# Towards Inherent Error-Resilient Voice Encoding Schemes in Audio Sensor Networks

Okan Turkes
Department of Computer Engineering,
Yeditepe University, Istanbul, TR
oturkes@cse.yeditepe.edu.tr

Sebnem Baydere
Department of Computer Engineering,
Yeditepe University, Istanbul, TR
sbaydere@cse.yeditepe.edu.tr

## ABSTRACT

Recently, multimedia utilization in Wireless Sensor Networks (WSN) has shown that robust encoding methods are imperative for any application requiring a certain level of quality. During ubiquitous data exploitation in these lossy networks, a data-conserving method coherent with coding and transmission scheme is essential. This study centers upon several basic reconstruction methods for unapprehended parts in the voice data gathered at the end-point of a real multi-hop Audio Sensor Network (ASN). Considering gathered voice signals, error concealment (EC) methods are inherently applied over lost packets in the testbed. Around 6,000 real single-path transmission tests are verified with an instrumental simulation. Besides, EC schemes are supported with a multi-path transmission in which data aggregation occurs at certain intervals. Nearly 300,000 qualitative results show that perceptual quality can be preserved promisingly with the utilization of low cost affordable correction techniques.

## Categories and Subject Descriptors

C.3 [**Special-purpose and Application-based Systems**]: Signal processing systems; C.2.1 [**Network Architecture and Design**]: Wireless Communication

## General Terms

Design, Experimentation, Measurement, Performance

## Keywords

Audio Sensor Networks, Voice Coding, Error Concealment, Voice Quality Assessment, Activity monitoring

## 1. INTRODUCTION

Recent advances in micro-electro-mechanical systems have given a rise to ubiquitous exploitation of multimedia data in sensor networks. Referencing any content format in these networks introduced the era of Wireless Multimedia Sensor Networks (WMSNs) [1]. Nevertheless, nature of multimedia coding and transmission introduces more demand in processing and energy. Even in resourceful networks, processing and transmission of multimedia data are nontrivial tasks. Techniques based on static resource reservation are mostly used in those networks. Yet, simplicity for a lightweight data encoding is a must in Audio Sensor Networks (ASNs). Because, nodes comprising relevant networks have limited capabilities for data reservation. To embrace a stringent Quality of Service (QoS), need for quality-potent multimedia transmission schemes is inevitable [11]. Particularly for ASNs, data quality assessment is significant during transmission. There are considerable number of inherent audio characteristics which affects intelligibility. Also, several impairments occur due to equipment quality, quantization, processing and transmission delays, etc. Their effects in quality are high when a counterbalance between a reasonable QoS and data handling is needed to be set. Sustaining a certain level in voice data validity is important in the shade of packet losses. Inherently, relevant parts inside a voice can pave the way for implementation of an error-resilient scheme. In this respect, error concealment (EC) approaches enable effective mechanisms that can recondition the distorted data as closely as the original without increasing the bandwidth demand [22].

In order to preserve data validity, smooth coding of the voice being recorded in the source node of an ASN has to be taken into account. Basically, voice data quality increases with higher sampling frequencies ($f_s$) and bit depths ($bd$). But, coding efficiency matter for convenient and continuous traffic through rest of the network. Besides, transmission of the samples gathered requires an efficient buffering management suitable for the adaptive network constructed. By this means, we have proposed a lightweight solution for voice coding and transmission in [18]. In pursuit of this study, we have shown that an error correction method on the received data can improve the voice quality [19].

Accordingly, this study analyzes several inherent error-resilient schemes in our voice coding and transmission model. Nearly 6,000 real multi-hop transmissions are conducted to analyze the effect of well-accepted EC methods in varying voice and network properties. With a conducive simulation, real testbed experiments are validated with around 300,000 sensitivity and random sampling analysis tests. Apart from single path transmission, effect of a data aggregation algorithm working on a multi-path network scheme is also presented. The results obtained from each transmission are


*WiNTECH'12,* August 22, 2012, Istanbul, Turkey.

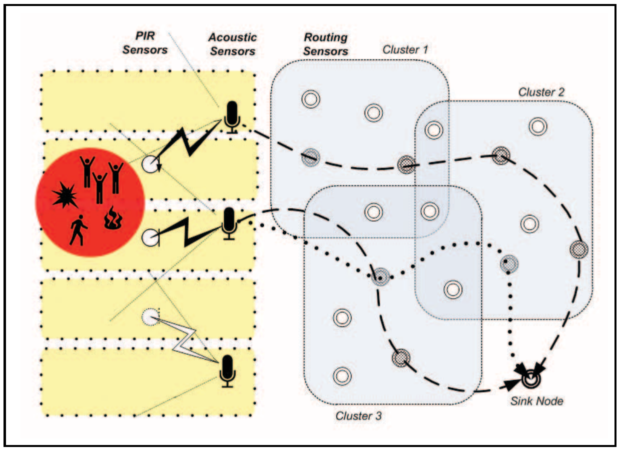



**Figure 1: ASN Activity Monitoring Scenario**

assessed with basic signal-to-noise ratio metric and a modified version of R-factor which can be mapped to a certain perceptibility. So, an extensive analysis of EC schemes in our network models is presented under both objective and perceptual basis. The promising enhancements obtained in perceptual voice quality are presented.

To evaluate the encoding schemes presented, we contemplate on an activity monitoring as a target application scenario, as demonstrated in Figure 1. Our main aim is to make disabled, helpless or elderly people living indoor benefit from the freedom of transferring information over a scalable and automated wireless medium. Crucial parameters of the schemes are determined by means of these necessities.

The rest of the paper is organized as follows. Section 2 renders the related work. The voice coding and transmission model is given in Section 3. Section 4 discusses on experimental setup. In Section 5, performance of the presented schemes are given. Conclusion is outlined in Section 6 with comments on forward plans.

