VOICE CODING AND TRANSMISSION IN WIRELESS MULTIMEDIA SENSOR NETWORKS

by

Okan Türkeş

Submitted to the Institute of Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science
in
Computer Engineering

Yeditepe University

2012

VOICE CODING AND TRANSMISSION
IN WIRELESS MULTIMEDIA SENSOR NETWORKS

APPROVED BY:

Prof. Dr. Şebnem BAYDERE ....
(Thesis Supervisor)

Assist. Prof. Dr. Dionysis GOULARAS ....

Assoc. Prof. Dr. Sedat ÖLÇER ....
(IBM Zurich Research Laboratory)

DATE OF APPROVAL: ..../..../2012

ACKNOWLEDGEMENTS

I would like to express my utmost gratitude to my supervisor Prof. Dr. Şebnem Baydere for her precious mentorship, guidance and support which have enwidened my academic viewpoint and knowledge throughout this endeavor. Any acknowledgement could never express my appreciation for the opportunity of benefiting from her experience and wisdom.

I convey my thanks to Assist. Prof. Dr. Dionysis Goularas and Assoc. Prof. Dr. Sedat Ölçer for their respectable and valuable advices during their serve for my graduation committee.

My special acknowledgements go to my office mate Kerem Irgan, since his rigorous exertions have been a tremendous source of inspiration to me. I appreciate him for introducing me about scientific dilligence, integrity, generosity and kindness. My colleague Özgur Altun and his indomitable spirit against the deadlocks in my studies become a chief mate and encouraged me a lot throughout my research. A world of thanks is for the true friendship of Suna Dündar, as well as, for her helpful and kind mood which makes the department more a home. I kindly appreciate Selçuk Bağcı not only for his definite friendship, but also for his ray of relieving energy to help me overcome difficulties. Any valuable words fall short to express my appreciation for Mehmetcan Uzun and his never-ending support, help and cooperation. Our eternal friendship is a true asset and it is great pleasure to share entire of my Yeditepe life with him. Many thanks to Necip Arğ for his dearworth collaboration and contribution to my thesis. I would also like to acknowledge the contribution of Hakan Gök. My heartfelt gratitude goes to Gözde Karakaş for her absolute support and precious sentiment. She has been, and is, the very breath of life to me.

Last and most of all, my deepest appreciation is towards to my beloved family whose endless encouragement is the source for all of my accomplishments. There are no adequate words for their countless sacrifices for me and unconditional love through all my life.

ABSTRACT

VOICE CODING AND TRANSMISSION IN WIRELESS MULTIMEDIA SENSOR NETWORKS

Multimedia data provide versatility to any kind of application concerned with information extraction, by rendering a phenomenon down to processable data. With the utilization of ubiquitous content forms in Wireless Sensor Networks (WSNs), the era of Wireless Multimedia Sensor Networks (WMSNs) is triggered. However, under-resourced sensor nodes of these networks have to struggle with multiple data features.

Audio encoding techniques matter to sustain a certain content validity to be preserved in the lossy nature of a WSN. Besides, handling bulk data gains importance in terms of computation and traffic overhead. Basic characteristics of a voice signal must be compromised with network properties to maintain an admissible quality.

This thesis analyzes voice coding and transmission in WMSNs in terms of data and network qualifications. With a comprehensive examination of the existing schemes relating to WMSNs and consideration of mass data delivery requirements in the shade of system constraints, several contributions are made: A priority-based data coding scheme is proposed, with a set of error-resilience methods. For this scheme, a voice transmission framework (VTF) is presented which deals with the robust delivery of partitions. A wide range of experiments over a data set is conducted in a real testbed environment with several network properties. A simulation is devised to weigh an affordable trade-off between data quality and network capabilities. Experiments are assessed with a modified version of an objective transmission rating factor which can be mapped to a certain perceptibility. With these attributes, VTF offers a promising quality improvement for WMSN applications.

ÖZET

KABLOSUZ ÇOKLU ORTAM DUYARGA AĞLARDA SES KODLAMASI VE İLETİMİ

Çokluortam verileri bir görüngüyü işlenebilir veriye dönüştürerek, bilgi özütlemeyle ilgili herhangi bir uygulama için çok yönlülük sağlar. Yaygın içerik biçimlerinin Kablosuz Duyarga Ağlar'da (KDA) kullanılmasıyla birlikte Kablosuz Çokluortam Duyarga Ağlar (KÇDA) dönemi başlamıştır. Gelgelelim, bu ağların düşük kaynaklı duyarga düğümleri verinin çoklu özellikleriyle mücadele etmek zorundadır.

Ses kodlama teknikleri bir KDA'nın kayıplı doğasında belli bir oranda korunması gereken içerik geçerliliğinin sağlanması açısından önem taşımaktadır. Bunun yanısıra, yığınsal verinin kotarılması işlem ve trafik işlem yükü bakımından önemli bir meseledir. Bir ses sinyalinin temel ayırıcı özellikleri makul bir kalitenin sağlanması için ağın özellikleriyle uzlaştırılmalıdır.

Bu tez çalışması KDÇA'larda ses kodlaması ve iletimini veri ve ağ niteliklerine göre analiz etmektedir. KDÇA'lar için var olan teknikler kapsamlı incelenerek ve sistem kısıtlarının gölgesinde yoğun veri dağıtımı için gereklilikler göz önüne alınarak birçok katkıda bulunulmuştur: Önceliğe dayanan bir veri kodlama sistemi birkaç hata iyileştirme yöntemiyle birlikte sunulmuştur. Bu sistemde, parçalara ayrılmış olan verilerin dayanıklı dağıtımını ele alan bir ses iletim çerçevesi (SİÇ) sunulmuştur. Bir ses havuzu kullanılarak ve birçok ağ özelliğine bağlı olarak gerçek bir test ortamında geniş yelpazede deneyler yapılmıştır. Veri kalitesi ve ağ yetenekleri arasındaki makul ödünleşimi ölçmek için bir simülasyon tasarlanmıştır. Belli bir algılanabilirliğe eşlenebilen nesnel bir iletim derecelendirme faktörünün tadil edilmiş bir uyarlamasıyla deneyler değerlendirilmiştir. Bu özellikleri sayesinde SİÇ, KDÇA uygulamaları için umut verici bir kalite değerlendirmesi önermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

a_d	Linear predictive coding vector of the distorted speech signal
a_i	Digitized amplitude value
a_x	Linear predictive coding vector of the original speech signal
A	Overall voice data
A_H	High prioritized amplitude set
A_L	Low prioritized amplitude set
A_{s_i}	Partition amplitude set
${}_VA$	Concatenated amplitude set
bd	Bit depth
B	Receive buffer
B_C	Concealment buffer
$d(n)$	Distorted voice signal
D_i^s	Voice sample data
f_s	Sampling frequency
I_d	Delay and echo effects factor
$I_{e,eff}$	Effective equipment impairment factor
I_s	Simultaneous impairment factor
L_x	Bark spectrum
m_{m_A}	Median of median values
m_{s_i}	Median value of a segment
M_t	Mask
n	Sample index
O	Number of critical bands
pL_H	Priority level label
p_w	Packet size
$\bar{\bar{P}}$	General pattern size
$\bar{\bar{P}}_{f_s,s_w}$	Pattern size of a data
Ppl	Packet loss probability

qdu	Quantization distortion unit
r_A	Range of amplitudes
R_0	Basic signal to noise ratio factor
R_i	Relay node
R_x	Autocorrelation matrices of the original speech signal
s_i	Segment
s_w	Segment size
S_i	Snooping node
S_i^A	Acoustic sensor node
S_{ij}^R	Router node
t	Duration
T	Transpose operation
$T_{f_s,bd}$	Threshold value of a data
$x(n)$	Original voice signal
$\mu_{1/2}^{m_A}$	Median of median values
$\mu_{1/2}^{s_i}$	Median value of a segment
σ_d^2	Linear predictive gain of the distorted speech signal
σ_x^2	Linear predictive gain of the original speech signal
ADC	Analog-to-Digital Conversion
ADPCM	Adaptive Differential Pulse Code Modulation
ARQ	Automated Repeat Request
BSD	Bark Spectral Distortion
CM	Control Message
CPU	Central Processing Unit
DCT	Discrete Cosine Transform
DM	Data Message
DMA	Direct Memory Access
DSP	Digital Signal Processing
DTW	Dynamic Time Warping

EC	Error Concealment
FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
FPR	Fading Packet Repetition
FR	Full-Reference
FT	Fourier Transform
HQL	High quality link
Hz	Hertz
IS	Itakura-Saito measure
ITU-T	Standards of International Telecommunication Union
JSCC	Joint Source Channel-Coding
LERP	Linear Interpolation
LLR	Log Likelihood Ratio
LPC	Linear Predictive Coding
LQL	Low quality link
MAC	Media Access Control
MELP	Mixed-Excitation Linear Prediction
MEMS	Micro-Electro-Mechanical Systems
MFCC	Mel-frequency Cepstral Coefficients
MOS	Mean Opinion Score
MSE	Mean Squared Error
NAIVE	Naïve Voice Encoding
NC	No Concealment
NR	No-Reference
HPPR	Highly Prioritized Packet Rate
PCM	Pulse Code Modulation
PESQ	Perceptual Evaluation of Speech Quality
PER	Packet Error Rate
PRR	Packet Reception Rate
PrSERVE	Priority Based Simple Error-Resilient Voice Encoding
PSNR	Peak Signal to Noise Ratio

QoS	Quality of Service
R-factor	Rating factor
RR	Reduced-Reference
RS232	Recommended Standard 232
RSSI	Received Signal Strength Indicator
SCCC	Source-Coding and Channel-Coding
SERVE	Simple Error-Resilient Voice Encoding
SNR	Signal to Noise Ratio
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
UART	Universal Asynchronous Receiver/Transmitter
USB	Universal Serial Bus
VoWSN	Voice Over Wireless Sensor Networks
VQA	Voice Quality Assessment
VQM	Voice Quality Metric
WSN	Wireless Sensor Network
WMSN	Wireless Multimedia Sensor Network
XSM	Xen Security Module

1. INTRODUCTION

Wireless Sensor Networks (WSNs) are large-scale distributed embedded systems typically consist of intentionally under-resourced sensing, processing and communicating devices, i.e. sensor nodes. Sensor nodes acting accordingly are densely deployed over an area of interest to cooperatively monitor physical phenomena and make envisioned inferences. Researches related to WSN gradually originate new ideas, leading to applications to be emerged. Thus, various life costs in many areas are being minimized by means of providing convenience. There are a number of different application fields for sensor networks such as environmental and industrial monitoring, biomedical health monitoring, vehicle tracking, military target tracking and surveillance [1, 2].

Since WSN nodes are supposed to work autonomously and unattended for long periods of time without any maintenance, their components are designed to consume relatively low power when compared to traditional sensors. This scheme, together with the defects inherited from the nature of wireless medium, poses some challenges specific to WSNs: high transmission error-rate, limited network lifetime, node failures, mobility, network partitioning, scalability issues and end-to-end delay.

With recent advances in micro-electro-mechanical systems (MEMS), multimedia utilization became feasible in sensor network technology and introduced the era of Wireless Multimedia Sensor Network (WMSNs) [3]. In principle, main target of WMSN applications is simple rendering of a particular content format information into processable data with resource-constrained nodes having multimedia sensing capabilities, during an event monitoring. Flexibility, accuracy and high quality services of the applications in these networks are widened with the integration of inexpensive hardware such as acoustics sensors, cameras, and other sensors producing multimedia content [4, 5].

Audio utilization in WMSNs is an inspiringly progressive subject in the literature. Audio, as a content form, is a multi-dimensional function that no measure itself can accurately evaluate all of its aspects. Voice signal, which is targeted in this dissertation as a specific audio data type also has multi-dimensionality. In a sensor network, a great deal of voice properties need to be analyzed promptly in order to identify and deduce from an event more accurately and descriptively. Since automated applications need an evaluation during recording and processing, requirements for these properties must be identified clearly.

Key challenges in WMSNs can be addressed by extensive researches on the coding and transmission of voices. In this thesis, a set of coding schemes based on light-handed transmission, error-resilience and prioritization of voice segments is proposed and examined in detail. Hybrid utilization of the proposed encoding schemes is coupled with a voice transmission framework which addresses the requirements of multimedia handling in WMSNs. Additionally, through a real homogeneously constructed multi-hop network testbed, the effect of channel conditions on the voice quality is investigated using an on-the-fly transmission rating factor. In the experiments, recorded signal segments sent from the source node are touched with the counterparts expected to be collected in each intermediate hop, thus quality derogations are evaluated comprehensively.

1.1. PROBLEM DEFINITION

The nature of multimedia delivery introduces more demand in processing and energy resources together with stringent Quality of Service (QoS) requirements. These requirements necessitate to revise the knowledge gained from WSNs to manage resources much more efficiently, since it is required to get the data within accepted quality and time range. To achieve expected lifetime for the application, energy efficiency is also an inevitable issue when satisfying these requirements. There is always a trade-off between delay, energy consumption, bandwidth usage and voice quality. As the overall network point of view, equilibrium between the sampling rate and comprehensible data needed for a specific deduction from the information gathered is tried to be analyzed. At this point, processing time and transmission capacitance of the nodes must be analyzed.

For any system or device in the field of telecommunications, need for the implementation of a quality-potent voice transmission scheme is inevitable to embrace a certain QoS satisfaction [6]. For any context in the domain, an evaluation of voice quality by means of quantifying the various data corruptions must be presented. In simple terms, quality of a voice refers to the intelligibility of the perceived voice by a listener. There is a considerable number of voice characteristics that affect voice quality; such as clearness, noise, timbre, pressure, loudness, etc. Also, factors related to the equipments used in a context can unconditionally distort the voice perception. Moreover, digital processing of the voice signals creates several analysis to be taken into consideration: quantization errors, data resolution, sampling frequency, etc.

Traditionally, voice quality measurements are carried out with a variety of time-consuming and expensive tests in a subjective manner. Over the recent decades, the results obtained with several evaluation measurements are used for determining standards, compiling specifications and comparing pieces of equipments. With an objective evaluation, several specific sequence testing methodologies are devised to automatically weigh factors affecting voice quality. Relating to the application areas, networks in any system architecture can facilitate these quality evaluation metrics in accordance to their capabilities. However, the challenging nature of sensor nodes circumscribes any implementation criteria in a strict manner. All of these factors make most of proposed well-known recipes for miscellaneous implementations related with multimedia in computer science hard to be applicable for WMSNs.

1.1.1. Voice Coding in Wireless Sensor Networks

Even in resourceful networks, setting a counterbalance between QoS requirements and multimedia data handling is a nontrivial process. Static resource reservation based methods are mostly used in those networks. Aside from techniques used in traditional networks, simplicity for a lightweight data encoding is a must in WMSNs. Because, WMSNs are vulnerable in terms of assuring the quality of links and node aliveness throughout data transmission. Therefore, static resource allocation techniques are not convenient for sensor networks.

Under separate cover, assessment of voice quality is another issue when devising transmission schemes for any level of communication stack. With an objective quantification of encoded and/or transmitted voices, determining a certain range in error tolerance for any scheme and methods proposed is the key purpose of this assessment. Subsequently, finding an optimal trade-off among resource utilization and application specific QoS requirements with an on-the-fly evaluation approach is a prominent difficulty. Performance criteria for audio signals are analyzed from a different angle of evaluation in several studies like [7–10]. However, there exists no standardization for voice quality evaluation in the literature. Several approaches which physiologically, acoustically, psychologically and perceptually take voice functions into consideration are mostly carried out by self-assessment surveys [11–13]. These kind of evaluation techniques are mostly applied offline regardless to the application implemented [14, 15] or offer a quality measure in a subjective manner [16, 17]. Nonetheless, automated applications need an evaluation during recording and processing in which a great deal of voice properties must be analyzed promptly. There are also several objective quality metering techniques that can evaluate several voice properties in company [18–20].

In one sense, smooth coding of the voice being recorded in the source node of WSN has to be taken into account to make prominent evaluation approaches available. Quality of a voice data basically depends on sampling rate and bit depth properties. However, coding simplicity and 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.

1.1.2. Voice Transmission in Wireless Sensor Networks

In WMSNs, applications like emergency response or activity monitoring require data transmission within a specific time bound. Therefore, one of the main objectives that has to be considered in the protocol design is to control or minimize the network delay. On the other hand, a WSN has to be capable against faults in order to collect the data at a certain quality.

1.1.2.1. Real-time Delivery

Sensor nodes need to be accommodated with plain services to manage the transmission of vast amount of voice data at a high pace. Considering tailored source and channel coding techniques, which are efficient in terms of processing power and required memory, can offer a solution for a light data traffic. At first sight, compression at the source may seem feasible to eliminate delay by reducing the data size radically. But, it is not necessarily effective in WMSNs because of the delay introduced by the computation time [21]. For that matter, implementation of a serviceable encoding scheme must be considered in order to handle mass data transfer.

1.1.2.2. Fault Tolerance

In WSN, transmission is vulnerable to the failures occurring due to harsh environmental factors. It is necessary to provide a robust scheme to discourage data losses caused by node malfunctions or link failures. For that matter, the protocol design for such networks should be fault-tolerant if it can deliver high portion of the data to the sink.

On the other hand, redundant and correlated parts inside a voice data can pave the way for the implementation of an error-resilient scheme. In this respect, error concealment (EC) approach has enabled effective mechanisms that can recondition the distorted voice data as closely as the original one without increasing the bandwidth demand [22]. Hence, the burden of retransmissions and consequent delays are avoided.

1.2. MOTIVATION & AIM

Resource-poor sensor nodes generally cannot meet the requirements of even the algorithms known as lightweight in traditional systems. Many of the researches on WMSNs proposing solutions for voice coding and transmission do not properly cover the facts of this field. Commonly, presented methods include several assumptions that cannot yield beyond a theoretical basis. For validation of theoretical solutions, a wide range of numerous experiments must be conducted on real testbed implementations [23].

Multi-dimensional properties of voice signals and under-resourced sensor nodes make voice processing and transmission a challenging task for WSNs. In this regard, voice coding techniques are at the heart of any solution in WMSNs. The validity and feasibility of these techniques matter for an efficient performance. On the other hand, voice quality evaluation of the signals gathered strictly related with network transmission quality. Link and transfer quality among each consecutive hop determines the success rating of the received samples. With interdisciplinary approaches, the challenges related with quality considerations in WMSNs can be overcome.

One of the intrinsic characteristic of voice data is correlation of segments to reconstruct lost parts [3]. Preferential parts of a voice data can be utilized to present reliable coding and transmission strategies and regulate network requirements. Hence, prioritization is one of the fundamentals that exploit relatively important part of voice can be used to adjust voice coding and transmission strategies to application and network requirements. Prioritization can address a fundamental solution to support QoS requirements.

As a different solution, multi-path transmission is another inspiring method for reducing the required bandwidth and minimizing delay during voice data transmission in WMSNs.

With these considerations, these techniques have to be analyzed comprehensively under different situations and environments. Moreover, findings acquired from the evaluation of existing methods can be inspiring when devising new ones. This thesis aims to propound a voice coding and transmission environment for WMSNs which can serve a lightweight transmission and preserve quality requirements of the applications. On the other hand, we aim to couple the experiences on still-image coding and transmission researches [24–26] with the ones on voice data, in order to put forward a general WMSN platform.

1.3. CONTRIBUTIONS

With regard to the previously outlined challenges in the research area, a number of contributions are made. A lightweight priority-based and error-resilient encoding scheme is presented with a voice transmission framework. According to this scheme, voice data segments are prioritized at the source based on the information they contain and labeled accordingly. Preferential data are preserved by considering several priority threshold values depending on common voice characteristics. The threshold levels are also taken into consideration for different segment sizes. In the course of transmission, a set of well-known error correction methods are applied to the lost partitions of voice data.

Implementation costs of the prioritization approach and error correction methods are evaluated on a real implementation on Telos [27] equivalent TMote Sky nodes. A real multi-hop testbed is established to analyze the effect of channel conditions on a voice data set encoded when well-accepted error concealment (EC) methods are applied. A wide variety of tests including 30,000 real transmissions are conducted for single-path transmission scenario. A preliminary study of this work is accepted as a publication [28]. Experiments are assessed with a modified version of an objective transmission rating factor that can be mapped to a certain perceptibility.

Moreover, the results obtained in a real testbed shown in Figure 5.7 are expanded and validated with a conducive simulation which depends on random sampling of transmission patterns. A set of sensitivity analysis are designated on several voice data properties and network schemes. According to this, nearly 1,6 million simulated transmissions offered an extensive analysis of voice coding and transmission in WMSNs.

1.4. ORGANIZATION OF THE THESIS

The rest of the thesis is organized as follows. Chapter 2 renders the researches in the literature related with the subject. The voice coding and transmission model is given in Chapter 3, embracing the voice data encoding techniques for transmission, error handling methods and priority-based delivery. Quality evaluations over the voice data are issued in Chapter 4. In Chapter 5 real testbed environment and the conducive simulation details are explained. Chapter 6 extensively discusses on the presented techniques, methods and schemes and presents the system performance and the testbed environment. The conclusion is outlined in Chapter 7 with comments on forward plans.

2. RELATED WORK

Sensor technology has been playing a vital role in our daily lives for the recent decade, and will inevitably become an irrepressible domain in computer science [1]. Comprehensive surveys on WSNs can be found in [1, 2]. Utilization of multimedia data has increased the functionality of the WSN applications. Since the arrival of WMSN in the literature, sensor network is being a powerful and affordable technology with each passing day. In the meantime, commercialization of this technology originates new ideas, leading to new scenarios to be emerged; thus various problems are waiting to be resolved in this area. There are several surveys about WMSNs which cover applications, challenges, proposed solutions and research issues. An intensive literature focus is given in [3]. In [4], several detailed comparisons among the applications relating to multimedia communication in this field are given. The existing research studies are thoroughly examined in terms of both mobile multimedia and WMSN domain. There are also some other surveys concentrate on different aspects of WMSNs. In [29], the authors examine multimedia streaming in WMSNs.

According to the mentioned surveys above, streaming a content format by taking the natural constraints of WSNs into consideration has to be characterized neatly. For that matter, multimedia coding and transmission in WMSN, as the fundamental focus of this dissertation in terms of voice, is an indivisible whole which requires concurrency and data integrity together. Another important implication from these surveys is that existing protocols from the mobile multimedia streaming do not consider the challenges of sensor nodes. By this means, in [30], authors state that there are a few number of studies relating to multimedia coding and transmission protocols in WMSNs.

In this chapter, the background and the literature related with our study are introduced. The first section provides an introduction to the audio/voice coding strategies proposed for WMSN. Right after, a detailed literature review on audio/voice data transmission over multimedia streaming sensor nodes is provided. The third section discusses the error recovery techniques whereas the fourth section focuses on audio/voice quality assessment methods. Also, real testbed setups with emphasis on audio handling in WMSN are explained.

2.1. VOICE CODING IN WMSN

Voice coding, as the lead-off phase of voice data handling in WMSNs, explicitly propounds a serviceable environment for the rest of the network. There exists a vast literature focusing on different aspects of audio/voice coding in the shade of resource constraints of any sensor node. The captured voice data have to be represented in an optimal manner to make reliable data delivery through the lossy medium possible. In order to decrease processing power, light-handed coding algorithms that can be accommodated on the sensor nodes have to be selected. Besides, these algorithms should provide a layout for robust data delivery which can preserve information integrity. For an efficient voice coding designated for sensor nodes, prominent targets in the literature can be summarized as follows:

i. Low complexity: Encoding algorithms going to be accommodated on sensor nodes have to be lightweight and form low-power in order to increase survivability. Besides, costs of a vocoder has to reserve an adequate space for other procedures of the application need further processing [31]. By providing a collaborative environment, simple algorithms at the mote level can make room for the network elements [9].
ii. Error resiliency: A robust and error-resilient coding of the source data need to be provided for further transmission. For a transmission which does not represent large communication or processing overhead, tunable voice coding techniques should be provided according to the application selected [32].
iii. Compression: Uncompressed voice signals may require high data rates and therefore consume excessive energy and bandwidth. A high ratio of compression has to be satisfied for sensor network in order to satisfy an enhancement in bandwidth and energy consumption [10].

In order to implement a simple voice coding algorithm at the mote level, data characteristics have to be analyzed intensively. There are several voice characteristics that can differentiate the voice coding technique selected for any application. For instance, Berisha et.al have focused on the signal energy and pitch features of the voice data to distinguish periods of any voice activity on the sensor nodes [33]. Pitch estimation of data segments tends for detection of speaker activities. However, this estimation can vary greatly due to lack of waveform periodicities. In [34], several characteristics of voice data, such as variation and irregularity of information, degree of noise, etc. are investigated in order to rater reliability in communication. In a separate study, several dimensions in voice data are explained [13].

In [18], very similar approach as this thesis presents is given for voice coding and transmission. Carvalho et.al state that packet loss concealment techniques are compatible with low packet error rates, but their performances decrease when consecutive losses occur.

Compression techniques as a target for voice coding in WMSNs can be grueling for the sensor nodes, if no extra Digital Signal Processing (DSP) chips are affiliated. Instead of using complex compression schemes like Mixed-Excitation Linear Prediction (MELP) and Linear Predictive Coding (LPC), more simpler algorithms can be preferred [35]. In most of the studies related to WSN, a well-known uncompressing method for voice coding is Pulse Code Modulation (PCM) in which the analog signals are mapped to a set of 8-bit digital representations according to the sample rate defined. There are several studies in WMSN that utilize PCM in the voice coding process [36]. A variant of PCM, Adaptive Differential Pulse Code Modulation (ADPCM) is utilized in several studies to compress the sampled data in order to allow further reduction of the required bandwidth [19, 37, 38]. The target of ADPCM is to reduce the size of quantization step to 4-bit [35]. In [39], authors study on the advantage of utilizing a suitable compression algorithm for voice encoding with particular attention on the trade-off between processing power consumption and quality of the voice.

2.1.1. Time Domain Approaches

A large number of WMSN studies concentrate on time-scale approaches for voice coding [40]. On the other hand, there are very few number of frequency domain based approaches in WMSNs, which basically concentrate on theoretical examinations.

Some of the WMSN studies rely on voice coding techniques are protocol-based investigations at the Media Access Control (MAC) layer most of the time. Because, simple methods on voice capturing and data quantization are necessary for a relieved data streaming and therefore voice coding and transmission in a WMSN should be considered in one piece. The features of the digitized signals should not tire the poor resources of a sensor node which has to race with synchronization issues. Besides, these signals have to be delivered with less errors throughout the network. In telecommunications and computer networks, as well as in WSNs, a channel access method provides several end-points connected to the same medium in order to make delivery possible over multi-points. By this regard, the capacity of the transmission medium is shared. In terms of channel access method for shared medium networks, the most-known strategies are Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA). However, a vast literature for TDMA based approaches in WSNs can be found whereas they do not emphasize FDMA strategies too much.

