshold based marking algorithm guarantees protection only for the most perceptually important speech packets, it does not guarantee that a specific speech flow achieves the desired protection percentage. This occurs because the thresholds have been selected among a certain number of speakers, so they are optimal in general. The packet marking algorithm pseudocode can be summarized as:

```
$Set\ p\ and\ M\ parameters$
$Threshold = T(p)$
while (Transmit packets) do
    $Evaluate\ P(X)$
    if (P(X) > Threshold) then
        $Packet\ marked$
    else
        $Packet\ not\ marked$
    end if
    if (Analyzed packets number=M) then
        $Update\ Threshold$
    end if
end while
```

Notice that every M packets the value of the threshold is updated. This is done empirically by tracking the percentage of protected packets in the last M packets. If such percentage is lower (higher) than the desired protection percentage p, then $T(p)$ is increased (lowered).

3. COOPERATIVE TRANSMISSION

The protection mechanism adopted to evaluate the effectiveness of the perceptual marking algorithm is based on cooperative transmission among network nodes [7].

In a wireless scenario, all devices share the same transmission medium and channel access is regulated by means of access protocols. When a packet is sent from a sender to a receiver it travels through the network hop by hop according to a previously created routing table determined by the adopted routing protocol. At each hop the device on the routing path receives the packet and attempts to transmit it toward the final destination. In addition, the broadcast nature of the wireless medium results in other nodes possibly receiving the packet as well, which gives rise to an opportunity for neighboring nodes to assist (cooperate) in packet delivery. This can help in mitigating the high Packet Loss Rate (PLR) at the receiver caused by wireless channel fluctuations.

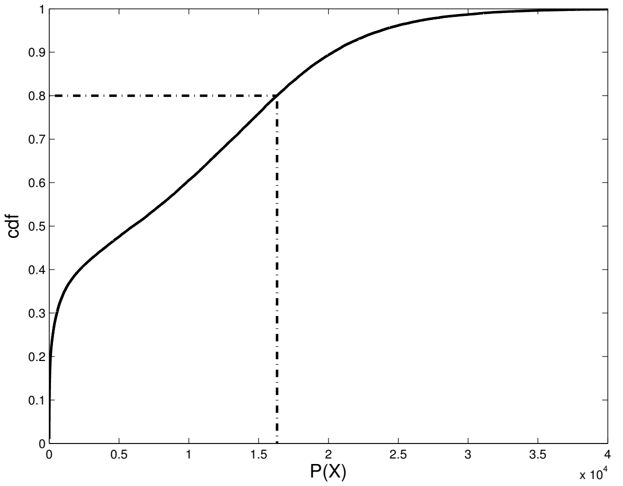

Fig. 2. $P(x)$ cumulative distribution function.

A basic cooperation system consisting of three entities, the source, the destination and relay node, is adopted. When a packet is sent from the source to the destination, a third device, usually placed between the source and destination node, can have access to the shared medium. The relay can thus forward a copy of the packet to the destination when the packet is marked as perceptually important, effectively lowering the PLR at the destination. The relay node provides a spatially distinct path for transmission in which the distribution of losses is different from the source to receiver link. The described mechanism requires a minimum device density in order to effectively provide QoS. However, relay nodes are available in WSNs because of the natural density required for wireless sensor networks to operate successfully.

4. EXPERIMENTAL RESULTS

In order to evaluate the perceptual marking algorithm performance we created a sensor network composed by three nodes, depicted in Fig. 3. The source nodes transmits speech packets compressed by means of ITU-T G.711 standard [6], A-law compression, every 20 ms, reaching a bit rate of 64 kb/s. The relay node only retransmits packets marked as perceptually important. All experiments have been conducted in an indoor scenario using the UTMOST platform while sending several speech traces, each one 3 minutes long, of the NTT Multilingual Speech Database.

Performance results of the proposed perceptual marking algorithm have been compared with random packet selection

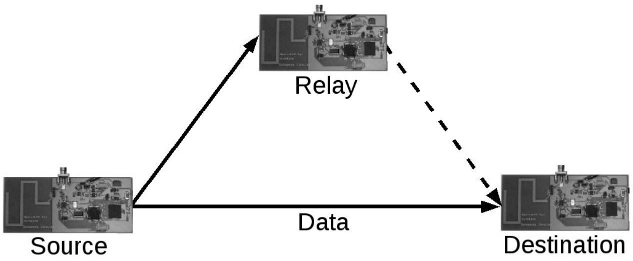

Fig. 3. Experimental data collection scenario.