## 2. RELATED WORK

Researchers found several techniques to deal with missing parts of a multimedia data due to lossy nature of WMSNs. In well-known surveys on EC [20, 3], authors present existing EC schemes for real-time audio and video transmission over traditional networks. However, streaming a content format has to be characterized neatly by taking the natural constraints of WMSNs into consideration. For that matter, error resilience in WMSN is an indivisible whole which requires concurrency and data integrity together.

Error-resilience approaches like Automated Repeat Request (ARQ) [9] relying on sender-side usually try to retransmit the lost data. Forward Error Correction (FEC) is also a popular technique where the data is encoded in a redundant way at the source that it can be later handled at the receiver[23]. However, these schemes are very costly and may not be suitable for WMSNs [7, 16]. Nevertheless, there are several studies focusing on these methods in which transmission schemes have to avoid any loss for compression. A packet redundancy protection is accommodated on preselected relay nodes responsible for retransmitting perceptually important lost packets [14].

Source-coder independent EC methods exploiting information only at the receiver side provide decent performance when packet losses are not under a certain level. These methods such as zero (white) substitution, packet repetition, interpolation are straightforward and therefore renowned strategies. Receiver-based EC approaches present a light-handed reconstruction because of their extremely low complexities. However, there is a possibility for amplitude and phase mismatches during reconstruction. When consecutive losses increase, effect of reconstruction for receiver-based strategies dramatically decreases. There are several modified versions of receiver-based EC algorithms to be exploited according to any application purpose. For instance, in [17], authors propose a pattern repetition algorithm which achieves a significant loss reduction with fading packet repetition (FPR) strategy. FPR employs two buffers to keep received packets and to monitor forthcoming gaps during reconstruction. In [6], a packet-based EC method on speech signals which eliminates losses caused by packet delay jitter is presented. Park et. al focus on a multiple adaptive codebook-based approach to reconstruct decoded signals [13].

A convenient transmission scheme selected for any application can also address transmission problems in ASN. By using multiple paths, error-resilience can be achieved [21]. MMSPEED [4] is the one of the representative of using multiple paths to achieve QoS for data and time domain. In [15], authors propose an EC method based on low parity checking in one part of their multi-path transmission. Also, bandwidth requirements can be satisfied by using multiple paths [12]. In [5], a node-disjoint set is presented for achieving bandwidth problems and maximizing data flow rate.

There is an absolute need for new criteria and mechanisms to handle multimedia streams in accordance with desired QoS. Yet, the number of research studies stressing on expected lifetime is very few. Studies which draw conclusion for multimedia data transmission generally focus on reducing bandwidth and provide an efficient data delivery in every part of a WMSN. It is quite evident that most of the presented methods include several assumptions that cannot yield beyond a theoretical basis. For theoretical validations, a numerous range of experiments must be conducted on real implementations [2].

## 3. VOICE CODING AND TRANSMISSION

In ASNs, mutual affection of voice coding and transmission is a stringent fundamental for any application design. Hence, strict interaction between coding techniques and QoS plays a significant role for the transmission scheme of this study. To make vigorous provisions against information dissipations, data handling is supported with error-resilience methods. Regardless of any topology, an affiliated node is designed to be bandwidth-efficient while bulk data delivery. An affordable memory which does not exceed resource limitations is provided.

### 3.1 Data Recording

An important prerequisite during voice recording is preserving data intelligibility. Also, analog to digital conversion (ADC) needs reasonable time to interpret collected streams. During sampling, determining a level for $f_s$ plays a key role for an operational ASN. The second prominent property during ADC is *bd*, which describes number of bits to represent a sample value. By increasing *bd* resolution, quantization errors in ADC are reduced.

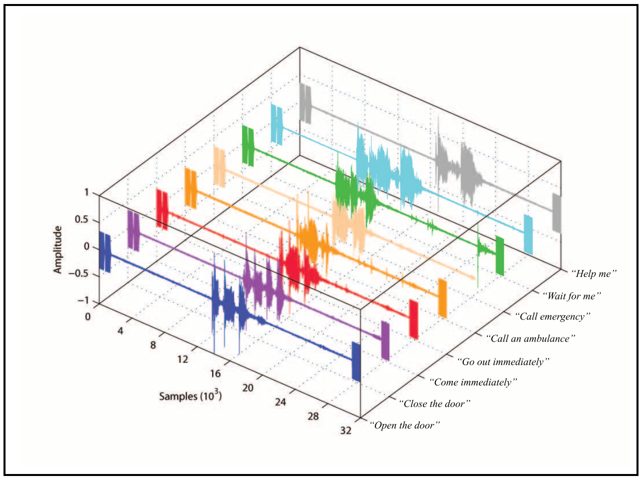



**Figure 2: Commands used in the testbeds**

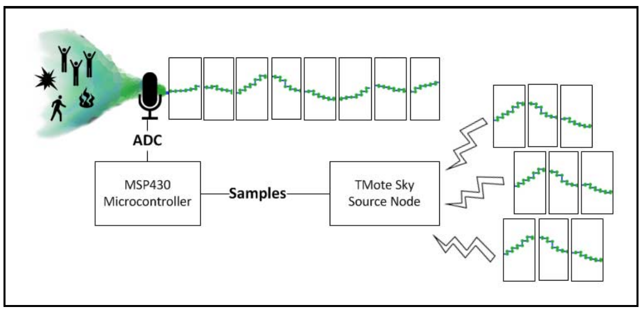



**Figure 3: Forwarding ADC samples to source node**

A voice sample set $D_i^s, i = 1, 2, \ldots, 8$ shown in Figure 2 is generated comprising of simple invocatory commands like "call emergency", "help me", etc. Voices are prerecorded with $f_s$ (KHz) $= \{4, 6, 8, 12, 16\}$, $bd = \{8, 16\}$ and have a duration of $t = 4$s. For source nodes, recording is hard while dealing with transmission. Nevertheless, they are considered as powerful devices in the literature [10]. Analogically, we record voices via an acoustic sensor on a micro-controller which is working in aggregate with the source. As illustrated in Figure 3, digital amplitudes $a_0$, $a_1$, $a_2$, $\ldots$, $a_n$ are generated by the micro-controller and sent to the source. In range of real numbers $r_A$=$\{-1, +1\}$, finite set of amplitude values constitutes the overall voice data $A$.