Most of the studies related with WMSNs deal with voice coding in the time domain. Time synchronization is significant in terms of increasing the sleep time interval for activities held in the sensor nodes. Furthermore, synchronization in time assures timeliness, throughput and network lifetime for end-to-end communication [37]. Coordination of delivery in a delay-bounded service is a must for any application in WMSN. There exist several studies dealing with TDMA protocols which try to ensure collision-free data transmissions. In [41], a traffic-adaptive medium access protocol is presented in terms of TDMA utilization. A similar study in [42] state that TDMA based protocols assist the nodes in a WSN to communicate without a collision and present a lightweight MAC protocol for WSNs which can provide a communication in dedicated time slots.

2.1.2. Frequency Domain Approaches

Sinusoidal models have been utilized in audio/voice processing for long ages. Many of these models are utilized to determine frequencies in the amplitude spectrum for modulation and demodulation techniques [43]. Identifying an exact frequency interval is an important issue for a specific application which stand to benefit from multiple waves in a signal. However, these methods are costly for sensor networks most of the time [44]. Many works dealing with frequency based approaches can only give theoretical solutions for WSNs.

A preliminary study in [45] proposes a simple sound recognition algorithm on sensor nodes by utilizing Mel-frequency Cepstral Coefficients (MFCC) and classification Dynamic Time Warping (DTW) algorithms. Recognition results have been achieved by using Mel-cepstrum formula implemented in Matlab. The filters created are transported to the sensor network. Introduced methods are promising for the future hardware implementations in sensor network technology.

Sheybani et.al try to decrease multidimensional properties of voice signals by directly embedding the filtered signals into wavelet forms [46]. By applying Fourier Transform (FT) on a time-domain signal set, they state that most of the voice characteristics are compressed into frequency representations. In their implementation, they claim distortions and noises due to transmission errors are eliminated. However, one basic disadvantage of FT is its dependence on time averaging over entire duration of the signal.

For the low levels of protocol stack in a WSN, approaches dealing with FDMA based transmissions are not considered thoroughly, since most of the sensors in the context are designed in accordance with TDMA based protocols.

2.1.3. Discussion

In many WMSN application scenarios in the literature, temporality of voice signals provide a simple voice coding environment. Hence, the data handling with time-scale approaches becomes easier. We see that the choice between time domain and frequency domain based approaches depend on the specific application. Research areas such as feature extraction, vowel production which adopt the sinusoidal properties of a voice data most of the time require high computation and memory usage. Simple vocoders dealing with temporal properties of voice signals can relieve transmission in a WMSN at a high extent. Nevertheless, a voice coding analysis in frequency domain needs time-scale strategies as a base and can be serviceable with a hybrid utilization of multiple access techniques.

2.2. VOICE TRANSMISSION IN WMSN

For the bursty nature of voice data, streaming disciplines are needed to be extensively investigated in WMSNs. Transmission robustness must be satisfied at a certain extent in order to operate sensor nodes under channel errors. Multimedia data delivery has to be based on QoS requirements of the overlaying transmission in a WMSN. Given the application necessities, a sensor network can provide support for different services required in multimedia transmission [3]. Some applications may only deal with event-triggered observations which require less traffic overhead. On the other hand, an affordable determination must be established between real-time or delay tolerant applications and voice data validity during transmission.

Incorporating the under-resourced properties of sensor networks into a transmission protocol design gains significance in order to efficiently utilize the limited resources of the network. Moreover, protocols should be as light-weight as possible. Many designs in the literature focus on specific protocols from the data-link layer up to transport layer [47].

Since WMSNs possess these unique properties, most of the existing solutions for the assessment of network utilization are not suitable for evaluating sensor network protocols [30]. When designing a data transmission protocol for WMSNs, several factors should be considered. One of the most significant issue for sensor network operations is energy-efficiency which imposes network lifetime, QoS requirements and data fidelity. WMSNs are different from traditional ad-hoc networks in terms of communication types. Communication channels often deal with mass data delivery between events and end-points of the network. The nodes of WMSNs are typically interested in an overall deduction of the environment with the utilization of multimedia data, rather than simple explicit readings sent from other nodes. Therefore, data transmission protocols at the application layer are typically data-centric. For this reason, data handling is a hard problem in WMSNs to be solved with several transmission approaches.

In terms of traffic utilization, [3] categorizes multimedia transmission protocols with their tolerances in terms of delay and/or losses. According to the presented protocols for

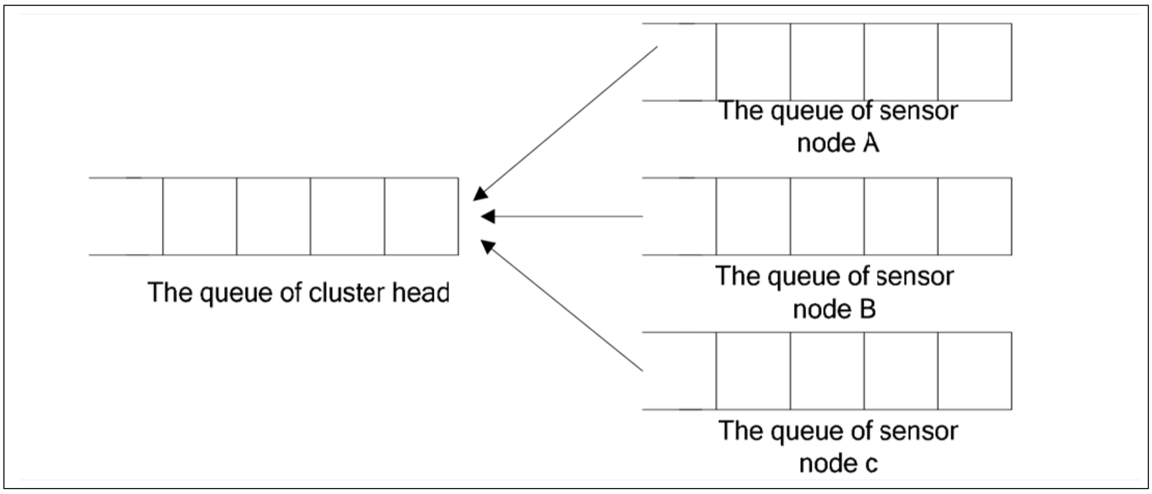


Figure 2.1. A cluster head implementation for delay-tolerance

WMSNs, processing overhead or bandwidth requirements of a snapshot or a data stream transmission can be compensated with either a delay-tolerant approach or a loss-tolerant approach. Solutions to construct a transmission model fulfilling real-time constraints and error-resilience at the same time is a very hard issue in WMSNs.

In [15], delays caused by mass data delivery in acoustic sensor networks are investigated with a Markov chain model. According to this study, mean waiting time of data units in a node mainly depends on the arrival rates from other ones. With the consideration of an affordable sensor node as a cluster head depicted in Figure 2.1, the delays can be minimized.

In one another study [37], authors try to eliminate the delays in order to provide a real-time transmission for voice over sensor networks. A time-synchronized sensor network platform for real-time data streaming across multiple hops is proposed. A high end-to-end throughput is wanted to be achieved by bounding the latencies and predictable lifetime of the network by means of decreasing the audio sampling quality at a certain level. In [48], a fair packet queueing method which addresses the concerns of delays related to multimedia data traffic is presented. In order to guarantee a serviceable traffic control, Biswas et.al proposes a data-centric attribute allocation mechanism [49]. In [47], authors conduct a study on particular transport protocols peculiar to WSNs. According to this study, existing transmission protocols merely attain the necessities of WMSNs in order to provide an efficient communication. In this regard, there is a need for new effective multimedia transmission techniques for WMSNs and a cross-layer design can pave the way for addressing challenges in WMSNs. Data handling issues are investigated by many researchers in three basic categories in WMSNs, as outlined below:

i. Maximizing the lifetime: QoS requirements have recently been considered with regard to multimedia delivery for WMSNs. In [50], an application admission control algorithm is proposed in which network lifetime subject to bandwidth and reliability constraints of the application is tried to be maximized. Another application admission control method is proposed in [51], which determines admissions based on the added energy load and application rewards. These approaches address application level QoS requirements by means of delay, reliability and energy consumption, as required in WMSNs. However, these strategies do not take tight balancing between multimedia data optimizations and in-network computations into consideration. In [17], a set of tests is conducted to observe the differences in energy levels during several transport approaches utilized for captured voice signals. Additionally, the technique provides basic security services such as authentication and confidentiality for control messages without incurring in additional energy consumption. Dasgupta et.al stress on maximizing the system lifetime by assigning several roles for a group of sensor nodes deployed in the monitoring area [52]. The authors explain that energy consumption is open to differentiation because of different deployment strategies.

ii. Balancing data handling: In [53], a linear programming approach is studied to balance data handling in WSNs. Study proposes a formulation to achieve a processing rate at the minimum level for every individual node in order to handle data during transmission. The authors also presents an optimal deployment for the relay nodes in a given transmission scheme to easily make the nodes available for data handling. [14] is also a promising study for adaptive voice data transmission in WMSNs.

iii. Maximizing data handling: There are several studies dealing with distributed computing approaches for sensor networks. The heuristic transmission model presented in [54] is an example for increasing the amount of the data transferred by means of distributed balancing. In a similar study [55], the authors possess on exploring an optimal trade-off between energy consumption and data transmission rate while trying to increase information level inside the data.

Except the protocols and methodologies designated at the application layer of sensor nodes, a convenient transmission scheme selected for any application can also address voice data transmission problems in WMSNs. The bandwidth requirements in voice transmission can be satisfied by using multiple paths [56]. Multi-path transmission also helps to overcome congestion problem in WMSNs [57]. By using multiple paths, that naturally occur in WMSNs, error-resilience is also achieved [58]. Another benefit of multi-path transmission is in timeliness domain, thus it helps to minimize delay in transmission. In [30], a node-disjoint multi-path transmission scheme is presented in order to solve the capacitance problem occurred in sensor nodes. In [53], a similar node-disjoint set is presented for achieving bandwidth problems and maximizing data flow rate. MMSPEED [59] is the one of the representative of using multiple paths to achieve QoS in both timeliness and reliability domain.

2.2.1. Prioritized Transmission

Prioritized transmission of different services is a commonly used technique in any kind of networks. The theoretical study [60] of Albanese et.al. shows the basis of priority encoding transmission. This paper is probably the most related one to our study. The studies about prioritization are mostly related to inter-prioritization of different data types, i.e. scalar data, audio, video etc. [61].

2.2.2. Discussion

There is an absolute need for new criteria and mechanisms to manage the multimedia stream handling according to the desired application-layer QoS. On the other hand, the number of research studies stressing on expected lifetime is very few. The studies in the literature which draw conclusion for multimedia data transmission generally focus on reducing bandwidth and provide an efficient data delivery in every part of a WMSN.

2.3. ERROR RESILIENCE

Sustaining a certain level in voice data integrity and validity is important in the shade of packet losses occurring in a sensor network. Researchers found several techniques to deal with missing parts of a multimedia data due to lossy nature of WMSNs. In very well-known surveys for error concealment (EC) [62, 63], authors presented the existing EC schemes for real-time audio and video transmission over traditional networks. Figure 2.2 summarizes the EC strategies in the literature.

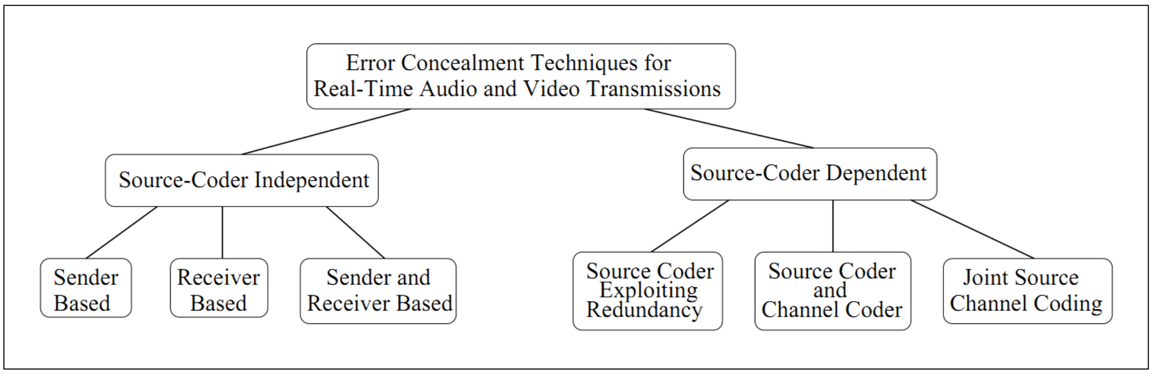


Figure 2.2. EC strategies for real-time audio and video transmissions

Source-coder independent EC methods exploiting information only at the receiver side provide decent performance when the packet losses are not under a certain level. The receiver-based methods are straightforward and therefore renowned reconstruction strategies such as white or zero substitution, previous packet repetition, interpolation. White or zero substitution methods are also considered as no concealment (NC) methods, which only suppresses a silence into lost partitions of the voice data. On the other hand, packet repetition and approaches similar to it present a light-handed EC environment because of their extremely low complexities and computational costs. However, there is a possibility for amplitude and phase mismatches during reconstruction process. On the other hand, when frequent and consecutive losses, the effect of reconstruction for source independent receiver-based strategies dramatically decreases. There are several modified versions of the receiver-based EC algorithms in order to be exploited according to the purpose of any application. For instance, in [64], authors propose a pattern repetition algorithm which achieves a perceptually audibility with a significant reduction for the silence gaps. In Figure 2.3, the original and the lost partitions are shown in a and b, respectively. In c, traditional packet repetition method is depicted whereas fading packet repetition (FPR) strategy is shown in d. FPR strategy employs time-scaled functions in which two buffers are utilized, one is to keep successfully received packets and the other one is to monitor forthcoming gaps in reproduction process.

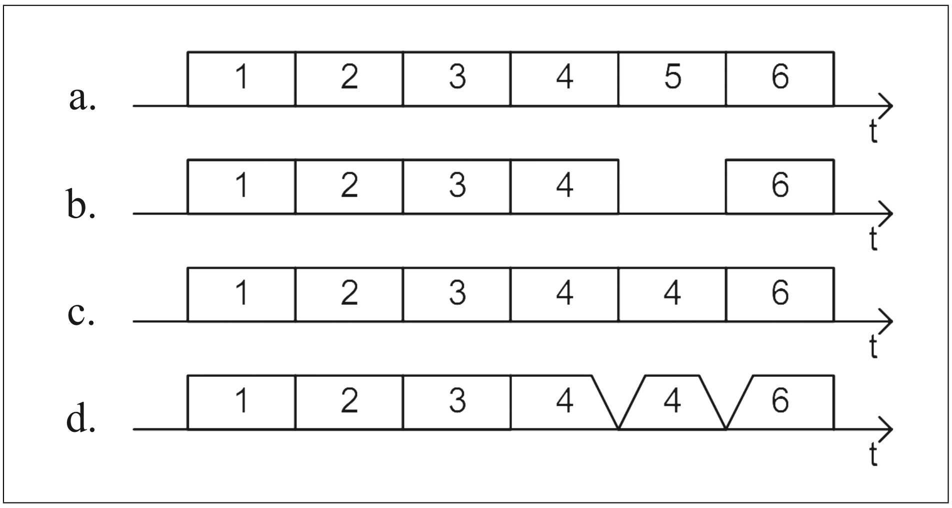


Figure 2.3. Error Concealment using the FPR approach

For the sender and receiver based methods, prioritization of the network packets plays a vital role since focus on reconstructing preferential data. The importance levels of the packets in priority-based strategies are determined by setting priorities which are based on several approaches like signal energy, difference to previous packets and voice onset or transition indicators. Transition indicators carry an evaluation information about the upcoming packets, so that they are prioritized accordingly.

The sender-based approaches usually try to retransmit the lost data in the absence of received information. Popular methods based on retransmissions are Automated Repeat Request (ARQ) [65] and Forward Error Correction (FEC) [66]. However, these concealment schemes are very costly and may not be suitable for WMSNs. ARQ scheme achieves efficient bandwidth usage but due to packet retransmissions, it cannot satisfy strict delay constraints. FEC is based on imposing redundant packets or appending erasure codes to the packets and so comes up with increased bandwidth which is already limited in WMSNs [4, 67]. Nevertheless, several studies dealing with the transmission schemes which have to avoid any loss in a compressed multimedia data focus on these methods. For instance, a packet redundancy protection algorithm is accommodated on preselected relay nodes responsible for retransmitting perceptually important lost packets [12].

The fundamental purpose of source-coder dependent EC methods is exploiting redundancy. The redundancy intrinsically contained in the nature of a voice data can be utilized to boost the robustness of the information transported. In the world-renowned pioneer work of Shannon [68], he states that:

> The redundancy must be introduced in the proper way to combat the particular noise structure involved. However, any redundancy in the source will usually help if it is utilized at the receiving point. In particular, if the source already has a certain redundancy and no attempt is made to eliminate it in matching to the channel, this redundancy will help combat noise.

According to the redundancy exploitation in the voice streams, the coder-specific schemes enable a high performance in EC. Most of the studies rely on utilizing redundant parts of multimedia data prefer prioritized transmission. A joint source channel-coding (JSCC) scheme minimizes transmission errors with a design of jointly integrated channel coders. To cope with noisy channels, they try to optimally partition bandwidth between the source and channel coders, depending on channel-loss status, normally characterized by some parameters [69]. However, this scheme is not affordable by sensor nodes.

Source-coding and channel-coding schemes (SCCC) extend traditional channel-coding methods like FEC for concealing bit errors in order to compensate packet losses and modify channel coding to suit different source coders. However, methods related with these kind of schemes are not applicable for WMSNs most of the time. The utilization of such schemes in voice transmission must be based of prudent decisions because sensor nodes already deal with high bandwidths.

2.3.1. Discussion

In regard to the information type exploited, some EC strategies are more promising than others in terms of reconstruction. However, any method for reconstruction of lost partitions in a WMSN has to be applicable in the resource-poor nature of the sensor nodes. It is seen that source independent receiver-based EC techniques are feasible for sensor networks. Besides, using the redundant data in a voice stream can foster the effect of reconstruction scheme selected.

2.4. VOICE QUALITY ASSESSMENT

WSN, by its very nature, is an unreliable networking paradigm. The most basic form of sensor networks does not assure delivery, reliability, flow control or error recovery. As we mentioned in previous sections, higher level protocols need to stress on QoS requirements of the overall system in a WMSN. On the other hand, voice data integrity and validity must be sustained during transmission.

Quality can be expressed as the result of the judgement of a perceived constitution of an entity with regard to its desired constitution [6]. Analogically, voice quality is a multi-dimensional term which can be considered as the consequence of a perception in which an evaluation is performed with regard to the relationship between the comprehended and the original, desired or expected voice signal.

There are several reasons to assess the voice quality. The most important one is to weigh the necessities of an application dealing with voice transmission in terms of its users' perceptional decisions. Another reason is end-to-end measurement of any impairment, caused by a network properties or its equipment features. It is very important that an application service has to offer VQM algorithms inside which provide voice quality evaluation in an autonomous manner and thereby reduce costs and enable a faster response for the application requirements. When evaluating the performance of an voice coding and transmission scheme, the quality of the degraded voice data at the receiving end has significant importance. Moreover, voice quality assessment (VQA) is utilized in quality

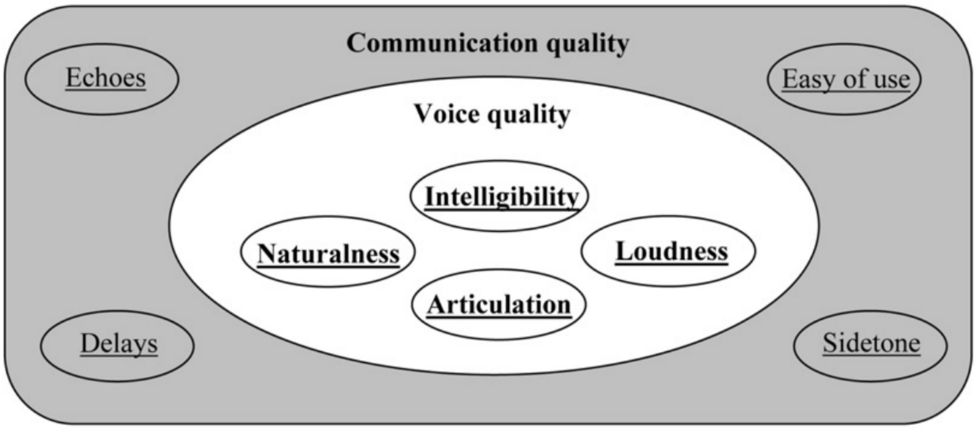


Figure 2.4. Psychological factors and aspects of voice quality

aware transmission schemes, when making rooting decisions. VQA is also beneficial for the adjustment of coding and transmission parameters. In many applications, transmission of the voices in full quality is not necessary. In those cases, intentionally reducing the quality of the voices extremely improves the network performance and the lifetime of the system. VQA is employed in several WMSN applications before the information extraction stage to figure out whether the received image is sufficient to extract the required information or not.

Voice quality is subject to multiple factors, as shown in Figure 2.4. These factors are unlikely to be independent. There is a vast number of voice quality metrics (VQMs) in the literature which focus on multi-dimensional properties of voice data. In this section, VQMs are categorized according to several assessment techniques.

2.4.1. Subjective Voice Quality Measures

Human auditory system has a great ability to evaluate quality of any voice [70]. For voice quality assessment, there are numerous perceptual voice quality test approaches which subjectively judge the received voice by integrating the evaluation process with the humans' auditory perception [34]. Subjective tests are conducted to make suggestions on perceived voice quality. However, the approaches are often very expensive, time-consuming, and labor-intensive [6]. According to these tests, judgments of the listeners are tried to be categorized or quantified. In these tests, there exist several methods to assess the subjective quality of voice signals. Most of the time, a set of predetermined voices are played to a group of audience and they are asked to rate the voice heard, normally using a 5-point quality scale. By categorizing and averaging the rates gathered, an average score is determined for each voice played in the tests. This evaluation technique is called mean opinion score (MOS). Telecommunication Standardization Sector of International Telecommunication Union (ITU-T) represented a set of recommendations [71] which provides an objective terminology to be used in conjunction with voice quality expressions in terms of MOS. There are several studies rely on MOS evaluation. In [12], authors define a perceptual based voice quality over a multi-hop WSN to evaluate packet losses. Another study focuses on generating a MOS-based formulation in order perceptually detect received voice quality [72].

2.4.2. Objective Voice Quality Measures

The purpose of objective measures is to predict the voice quality autonomously. These measures ascertain quantitative ratings and some of them may be mapped to the subjective ratings obtained by MOS results. Objective VQMs are examined in three different categories according to availability of the original voice data.

i. Full-Reference (FR) Metric; Original voice is available.
ii. No-Reference (NR) Metric; Original voice is not available. It is also called as "single-ended" and "blind".
iii. Reduced-Reference (RR) Metric; Original voice is partially available or information about the features of the voice is available for VQA purposes.

FR metrics, namely intrusive metrics or input-to-output metrics, are the most commonly used metrics. They may be used for evaluating the performance of voice related methods but they cannot be used dynamically in a communication system because the reference image is not available at both intermediate nodes and the sink. The basic structure of a FR objective VQM is demonstrated in Figure 2.5. On the other hand, NR metrics, namely non-intrusive

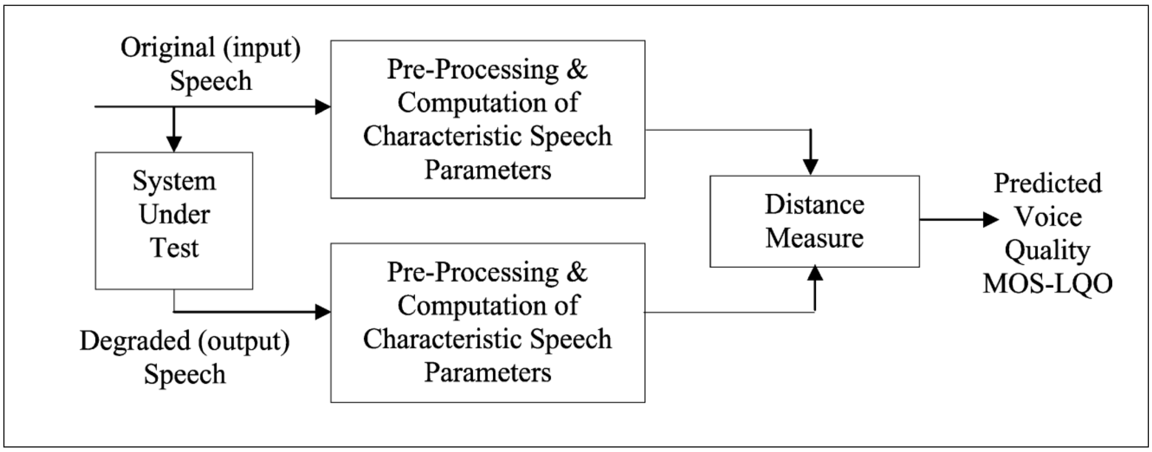


Figure 2.5. Basic structure of a FR objective voice quality measure

metrics, may be used for this purpose but NR VQA is extremely difficult for an automated algorithm. Besides, RR metrics provide a good compromise between FR and NR metrics for image communication.

Objective VQMs can be classified according to the domain they are investigated. In Figure 2.6, these domains are shown.

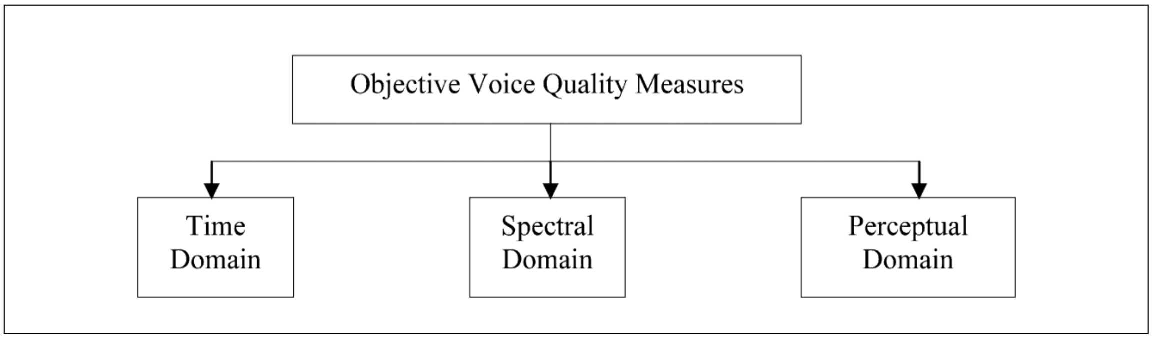


Figure 2.6. Classification of objective voice quality measures

Although there are lots of proposed VQMs, there is no universal VQM which can be used in all situations and all types of voice data. Because of this, VQMs should be chosen in accordance with the aim of the application by conducting experiments over the possible voice data which would be faced throughout the application. Since, we use VQMs for only the evaluation of our methods, we further focus on objective FR VQMs.