	Perceptual protection					Random protection				
ΔMOS	MOS	Protected packets (%)	Target protec. (%)	SD link PLR (%)	Receiver PLR (%)	MOS	Protected packets (%)	Target protec. (%)	SD link PLR (%)	Receiver PLR (%)
0	3.067	0.00	0.00	5.70	5.70	3.067	0.00	0.00	5.70	5.70
0.169	3.311	9.36	10.00	5.70	4.94	3.142	10.00	10.00	5.70	5.13
0.258	3.483	19.46	20.00	5.70	4.31	3.225	20.00	20.00	5.70	4.58
0.390	3.709	29.25	30.00	5.70	3.60	3.319	30.00	30.00	5.70	3.99
0.532	3.956	39.42	40.00	5.70	2.91	3.424	40.00	40.00	5.70	3.40
0.636	4.174	49.56	50.00	5.70	2.24	3.538	50.00	50.00	5.70	2.84
0.622	4.279	59.49	60.00	5.70	1.78	3.657	60.00	60.00	5.70	2.27
0.567	4.354	69.43	70.00	5.70	1.27	3.787	70.00	70.00	5.70	1.72
0.429	4.374	79.29	80.00	5.70	0.77	3.945	80.00	80.00	5.70	1.16
0.247	4.378	89.27	90.00	5.70	0.49	4.131	90.00	90.00	5.70	0.58
0	4.381	100.00	100.00	5.70	0.02	4.381	100.00	100.00	5.70	0.02

Table 1. MOS performance results for both perceptual and random protections.

by means of objective speech quality measures converted to the MOS scale according to [8]. MOS values are in the range from 1 to 5, corresponding to Bad, Poor, Fair, Good and Excellent speech quality. In Table 1 we report speech quality results as a function of the protection percentage both for perceptual and random packet selections. The results refer to a collected sender to destination (SD) loss trace with a PLR equal to 5.70%, the overall MOS value with no losses is equal to 4.382. Results are also depicted in Fig. 4.

The proposed perceptual marking algorithm reaches better speech quality values with respect to random selection for every imposed protection percentage. The performance improvements range from a minimum of 0.169 to a maximum of 0.636 reaching the Good speech quality level, MOS value equal to 4, with the half number of protected packets (40% perceptual against 80% random). The most important results of the performed analysis is the lower number of protected packets required to reach the same speech quality. In fact, for the above example, the perceptual algorithm requires 40% less packets to be protected, thus reducing the relay bandwidth requirements and energy consumption.

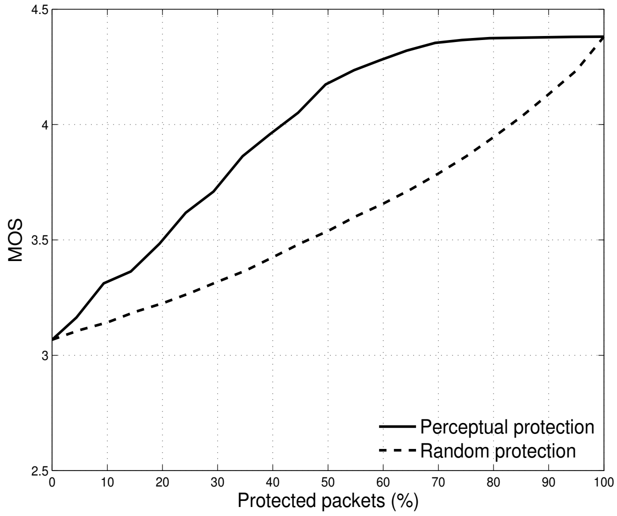

Fig. 4. Perceptual and random MOS comparison.

5. CONCLUSIONS

In this paper we proposed a low-complexity perceptual based packet selection algorithm suitable for implementation on tiny mote devices for speech transmission over WSNs. The algorithm allows wireless sensor nodes to select which packets to protect during transmission in order to increase speech quality while expending a small amount of energy. Experimental results based on a cooperative transmission among devices show it is possible to obtain speech quality improvements with respect to a random selection for the same percentage of protected speech packets. Moreover, our results show that the proposed algorithm enables achieving a good speech quality level while saving up to 40% of protected packets, compared to random selection, thus reducing the required additional bandwidth and devices energy consumption.

6. REFERENCES

[1] I.F. Akyildiz, T. Melodia, and K.R. Chowdury, “A survey on wireless multimedia sensor networks,” *Computer Networks (Elsevier)*, vol. 51, no. 4, pp. 921–960, March 2007.

[2] R. Mangharam, A. Rowe, R. Rajkumar, and R. Suzuki, “Voice over sensor networks,” in *Proc. IEEE Real-Time Systems Symposium*, Rio de Janeiro, Brazil, December 2006, pp. 291–302.

[3] ITU-T Recommendation G.726, “40, 32, 24, 26 kbit/s adaptive differential pulse code modulation (ADPCM),” 1990.

[4] ITU-T Recommendation P.800, “Methods for subjective determination of transmission quality,” 1996.

[5] G. Litovsky, M. Tacca, and A. Fumagalli, “UTD mote system (UTMOST): Improving functionalities in wireless sensor nodes,” Tech. Rep. UTD/EE/10/2008, The University of Texas at Dallas, January 2008.

[6] ITU-T Recommendation G.711, “Pulse code modulation (PCM) of voice frequencies,” 1988.

[7] N. Agarwal, D. ChanneGowda, L. Narasimhan Kannan, M. Tacca, and A. Fumagalli, “IEEE 802.11b Cooperative Protocols:A Performance Study,” in *IFIP/TC6 NETWORKING 2007*, Atlanta, GA, USA, p. 415.

[8] ITU-T Recommendation P.862.1, “Mapping function for transforming P.862 raw result scores to MOS-LQO,” 2003.