## 3.2 Data Segmentation

Partitioning of the recorded streams into data segments includes low-cost steps in an effort to provide negligible delay during packet delivery. In a unique test, nodes can have a segmentation size $s_w = \{20, 40, 80\}$ which indicates the number of amplitude values inside. A segment $s_i, i = 1, 2, \ldots, n(s)$ comprised of $s_w$ amplitude values inside is created by dividing $A$ into several partitions. Partition sets $A_{s_i}, i = 1, 2, \ldots, n(s)$ are encapsulated with corresponding indices $i$ to form network packets $p_i, i = 1, 2, \ldots, n(p)$. With a corresponding $f_s$, $n(p)$ differs according to $s_w$ determined, where dense samples may yield in end-to-end delay. $bd$ is also an influential fundamental to determine $s_w$. Therefore, segmentation and subsequently generation of network packets are tightly interconnected actions.

## 3.3 Data Delivery

In order to avoid congestions and delay-jitter, a simple buffer management is utilized to minimize processing delay during pre-transmission. A node starts to fill its buffer $B$ out until it gets full. In this mechanism, two kind of messages are used: control messages (CM) and data messages (DM). Since the buffered DM in the nodes are transmitted without any reliability, no delay is introduced in terms of acknowledgments. But, CM are transmitted with a reliable protocol. So, consecutive nodes easily synchronize the transmission. The occupancy level of $B$ is kept track with the local counter $l = 1, 2, \ldots, 10$. Buffer cells also hold real packet indices of the perceived data to provide an effortless signal reconstruction in advance.

## 3.4 Error Concealment

We focus on well-known feasible EC methods in a comprehensive manner. As illustrated in Figure 4, underlying logic in all of these algorithms is handling errors caused by packet losses in a particular segment with the help of its neighbors. Under favor of the simplicity of the methods and the buffering management, tolerating consecutive losses becomes possible in a specified extent. Successfully gathered packets in $B$ are utilized when a packet loss is identified.

Although the methods differ from the point of assessment, there are several common steps in their implementation. Figure 5 shows that an EC method begins whenever $B$ gets full. A lost packet index is determined by considering the real indices of the successfully received packets in $B$. If a hole in the transmission is detected, an error list is utilized to keep track of every lost packet indices. For every lost packet indices, EC algorithm is run. Reconstructed packets are set as modified and pushed in a new buffer $B_C$ created for the concealed signals. $B_C$ arranges the real indices of the reconstructed signals, thus concealed data is prepared for transmission. Data delivery is retained until reconstruction for each lost packet is completed. With the completion of EC in a single transmission step, trigger for continuation of the transmission is provided.

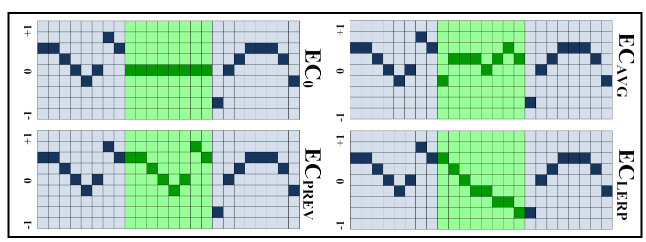



**Figure 4: EC techniques illustrated**

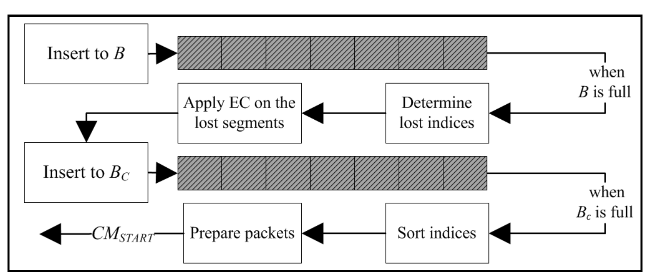



**Figure 5: Buffer management during EC handling**

EC methods are appraised with the encoding scheme implemented in accordance to the transmission fundamentals. Since reconstruction mechanisms retain transmission for a specified time, a spatial extent is determined for error handling. According to this, EC methods behave mild for the losses in the boundaries of the buffer authority. Size of $B_C$ is same with of $B$, and number of the concealed packets cannot exceed it. Thus, undetected lost packets are treated with no concealment (NC) in order not to pulse delay-jitter.

### *3.4.1 EC with Silence Substitution ($EC_0$)*

The aim of $EC_0$ is to directly replace real positions of the lost sample values with zero amplitude values. As a raw reconstruction method, $EC_0$ can be considered as adding silence in place of pattern holes. Most of the studies refer this method as NC method [20, 19].

### *3.4.2 EC with Repetition ($EC_{PREV}$)*

One of the simplest error resilience technique is substituting the lost segments with their previously received neighbors [20], which is the exact idea in $EC_{PREV}$. Reconstruction for the losses are considered for the past direction in time and future information is discarded.

### *3.4.3 EC with Averaging ($EC_{AVG}$)*

More correlated information reconstruction can be obtained by utilizing not only the preceding signals in time, but also the following signals. In $EC_{AVG}$, mean values of the amplitudes taken from the successfully transmitted neighbors of a lost segment form up the reconstructed segment.