2.4.2.1. Signal-to-Noise Ratio

Signal-to-Noise Ratio (SNR) is a general time-domain VQM which weighs the differences in signals. It is used mainly to assess distortions of the signals caused by external sources. It is commonly used in audio processing to compare the performance of EC algorithms.

SNR is based on the ratio between the original signal and its noise formed after a transmission, which is defined as

$$SNR = 10\log_{10}\frac{\sum_n x^2(n)}{\sum_n (x(n)-d(n))^2} \tag{2.1}$$

where $x(n)$ represents the original voice signal, $d(n)$ represents the distorted voice reproduced by a voice processing system and n is the sample index. There are modified versions of SNR, like SNR_{seg} and PSNR in the literature.

2.4.2.2. The Log Likelihood Ratio

The Log Likelihood Ratio (LLR) is a frequency-domain metric which is frequently presented in terms of the autocorrelation method of linear prediction analysis, in which the voice signal frames is windowed with the length of 15 to 30 ms. The LLR measure can be written as

$$d_{LLR}(a_x, a_d) = \log\left(\frac{a_d R_x a_d^T}{a_x R_x a_c^T}\right) \tag{2.2}$$

where a_x is the linear predictive coding (LPC) vector of the original speech signal, a_d is the LPC vector of the distorted speech signal, R_x is the autocorrelation matrices of the original speech signal, and T denotes a transpose operation.

2.4.2.3. Itakura-Saito Distortion Measure

The Itakura-Saito measure (IS) is a variation of the LLR that includes in its computation the gain of the all-pole LPC model, and is defined as

$$d_{IS}(a_x, a_d) = \frac{\sigma_x^2}{\sigma_d^2}\left(\frac{a_d R_x a_d^T}{a_x R_x a_c^T}\right) + \log\left(\frac{\sigma_d^2}{\sigma_x^2}\right) - 1 \tag{2.3}$$

where σ_x^2 and σ_d^2 are the LPC gains of the original and distorted speech signals, respectively. Linear prediction coefficients (LPC) can also be used to compute a distance measure based on cepstral coefficients known as the cepstral distance measure.

In [73], a generalized wavelet filter utilization is evaluated by IS metric.

2.4.2.4. Bark Spectral Distortion Measure

Bark Spectral Distortion Measure (BSD) is a perceptual objective VQM which calculates signal distortions based on the quantifiable properties of auditory perception. The overall BSD measurement represents the average squared Euclidean distance between spectral vectors of the original and coded utterances. The main aim of the measure is to emulate several known features of perceptual processing of speech sounds by the human ear, especially frequency scale warping, as modeled by the bark transformation, and critical band integration in the cochlea; changing sensitivity of the ear as the frequency varies; and difference between the loudness level and the subjective loudness scale. BSD formula is expressed as

$$BSD = \frac{\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{O}[L_x^{(m)}(i) - L_d^{(m)}(i)]^2}{\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{O}[L_x^{(m)}(i)]^2} \tag{2.4}$$

where M is the number of frames (voice segments) processed, O is the number of critical bands, $L_x^{(m)}(i)$ is the bark spectrum of the m^{th} critical frame of original speech, and $L_d^{(m)}(i)$ is the bark spectrum of the m^{th} critical frame of coded speech. BSD works well in cases where the distortions in voice regions represent the overall distortion because it processes voiced regions only. Hence, voiced regions have to be detected.

2.4.2.5. Perceptual Evaluation of Speech Quality

Another perceptual-based objective VQM is Perceptual Evaluation of Speech Quality (PESQ) which takes into account the subjective nature of clarity and human perception. In [72], according to the perceptual features of the human auditory system, they formulate PESQ coefficients in order to map the results to MOS.

2.4.2.6. R-Factor

More recently, a number of non-intrusive voice quality measures based mainly on statistical models have been presented by ITU-T. R-factor is an objective VQM represented by ITU-T which enables the prediction of the quality of a network degraded speech stream to be made in a non-intrusive way. Moreover, its results can be directly mapped to MOS. However, although good results were reported, the technique suffers from the following drawbacks: 1) its performance seems to be affected by the gender of the speaker, 2) its application is limited to voice signals with a relatively short duration in time, 3) its performance is influenced by distorted signals with a constant level of distortions, and 4) the vocal-tract parameters are only meaningful when they are extracted from a speech stream that is the result of glottal excitation illuminating an open tract.

In [19], authors adopt R-factor and measure its parameters for their platforms using real voice samples. Another study [18] focuses on R-factor in order to automatically determine MOS results.

2.4.3. Discussion

Currently, various tools to objectively and subjectively measure voice exist. Perceptual evaluation by a trained listener, acoustic measures, physiologic measures, and self-assessment voice handicap surveys offer different perspectives on describing vocal function. However, they are costly and time consuming. On the other hand, there is a vast number of objective metrics in the literature. But, no standard measure of voice function is currently available. This lack of standard voice measure compromises the evaluation of different treatments and the evaluation of their outcomes.

2.5. TESTBEDS

Real testbeds are superior to the theoretical studies or simulators since they work with real elements. Although the studies mentioned above include notable theoretical solutions, many of them are not validated by real testbed experiments. Recent testbeds on WMSNs are

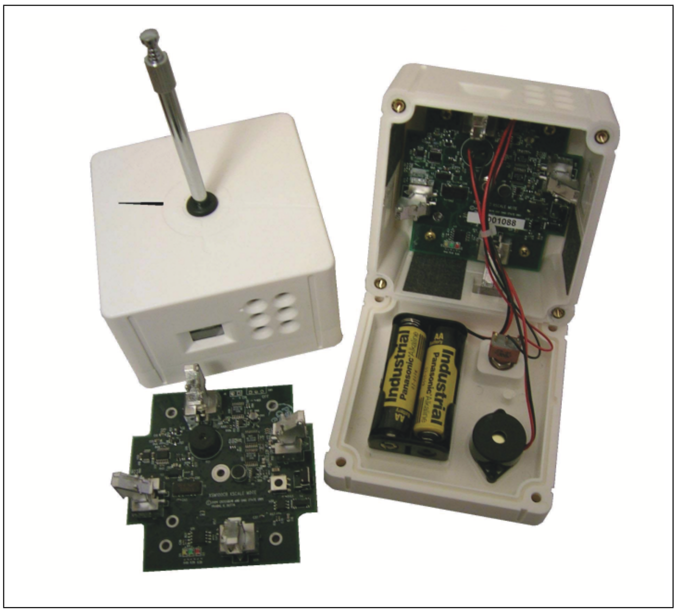


Figure 2.7. The eXtreme Scale Mote and the XSM circuit board

exposed, and current applications and hardware platforms are revealed in [23]. But most of the studies stressing on real implementations by-pass audio/voice utilization in WMSNs since they believe that necessities for video or image utilization in these networks can already be provided for audio/voice data utilization.

In [74], a large-scale testbed is constructed to develop and test middle-ware algorithms. They focus on designing ExScal motes equipped with cheap microphones, represented in Figure 2.7 and generated a mass node programming environment to decrease manual programming costs throughout the network. The nodes facilitate XSM boards which use the Atmel ATmega128L micro-controller. This hardware, like most inexpensive and low-power micro-controllers, does not provide privileged instructions or memory protection, it is possible for an application to take control of the hardware and, in the worst case, render it remotely inaccessible.

Another study focuses on utilizing microphone as an acoustic device to demonstrate the reconstruction of a high-resolution voice signal by using six MicaZ motes placed randomly around an acoustic source [36]. With a distributed data streaming and signal reconstruction, they achieve voice transmission over a WSN.

In [31], authors design a real testbed environment for voice sensing with Zigbee chip CC2430 radio as the wireless communication hardware. The voice data captured are coded with ADPCM technique in order to reduce bandwidth and data transmission is handled by separating the uplink and downlink with Time Division Duplex (TDD) technique.

2.5.1. Discussion

The limitations and requirements faced during the development of an application for a WSN increase the tendency in the literature to simulation or theoretical based studies in which the results may not be accurate due to lack of realistic models for radio, mobility etc. It is important to test the effectiveness and the feasibility of algorithms and protocols in a controlled environment.

3. VOICE CODING AND TRANSMISSION

In audio sensor networks, mutual affection of voice coding and transmission is a stringent fundamental for any application design. Strict interaction between coding techniques and providing QoS support plays a significant role for any transmission scheme. Coding mechanisms and the attributes of the voices encoded should be suited to a reliable transmission protocol. A coding scheme in the lossy nature may result in some obstructions caused by heavy information delivery. On the top of fundamentals for simple encoding, the implementation of the nodes for the delivery has to reckon with the unperceived signals to sustain a reasonable data quality. To make intent provisions against information dissipations, data handling in the nodes must be fit up with error-resilient encoding techniques. While paying regard to an efficient transmission, these techniques have to be simple and suitable for the network structured. Regardless of the network topology constructed, an affiliated node has to be bandwidth-efficient in order to strive against bulk data transmission. Besides, it has to require an affordable memory and a reasonable bandwidth which do not exceed the resource limitations while dealing with both transmission and error concealment.

After arranging a coding scheme appropriate for the transmission, the reliability of the network has to be projected. Assessment of a reasonable voice quality can be provided not only with concealment techniques, but also with a different transmission scheme. Because, methods designated for error concealment may yield in suboptimal quality results in the course of the transmission. To achieve this, one definite approach is multi-path delivery of the information.

In this approach, voice signals are sent multiple through different paths in order to assure the expected quality. However, issues related with end-to-end delay and synchronization must be beared in mind. Moreover, applications utilizing single source and sink must be able to handle mass data cumulated at the starting and end points. To overcome this time efficiency problem, the transmission scheme should extract and manage available paths effectively. For instance, the information gathered from different links can be fused in pivot nodes which are located in certain spatial intervals. Nevertheless, fusion implementation must be suitable

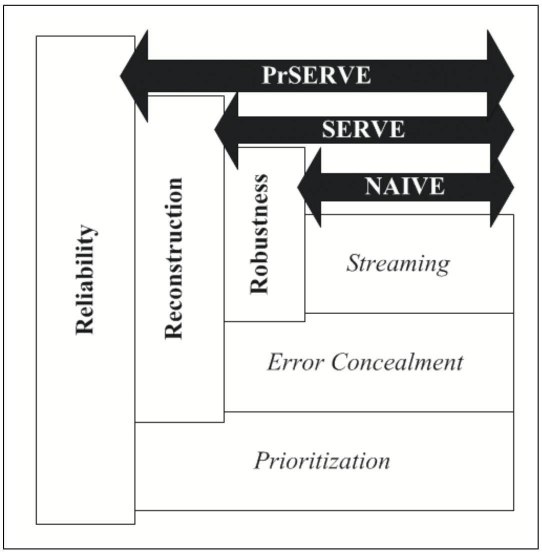


Figure 3.1. Voice Encoding Schemes

for the coding and transmission scheme in order to have a low-cost implementation. Paths may unconditionally have different link qualities which may affect the fusion performance. Under the seal of real-time constraints, a network topology can meet the incapabilities of an encoding scheme. Along with the transmission scheme, cohesive approaches for both coding and transmission may relieve each other. Parts considered necessary in the overall data can be extracted and prioritized at a certain extent. Instead of sending all of the information at once or many times, effectiveness in delivery time can be achieved with prioritization. After preferential delivery, approaches and their implementation costs for voice quality assessment may differ according to the network topology constructed.

In this chapter, a set of voice encoding schemes are explained in a bottom-up manner, as shown in Figure 3.1. In the first approach, an elementary encoding scheme which deals with data handling by means of reliable transmission is presented. The second approach supplements the presented methods in the first encoding scheme with error-resilience. The last scheme covers all of the techniques discussed in the previous approaches with a priority-based transmission. So, from the source to the end, the steps for voice encoding and transmission paying regard to robust transmission are expressed. Finally, a voice transmission framework which is designed to be coupled with priority encoding, is proposed and explained in detail.

3.1. NAIVE VOICE ENCODING (NAIVE)

We present a simple voice encoding scheme, called *Naïve Voice Encoding (NAIVE)*, which deliberately rejects sophisticated techniques contrary to the nature of tiny sensor nodes. This scheme creates an infrastructure for the further schemes analyzed in this thesis. In this section, basic steps of NAIVE in terms of data coding and transmission are explained in detail, from recording at the source through the transmission to the evaluation at the sink.

3.1.1. Data Recording

In order to interpret audio streams collected via an acoustic sensor, the conversion method of analog signals to digital representatives needs reasonable time for the rest of related applications in a WMSN. Audio is a typical content form that no measure can accurately evaluate all of its aspects. An equilibrium between the voice data characteristics and system load has to be constructed to make a remarkable amount of deduction.

The fundamental property of voice processing is sampling frequency (f_s). Converting analog voice signals continuous in time to digital values at a defined rate is required for sampling. In this point, determining a level for f_s to be able to make inferences play a key role for an operational audio sensor network. For that matter, the evaluation of data gathered must race within the time. In this thesis, a range of f_s is determined for the voice signals dealt. As a second prominent property during analog to digital conversion (ADC) is audio bit depth (bd), which describes the number of bits used to represent a sample value. In a set of digitized voice data, bd accounts for data resolution. By increasing the resolution, quantization errors in ADC can be reduced. However, in this thesis, affordable bd levels are determined during ADC, since transmission capacitance of the nodes are important.

Computations needed for recording and processing voice data may be hard for a source node while already dealing with transmission issues. Nevertheless, source nodes are considered as powerful devices in the literature [75]. Analogically in this thesis, voices to be transmitted are recorded via an acoustic sensor on a conducive micro-controller which is cooperatively working in aggregate with the source node. After applying ADC on a voice data gathered,

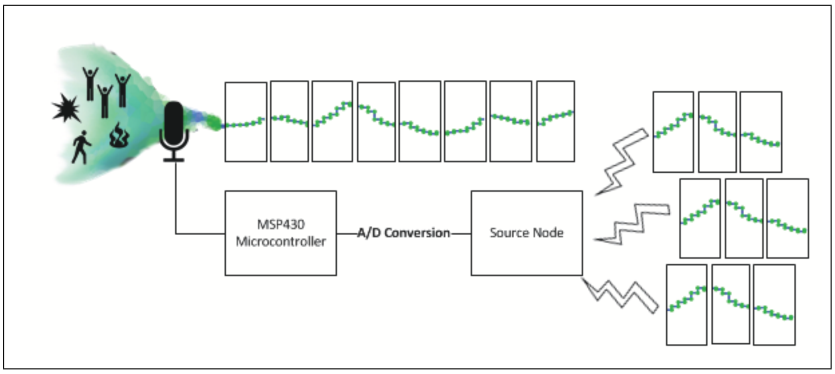


Figure 3.2. Forwarding ADC processed samples to source node

represented in Figure 3.2, digital amplitude representations a_0, a_1, a_2, $\ldots$, a_n are obtained by the micro-controller and sent to the source node. Due to storage limitations, captured voices are assumed as finite. Duration $t = 4s$ is same for all of the voice samples utilized in the framework, which will be comprehensively discussed later. Hence, the finite set of amplitude values, varying in the range of real numbers from r_A={ $-1, +1$ } constitutes the overall voice data A.

3.1.2. Data Segmentation

Partitioning of the recorded voice streams into data segments has to include low-cost steps in an effort to provide negligible processing delay during the transmission of the packets. Additionally, size of a partition transmitted in a single network packet should be maximized, so that the network can handle the intensity of continuous content traffic in real time. However, partition size rigidly interrelates with environmental factors, and hence it has to be determined according to the application area chosen. A data partition is to be called as *segment* in the rest of the thesis.

At the beginning of data delivery, the amount of segmentation will be determined with both the duration (t) of a voice data being sent and the segmentation size. For the sake of example, the number of segments will be higher when size of a segment decreases or the duration of the voice is too long. In this thesis, a segment size (s_w) at the source is treated as same as the packet size (p_w) in a transmission. The nodes are built up to be able to hold $p_w = \{20, 40, 80\}$

amplitude values of voice samples in a transmission packet. On the other hand, voices being delivered has a finite duration, as stated above. Hence, the amount of segmentation is ensured in order not to obstruct the computation overhead. The exact number of segments, where it identically gives the packet number, needed to be prepared by the source node is simply determined by s_w chosen in a particular transmission, as evinced in Equation 3.1. A segment $s_i, i = 1, 2, \ldots, n(s)$ comprised of a sample set having s_w amplitude values inside is created by splintering a voice signal A into partitions. These partition sets $A_{s_i}, i = 1, 2, \ldots, n(s)$ are encapsulated with their index values i to form network packets $p_i, i = 1, 2, \ldots, n(p)$ whenever received at the source.

$$n(s) = n(p) = \frac{f_s \times t}{s_w} \tag{3.1}$$

From the Equation 3.1, it can be clearly noticed that $n(p)$ needed to be created is directly proportional to f_s, in which higher sampling rates may yield in end-to-end delay throughout the network transmission. This consideration is also demonstrated in Figure 3.3. Number of packets with corresponding sampling rates differ according to the segment sizes. Beside determining a reasonable f_s for a robust delivery, an increase in p_w can also be considered to handle transmission delay. However, nodes have data memory limitations and data resolution is also an influential fundamental to determine segmentation size. In light of these considerations, disjunction of voice data into segments and subsequently molding these segments into network packets gains importance in case of the judgement of message losses in the environment. We aim to examine how segment width has an effect on data intelligibility when voice characteristics differ and packet losses occur in the environment.

3.1.3. Data Delivery

Heavy number of data packets passed over to each hop of the network engender transmission delay and bandwidth. To overcome this, source nodes S_i encapsulate sample units as segments according to the packet size determined. Data segments generated are buffered in data memory with their packet indices in order to minimize processing delay. Same buffering structure is also constructed at router nodes S_{ij}^R in order to minimize the relay

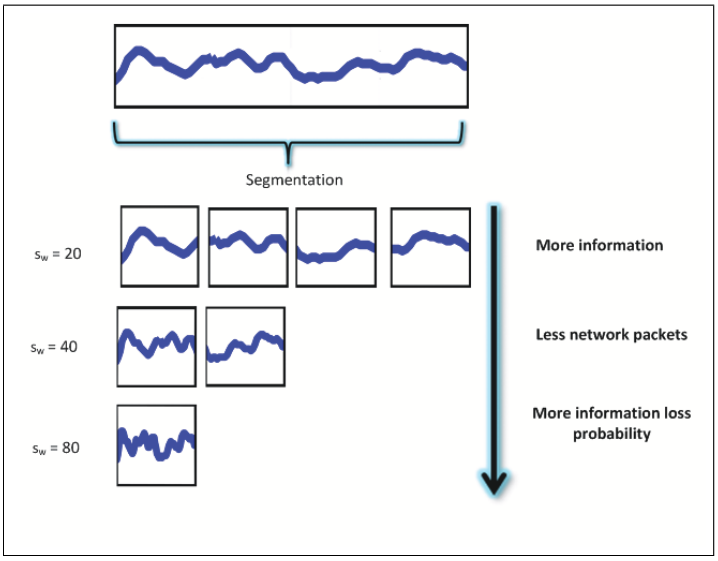


Figure 3.3. An illustration for segmentation process

time. Buffer sets are transmitted in a bursty and regulated manner to avoid congestions and delay-jitter. Basically, a routing node starts to get packets into its buffer until it is full. In this mechanism, two kind of messages are used, namely 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. However, CM are transmitted with a reliable protocol, so that consecutive nodes can easily synchronize the transmission.

The occupancy level of the buffer B is kept track with the local counter $l = 1, 2, \ldots, 10$. After a successfully perceived packet is stored in a cell of buffer B which contains p_w samples inside, l is incremented until the queue is full. The cells also structured to hold the real packet indices of the perceived data in order to provide an effortless reconstruction mechanism in advance.

3.1.4. Discussion

Since the methods served for data segmentation, delivery and error handling in NAIVE are tightly connected in each step of the implementation, a viable settlement on the nodes is applied for coding and transmission. All in all, NAIVE is a simple coding and transmission scheme. Each step in the implementation lifts the burden for the network regularity.

3.2. SIMPLE ERROR-RESILIENT VOICE ENCODING (SERVE)

We propose an applicable coding scheme for reliable voice transfer, called *Simple Error-Resilient Voice Encoding (SERVE)*, which enframes reliable data handling approaches. It puts account the data handling schemes presented in NAIVE.

3.2.1. Error Handling

One of the decent methods for quality assessment of voice signals is error concealment (EC). The idea behind the EC is using correlated voice data for reconstruction of the lost partitions. The correlated parts which are pertinent to the lost partitions benefit to a quality enhancement of the overall data without increasing required bandwidth. With the utilization of EC methods, the necessity of using reliable transmission is removed to some extent, hence no additional delay is introduced. We focused on well-known and straightforward EC methods discussed in a comprehensive manner below.

3.2.2. Error Concealment Methods

For error-resilience in sensor nodes, feasible algorithms are selected. The 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 buffering management implemented on the nodes, tolerating consecutive losses becomes possible in a specified extent. According to this mechanism, successfully gathered packets in the buffer B of a sensor node 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. An EC method begins whenever B gets full during the transmission. 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, the 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. Size of B_C is as same as of B, and number of the concealed packets cannot exceed it. For the remainder, no concealment is applied in order not to increase the overhead. Data delivery is suspended until the reconstruction for each lost packet is completed. With the completion of the error-resilience mechanism in a single transmission step, the trigger for the continuation of the transmission is provided. Concealed data segments are transmitted just after the healthy segments stored in B are emitted.

3.2.2.1. Error Concealment with Silence Substitution (EC$_0$)

The aim of EC$_0$ is to directly replace the real positions of the lost sample values with zero amplitude values. So, it deals with the lost patterns during the transmission at a raw level. The flow of the algorithm is as shown in Algorithm 3.1. EC$_0$ can be considered as adding silence in substitution for the pattern holes. Most of the research studies refers this method as a no concealment (NC) method [62]. In our study [76], we also referred EC$_0$ as NC. However, lost patterns are sent as silent segments in order to ensure real duration of the voice transmitted through the instrumentality of this method.

3.2.2.2. Error Concealment with Repetition (EC$_{PREV}$)

One of the simplest error-resilience technique is substituting the lost segments with the previous received ones [62], which is the exact idea brought into existence in EC$_{PREV}$. The repetition of the same signal perceived for the lost segments is proposed, as disclosed in Algorithm 3.2. In this method, correlated parts of the voice data are only considered in the past direction in time. The emphasis of the correlated future information is discarded.

3.2.2.3. Error Concealment with Averaging (EC$_{AVG}$)

For the lost patterns, one rational step further for EC$_{PREV}$ can be utilizing not only the preceding signals in time, but also their followings for reconstruction. EC$_{AVG}$, as a well-known straightforward EC technique, is implemented to boost the advantage of the correlated parts in the voice data. Implementation steps of this method is as in Algorithm 3.3. Mean values of the signals summed up from the successfully perceived packets may sometimes result in uncorrelated information generation with regard to its neighbors. For this reason, a connective method can be rational to preserve the data quality and integrity.

Algorithm 3.1. EC$_0$ Algorithm

Read the buffer B and determine indices i of the lost packets p_i in ascending order.
Create an error list for the lost packets.
repeat
  Find packet index i going to be reconstructed.
  Create B_C for concealed packets, with equal size of B.
  Create each element in B_C with a size of p_w.
  Push a zero vector, with a size of p_w, in B_C.
  Associate the last element of B_C with $A^L_{p_i}$.
  Set the modified bit of p_i in the error list.
until Each unmodified lost packet $A^L_{p_i}$ in the list
return

Algorithm 3.2. EC$_{PREV}$ Algorithm

Read the buffer B and determine indices i of the lost packets p_i in ascending order.
Create an error list for the lost packets.
repeat
  Find packet index i going to be reconstructed.
  Create B_C for concealed packets, with equal size of B.
  Create each element in B_C with a size of p_w.
  if $i = 1$ **then**
    Push a preceding neighbor in B_C, as a zero vector having a size of p_w.
  else if $| B^{p_{l-1}}.i - i| > s(B)$ **then**
    Push a zero vector, with a size of p_w, in B_C.
  else
    For p_i, determine its preceding healthy neighbor p_{l-1} in B.
    Get the content of the buffer cell $B^{p_{l-1}}$.
    Push $B^{p_{l-1}}$ in B_C.
  end if
  Associate the last element of B_C with $A^L_{p_i}$.
  Set the modified bit of p_i in the error list.
until Each unmodified lost packet $A^L_{p_i}$ in the list
return

Algorithm 3.3. EC$_{AVG}$ Algorithm

Require: $1 \leq n \leq p_w$
Read the buffer B and determine indices i of the lost packets p_i in ascending order.
Create an error list for the lost packets.
repeat
  Find packet index i going to be reconstructed.
  Create B_C for concealed packets, with equal size of B.
  Create each element in B_C with a size of p_w.
  if $i = 1$ **then**
    Set the preceding neighbor as a zero vector having a size of p_w.
  else if $i = n(p)$ **then**
    Set the following neighbor as a zero vector having a size of p_w.
  end if
  if $|B^{p_{l-1}}.i - i| > s(B)$ **then**
    Push a zero vector, with a size of p_w, in B_C.
  else
    For p_i, find the preceding healthy packet p_{l-1} in B.
    For p_i, find the following healthy packet p_{l+1} in B.
    Get the content buffer cells $B^{p_{l-1}}$ and $B^{p_{l+1}}$.
    for all a_n in $B^{p_{l-1}}$ and $B^{p_{l+1}}$ **do**
      $temp[n] \leftarrow (B^{p_{l-1}}[n] + B^{p_{l+1}}[n]) \div 2$
    end for
    Push $temp$ in B_C.
  end if
  Associate the last element of B_C with $A^L_{p_i}$.
  Set the modified bit of p_i in the error list.
until Each unmodified lost packet $A^L_{p_i}$ in the list
return

3.2.2.4. Error Concealment with Linear Interpolation (EC$_{LERP}$)

Interpolation is an analysis which fills the numerically indefinite spaces in a sample set with a determined anticipation. One of the simplest determination for interpolation is fitting the spaces with a linear approach. In EC$_{LERP}$, a straightforward first-order curve fitting is applied on the lost segments. For a lost packet, EC$_{LERP}$ links the last amplitude value of the preceding packet with the first one of the following signal. The details of the algorithm is given in Algorithm 3.4.

Algorithm 3.4. EC$_{LERP}$ Algorithm

Read the buffer B and determine indices i of the lost packets p_i in ascending order.
Create an error list for the lost packets.
repeat
  Find packet index i going to be reconstructed.
  Create B_C for concealed packets, with equal size of B.
  Create each element in B_C with a size of p_w.
  if $i = 1$ **then**
    Set the preceding neighbor as a zero vector having a size of p_w.
  else if $i = n(p)$ **then**
    Set the following neighbor as a zero vector having a size of p_w.
  end if
  if $|B^{p_{l-1}}.i - i| > s(B)$ **then**
    Push a zero vector, with a size of p_w, in B_C.
  else
    For p_i, find the preceding healthy packet p_{l-1} in B.
    For p_i, find the following healthy packet p_{l+1} in B.
    Get the content buffer cells $B^{p_{l-1}}$ and $B^{p_{l+1}}$.
    $incVal \leftarrow |B^{p_{l-1}}[p_w] - B^{p_{l+1}}[1]| \div p_w$.
    for $n = 1 \rightarrow p_w$ **do**
      $temp[n] \leftarrow incVal + B^{p_{l-1}}[p_w]$
      $incVal \leftarrow incVal + incVal$
    end for
    Push $temp$ in B_C.
  end if
  Associate the last element of B_C with $A^L_{p_i}$.
  Set the modified bit of p_i in the error list.
until Each unmodified lost packet $A^L_{p_i}$ in the list
return

The linear interpolant values calculated are replaced in the place of the lost signals. With the implementation of EC$_{LERP}$, the possibility of incoherent reconstructions that EC$_{AVG}$ may cause is wanted to be foreclosed.

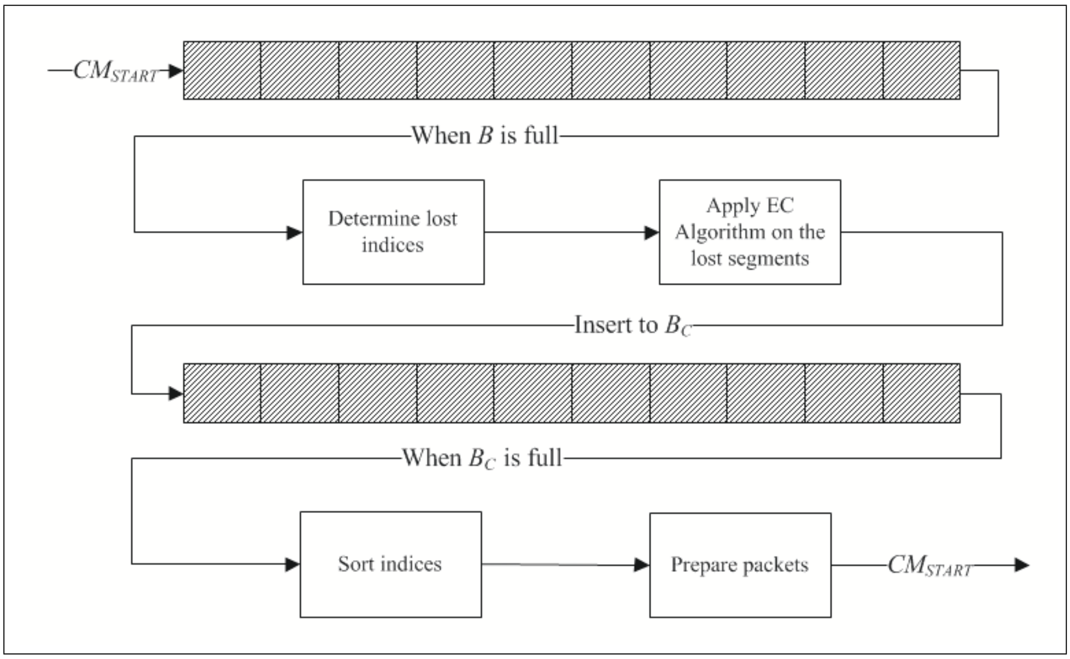


Figure 3.4. Buffering management during EC handling

3.2.3. Discussion

Methods selected for EC are appraised with the encoding scheme proposed and implemented tenderly in accordance to the data transmission fundamentals. Since reconstruction mechanisms on the nodes retain transmission for a specified time bound, a spatial extent for the error handling is determined.

According to this extent, EC methods behave mild for the packets lost in the boundaries of the buffer authority. But, consecutive segments forming pattern holes bigger than B are treated with NC in order not to pulse the delay-jitter.

To conclude, SERVE provides an error-resilient data transmission by utilizing the NAIVE encoding scheme.

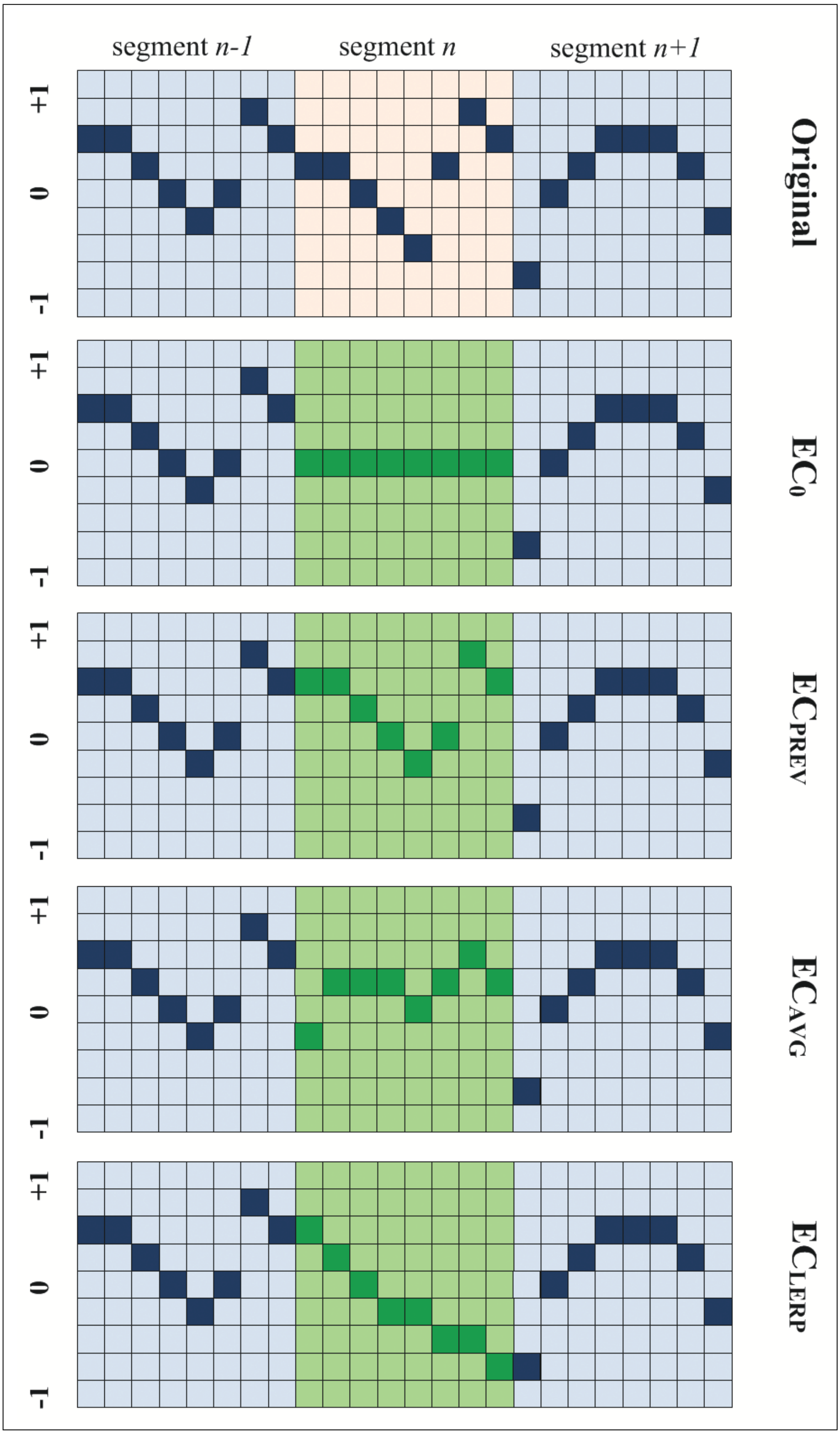


Figure 3.5. EC techniques illustrated for a lost segment

3.3. SERVE WITH PRIORITIZATION (PrSERVE)

Main purpose of a WMSN application for gathering voices from a phenomenon is deducing a result while engaging in usable information. In conjunction with this, the partitions of a voice that include structural information are more useful for any application. Setting priority levels for voice data is beneficial for resource utilization and data-centric transmission schemes. Since the voice data are not homogeneous, the temporal relation in the information makes some partitions more valuable in terms of reconstruction and information extraction.

In the previous chapters, the benefits of a prioritized encoding of the favorable data parts are mentioned. Prioritization of packets for different kind of network traffics and services is a frequently used method for QoS support [61, 77]. On the other hand, key facts of an EC mechanism accommodated to the sensor nodes are expressed. Subsequently, a comprehensive scheme for robust voice transmission is introduced in SERVE, which comprises several reconstruction approaches for data losses.

Here, *SERVE with prioritization (PrSERVE)* is put forward as a combination of error-resilience and segment prioritization. PrSERVE, as the name implies, aims to "preserve" the preferential information during voice data delivery. It also serves same coding and transmission routines of NAIVE and error recovery routines of SERVE to increase solidity in quality. Combining these two orthogonal approaches improves the perceptual quality of the received voice along with an increased network lifetime. In this scheme, the information going to be sent are priority encoded just before the transmission. Then, reconstruction performance is increased with the EC applied for the voice partitions lost.

3.3.1. Priority Measure

The significance of determining a priority level for a voice data segment should be evaluated quantitatively. There is a single-layer of prioritization for the data segments encapsulated in network packets. According to the threshold value obtained at the hands of the priority measure, packets before the delivery are simply considered as preferential or not.

To measure a priority level, handling full voice data can be hard for a sensor node that has limited memory and processing power. Hence, gradation of the data on the source node should be done without acquiring the whole information. In this study, the priority measure proposed to weigh the importance of the data segments is evaluated with a set of preliminary tests. The details of the prioritization measurement is explained below.

3.3.1.1. Threshold Determination

To specify a threshold value, one of the simplest and fair approaches is selecting the halfway point of the samples inside data segments. To put this approach into practice, median is selected for the separation of the higher half of a sample set from the lower half. In this manner, segments having amplitude values greater than the threshold determined are set to high priority.

As a preliminary evaluation, only the parts contain voices in a set of data utilized in the experiments are taken into account. In the process of threshold determination, the parts including voices inside this set are extracted and concatenated in a new file. After partitioning this file into s_w sized segments, median values of them are calculated.

In a similar approach, the halfway point of these median values, which gives the threshold value, is found. To specify this threshold value accurately, two basic points lay the groundwork for mean value determination are expressed as follows:

i. We have to take the absolute values of the amplitudes inside the segments. Since, $r_A = \{-1, +1\}$ for the amplitudes of the concatenated voice data set ${}_VA = \{a_0, a_1, a_2, \ldots, a_n\}$, the ones nearer to the margins of r_A are assumed to have importance and to be thereby more valuable. With this basic shallow prediction, significance of the amplitudes are evaluated in the domain of natural numbers.

ii. Sorting the absolute amplitude values can give us an order of importance inside a segment. The mean value at the center point of the sorted segment is determined, which is exactly $({}^{s_w}/_2 + 1)^{th}$ sample in the ordered list.

A median value of a segment can be expressed in a nutshell as in Equation 3.2.

$$m_{s_i} = \mu_{1/2}^{s_i} = \mu_{1/2}(|A_{s_i}|) \tag{3.2}$$

Every median value for the corresponding segments are stored in the medians array $m_A = \{m_{s_1}, m_{s_2}, \ldots, m_{s_{n(s)}}\}$. Median of medians, for the overall data set, is determined as in Equation 3.3.

$$m_{m_A} = \mu_{1/2}^{m_A} = \mu_{1/2}(m_A) \tag{3.3}$$

The value m_{m_A} represents a raw threshold value for the data containing only voices. However, with regard to m_{m_A}, prioritization is applied on every transmitted segments of any voice data. For that matter, the amount of highly prioritized segments can differ in this approach. Therefore, thresholds for different kinds of voice data have to be determined before transmission in order to know the level of prioritization.

3.3.2. Priority Levels & Thresholds

In a prioritization scheme, priority levels and the corresponding thresholds should be determined. These are crucial parameters which must be defined prior to a network deployment. Depending on the network architecture and the requirements of the application, amount of preferential data can differ. With regard to network architecture, the density of the network, available links between nodes and environmental conditions that effect the channels are determinant. On the contrary, requirements of applications can differ and therefore the possible voices that would be encountered in the field of application are decisive. From the application's point of view, the most significant factor is the information that to be exploited from the voice. Another issue is determining the threshold values for associating segments with the priority levels. The thresholds to classify segment weights obtained by the priority measures are determined in such an activity monitoring application.

According to a threshold value proposed, segments which are turned into network packets at the source are set as high or low priorities. In Equation 3.4, a label pL_H is shown, which keeps track of the priority level for each of the packets emitted from the source. According to this, segments which have lower mean values than the threshold value T are labeled with low priority. Vice versa for the ones having higher mean values than T.

$$pL_H = \begin{cases} 0, \; if \; m_{s_i} < T \\ 1, \; if \; m_{s_i} \geq T \end{cases} \tag{3.4}$$

When $T = m_{m_A}$ at a certain f_s and bd, it is observed 15-18 per cent of the overall packets have pL_H values set as 1, for all of s_w presented. In other words, the highly prioritized voice segments constitute nearly one-sixth of the overall data. However, according to the application domain, the amount of highly prioritized packet rate (HPPR) can vary. For this reason, we have determined several thresholds of various priority proportions, $T_{f_s,bd}$, for f_s (KHz) $= \{4, 6, 8, 12, 16\}$ and $bd = \{8, 16\}$, as demonstrated in Table 3.1.

Table 3.1. Thresholds with respect to sample frequency and bit depths

$T_{f_s,bd}$	5%	10%	15%	20%	25%	30%
$T_{4,8}$	0.2381	0.1577	0.0711	0.0303	0.0147	0.0088
$T_{4,16}$	0.2389	0.1635	0.0763	0.0308	0.0151	0.0101
$T_{6,8}$	0.2407	0.1589	0.0728	0.0304	0.0149	0.0091
$T_{6,16}$	0.2431	0.1668	0.0780	0.0311	0.0153	0.0107
$T_{8,8}$	0.2411	0.1612	0.0745	0.0328	0.0155	0.0097
$T_{8,16}$	0.2438	0.1675	0.0793	0.0333	0.0156	0.0116
$T_{12,8}$	0.2455	0.1634	0.0791	0.0332	0.0167	0.0104
$T_{12,16}$	0.2494	0.1653	0.0808	0.0339	0.0171	0.0120
$T_{16,8}$	0.2477	0.1647	0.0801	0.0349	0.0187	0.0112
$T_{16,16}$	0.2497	0.1673	0.0815	0.0355	0.0193	0.0131

3.3.3. Discussion

Prioritized transmission of a specific service is a conventional method in any kind of networks. It provides preserving the important parts of a data stream. Thresholding, is one of the simplest approach to identify priority levels in a data. By considering the system necessities, we have presented a lightweight priority-based transmission scheme.

3.4. VOICE TRANSMISSION FRAMEWORK

Transmission of voice signals is a crucial task in WMSN. Certain perspectives of the network and the nature of voice data should be taken into consideration in any approach suggested for data delivery. First and foremost, links of WMSNs have very poor qualities when compared to traditional networks due to erratic characteristics of the sensor nodes. Besides, broadcast streaming in a WSN should be brought under control. Several number of influences occur in the design of transmission schemes with multimedia data utilization. Complications such as end-to-end delay, delay-jitter, additional bandwidth requirement and congestions caused during bulk data streaming and QoS support have to be smartly issued in terms of data coding and transmission.

In view of these realities, a new voice transmission framework, in conjunction with SERVE and PrSERVE, is structured. It is based on the dense but resilient transmission of voice partitions via multiple paths.

This framework affords the required memory and bandwidth needed for the bulk data handling with an abutment of scalability. In a feat of satisfying adjustments for the transmission parameters, QoS requirements for the application are provided in a breeze. A regulated delivery is adapted by means of buffering mechanism devised, thus congestions at inter-hops are avoided. This mechanism also helps to overcome delay-jitter. With the utilization of buffering management, robustness is imposed upon the delinquent nodes.

In the framework, there are disjoint multi-paths constructed with a chained topology which have different link qualities. At certain spatial intervals, the paths are engaged at several specialized intermediate nodes implemented with a fusion design. For both SERVE and PrSERVE schemes, the network structured can be easily employed. There are four different kind of sensor nodes that facilitates the proposed encoding schemes: source, routing, fusion and sink nodes.

3.4.1. Node Types

A *source node* is the point where the first phase of the transmission occurs, as shown in Figure 3.2. In the majority of the hardware platforms, voice data exceed the memory of the nodes. Therefore, the voices are acquired from the acoustic sensor via serial bus as segments that can fit in available memory.

For the transmission scheme presented in SERVE, multiple copies of a received segment from the acoustic sensor are transmitted to the all relevant neighbors with broadcasting approach.

For PrSERVE scheme, segments forwarded from the acoustic sensor are prioritized according to predetermined priority levels. In this process, each segment is labeled as preferential or not in the wake of pL_H value determination, as discussed in Section 3.3.2. Then, segments are inserted into their corresponding priority buffers. The information about links are gathered and fed to a *cost function* which weighs available links.

The collection of the link and neighborhood information is assumed to be done by making use of the broadcast nature of the medium. With the weights calculated by the cost function, the priority levels are mapped to the available links. The highest priority buffer sets are to be sent via the best available link. Algorithm 3.5 shows the steps for these operations.

A *routing node*, which aims to transfer the received data at a high pace is also responsible for the rest of the network and has to render an error-resilient account of the unperceived data. The basic steps of data delivery is discussed with details in Section 3.1.3. Both SERVE and PrSERVE schemes have the same implementation for routing nodes.

A *fusion node*, which is a specialized version of a *routing node*, merges the buffers gathered from different paths in SERVE whereas the buffers of different priority levels defined in PrSERVE are concatenated. The logic in the implementation inside this node type is to decrease the amount of data derogations, occurred in different disjoint paths, for the rest of the network.

A *sink node* is the end point of the framework where the transmitted data are reunited. Regardless of the transmission scheme selected, data handling steps are similar to a *fusion node* implemented for PrSERVE scheme. It tries to reconstitute the overall received data according to the real indices of the packets received.

Algorithm 3.5. Source Node Algorithm

Require: Number of levels $n(pL_n)$ in prioiritization
Require: Threshold values T_{pL_n} for each pL_n
1. Reset the packet index i.
2. **for all** Priority levels pL_n **do**
3. Create link buffers B_{PL_n}
4. **end for**
5. **while** $Receivedmessage \neq CM_{START}$ **do**
6. Wait for a voice data transmission
7. **end while**
8. **for all** Data packets p_i **do**
9. Calculate the priority value of the overall amplitudes.
10. Prioritize the packet.
11. Map to the corresponding transmission link.
12. **while** $Any\ link\ buffer\ B_{PL_n} \neq full$ **do**
13. Send CM_{START} to the next router node of the corresponding link.
14. Send B_{PL_n} afterwards.
15. Clear B_{PL_n}.
16. **end while**
17. **end for**
18. **return**

3.4.2. Discussion

This framework presents an overall evaluation environment for data degradations due to packet losses occurring in the links of disjoint paths. For every data transmission, it provides intermediate hops to keep track of their lost segments. So, an augmented amount of interpretation becomes possible when a large variety of loss rates obtained by different link qualities. When the number of disjoint paths increases, computational requirements for source, fusion and sink nodes will increase. Several analysis will be made in Chapter 6, such as the impact of fusion nodes, prioritization, error concealment methods, etc. on data quality enhancement.

4. VOICE QUALITY EVALUATION

In this thesis, voice quality is valuated by a simplified version of transmission rating factor (R-factor) of ITU-T which is defined as below:

$$R = R_0 - I_s - I_d - I_{e,eff} + A \tag{4.1}$$

where it is applied between the sample values in a segment of the original at the sender side and reconstructed at the receiver side voice data, respectively.

i. R_0 expresses the basic signal-to-noise ratio, comprising circuit and acoustic noise source parameters inside, relative to received speech level.
ii. I_s combines all impairments which may occur more or less simultaneously with the voice transmission, such as too loud speech level, non-optimum sidetone, quantization noise, etc.
iii. I_d sums all the impairments caused by delay and echo effects.
iv. $I_{e,eff}$ is effective equipment impairment factor which represents impairments caused by low bit-rate codecs. Moreover, impairments due to packet losses of random distribution are examined by this factor.
v. A is the advantage factor which allows for an advantage of access for certain systems relative to conventional systems, trading voice quality for convenience.

Since we want to assess the quality as on-the-fly, the metric chosen for this valuation has to be suitable to be accommodated in the memory of the sensor nodes. However, the R-factor is applied after packets are received at each inter-hop of the testbed in order to examine the performance; thus we want to reveal if the metric embraces our presupposition or not.

The recommended input parameters for loudness ratings, circuit and room noises at the sender and receiver sides of the network which defined in R_0 are all referred to default values, since we only send a prerecorded data through the network in each test. Parameters defined in I_s are also set to their recommended values. For the impairments caused by delay and echo in the stream, I_d parameters associated with transmission time are also calculated with the prescribed values.

The audio transmission loss rates are directly set to Ppl whereas simultaneous impairments caused by quantization is associated with fs. This study treats the random packet loss probability (Ppl) as the main parameter of particular impairment factor under random packet loss—$I_{e,eff}$. Thus, a relationship between voice quality and transmission success rate is wanted to be revealed. Apart from the transmission quality of the network, data quality also depends on rates of the samples transferred in a particular time. By this means, diversity in sampling frequencies of streams transmitted to the end point is directly pertinent to simultaneous impairments in the quality. Effects of quantization density to R-factor are investigated with quantizing distortion unit (qdu) defined in I_s. The permitted interval for qdu starts from value 1, meaning that a complete quantization is supplied in the network. Just the contrary, when the quantization quality is at the lowest, it ends at value 14. For specifying a scale between sampling frequency and qdu, we assume the maximum sample rate—16 KHz utilized in the tests has the complete quantization whereas the minimum audible frequency of a human ear can sense—3 KHz is set as the maximum distortion unit permitted.

For each of the frequencies used in our voice data pool, qdu grades are determined and corresponding I_s values are generated as shown in Table 4.1.

Table 4.1. qdu and I_s values according to sampling frequencies

fs (Hz)	qdu	I_s
16000	1	1.4136
11025	6.025	10.0949
8000	9	17.1918
6000	11	22.0516
4000	13	26.4005

The equation between sampling frequencies and simultaneous impairment values is fit to a straight trend-line. Then, by setting the parameters related to other factors to their recommended default values, R-factor equation is simplified to the following function of sampling frequency and packet loss rate:

$$R(fs, Ppl) = 58.9843 - 95\frac{Ppl}{Ppl + 1} + 2.0714 \times fs \tag{4.2}$$

As demonstrated in Table 4.2, a value obtained with R-factor can be mapped to a Mean Opinion Score (MOS) which is a widely used subjective test metric in telecommunications.

Table 4.2. R-Factor to MOS mapping

R-Factor	MOS	User Experience
90	4.3	Excellent
80	4.0	Good
70	3.6	Fair
60	3.1	Poor
50	2.6	Bad

MOS of an audio is simply 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. In this way, R-factor gives the advantage of evaluating voice data received in both objective and subjective manners. Figure 4.1 demonstrates the listening quality scale.

As a different evaluation measure, the correlation among each inter-hop link quality is traced with the following signal-to-noise ratio (SNR) metric in Equation,

$$SNR(dB) = 10\log_{10}\frac{|A_{signal}|}{|A_{noise}|} \tag{4.3}$$

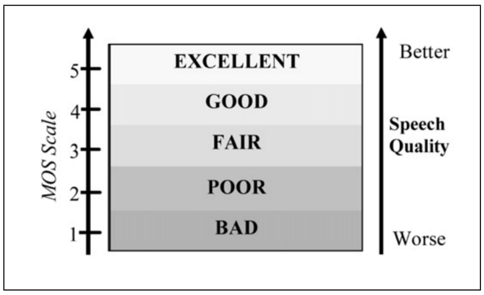


Figure 4.1. The ITU-T listening quality scale

where $|A_{signal}|$ is the sum of total absolute values of the original amplitudes and $|A_{noise}|$ is the sum of the overall absolute values of the subtracted unperceived redressed amplitudes from the original amplitudes in a transmission.

SNR is utilized to examine the link quality estimation in the chain topology and to seek if an unexpectedness occur among each consecutive hops. Besides, effect of bit depth of the data sent and the packet width on the overall transmission quality are examined with this metric. For the quality evaluation of both the network and voice is examined by the comparison of SNR and R-factor.

5. EXPERIMENTAL SETUP

In this chapter, steps in constructing an experimental network setup are explained. For the evaluation of the encoding schemes presented in Chapter 3, we contemplate on an activity monitoring as a target application scenario. Activity monitoring in WMSNs is such an application which aids in beholding the activities of a target population in a specific location, through the instrumentality of sensor nodes equipped with low-power acoustic sensors. Intended population of these applications is disabled, helpless, or elderly people living indoor and the main target is to make them benefit from the freedom of transferring information over a scalable and fully automated wireless medium. Crucial parameters for the discussed schemes are determined by means of the necessities of this application.

5.1. GENERAL WMSN SCENARIO

For the voice transmission system, considered WMSN scenario is depicted in Figure 5.1, which is the inclusion of a set of sensors; Type 1 $S_i^A, i = 1, 2, \ldots, l$, sensor equipped with an auxiliary micro-controller that has an acoustic sensor on it and Type 2 $S_{ij}^R, i = 1, 2 \ldots k_1, j = 1, 2, \ldots, k_2$, simple routing sensors, and a sink. In an emergency situation, the voice signals in the environment are recorded via acoustic sensor and then transmitted to the network end-point via routing sensors S^R. Depending upon the transmission parameters, sensors S_i^A partition the captured voices into s_w sized segments. These segments are serially conveyed to the network layer for encoding.

Experimental setup ensures the survivability of the sensor nodes. With a reliable transmission protocol, communication path between sensors S_i^A and the sink is established. Enforcement of path reliability is not a target of this thesis and is not studied further.