### *3.4.4 EC with Linear Interpolation ($EC_{LERP}$)*

One of the simplest determination for interpolation is fitting the spaces with a linear approach. In $EC_{LERP}$, a first-order curve fitting is applied on the lost segments. For a lost packet, the method links the last amplitude value of the preceding packet with the first one of the next packet. Interpolated values are replaced in the place of the lost signals. So, possible incoherent reconstructions are foreclosed.

## 4. EXPERIMENTAL SETUP

System model devised for the testbed is constructed with regard to the general scenario. ASN transmission scheme is shown in Figure 6, which consists of a set of sensors; Type 1 $S_i^A, i = 1, 2, \ldots, l$, sensor equipped with a microcontroller that has an acoustic sensor on it, Type 2 $S_{ij}^R, i = 1, 2 \ldots k_1, j = 1, 2, \ldots, k_2$, simple routing sensors, and a sink.

Transmission ways are twofold: 1) single path and 2) disjoint multi-paths constructed with a chained topology which have different link qualities. At certain spatial intervals, paths are engaged at specialized intermediate nodes implemented with a fusion design.

Single-path transmission is implemented with a real experimental setup. Besides, a simulation is devised to generalize the outcomes of single-path transmission. Multi-path transmission with fusion is implemented within the simulation.

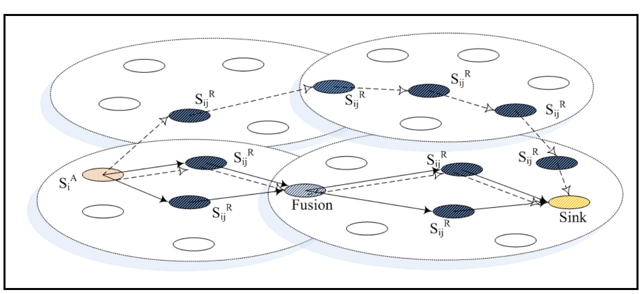
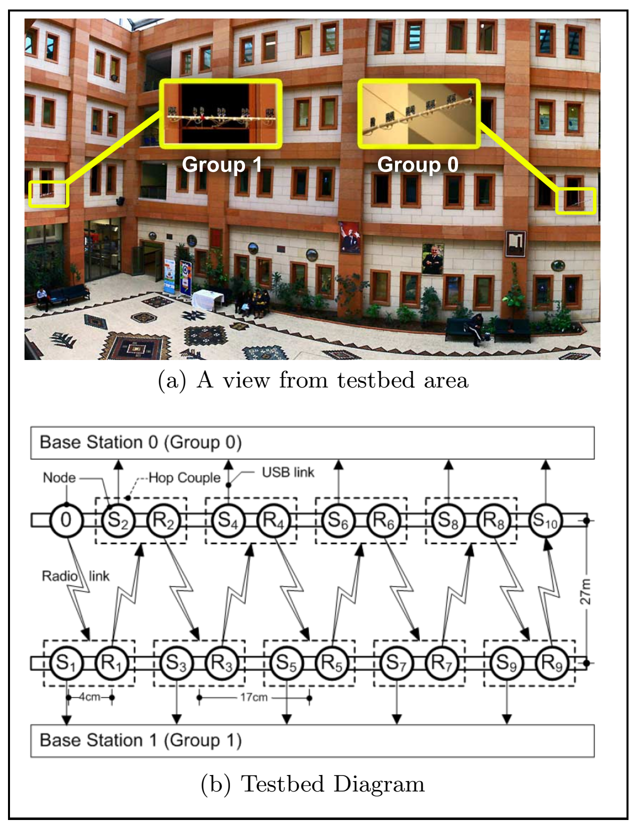



**Figure 6: General ASN Transmission Schemes**

### 4.1 Real Setup

Real tests are conducted inside a large atrium of a building as illustrated in Figure 7(a) For sharing a common communication medium open to implicit environmental factors, nodes are homogeneously deployed at a same communication range. A 10-hops network is set up by using 20 TMote Sky sensor nodes which are homogeneously deployed at a same communication range. In the experiments, TinyOS v2.1 with nesC v1.3 [8] is utilized to realize a simple voice transfer over a single-path chain topology. Nodes are partitioned into two groups named Group 0 and Group 1. They are lined up vertically on thin linear sticks which are approximately 5m horizontally above from ground, with no obstacles between them. As Figure 7(b) depicts, the sticks equipped with the sensors nodes are positioned parallel to each other to complete a hypothetical rectangular area. Distance between two groups is measured as 28m. Output power of the nodes is set to -7dBm. Each node is connected to a base station computer in order to get results from intermediate hops.

The groups consist of five "*hop couples*" with intra and inter couple spacings of 4 and 17cm, respectively. 2 nodes




(a) A view from testbed area

(b) Testbed Diagram

**Figure 7: Real Testbed Environment**

in a couple are given the same node ID. One of the nodes, called relay node ($R_i$, i=1, 2 . . . 9), is used to send the incoming data to the next hop couple with consecutive node ID via radio link. The other one, called snooping node ($S_i$, i=1, 2 . . . 9), is used to send the data to the base station computer via USB link. So, intermediate results are recorded in each hop. Base station computers at each side record the data along with corresponding loss patterns and packet error rates (PERs).

The voice transmission scheme can be cascaded into several steps as follows: A voice data having $t = 4$s is partitioned into $s_w$ sized segments at the $S_i^A$ node. Each segment is encapsulated into network packets and conveyed towards the sink over a single-path multi-hop network. In anticipation of generating a homogeneity by means of data variety with an agile transmission environment, voice recording process for each test is by-passed. $S_i^A$ is already supposed to be more capable, therefore its role is accomplished by a computer via functions implemented under Matlab and Java. A voice data file is prerecorded via the acoustic sensor of the micro-controller, having properties as $f_s = 8$KHz and $bd = 8$ to be utilized in the real tests. Different $s_w = \{20, 40, 80\}$ are utilized as the distinctive in-network parameters in different tests. So, data packets are generated with size of 20B, 40B and 80B, respectively with extra 2B used for the segment offsets. The packets are transmitted from the computer to the source node via USB links. Then, data packets are transmitted through $S_{ij}^R$ nodes to the sink. EC algorithms with buffering management are applied on the lost segments. Back-channel mechanism of the snooping nodes is used to collect the data at each hop. For each test, base extracts the reconstructed lost patterns and their indices from the received voice. Moreover, transmission details and information of the hops are saved as mask patterns for further sensitivity analysis in the simulation.

## 4.2 Simulation Setup

Single-path transmission results are expanded to determine best voice and in-network parameters with the simulation. Besides, a multi-path delivery with data fusion, which is not ruminated over the real testbed setup, is also implemented. For all EC schemes, the voice set in which each data has same $t$, but with several $f_s$ and $bd$ versions is dissected into $s_w$ sized segments. Segments are tuned into network packets and transmitted over several paths of designated topologies.

Despite the convenience of an extensive setup in a simulation, several aspects in the implementation process require a complicated waterfall model consideration. For this reason, a number of implementation criteria is presupposed in this simulation design. On the other hand, we aim a target-driven application which only focuses on loss pattern generation on different network topologies. Therefore, a high level of abstraction is thought for the protocol stack used in WSN. Physical and link layer specifications and requirements are substantially discarded. Nodes are assumed to be perfect in terms of survivability and maintenance. Network layer is implemented in terms of packet transmission, and by-passes the calculations for achieved net average bit rate in the goodput. The packets are transferred to the application layer with no delay occurrence. In the application layer, some limitations related with processing and memory requirements are taken into consideration with simplistically.

As the testbed setup depicted in Figure 8, we setup a network topology consisting two disjoint multi-hop paths in Matlab. In each test, transmission reliability of the paths differ according to link quality definitions. There are 10 hops for each path. On one level, the disjoint paths can be considered as the simulated instances of single-path. On another level, starting from the source, paths knot with each other in every two hops, so there exists 5 fusion nodes. The knitting nodes correct the unperceived packets of disjoint paths by aggregating the data received from different paths.

Original voices $D_i$, $i$=1, 2 . . . , 8 with $t = 4$s, $bd = \{8, 16\}$ and $fs$ (KHz) $= \{4, 6, 8, 12, 16\}$ are read and partitioned into segments having $s_w = \{20, 40, 80\}$. So, a data set with 240 distinct structure elements is formed up in a single test. For each test, an initial pattern, as a 1's-array at the beginning, is transformed into devalued versions at each hop by setting the lost packet indices to 0 in each link transfer. Size of the initial pattern, so as its ensuing versions with regard to a stochastic approach, matter in order to be applicable for every $f_s$ and $s_w$. Cardinality of the initial pattern expressed in Equation 1 must be divisible for all $f_s$ and $s_w$ versions of the segmented data set in order to hold the tests at once, hence it must have a general cardinality for every $n(p)$. So, we utilized a 2D array to hold every $n(p)$.

$$\bar{\bar{P}}_{f_s, s_w} = n(p) \tag{1}$$

General pattern cardinality $\bar{\bar{P}}$ for all $f_s$ and $s_w$ defined is calculated as 9600 by the least common multiplier of all cardinalities in the set, as expressed in Equation 2.

$$\bar{\bar{P}} = LCM(\bar{\bar{P}}_{6,20}, \bar{\bar{P}}_{6,40}, \bar{\bar{P}}_{6,80}, \ldots, \bar{\bar{P}}_{16,40}, \bar{\bar{P}}_{16,80}) \tag{2}$$

New patterns are formed as replicas of the initial pattern while transmitting the packets on their paths. For the fusion nodes, data received from disjoint paths are aggregated with regard to their packet indices. During fusion, if packet indices coincide, one of them is taken into account. In case of no coincidence, both of the packets are reconstructed. Data aggregation is used to lower PER at some extent and to ensure hop-by-hop reliability. Pattern generation is continued by transmitting the fused patterns over the paths. 25 general patterns are formed in a unique test. Corresponding masks $M_t$, $t$=1, 2 . . . , $n$ are generated from the loss patterns generated for every $s_w$ and $f_s$. In each test, they are applied on the segmented data set in each hop. For every $\bar{\bar{P}}_{f_s, s_w}$, $\bar{\bar{P}}$ is down-sampled to actual $n(p)$ of a unique data. When projection of a single data is over, lost data segments are determined. In each test, a total number of 6,000 masks are projected on the data set.

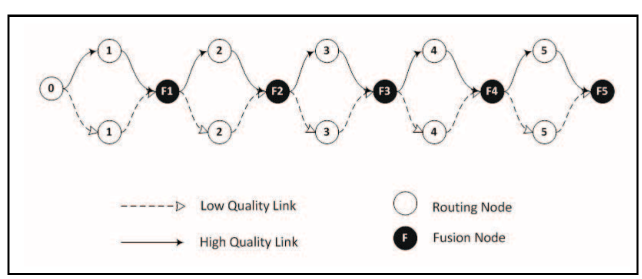



**Figure 8: Simulation Testbed Diagram**

### 4.3 Voice Quality Assessment

Assessing the quality of the reconstructed data at the receiving end of an error-resilient network has significant importance. Voice quality is subject to multiple factors which are unlikely to be independent. Metrics focusing on multi-dimensional properties of voice can be as complex as possible. However, to assess voice quality on-the-fly, any measure chosen has to be suitable to be accommodated in the memory of the sensor nodes. In this study, a modified version of R-factor [19] as shown in Equation 3 is applied for voice quality assessment.