As discussed in the following section, a set of voice commands are utilized in the experimental setup. The results obtained from the quality diversity of these data are estimated with regard to different data segment sizes determined in the tests.

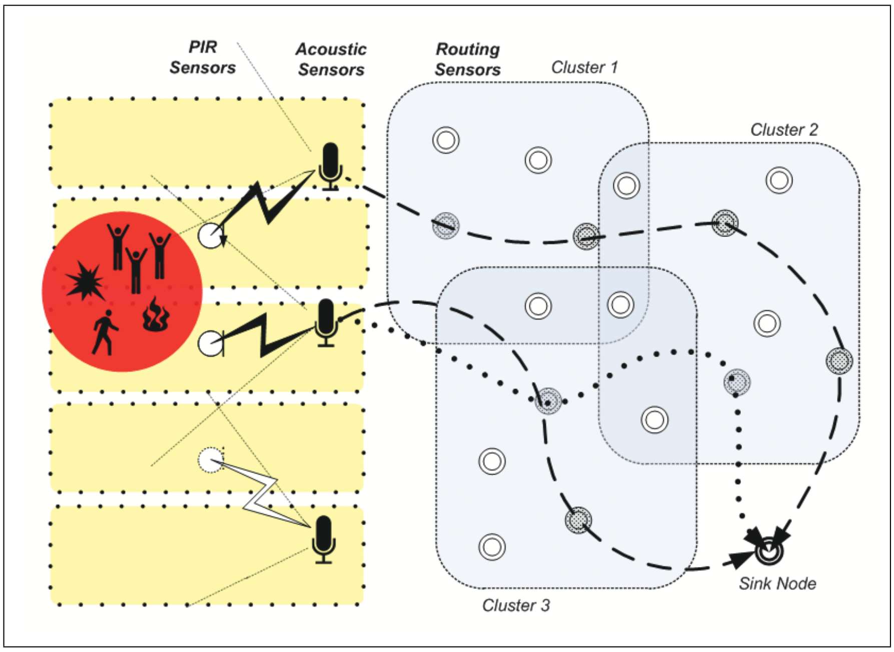


Figure 5.1. General WMSN activity monitoring scenario

5.2. VOICE DATA ANALYSIS

An important prerequisite for voice coding and transmission is the intelligibility of the data during voice recording. In terms of recording and transmission process of the voice data, intelligibility indicates how much information can be clearly extracted from the data. Voice intelligibility depends on a large variety of factors. For example, certain frequency intervals are more important to comprehend the information inside the voice data. In this section, we stress on voice data analysis and explain the steps in determining our data set utilized in voice coding and transmission in WMSNs.

5.2.1. Voice Capture

For the implementation of data recording, a real single-hop network testbed is set up to analyze the possibility of sending short voice commands over resource constrained WSNs. The testbed, as demonstrated in Figure 5.2, consists of an auxiliary MSP430FG4618 micro-controller [78] with acoustic sensor and two TMote Sky [79] nodes. The nodes are programmed using TinyOS v2.1 libraries and tool-chain [80]. The intention of utilizing a

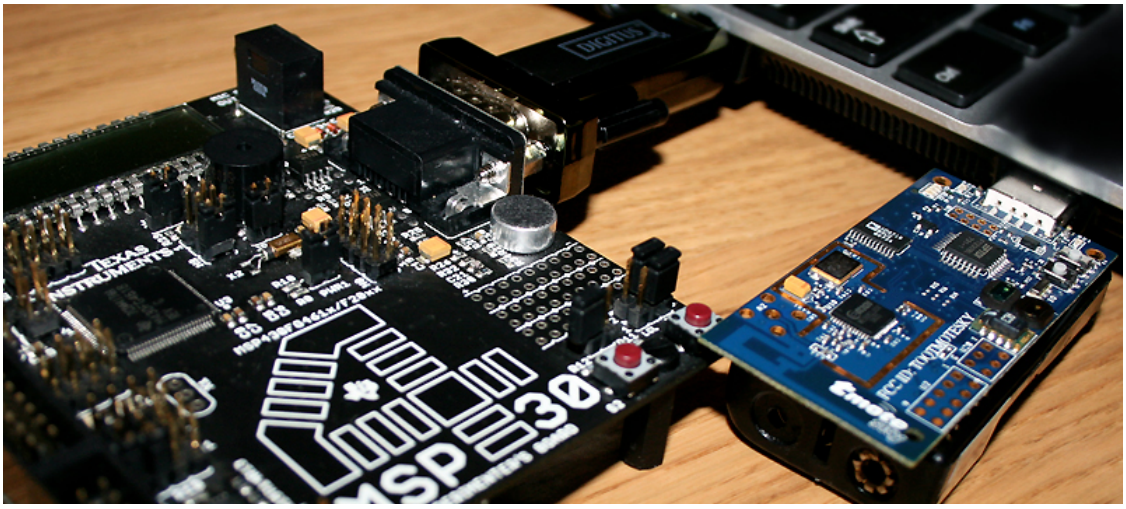


Figure 5.2. A micro-controller affiliated with the source node

conducive component for a source node S_i^A is reducing the resource requirements in favor of make it possible to deal with other operations.

The recorded signals are stored to the direct memory access (DMA) of the micro-controller. A DMA can be exploited as reading/writing a data from/to memory without using central processing unit (CPU), so storing a recorded information can be faster. Also, any other operations are not interrupted during data storage.

Analog voice data is converted to its digital representatives by means of Analog-to-Digital Converter (ADC) functionality of the micro-controller and simultaneously forwarded to DMA. Then, Universal Asynchronous Receiver/Transmitter (UART) configurations are done to transmit the bits. At this point, the problem is that; ADC has recorded data in 12-bit whereas the DMA allocates each information in 16-bit data model representation. So, a memory block has to host the digitized data in its least significant 12 bits. Other bits are set to zero values. Besides, the UART component utilizes a buffer that has 8-bit of size. Therefore, the data, that is on the memory, is divided into two parts. Initially, the most significant 8 bits and just after the least significant 8 bits are handled during serial transfer. For each step, the bits, that are in the buffer of UART, are transferred to a computer by means of serial connection, straight through Recommended Standard 232 (RS232) port.

The collected information on the computer are segmented and again forwarded to the source node as network packet in terms of Universal Serial Bus (USB) connection. As a consequence, voice capturing on a source node is put into existence in the study called *Voice Over Wireless Sensor Networks (VoWSN)* [81]. The transmission of voice signals throughout a wireless network in which every its node is capable of sensing and interpreting events and processing jobs collaboratively with other nodes is called VoWSN. With VoWSN, voice data captured with an event trigger can be transferred in an automated way. The advantage of VoWSN to similar technologies is flexibility and scalability of the network.

By this means, we aim to reveal that voice capture and transfer on a source node with limited resources, with the help of an auxiliary device, become possible. However, literal integration of the sensor node and the micro-controller is a further study and is not mentioned further.

5.2.2. Data Set Generation

The feasibility of voice capture on a sensor node is represented above. However, for ensuring a reliable network model and data transfer for any application selected, we aim to determine the voice characteristics in a decisive manner. To encourage simplicity of the designs and implementations, tangible properties must be identified.

We have prerecorded a voice sample set $D_i^s, i = 1, 2, \ldots, 8$ via the acoustic sensor, comprising of simple invocatory commands inside, each having a same fixed duration of $t = 4s$. Each of the voices is generated with varying sampling frequencies f_s (KHz) $= \{4, 6, 8, 12, 16\}$ listed in Table 5.1 and bit depths $bd = \{8, 16\}$. The effects of these voice characteristics are demonstrated in Figure 5.4.

As demonstrated in Table 5.1, numbers of packets with corresponding sampling rates are same for all of the voices in the pool, since duration *(t)* of all data is equal to each other. Regardless of sampling frequency of a voice data, the nodes are built up to be able to hold s_w = {20, 40, 80} sampling units in a transmission packet. In this point, segmentation of voice data into network packets gains importance in case of the judgement of message losses in the environment.

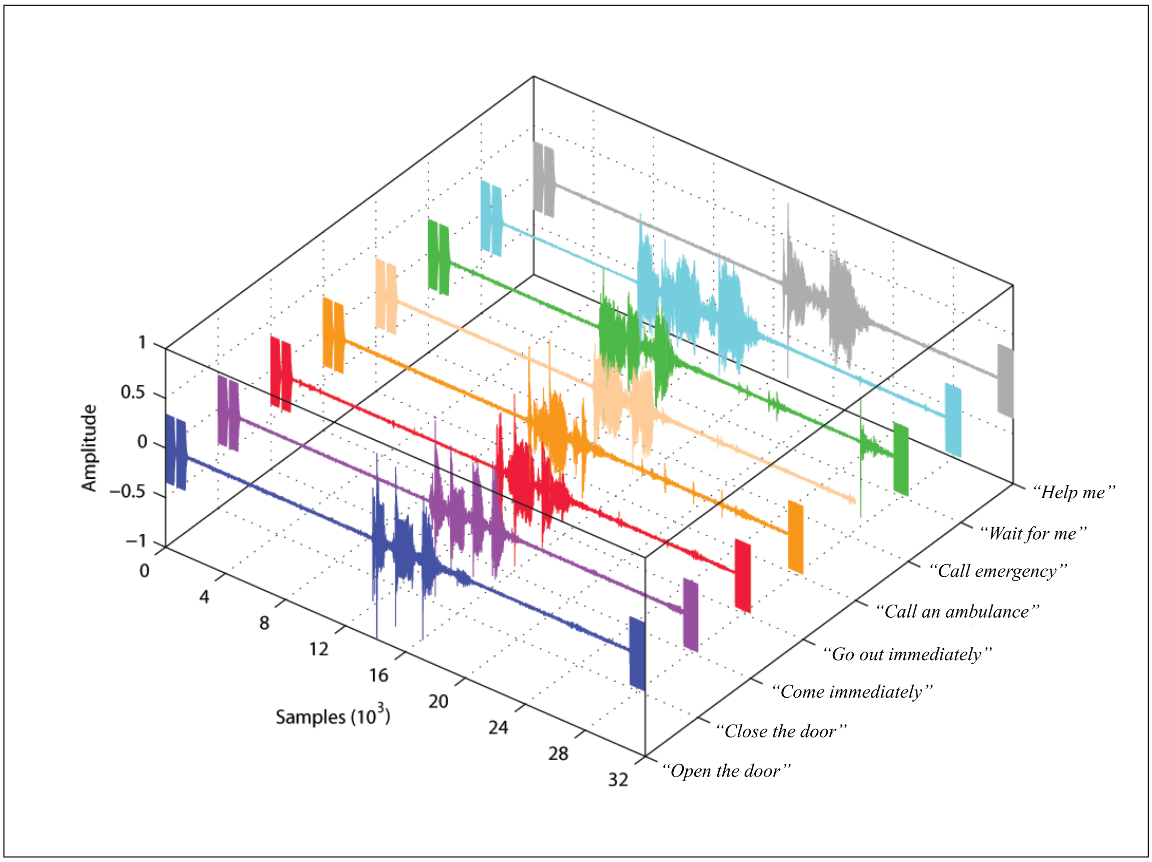


Figure 5.3. Commands used in the testbeds

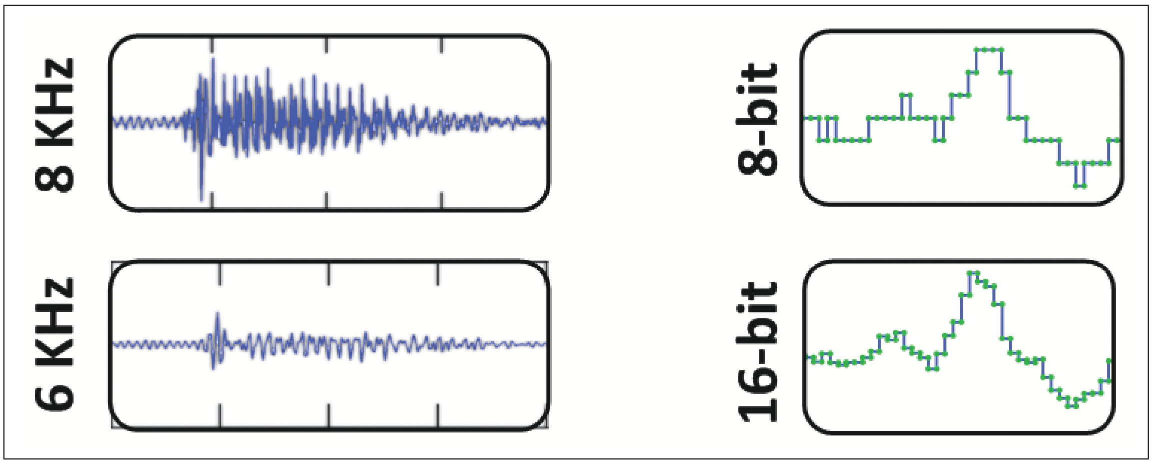


Figure 5.4. An illustration for conventional voice characteristics

Table 5.1. Number of segments according to segment size and sample rate

f_s/s_w	4000	6000	8000	11025	16000
20	800	1200	1600	2205	3200
40	400	600	800	1103	1600
80	200	300	400	552	800

We want to examine how the width of a segment determined affects the intelligibility inside an overall data when data packet losses occurs in the environment. Distinct data packets demanded to be passed over to each inter-hop of the network simply relate with network transmission delay and bandwidth according to their sample units inside.

5.3. REAL TESTBED

We have set up a real testbed in order to survey the channel conditions while transmitting the voice data through the sensor nodes accommodated with the principles of NAIVE scheme. On the other hand, the feasibility of SERVE scheme is also analyzed in this testbed. For both of the schemes, numerous tests are held at certain times and on different days.

5.3.1. System Model

The system model devised for the testbed is constructed with regard to the general WMSN scenario and the voice transmission system discussed in Section 5.1. However, the model does not include the multiple path transmission.

This system model creates the infrastructure for the encoding schemes NAIVE and SERVE, presented in Chapter 3. In all schemes, 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 multi-hop network.

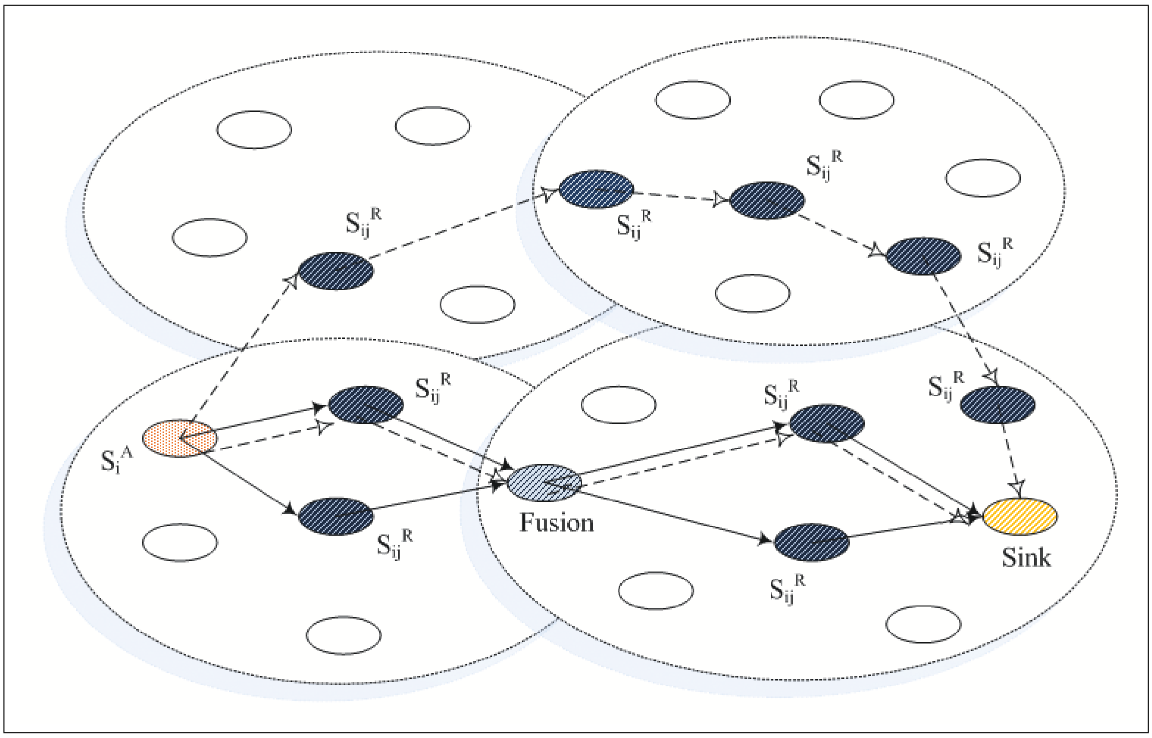


Figure 5.5. General WMSN schemes

5.3.2. Methodology

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 $f_s = 8KHz$ and $bd = 8$ and utilized in the tests. The segments constituting the file data are transferred from the computer in terms of serial connection. They are tuned into data packets which are sent to the source node of the established WMSN testbed. The testbed consists of TMote Sky sensor nodes. The nodes are programmed using TinyOS v2.1 libraries and tool-chain.

Over a chain topology, the data packets are transmitted through S_{ij}^R nodes to the sink node. In order to collect received voice data at each hop, a back-channel mechanism is devised. With the help of this mechanism, received data with the indicators at each hop are sent to the computer serially. For each test, computer extracts the loss patterns with the lost indices

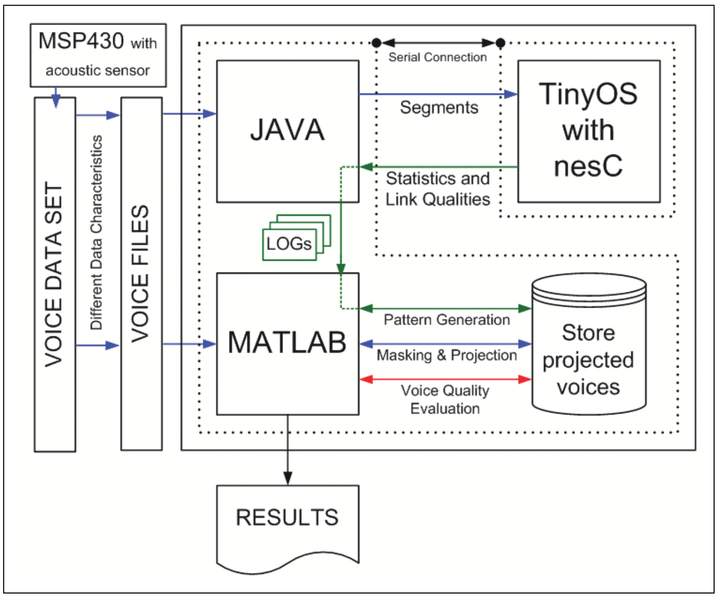


Figure 5.6. Overall software diagram of the testbed

from the received voice. Moreover, transmission details and information of the hops are saved within the pattern as Matlab structure variables. Overall software diagram of the setup is depicted in Figure 5.6.

5.3.3. Testbed Setup

An indoor testbed environment comprising a 10-hops network is conducted by using 20 TMote Sky sensor nodes. In the experiments, TinyOS v2.1 with nesC v1.3 [80] is utilized to realize a simple voice transfer over a chain topology. As a back-channel, ten of the nodes are utilized for data collection at the intermediate hops of the network. The Ten of them are used for data collection as backchannel. The details of the testbed are given in the following sections.

5.3.3.1. Node Deployment

Tests are conducted inside a large atrium of a building as illustrated in Figure 5.7. For sharing a common communication medium open to implicit environmental factors, nodes are homogeneously deployed at a same communication range. Moreover, the nodes are partitioned into two groups named Group 0 and Group 1. They are lined up vertically

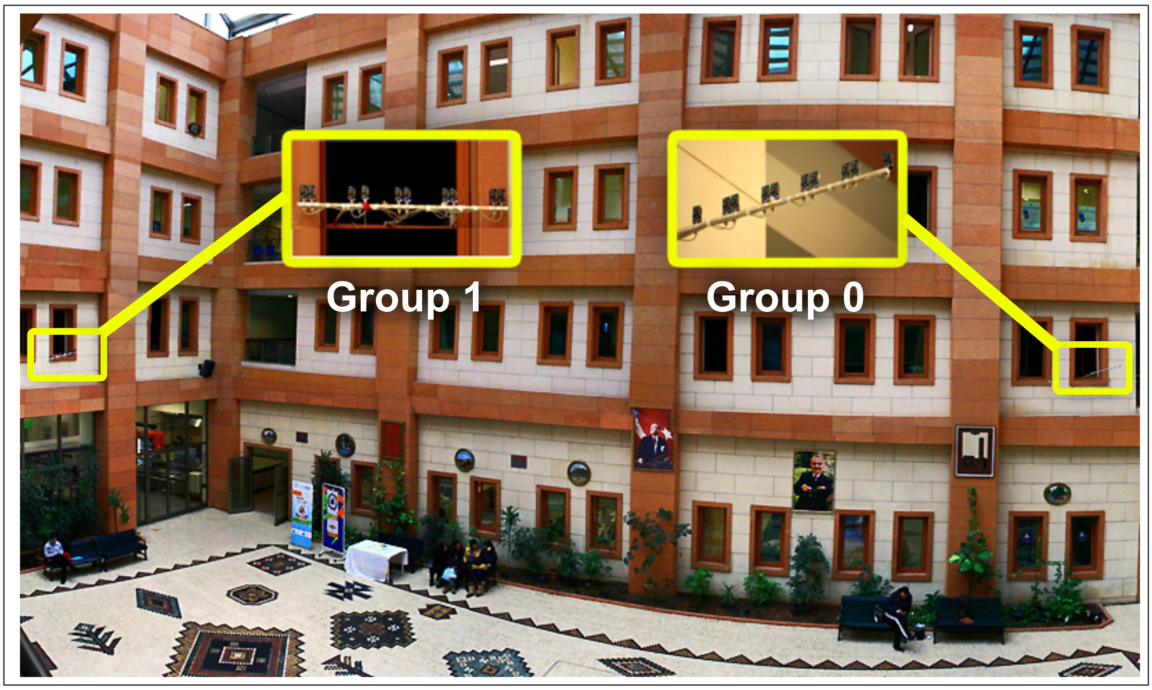


Figure 5.7. A view from testbed area

on thin linear sticks, which are horizontally pointed out from the windows on the first floor, approximately 5m above from ground, with no obstacles between them. The distance between the groups is measured as 28m. The groups consist of five "hop couples" with intra and inter couple spacings of 4 and 17cm, respectively.

As depicted in Figure 5.9, the sticks equipped with the sensors nodes are positioned parallel to each other to complete a hypothetical rectangular area, when viewed from above. The output power of nodes is set to -7dBm. Each group is connected to a base station computer via self-powered USB hubs in order to avoid performance variation due to power differences when run on batteries.

5.3.3.2. Transmission Scheme

The voice transmission scheme can be cascaded into several steps as follows: At the computer side, a prerecorded voice data of $t = 4$ s duration is partitioned into segments. The data has a sampling rate of $f_s = 8$KHz and a bit depth of $bd = 8$. Previously presented segment sizes $s_w = \{20, 40, 80\}$ are utilized as the distinctive in-network parameters in

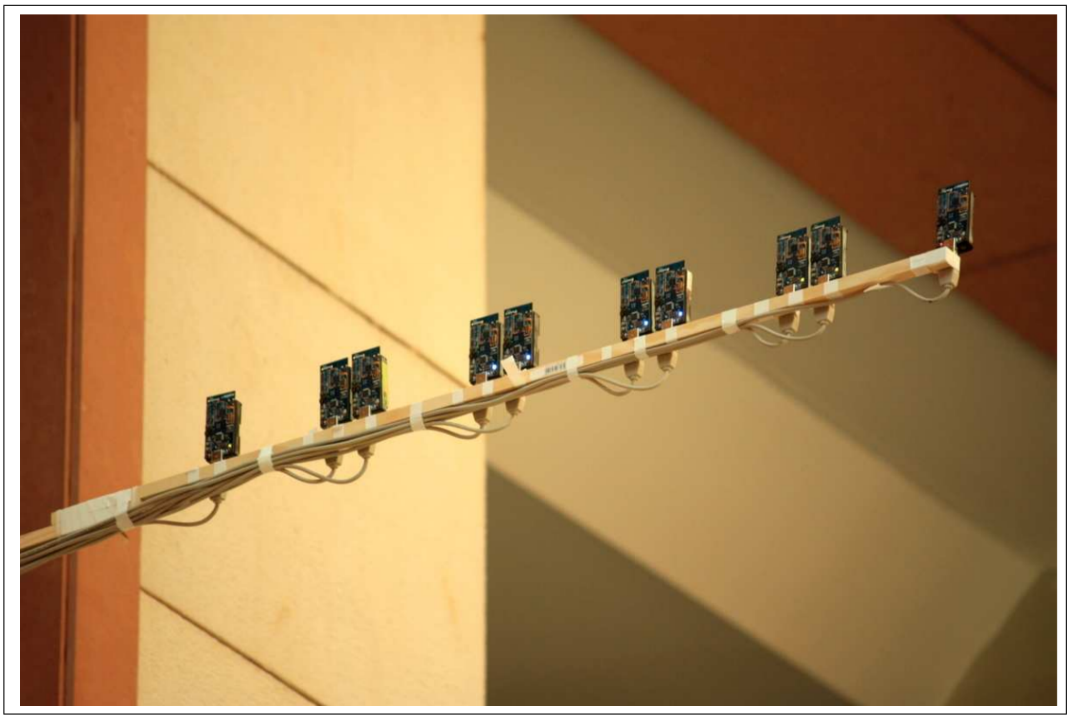


Figure 5.8. Node Group 0

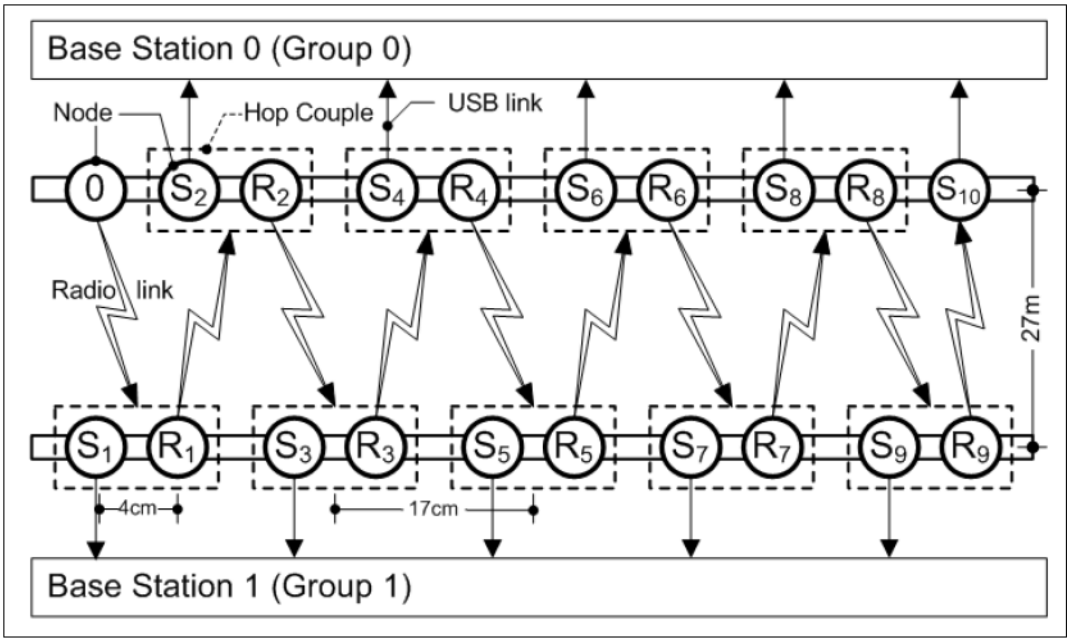


Figure 5.9. Testbed Diagram

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 serially via USB links. Then, the source node sends the packets to the sink node over ten hops with best-effort delivery.