$$R(fs, Ppl) = 58.9843 - 95\frac{Ppl}{Ppl+1} + 2.0714 \times fs \qquad (3)$$

where $Ppl$ is treated as packet loss impairment factor. Thus, effect of PER on voice quality is wanted to be revealed. Since voice quality also depends on sample rate generated in a particular time, $f_s$ is set as simultaneous impairment factor. Any R-factor value can be mapped to a Mean Opinion Score (MOS) which is a perceptual grading determined by an experimental group of audience. With their perceptual grades for voice quality, ranging from bad to excellent, MOS is identified among a numerical quality scale from 1 to 5, respectively. Thus, R-factor gives the advantage to evaluate received data in both objective and subjective manners. As a second metric, signal-to-noise ratio (SNR) is used as

$$SNR(dB) = 10\log_{10}\frac{|A_{signal}|}{|A_{noise}|} \qquad (4)$$

to examine the data quality estimation over the chain topology, where $|A_{signal}|$ is the sum of total absolute original amplitude values and $|A_{noise}|$ is of the difference between reconstructed amplitudes from the original ones. Effects of $bd$ selected and $s_w$ of the data are examined with SNR. To assess both network and voice quality, R-factor and SNR values are compared.

## 5. PERFORMANCE ANALYSIS

Several qualitative evaluations are intended to give insights into our error-resilient voice coding and transmission. Voice characteristics and in-network properties are analyzed with a set of sensitivity analysis. Hereafter, a quantitative analysis of the results is conducted to estimate the quality obtained for the proposed schemes. An extensive evaluation of EC methods over the real and simulated setup is supported with random sampling tests. Contribution of data fusion over the multi-path transmission is also investigated.

### 5.1 Sensitivity Analysis

For a subtle evaluation of our design inputs, a sensitivity analysis is devised by experimenting a set of inputs as constants of the transmission while others are deliberately differ in a large variability. In each experiment, a single $f_s$, $bd$ or $s_w$ is set as the constant. Around 4,000 real SNR results according to the hop numbers and different $s_w$ of the received voices sampled at $f_s$ = 8KHz are shown in Figure 9. It is clearly seen that, correlation among SNR values and each $s_w$ offered in the tests slowly declines. However, we can evenly state that there is no big difference between the results of different $s_w$. Also, differences in $bd$ do not alter SNR values of data with same $f_s$ too much. Therefore, $bd$ = 8 and $s_w$ = 40 are determined as suggested constant voice properties.

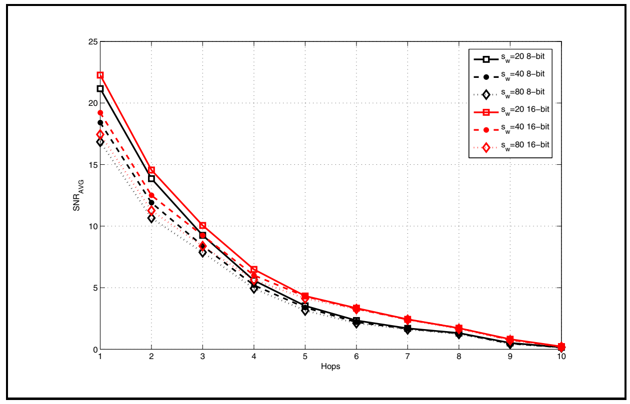



**Figure 9: EC$_0$ SNR values for different $s_w$ and $bd$**

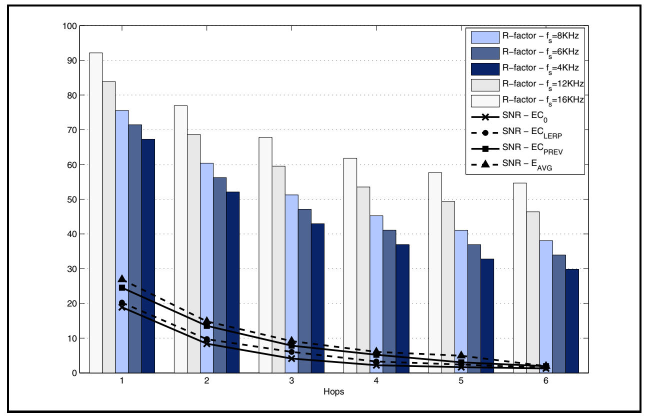



**Figure 10: Results for EC methods with different $f_s$**

Over 6 hops, effects of $f_s$ and PER on the reconstructed signal quality and intelligibility are shown in Figure 10 which demonstrates nearly 15,000 simulation test results. For $f_s$ = 8KHz, a quality-potent data cannot be assured after the fourth hop and need for error-resilient scheme is thereby inevitable.

### 5.2 Random Sampling Analysis

Differences between EC methods are analyzed with a set of random processes. In the real transmissions, a data sampled at 8KHz/8bit is transmitted with $s_w$ = 40. For each EC method, 450 real transmissions are conducted. Nearly 6,000 corresponding loss patterns are stored in each hop for further evaluation with the simulation.

Correlation between SNR values and PER is depicted in Figure 11. SNR values exponentially decrease when PER increases. Results explicitly indicate that EC$_{AVG}$ has the maximum performance among other techniques. The second successful gain is obtained by EC$_{PREV}$ method. Unexpectedly, the results of EC$_{LERP}$ are unpromising even than of EC$_{PREV}$. As expected, EC$_0$ has the worst performance among other reconstruction strategies.

Running a set of Monte Carlo simulations for 1,000 times, we obtained 300,000 transmitted data with uniformly distributed PERs. Promisingly, real test outcomes are quite similar to the simulated results. Strong correlation between them is shown in Figure 12.

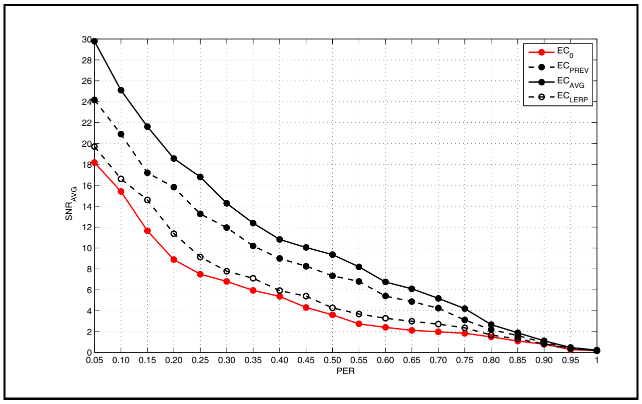



**Figure 11: Real testbed results for EC methods**

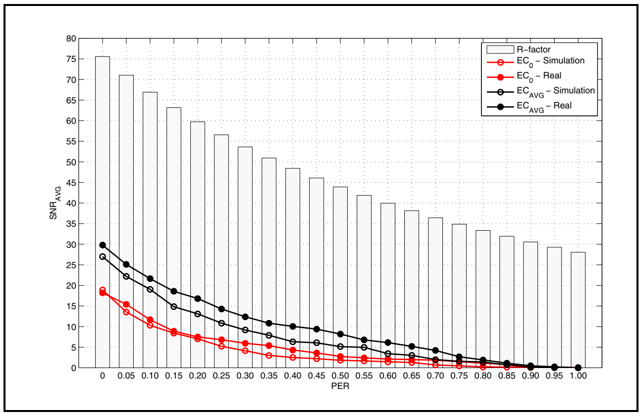


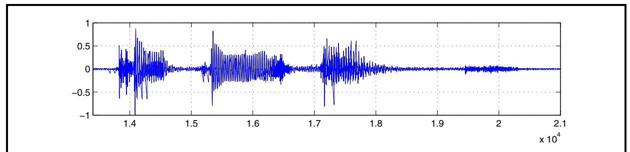

**Figure 12: Real and simulated transmission results**

## 5.3 Implementation Costs

Implementation costs of the presented EC techniques are given in Table 1. They are implemented on real sensor nodes formed of TMote Sky nodes. They are programmed in TinyOS v2.1 with nesC v1.3. In the implementation, data is transmitted in packets that hold $s_w = 20$ amplitude values. Buffer utilization is at the minimum level for $EC_0$. Besides, other EC techniques utilize exactly same buffer size when the worst case is taken into consideration. According to this, $B$ is always filled with its entirety during packet transmission for any EC method. For $B_C$, we calculate the costs as if all of the buffer cells are utilized to reconstruct the packets. $B$ and $B_C$ constitute the send buffers. In terms of program size, memory utilization, energy consumption and time, EC techniques do not differ so much from each other. Alike as the real and simulation results obtained, the performance of $EC_{AVG}$ is also notable in Figure 13.

**Table 1: Implementation Costs**

| Method | CPU Cycle | Program Memory (bytes) | Data Memory (bytes) | Network Buffer (bytes) | Energy (mJ) |
|---|---|---|---|---|---|
| $EC_0$ | 5720 | 630 | 82 | 2 | 2.15 |
| $EC_{PREV}$ | 6042 | 632 | 82 | 250 | 2.27 |
| $EC_{AVG}$ | 6120 | 632 | 82 | 250 | 2.30 |
| $EC_{LERP}$ | 6204 | 632 | 84 | 250 | 2.32 |




(a) A section from original voice

(b) The section $EC_0$ applied when PER=30%

(c) The section $EC_{AVG}$ applied when PER=30%

**Figure 13: A section when $EC_0$ and $EC_{AVG}$ applied**

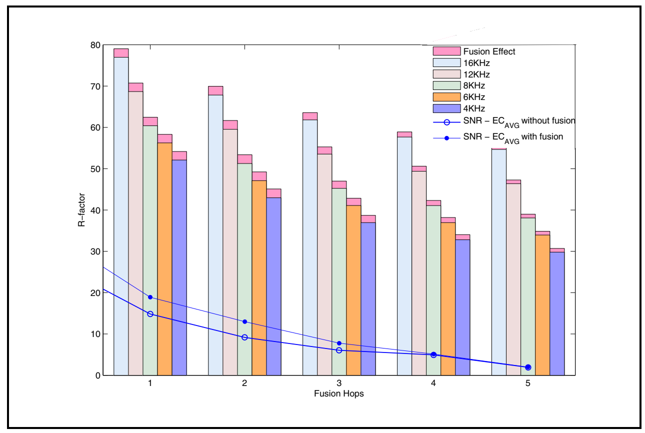



**Figure 14: Data fusion effect in a multi-hop delivery**

## 5.4 Effect of Multi-path Transmission

We also demonstrate the contribution of data aggregation algorithm on the knitting nodes by comparing R-factor values. For each of the fusion hops, performance of data aggregation is evaluated by comparing the same PER results of a single hop transmission. As R-factor values in Figure 14 demonstrate, multi-hop transmission with data fusion has a notable contribution to voice quality. Nevertheless, the performance of the data aggregation decreases when the level of fusion in the network increases. Regardless of $s_w$ or $bd$ of a data transmitted, R-factor values only depend on the overall PER and $f_s$ of the voice. Therefore, we also use SNR to understand if there is a difference between normal and reconstructed signals. It is clear that fusion has a remarkable contribution to data quality.