In our voice transmission scheme, each hop consists of two nodes called "*hop couple*". They have given the same node ID. In each hop couple, one of the nodes, called relay node (R_i, i=$1, 2 \ldots 9$), is used to send the incoming data to the other hop couple with the consecutive node ID via radio link, while the other node, called snooping node (S_i, i=$1, 2 \ldots 9$), is used to send the incoming data to the base station computer via USB link. By this way, while transferring a voice over ten hops, the intermediate results are recorded in each hop by using these snooping nodes. There are only two single nodes numbered with 0 and 10, as the source and the sink node respectively. The base station computers at each side records the voice data along with loss patterns and their PERs.

5.3.4. Examination

To make hop-based comparisons accurately and equitably, it is necessary to get the PERs occurred in each hop concurrently. As conducting the tests on different time periods, packet losses dramatically fluctuate due to changing environmental conditions. The result of one-day-long transmission tests for ten hops is given in Figure 5.10 as an example of this situation. As the figure demonstrates, in the noon, high PERs occur due to the noise induced by the crowd in the atrium. However, in the midnight, low PERs are obtained. While transferring the data, the intermediate results in each hop are recorded by snooping nodes.

864 preliminary tests had been conducted for each hop individually at several periods, hence different channel conditions are obtained. In Figure 5.11, PERs results are depicted for $s_w = 20$, $s_w = 40$, $s_w = 80$, respectively. In these tests, it was observed that circumstantially, the PERs obtained for hop 9 and hop 10 dramatically break out the transmission most of the time for all s_w experimented. For $s_w = 80$, it is observed that μ PER = 0,278 whereas μ PER = 0,124 and μ PER = 0,183 for $s_w = 20$ and $s_w = 40$, respectively.

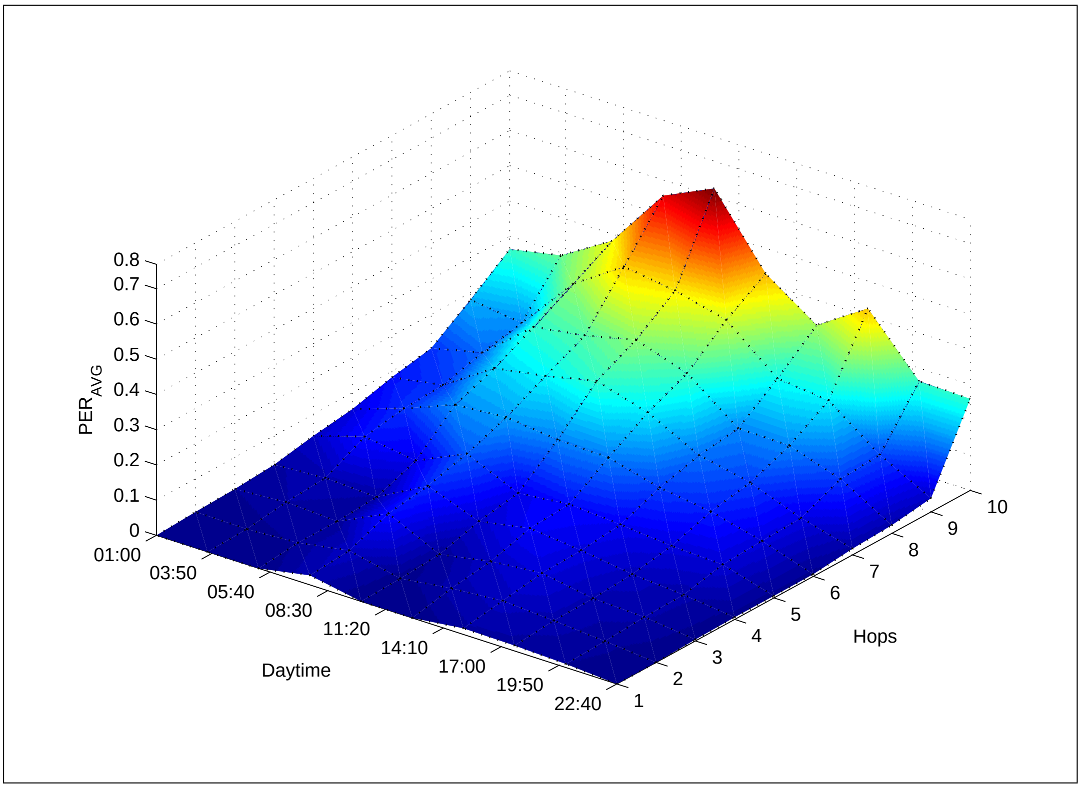


Figure 5.10. Daytime transmission results

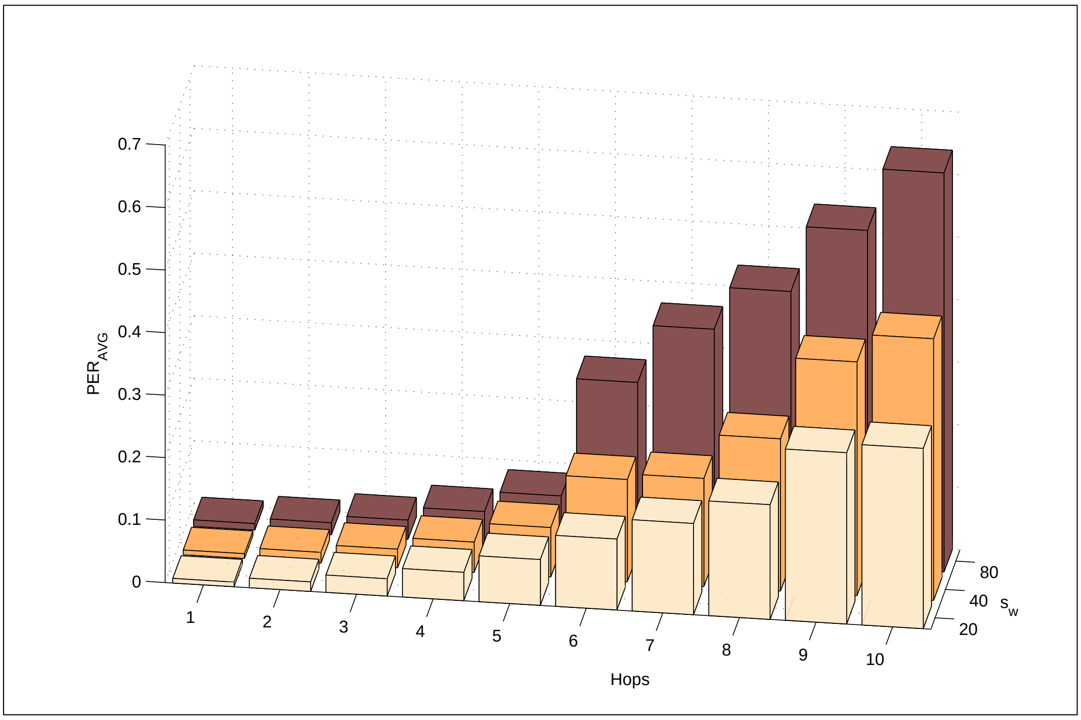


Figure 5.11. Average PER according to the segment sizes

5.3.5. Discussion

The testbed results gained from indicate attained from the tests indicate the vulnerability of raw transmissions. A transmission scheme with $s_w = 20$ has the least average PER. However, a decent data transfer can also be possible when $s_w = 40$. Because, average PER values for $s_w = 20$ and $s_w = 40$ are nearly same. On the other hand, the number of packets need to be transmitted will two times less when s_w is doubled. Therefore, the end-to-end delay will be reduced twice. For NAIVE and SERVE encoding schemes, we can thereby state that this transmission scheme can gracefully operate in terms of a less packet loss probability and a short end-to-end delay when $s_w = 40$ is utilized.

5.4. SIMULATION

In computer science, a system or model can be implemented beyond the normal course of reality by defining a set of specifications. With the determination of the elements and their necessities in an exact manner, constituting virtual analogies can be a candidate solution to validate a real problem laying inside a system or model. For any application, a nominal implementation of a real feature, action or process can be needed in order to resolve the demands which cannot be engaged in the implementation or analyzed thoroughly. For instance, there are quite a number of exemplary applications in WSN which cannot go beyond a theoretical base since running real experiments is costly and time consuming most of the time. By this means, a simulation can be devised to eventually present the literal effects of particular conditions and courses of action.

A simulation environment opens up an opportunity for conducting several analysis for numerous times, with a predefined set of system characteristics and features. So, a large set of sensitive evaluation results can be obtained. Functions or systems relating to a probability method can be fit and scaled easily in a specific solution accordingly.

However, gathering concrete and reasonable results from a virtual implementation is not a trivial task. Acquisition of the relevant characteristics and behaviors of a courted property matter in a simulation design. Moreover, approximations and assumptions considered in a virtual implementation affect the fidelity and the validity of the outcomes. A model can be based on a strict set of considerations, however it is possible to design a simulation with a high abstraction level. For the sake of example, homogeneity of a specific layer can be affected by many factors in a real WSN testbed environment. By excluding the unexpected situations outside of the simulation in this field, a high level of abstraction at a certain extent can be an essential solution to examine the results in a feasible way.

In this section, considering the advantages of and the challenges in a nominal implementation, the simulation environment served for the encoding schemes is explained and analyzed.

5.4.1. System Model

General WMSN scenario represented in Section 5.1 lays out exactly the same transmission system model intended for this simulation. The necessity of configuring a multi-path transmission, which is not ruminated over the real testbed setup explained in Section 5.3, is also taken into consideration in this instance. The system of the simulation is an inclusive model for all of the encoding schemes portrayed in Chapter 3. For all of the schemes, a voice data set in which each data with the same duration in common with others, but with several f_s and bd versions is dissected into s_w sized segments at the S_i^A node. Each segment is tuned into network packets and conveyed towards the sink over several paths of designated topologies.

5.4.1.1. Radio & Channel Model

Radio propagation modeling is one of the most restraining factor for WSN simulators to obtain accurate enough seamless models consisting of numerous number of nodes [82]. In our simulation, the testbed setup does not consist of a large scale of sensor nodes, as discussed further. Moreover, since we aim to investigate the effects of packet losses on the voice quality, the simplest models are considered for implementing radio and channel conditions straightforwardly.

With these considerations, free space propagation model is selected as the propagation model.

It assumes there are neither absorbing obstacles nor reflecting surfaces in any circumstance in the environment. According to this model, the environment is assumed to be empty in terms of many type of noises related with the transmission. However, when considering the under-realistic propagation model, the signal diverseness is also assumed to be low and packet losses thereby are at the minimum level [83]. For this reason, we postulate that packet losses occur due to radio channel errors or collisions are mainly located near the intersections. With this primitive supposition, channel models are implemented according to the indicated packet loss probabilities. Therefore, a channel is stochastic, i.e. the propagation loss is a random process that changes over time.

5.4.1.2. Assumptions

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 deals with the 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. Their interaction and behavior between the shared system medium in terms of energy and radio propagation are totally ruled out. Therefore, for a designed node, no search for an optimal neighbor is performed. According to the selected transmission protocol, the 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.

5.4.2. Methodology

The footsteps in the designing process of the procedures can be sequenced in three basic phases: voice data set utilization, designing a topology with pattern generation and transmission, as outlined below:

i. For the determination of the voice data set to be utilized in the simulation tests, four kind of vectorsused in order to hold the voice data, set of sampling frequencies, data resolutions and segment sizes for a variety of analysis. In these vectors, any data having different characteristics and properties are welcomed. Thus, data utilization can be adapted to any information and/or characteristics going to be used. Moreover, different set of sizes can be set in the vector holding segmentation information. This segmentation property can also be utilized in the analysis of the voice transmission.

   As the first phase of the simulation methodology, data in the vector associated with the voices are read with different sampling frequencies and bit depths defined in the corresponding vectors. They will have different number of samples inside according to their duration and sampling frequencies. The samples of each voice data are stored in a data pool structure. The structure including all of the data is sent to the segmentation process. Created segments are associated with their data and stored in $s_w \times n(s)$ sized arrays. As discussed later, the segments are also prioritized for PrSERVE scheme, with regard to the priority measure detailly explained in Section 3.3.1.

ii. Generation of patterns through a network topology is designed as a second phase. An initial pattern is generated with a cardinality of general pattern. A general pattern cardinality ($\bar{\bar{P}}$) is calculated according to the sampling frequencies and segment sizes utilized in the first phase. As discussed in Section 5.2.2, the number of packets $n(p)$ vary according to the sampling frequency f_s and packet size p_w in a network transmission. Since we utilize a vector of f_s in the simulation, we need to create a general pattern that can be down-sampled to the exact pattern cardinality of the corresponding data in any test. The details of determining $\bar{\bar{P}}$ is given in Section 5.4.

   Just after the initial pattern is created, it is sent over a procedure, called *transmitPattern*, which takes the pattern and a path loss probability as the main arguments. After the execution of this procedure, a new pattern is generated with an overall packet loss rate is equal to the path loss probability defined. By this means, the initial pattern is turned into a result vector as if it is routed one-hop forward. Similarly, other hops can be created with an analogical manner. Thus, any network topology with holding the patterns created at each hop can be designed in a scalable and flexible manner. The actual topology designated is explained in Section 5.4.3.

iii. Thirdly, the pattern sets are applied on each of the segmented data sets. According to the properties and characteristics of voice data, the data are constructed according to the encoding schemes. A comprehensive evaluation of the reconstruction processes for the schemes are given in Chapter 6.

5.4.3. Testbed Setup

As the testbed setup, we implemented a network topology consisting two disjoint multi-hop paths in Matlab. The transmission reliability of the paths can differ according to the definitions in the link qualities in any test. There are 10 hops for each path. On one level, these disjoint paths can be considered as the simulated instances for the topology presented of the real testbed setup, discussed in Section 5.3. On another level, starting from the source, paths knot with each other in every two hops, so there exists 5 fusion nodes in the topology. In this knitted multi-path transmission, a fusion algorithm is accommodated on the knitting nodes which corrects the unperceived packets of disjoint paths by either concatenating or merging the data received from different paths according to the encoding scheme selected.

The testbed consists of two major routines in terms of voice data handling: pre-processing and post-processing of the voice data set, in which the essential steps are given below.

5.4.3.1. Data Set Segmentation

The original voices D$_i$, i=$1, 2 \ldots, 8$ with fs (KHz) $= \{4, 6, 8, 12, 16\}$, $bd = \{8, 16\}$ and $t = 4$s are read and partitioned into segments according to $s_w = \{20, 40, 80\}$. So, a data set with 240 distinct structure elements is formed up in a single test. In this set, the segments of the corresponding data are hold with their indices and then treated as network packets in order to further apply the loss patterns over in any test.

5.4.3.2. Loss Pattern Generation

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. The size of the initial pattern, so as its ensuing versions with regard to a stochastic approach, matter in

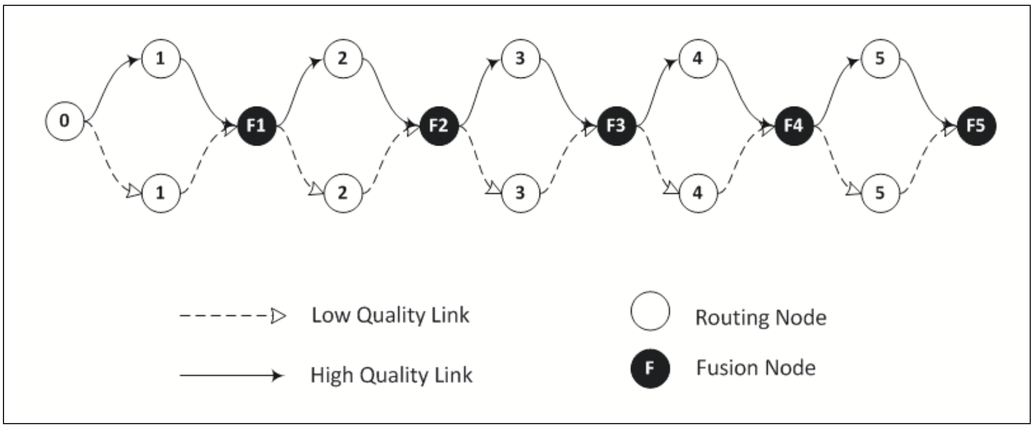


Figure 5.12. Simulation Diagram

order to be applicable for every f_s and s_w determined. Clearly saying, the cardinality of the initial pattern must be general for all f_s and s_w versions of the segmented data set in order to hold the tests at once, that is it must has a divisible cardinality for every $n(p)$. So, we also utilized a 2D array to hold every $n(p)$ in the transmissions as briefed in Equation 5.1.

$$\bar{\bar{P}}_{f_s,s_w} = n(p) \tag{5.1}$$

General pattern cardinality $\bar{\bar{P}}$ for all f_s and s_w defined in the vectors is calculated as 9600 according to the least common multiplier of all cardinalities in the set, as expressed in Equation 5.2.

$$\bar{\bar{P}} = LCM(\bar{\bar{P}}_{6,20}, \bar{\bar{P}}_{6,40}, \bar{\bar{P}}_{6,80}, \ldots, \bar{\bar{P}}_{16,20}, \bar{\bar{P}}_{16,40}, \bar{\bar{P}}_{16,80}) \tag{5.2}$$

While transporting the packets on the paths, new patterns are formed as the replicas of the initial pattern. For the fusion nodes, the preceding patterns which may contain redundant but different packet indices are fused. A fusion node, according to the encoding scheme selected, decides to recover lost packet indices in the patterns by either concatenating or merging them. So, also the fused patterns are generated in the simulation. This fusion technique is utilized to decrease the packet loss probability and to ensure hop-by-hop reliability. The pattern generation is continued by transmitting the fused patterns over the paths. A total of 25 general patterns are formed in a unique test.

5.4.3.3. Masking & Projection

Corresponding masks M_t, t=$1, 2 \ldots, n$ are generated from the loss patterns generated, for every s_w utilized in the experiments and f_s in the pool. In each test, they are applied on the segmented data set for every hop with respect to their corresponding packet indexes. For every $\bar{\bar{P}}_{f_s,s_w}$, $\bar{\bar{P}}$ is down-sampled to the actual cardinality $n(p)$ of a unique voice data. After the projection of a single voice data is over, unperceived data segments are determined, as listed in Figure 5.14.a. In Figure 5.14.b, a cross-section of a voice data is shown within the mask generated, with respect to losses occurred during an ordinary data transfer. Unperceived signals in the mask are shown in whites. For the schemes SERVE and PrSERVE, the lost packets are reconstructed with EC algorithms presented in Section 3.2. For NAIVE, EC_0 is applied in order to preserve the data integrity. A total number of 6,000 masks are projected on the data set.

5.4.4. Discussion

A data-centric wireless sensor network simulation environment is devised. It is adaptive to different sampling frequencies, bit depths of any voice selected for testing. Besides, data transfer can be dynamically set for different segment sizes. Number of tests being held can be determined easily. The simulation is flexible for implementing a required network topology. The number of nodes, hops and paths can be identified with different link qualities and packet loss probabilities. This simulation is convenient to support the outcomes of a real simulation by utilizing random sampling processes.

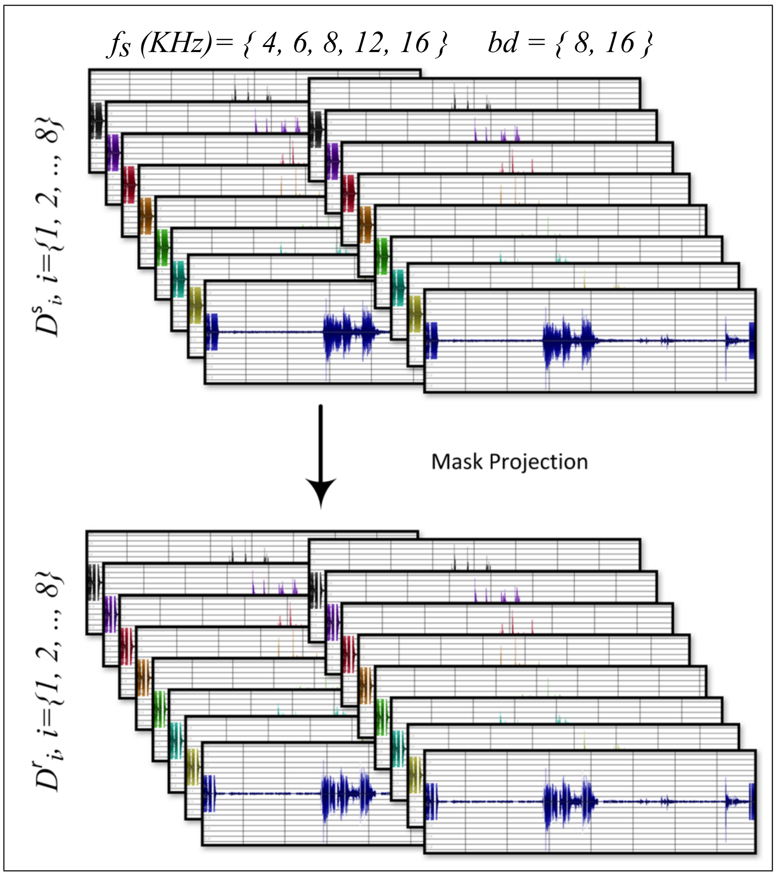
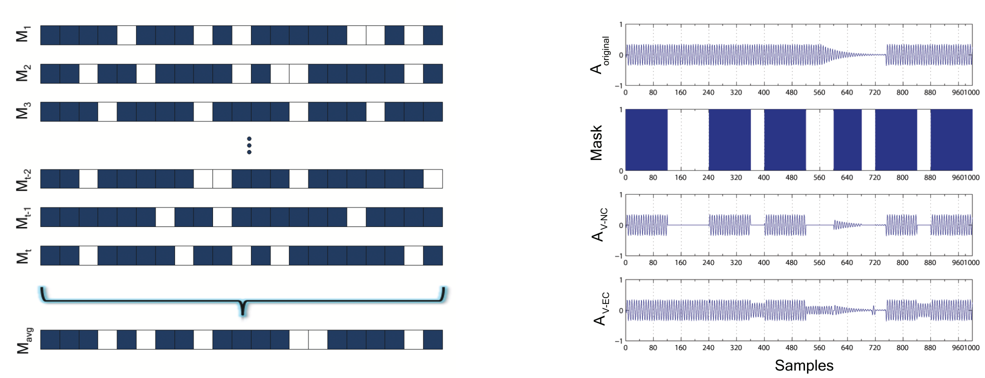


Figure 5.13. An illustration for received data generation in the simulation



a. An illustration for the determination of the average mask file

b. An instance of construction schemes being applied on the lost segments

Figure 5.14. Masking & Projection processes

6. PERFORMANCE ANALYSIS & EVALUATION

Throughout this dissertation, several qualitative observations about the voice data characteristics and network properties have been presented. This chapter will report upon the experimental evaluations mentioned in the previous chapters. Since our work is intended to give insights into voice coding and transmission in WMSN, some conclusions have been anticipated in the qualitative evaluations. Hereafter, a quantitative analysis of the results is conducted in order to estimate the quality obtained for the proposed of techniques and models.

6.1. NAIVE EVALUATION

This section gives a qualitative assessment of voice characteristic over a real experimental setup. Additionally, the testing process is corroborated with a set of different inputs manipulated in a conducive simulation environment. The technicality for the evaluation of systematically varying inputs is examined in detail.

In this evaluation, the system model utilized in a real testbed environment is expressed in Section 5.3.1. The elements in the system model can be consulted in order to understand the methodology and the performance analysis of this scheme broadly.

6.1.1. Methodology

In the voice data set, the original files D_i, i=$1, 2 \ldots, 8$ sampled according to the different voice characteristics are transferred to the sink node in segments. The amplitudes are encapsulated into network packets according to the predefined segment sizes. In each group of segmentation tests, these files are partitioned into $s_w = \{20, 40, 80\}$. For any file D_i, we consider the original amplitude values in the overall data as A_0. On the other hand, the entirely gathered data packets at the intermediate nodes and at the sink form up corresponding A_R, which is the set of the received amplitude values.

According to the descriptions of the transmission steps explained in Section 5.3.2, the data gathered are reconstructed according to the packet indices at every network hop. The unperceived packet indices are determined and lost amplitudes are substituted with silence. By this way, EC_0 is applied to the gathered data.

We have investigated the voices transmitted in real with different conditions in order to compare with the simulation outcomes. In the real transmissions, a voice file sampled at 8KHz/8bit is transmitted with all of the segment sizes presented and its loss patterns for all transmissions are stored. For each segment size, 288 real transmissions are performed. Since there are 10 hops, a total of 8640 loss patterns are generated. These patterns are utilized and expanded with a random sampling process in the simulation. According to this, a set of controlled experiments are held to identify a trade-off between segment sizes utilized and the voice characteristics. While a loss pattern generated in a simulation remains constant, the replicas of the voices sampled at different fs and bd are transferred in the simulation environment and silence substitution is applied on the lost segments just like in a similar approach expressed in the real experimental setup.

The voice quality at each of the end points of the network is assessed with regard to the evaluation metrics expressed in Chapter 4. R-factor values are obtained according to the overall packet loss rates in an overall transmission. Moreover, the sampling frequency preset for a unique test is also integrated into the formulation. For SNR calculation, the absolute difference between A_0 and A_R is taken into consideration.

6.1.2. Performance Analysis

In this scheme, the outcomes of the experimental setup is analyzed to observe the effects of voice characteristics in the transmission. The transmission property only vary according to the segment size selected whereas different sample rates and bit depths are taken into consideration for voice data analysis.