# 6. CONCLUSION

The main focus of this study is to set up a coding and transmission model with inherent error-resilience schemes in a real multi-hop ASN. With consideration of intrinsic system and data characteristics, a set of well-accepted EC techniques are presented with extensive performance assessments. Results prove a decent improvement in voice quality unless PER dramatically increases. Hence, a simple data aggregation technique based on a multi-path transmission is also presented for quality improvement. We prove an error-resilient encoding scheme can be provided by exploiting inherent properties of the voice transmitted over a specialized multi-hop ASN. As future work, we focus on importance levels of the packets in priority-based strategies where preferential data can be determined with more inherent data characteristics like signal energy, difference to previous packets and voice onset or transition indicators. With a particular network protocol, these importance levels can be exploited in a straightforward manner. We aim to conclude our coding and transmission scheme with a complete transmission framework for WMSNs.


## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial assistance of Turkcell Akademi for this research progress.


## 7. REFERENCES


[1] I. Akyildiz, T. Melodia, and K. Chowdhury. A survey on wireless multimedia sensor networks. *Computer Networks*, 51(4):921–960, 2007.

[2] I. Akyildiz, T. Melodia, and K. Chowdhury. Wireless Multimedia Sensor Networks: Applications and Testbeds. *Proceedings of the IEEE*, 96(10):1588–1605, 2008.

[3] V. Bhute and U. Shrawankar. Error concealment schemes for speech packet transmission over IP network. In *Systems, Signals and Image Processing, 2008. IWSSIP 2008. 15th International Conference on*, pages 185–188. IEEE, 2008.

[4] E. Felemban and E. Ekici. MMSPEED: multipath Multi-SPEED protocol for QoS guarantee of reliability and. Timeliness in wireless sensor networks. *IEEE Transactions on Mobile Computing*, 5(6):738–754, June 2006.

[5] P. Floréen, P. Kaski, J. Kohonen, and P. Orponen. Exact and approximate balanced data gathering in energy-constrained sensor networks. *Theoretical computer science*, 344(1):30–46, 2005.

[6] D. Florencio and L.-W. He. Enhanced adaptive playout scheduling and loss concealment techniques for voice over ip networks. In *Circuits and Systems (ISCAS), 2011 IEEE International Symposium on*, pages 129 –132, may 2011.

[7] E. Gurses and O. Akan. Multimedia communication in wireless sensor networks. In *Annales des Télécommunications*, volume 60, page 872. Presses Polytechniques Romandes, 2005.

[8] P. Levis, S. Madden, J. Polastre, R. Szewczyk, K. Whitehouse, A. Woo, D. Gay, J. Hill, M. Welsh, E. Brewer, and D. Culler. *TinyOS: An operating system for sensor networks*. Cambridge University Press, 2006.

[9] S. Lin and D. J. Costello. *Error control coding: fundamentals and applications*. Prentice-hall Englewood Cliffs, NJ, 1983.

[10] K. Lorincz, D. Malan, T. Fulford-Jones, a. Nawoj, a. Clavel, V. Shnayder, G. Mainland, M. Welsh, and S. Moulton. Sensor Networks for Emergency Response: Challenges and Opportunities. *IEEE Pervasive Computing*, 3(4):16–23, Oct. 2004.

[11] a. Mahdi and D. Picovici. Advances in voice quality measurement in modern telecommunications. *Digital Signal Processing*, 19(1):79–103, Jan. 2009.

[12] M. Maimour, C. Pham, and J. Amelot. Load repartition for congestion control in multimedia wireless sensor networks with multipath routing. In *Wireless Pervasive Computing, 2008. 3rd International Symposium on*, pages 11 –15, 2008.

[13] N. I. Park, H. K. Kim, M. A. Jung, S. R. Lee, and S. H. Choi. Burst packet loss concealment using multiple codebooks and comfort noise for celp-type speech coders in wireless sensor networks. *Sensors*, 11(5):5323–5336, 2011.

[14] M. Petracca, G. Litovsky, a. Rinotti, M. Tacca, J. De Martin, and a. Fumagalli. Perceptual based voice multi-hop transmission over wireless sensor networks. *2009 IEEE Symposium on Computers and Communications*, pages 19–24, July 2009.

[15] S. Qaisar and H. Radha. Multipath multi-stream distributed reliable video delivery in wireless sensor networks. In *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*, pages 207 –212, march 2009.

[16] P. Sarisaray, G. Gur, S. Baydere, and E. Harmanci. Performance Comparison of Error Compensation Techniques with Multipath Transmission in Wireless Multimedia Sensor Networks. *15th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems*, pages 73–86, October 2007.

[17] N. Tatlas, A. Floros, T. Zarouchas, J. Mourjopoulos, and U. Patras. An error âĂŞ concealment technique for wireless digital audio delivery. *Simulation*, pages 181–184, 2006.

[18] O. Turkes and S. Baydere. Voice Quality Analysis in Wireless Multimedia Sensor Networks: An Experimental Study. In *The International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pages 317–322. IEEE, December 2011.

[19] O. Turkes and S. Baydere. Voice Quality Improvement with Error Concealment in Audio Sensor Networks. In *Proceedings of 10th IFIP International Conference on Wired/Wireless Internet Communications (WWIC)*, pages 307–313. Springer, LNCS, June 2012.

[20] B. Wah. A survey of error-concealment schemes for real-time audio and video transmissions over the Internet. *Proceedings International Symposium on Multimedia Software Engineering*, pages 17–24, 2000.

[21] H. Wu and a. Abouzeid. Error resilient image transport in wireless sensor networks. *Computer Networks*, 50(15):2873–2887, 2006.

[22] Q. F. Zhu and Y. Wang. *Error Control and Concealment for Video Communication*, page 163. Marcel Dekker, Inc., New York, 1999.

[23] M. Zorzi. Performance of FEC and ARQ error control in bursty channels underdelay constraints. In *48th IEEE Vehicular Technology Conference, 1998. VTC 98*, volume 2, 1998.