6.1.2.1. Sensitivity Analysis

A property or characteristic of voice data coding in WMSN is subject to its several properties or of other elements in the system. A model selected for the application design have to cope with natural intrinsic variability of the mechanisms inside. For a subtle evaluation of the inputs inside our design, a sensitivity analysis is devised by experimenting a set of inputs as the constants of the data transmission while others are deliberately differ in a large variability. In each experiment, a single f_s, bd or s_w is determined as the constant input. So, the robustness of the voice quality against voice characteristics and segment sizes are evaluated in an unprimed manner. According to this evaluation, the tests are conducted with silence substitution for all lost partitions of the overall data utilized in the experiments. Therefore, by only the setting an input as constant in a test, the variation between sampling frequency f_s, data resolution bd and size of the segmentation s_w is tried to be comprehended. Moreover, a loss pattern generated in a single test is projected to all of the A_R generated with these controlled data and network property inputs. By this means, the differentiation between several characteristics of the same data replicas of a voice is determined distinctively.

The R-factor values for each sampling frequency in Figure 6.1 depict the quality degradations of the voice data. Anticipated outcomes demonstrate that a comprehensible data can be obtained at the 5^{th} hop for $f_s = 12$ KHz. When $f_s = 16$ KHz, the hop number that can handle intelligibility can go further to 10.

The real SNR results according to the hop numbers and different segment sizes of the received voices sampled at $f_s = 8$KHz and $bd = 8$ are shown in Figure 6.2. It is clearly seen that, the correlation among SNR values between each segment size offered in the tests slowly declines. The reason for that is the amount of the information carried in the lost segments dramatically increases when PER increases for each hop. However, we can evenly state that there is no big difference between segment sizes utilized. Moreover, according to the controlled experiments, the data resolution differences do not alter the SNR values of a same data sampled with 8-bit or 16-bit too much.

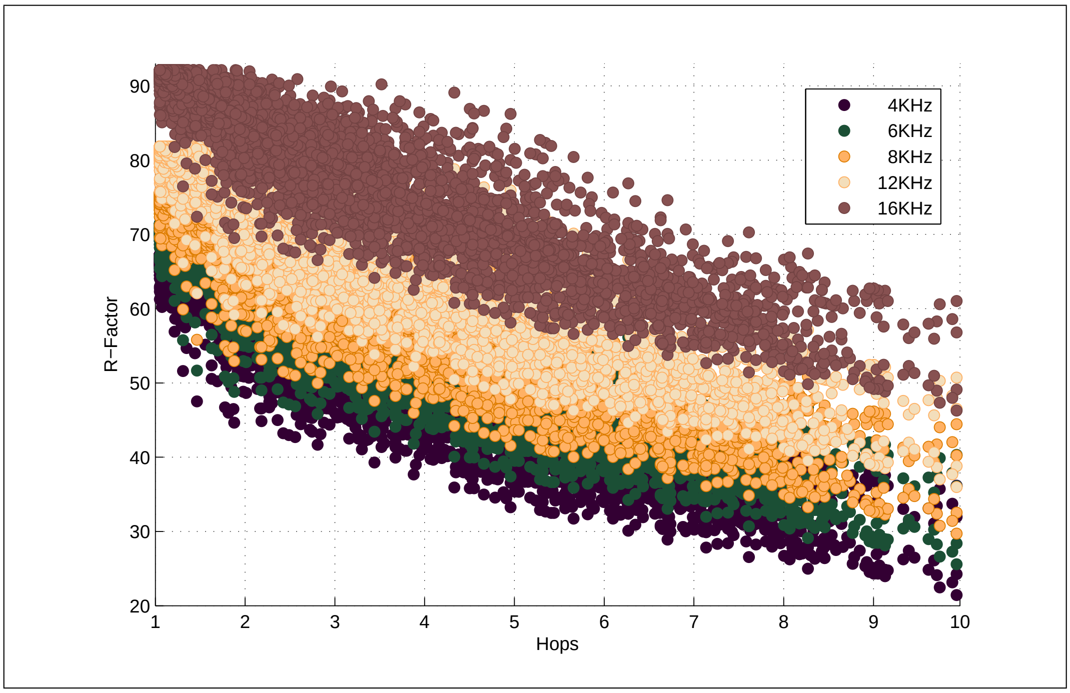


Figure 6.1. R-factor results according to sampling frequencies and network hops

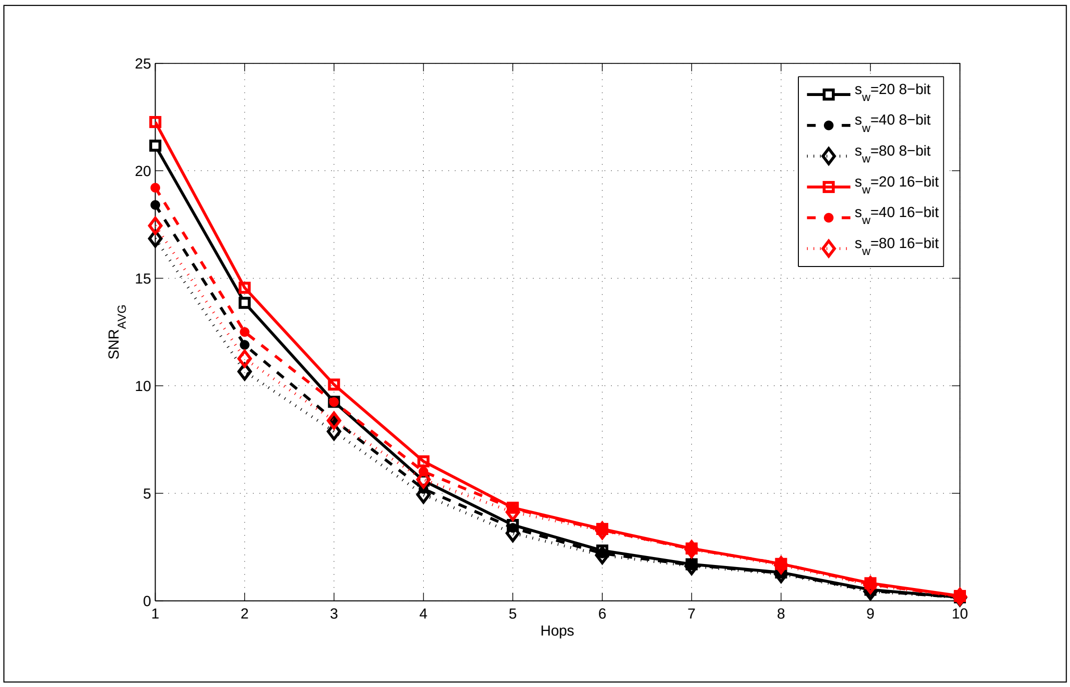


Figure 6.2. SNR values of different segment sizes and bit depths

6.1.3. Discussion

Regardless of PERs, a decrease in sampling frequency directly decreases the R-factor values of a received data. Also, the results show that PER of a data delivery negatively affects the data intelligibility. When the segment size of a transmission as the unique in-network qualification of this study is taken into consideration, we can say that network packet sizes including different voice amplitudes do not differ much from each other. However, when environmental factors are taken into account, the reasonable segment size is discussed as $s_w = 40$ in Section 6.1.4. Hence, we can offer for an optimal, efficient and reliable transmission in terms of intelligibility for a data sampled at 8KHz/8bit and the size of the network packets implemented as 40.

6.2. SERVE EVALUATION

In this section, an evaluation of EC methods over a real experimental setup which is supported with random sampling tests. With the help of a simulation setup, effects of fusion algorithm accommodated on fusion nodes in the system is also investigated. The details and the outcomes of the performance assessment are summarized below.

We establish an experimental setup for the analysis of SERVE with respect to the system model given in Section 5.4.1.

6.2.1. Methodology

All of the files D_i, i=$1, 2 \ldots, 8$ in the voice data set sampled according to different voice characteristics are read and segmented into $s_w = \{20, 40, 80\}$ sized partitions. Amplitude values inside any file D_i, form up the original data A_0, after their serially transport to the source node is accomplished. In each transmission performed, successfully conveyed amplitudes with the lost segments constitute the entire received data A_R at each hop. The packets are transmitted as in the steps explained in Section 5.4.2. EC algorithms are applied on the lost packets according to their offsets.

A similar approach as in Section 6.1 is followed for the quality evaluation of the gathered voices. In the real transmissions, a voice file sampled at 8KHz/8bit is transmitted with $s_w = 40$ and its loss patterns for all transmissions are stored. For each EC methods, 150 real transmissions are performed. In every hop, nodes count the number of lost packets and applies one of the selected EC algorithm. So, new packets with their exact offsets are reconstituted and they are set as modified. In the reception process of the packets at the end points of the network, total number of lost and modified packets are separately identified.

For the fusion nodes, data transported from disjoint paths are merged according to their packet indices. In the fusion process, these packet indices may coincide wit each other. In this situation, one of them is taken into account. In the case of no coincidence, both of the packets are recovered. Hence, the effect of the fusion algorithm over error-resilience is expected to be increased.

6.2.2. Performance Analysis

In the real experiments [84], a total of 6,000 reconstructed voices are stored for further evaluation. On side of the simulation, numerous loss patterns generated for random sampling are projected over the data set read and segmented. For each hop, the voice quality evaluation is objectively performed with the SNR metric explained in Chapter 4. The absolute difference between A_0 and A_R is taken into consideration in SNR calculation.

For the experiments conducted in the real testbed, the correlation between SNR values and PER for each EC techniques is depicted in Figure 6.3. In each test, a voice file sampled at 8KHz/8bit is utilized. The size of the network packets are set as 20. As seen in the graph, SNR and PER values are changing as inversely proportional for all EC methods. When PER increases, SNR values exponentially decrease. Also, the results explicitly indicate that EC_{AVG} has the maximum performance among other techniques. The second successful gain is obtained by EC_{PREV} method. As expected, values for EC_0, namely NC method, are the worst. Unexpectedly, the results for EC_{LERP} are less than the proposed techniques.

6.2.2.1. Random Sampling Analysis

We have conducted a set of Monte Carlo simulations to compare the real testbed results of the EC methods presented in SERVE. Among 1,000 transmissions, to make reliable comparisons, we fed the link qualities in the simulated network with different PERs and HPPRs. Running the simulations 1,000 times, we obtained 1,6 million different transmitted voice data with uniformly distributed HPPRs and PERs. The simulation results of the same techniques are shown in Figure 6.4. Promisingly, the outcomes of the real tests are very similar to the simulation results. In each test of the simulation, all of the data with different characteristics in the voice data set are utilized. The strong correlation between the real results and the simulation results are shown in Figure 6.5. There is only a slight inconsistency between the real and simulated values of EC_{AVG} for low PERs. This is because, the infinite values obtained due to no difference between the original and the distorted signals are discarded in the real evaluation whereas in the simulation infinites are set to a high SNRs.

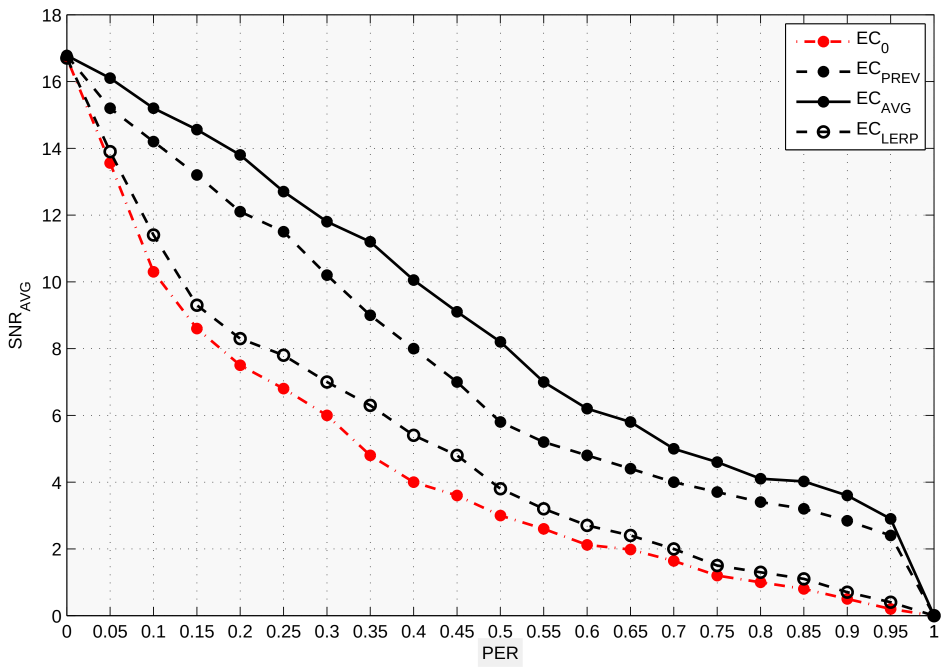


Figure 6.3. Real testbed results for EC methods

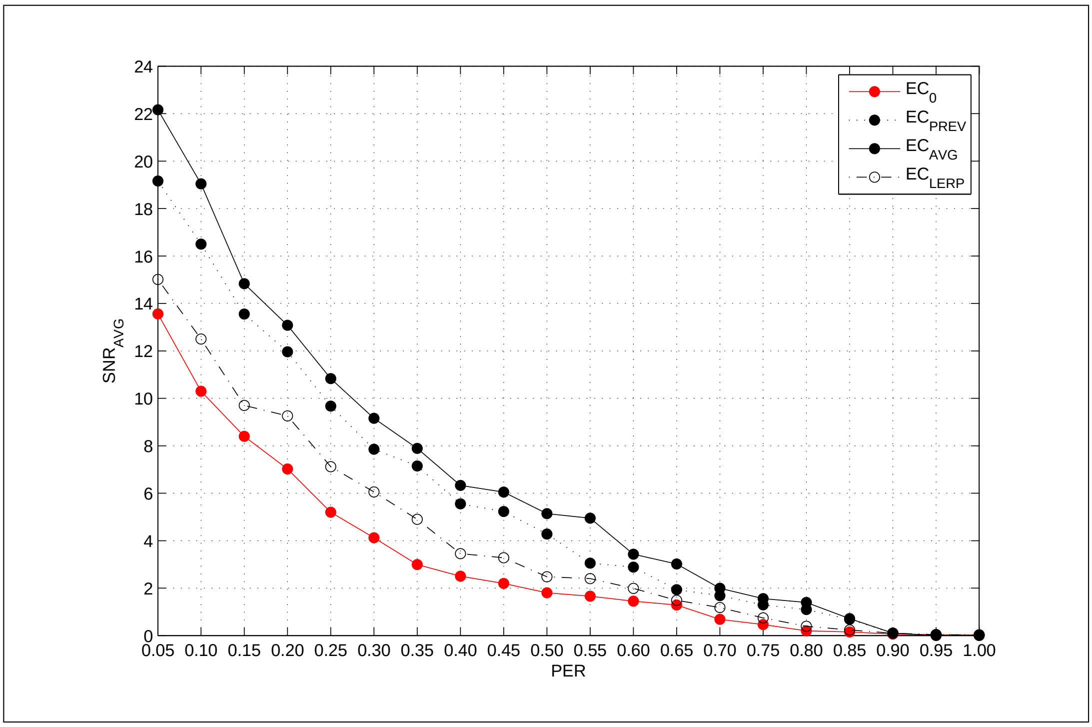


Figure 6.4. Simulation testbed results for EC methods

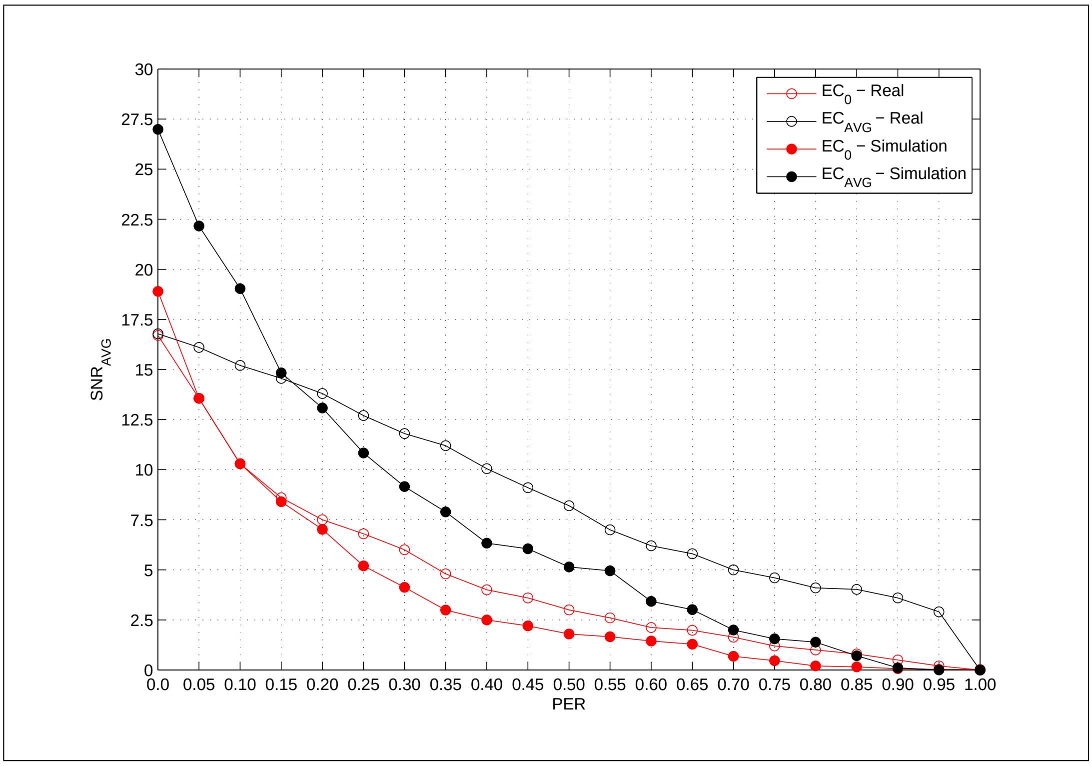


Figure 6.5. Real transmission results vs simulation results

We have gathered 22,800 voice loss patterns via 10 hops. The graph, which shows R-factor and transmission success rate relation in Figure 6.6, includes all the results projected to eight different voice data. Also, all of the sampling frequencies utilized in the study is shown altogether in the graph. The effect of all sample rates over the R-factor metric values are excluded to clearly demonstrate that any EC algorithm will not contribute on the R-factor value. Although we facilitate R-factor to weigh the quality of a voice perceptually, we cannot decide on the effect of an EC technique by this evaluation metric. Because, the formulation does not include impairments or reconstructions performed by any EC algorithm. Regardless of packet size or sample rate of a data transmitted, the R-factor values only depend on the overall transmission success rate. For this reason, we utilize SNR metric to weigh the effects of EC strategies. To make a perceptual voice quality evaluation, we correlate the values gathered from both R-factor and SNR test results.

The reflection of applying EC_{AVG} on the data gathered in comparison to EC_0, or simply NC, can be clearly seen in Figure 6.7 which consists of nearly 18,000 SNR values calculated for all of the segment sizes offered. SNR values generated for both EC_0 and EC_{AVG} are shown in color gradient in which colors get dark when hops increase.

The distinction between values gathered for each network hop can be noticed easily. When data losses are at minimum, SNR values are at the highest when s_w=20 in comparison to other segment sizes. For s_w=40, EC_{AVG} can barely assess the noise ratio value computed when EC_0 applied. For s_w=80, when overall data loss is nearly between 3 and 7 per cent, some values of EC_{AVG} is less than of EC_0. This can be thought as due to big information loss when data segment width gets bigger. Despite the inconsistencies in concealment at the highest transmission rates, an assessment is supplied on the whole data set. When data packet losses increase with regard to data hopping over the hops, SNR values approaches to 1. For all of the segment sizes, values indicate that concealment on the lost segments notably increases the signal quality. When s_w=40 and s_w=80, the effects of amelioration is nearly same. The difference between EC_{AVG} and EC_0 is higher than other segment sizes when s_w=20. Same information for reconstruction schemes applied on the gathered data is also depicted in Figure 6.8 in a different manner. Average SNR values at each hop are listed. The biggest effect of EC applied can be seen when data is transmitted with s_w=20.

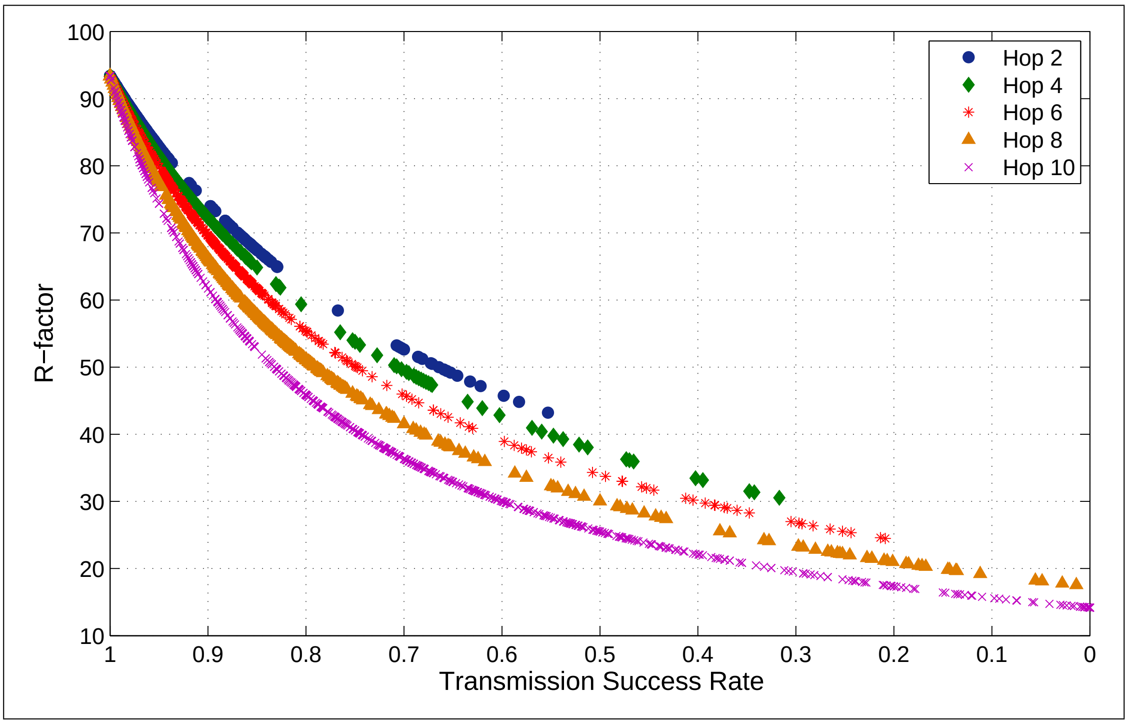


Figure 6.6. Real transmission R-factor results according to PER

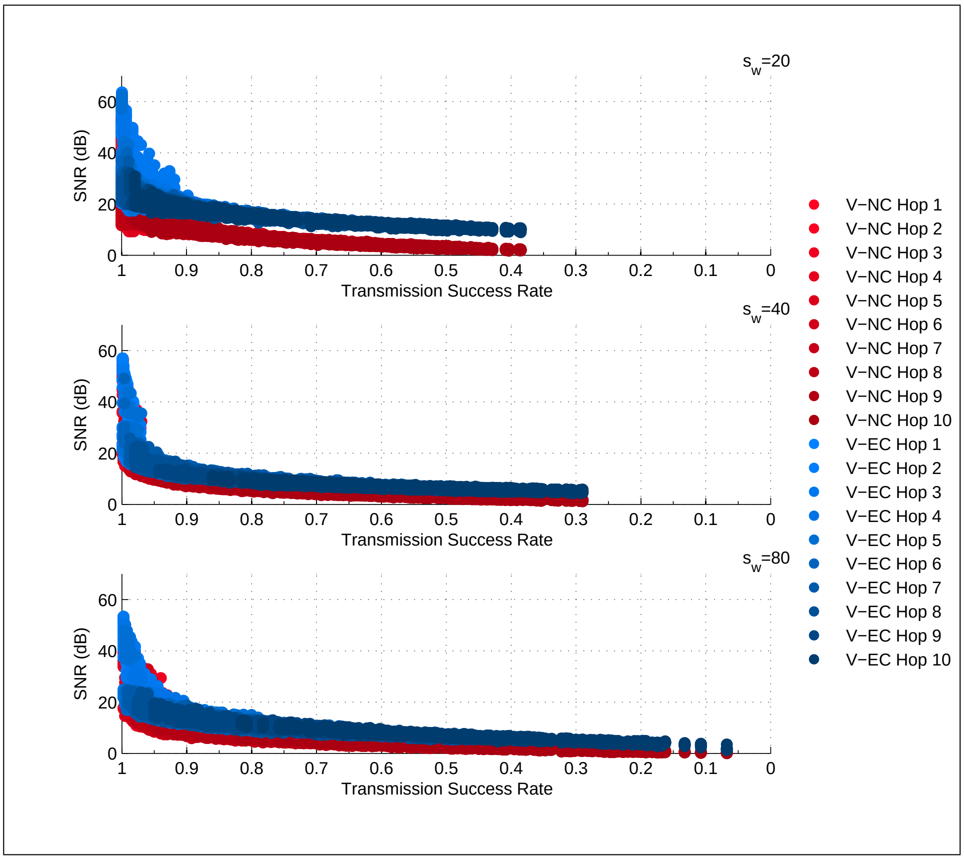


Figure 6.7. Real SNR values of EC_0 and EC_{AVG} with respect to segment size

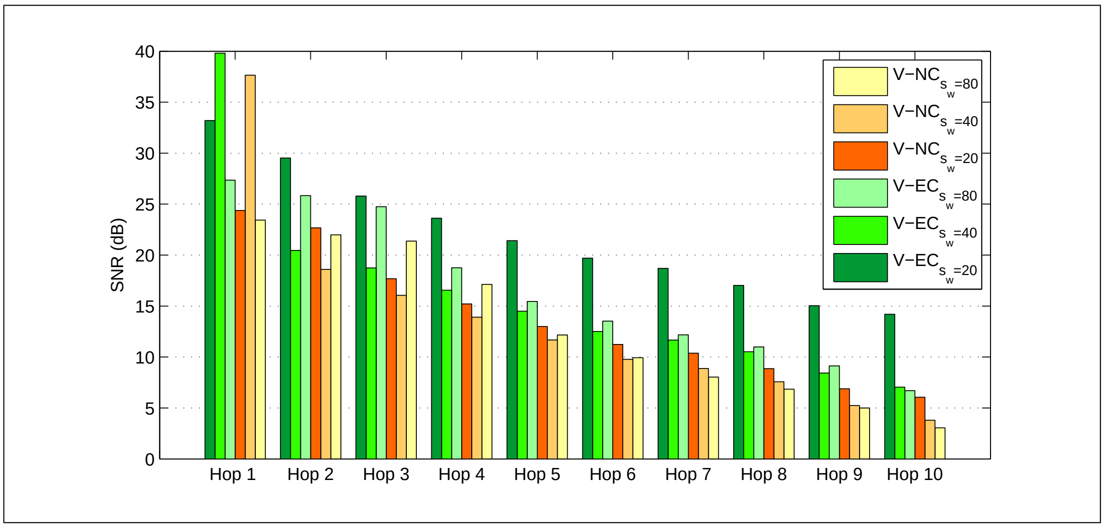


Figure 6.8. Real hop-based SNR comparison of EC$_0$ and EC$_{AVG}$

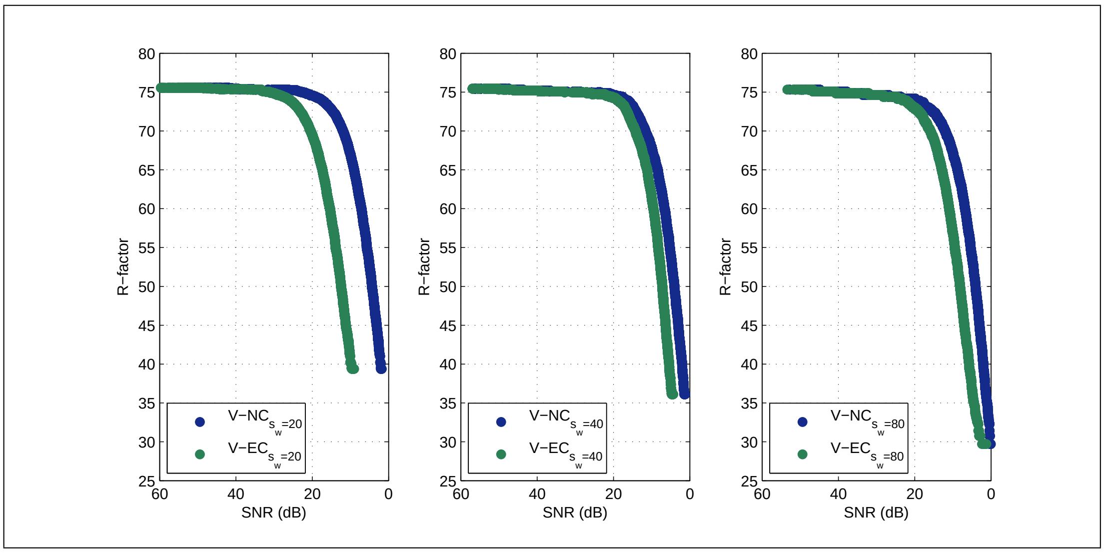


Figure 6.9. SNR and R-factor comparison with regard to segment size

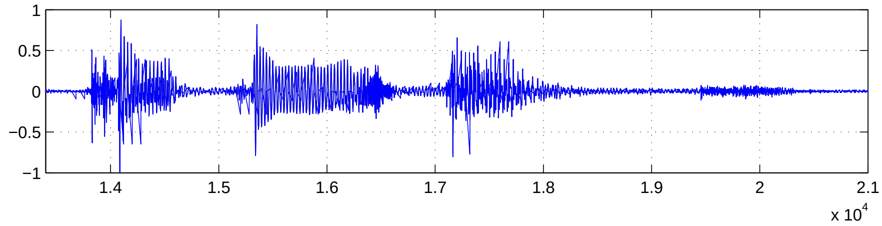
The graph, which shows R-factor, SNR and transmission success rate relation in Figure 6.9, includes all the results projected to eight different voice data with all sampling frequencies and bit depths. The correlation between R-factor and SNR is depicted for EC$_{AVG}$ and EC$_0$ when fs=8KHz. For both 8bit and 16bit, SNR values for concealment algorithm on the overall data shows resemblance. The effect of segmentation width can be smoothly seen on the data set. When s_w=20, increase in EC$_{AVG}$ values are at maximum. Quality metrics for network and data intelligibility—R-factor and SNR visibly relate with each other.

6.2.3. Discussion

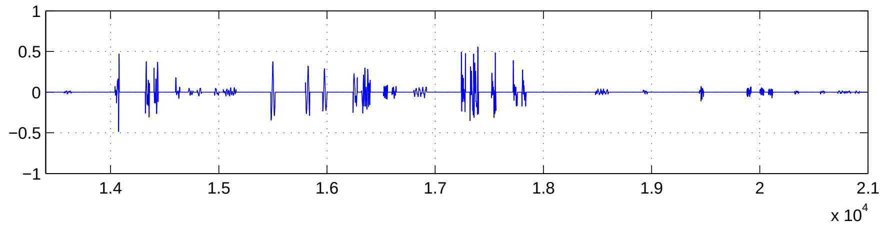
Among the error robustness results of the EC algorithms, the effect of EC$_{AVG}$ is notable. In this encoding scheme, with the utilization of correlated parts in the data, delays caused by retransmissions included in reliability mechanisms such as ARQ are avoided. Moreover, extra bandwidth required for FEC algorithms are also avoided. FEC schemes append extra data to the transmitted packets for error correction. It is necessary to mention that presented EC algorithms can be used in multi-tiered systems in which source nodes have more processing and energy resources. However, by means of the multi-tiered architecture, much simpler intermediate nodes would be used and overall lifetime of the system would be increased. Therefore, the ratio of the source nodes capturing voice and the intermediate nodes is determinant to employ the EC algorithms. In Figure 6.10, an overview of the EC techniques is given with an application over a voice section sampled at 8KHz/8bit.

6.3. PrSERVE EVALUATION

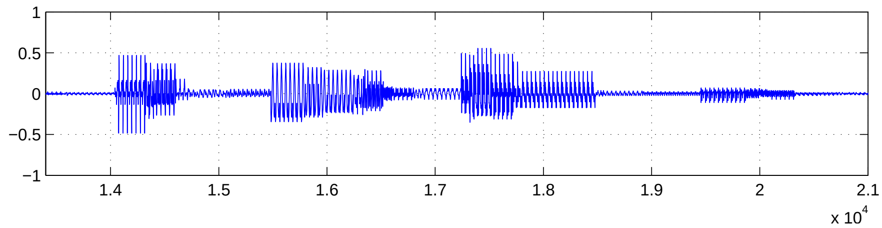
PrSERVE is a good example of a combination of two different coding schemes. This method is applied to the voices by coding them with a priority-based approach and different EC algorithms. Critical parameters in PrSERVE, such as thresholds and fusion, are taken into account in the overall evaluation. Voice signals of the commands or words spoken by a person in a particular activity monitoring application environment need to be rendered to a processable data. When gathering the information, some parts like silence have less importance than the voice signals. In other words, voice data parts containing in an information gathered are valuable in order to assist people needing urgent help. Therefore, their successful transmission to the sink node is more important than transmitting the silence. Since there are two classes of data, i.e. speeches, commands, screams, etc. related to a human voice and silences in the data, for this application, it is suitable to employ two priority levels as "important" and "not-important". The thresholds for different fs, bd and p_w obtained by the priority measure determine the importance level of the packets at the network. The system model of this scheme is exactly same of the simulation designed in Section 5.4.1.



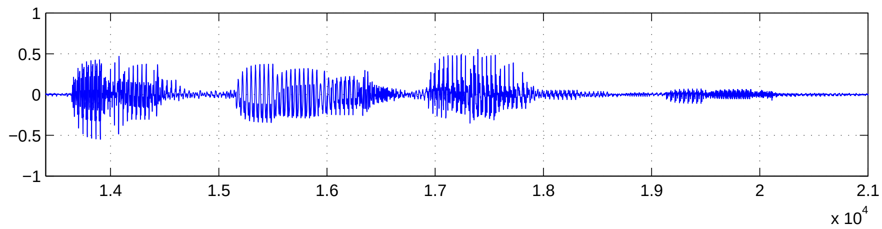
a. A section from original voice



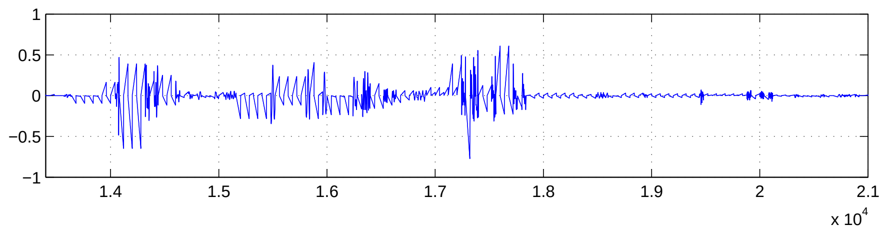
b. The section EC_0 is applied when 30% of the data is lost



c. The section EC_{PREV} is applied when 30% of the data is lost



d. The section EC_{AVG} is applied when 30% of the data is lost



e. The section EC_{LERP} is applied when 30% of the data is lost

Figure 6.10. A section from a reconstructed voice with different EC techniques

6.3.1. Methodology

The files D$_i$, i=$1, 2 \ldots, 8$ in the voice data set with different f_s and bd are read and segmented into $s_w = \{20, 40, 80\}$ sized partitions. In the segmentation phase, the mean values m_{s_i} of the amplitude values inside a partition is determined and compared with the threshold value $T_{f_s,bd}$ determined. The threshold value varies according to the HPPR selected in a unique test. The segments having m_{s_i} greater than $T_{f_s,bd}$ are labeled as highly prioritized. For the remaining segments, vice versa. After the labeling process, the original voice data A is separated into two groups: preferential data A_H and non-preferential data A_L. These data are sent from separate paths; A_H segments are sent over a high quality link whereas segments of A_L are conveyed over a low quality link. In every two hops, these data are concatenated in fusion nodes.

6.3.2. Performance Analysis

A vast number of Monte Carlo simulations are conducted to compare the results of PrSERVE scheme with others. To make reliable comparisons, we fed the link qualities in the simulated network with different PERs and HPPRs. Running the simulations 1,000 times, we obtained 1,6 million different transmitted voice data with uniformly distributed HPPRs and PERs. For each of the fusion hops, the performance of prioritization is evaluated by comparing the results with a single hop network without prioritization. In this approach, total PER in a fusion node for PrSERVE is calculated and used for the single hop transmission. PER calculation for a PrSERVE fusion node is given as

$$PER = HPPR \times PER_{HQL} + (1 - HPPR) \times PER_{LQL} \tag{6.1}$$

where PER$_{HQL}$ and PER$_{LQL}$ are the PERs of the high and low quality path links, respectively.

In Figure 6.11, a comparison between PrSERVE and SERVE is given in terms of NC is applied on the lost segments. The outcome of 10,000 tests are assessed with SNR metric. The pure performance of the prioritization can be clearly seen when HPPR is 25 per cent.

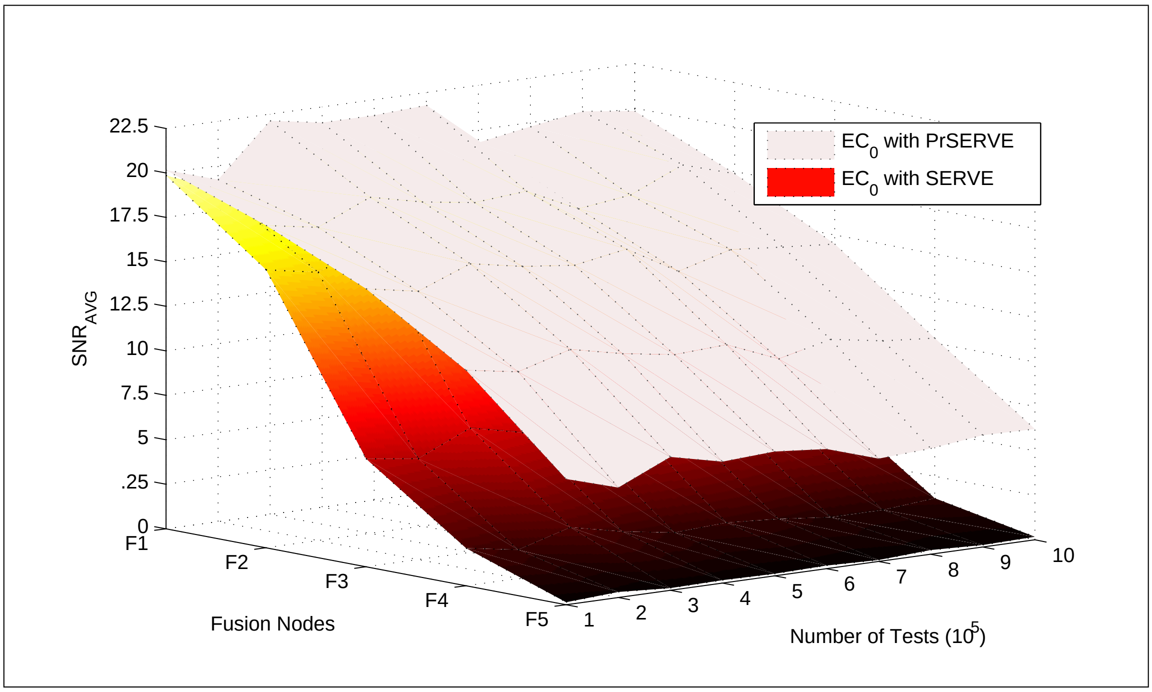


Figure 6.11. Comparison between PrSERVE and SERVE when EC_0 is applied

In the non-prioritized approach, the SNR values dramatically decrease in every fusion node. Meanwhile, for the same PERs, PrSERVE scheme can graciously preserve the quality of the voice.

Figure 6.12 demonstrates the effect of prioritization coupled with EC techniques. Analogically, same effect for each of the EC methods presented for SERVE is assessed and fostered with prioritization. The performance of the PrSERVE scheme can be increased by an increase in HPPR, as shown in Figure 6.13. When HPPR is 50 per cent, a SNR value of nearly 60 is obtained for EC_{AVG} technique. Even for low HPPRs, the SNR values obtained for EC_{AVG} in SERVE are improved in PrSERVE. The overall comparison between SERVE and PrSERVE is given in Figure 6.14. Effectiveness of PrSERVE scheme is explicitly notable for each EC algorithm offered.

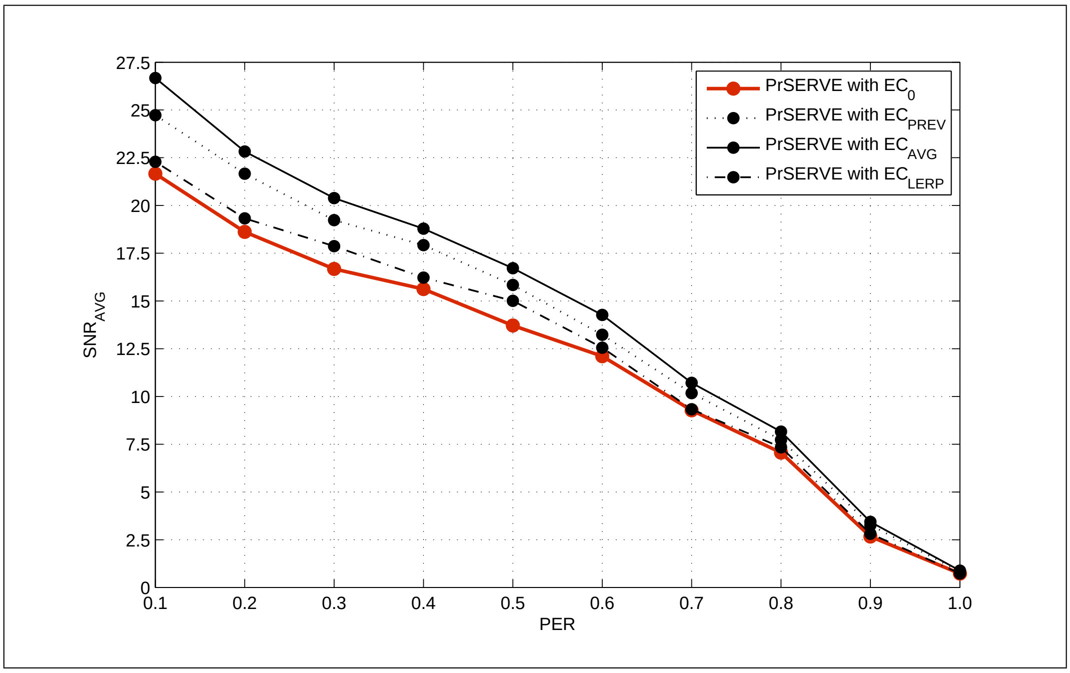


Figure 6.12. Comparison of EC methods in PrSERVE scheme

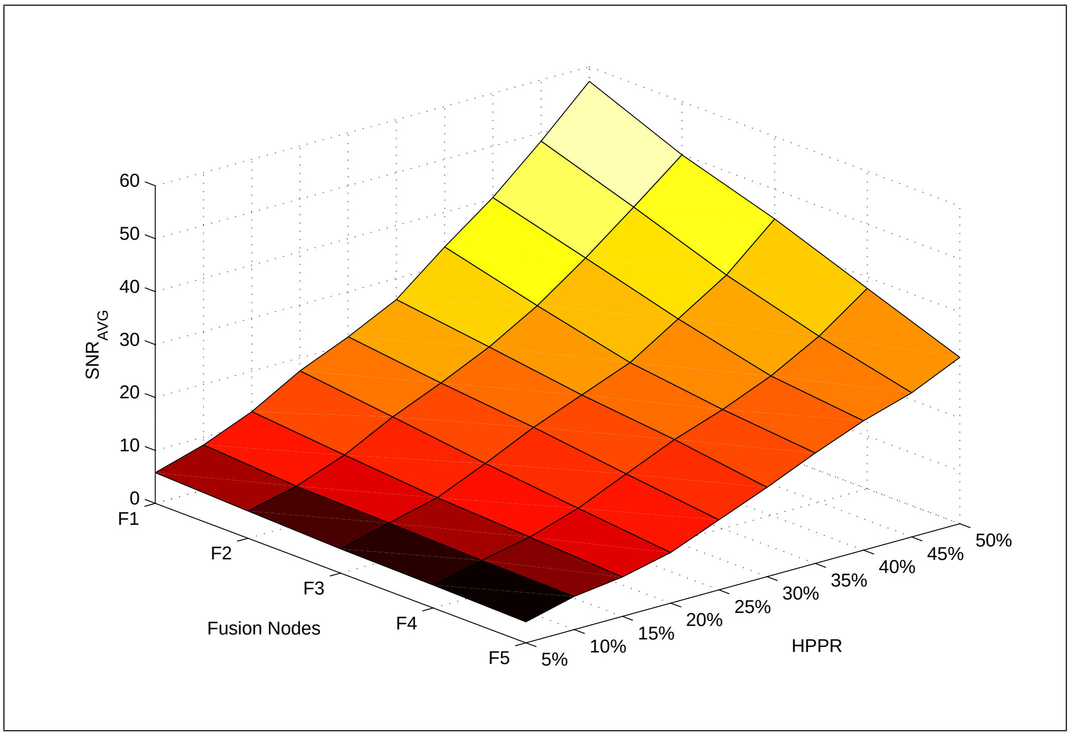


Figure 6.13. Effect of highly prioritized packet rate

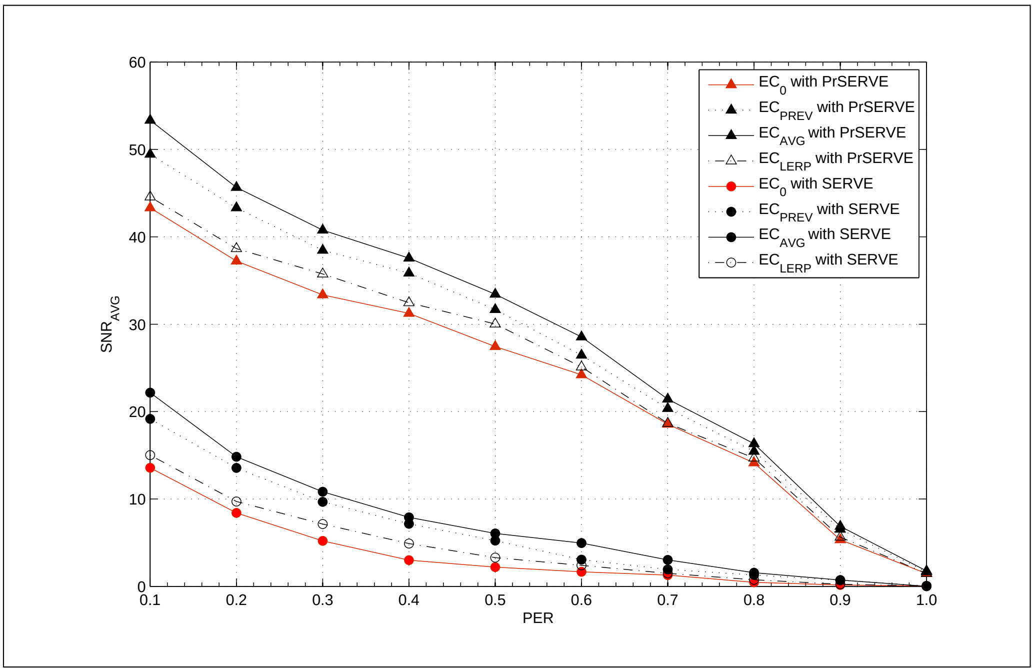


Figure 6.14. Overall comparison between SERVE and PrSERVE

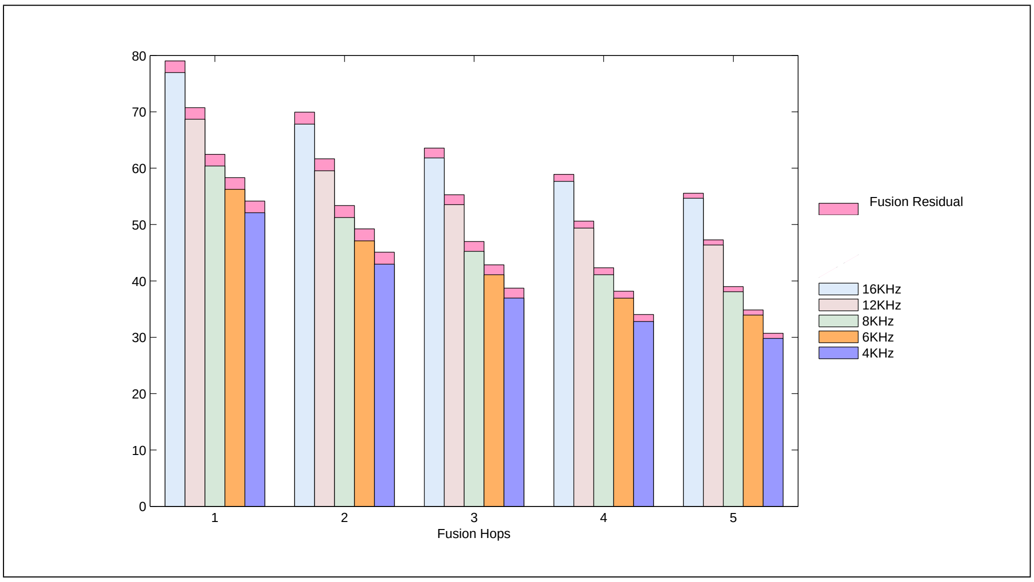


Figure 6.15. Effect of fusion nodes in quality

We also demonstrate the effect of fusion algorithm on the nodes by comparing R-factor values with a transmission the knitting nodes are included. As can be seen in Figure 6.15, the performance of the fusion decreases when the level of fusion in the network increases.

6.3.2.1. Implementation Costs

The implementation costs of the presented reliable transmission techniques are given in Table 6.1. EC techniques and the prioritization on the source 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, the voice data is transmitted in packets that hold twenty amplitude values. The network buffer utilization is at the minimum level for EC_0. Other EC techniques utilize same buffer size when the worst case is applied. According to this, the transmission buffer B in any EC method always filled with its entirety during packet transmission. For B_C, we calculate the costs as if all of the buffer nodes are utilized to reconstruct the packets. These two buffers constitute the send buffers of the implementation. In terms of program size, memory utilization, energy consumption and time, EC techniques do not differ so much from each other. We can clearly say that priority-based approach is quite costly than the EC algorithms since more computation is required for median calculation.

For a prioritization with thresholding, the resolution of a voice data can suffer to obtain a required HPPR. As shown in Figure 6.16, for 8-bit voice data, the priority algorithm presented in Section 3.3.1, cannot handle the amplitude values inside a segment since the resolution is not sufficient for median calculation. However, for 16-bit data, the resolution inside the data can afford to determine a HPPR correctly.

Table 6.1. Implementation Costs

Technique	CPU Cycles	Program Memory (bytes)	Data Memory (bytes)	Network Buffer (bytes)	Time (ms)	Energy (mJ)
EC_0	5720	630	82	2	0.72	2.15
EC_{PREV}	6042	632	82	250	0.76	2.27
EC_{AVG}	6120	632	82	250	0.77	2.30
EC_{LERP}	6204	632	84	250	0.78	2.32
Pri-based	10668	764	266	516	1.34	3.99

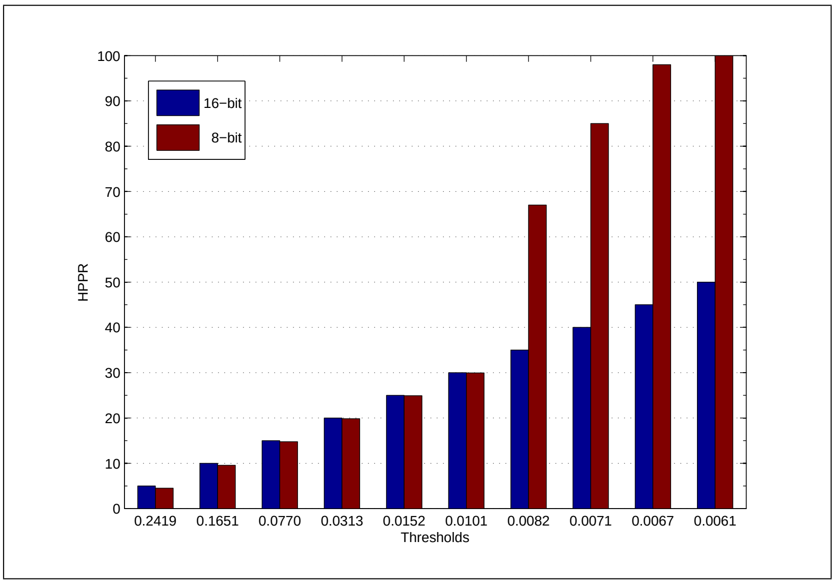
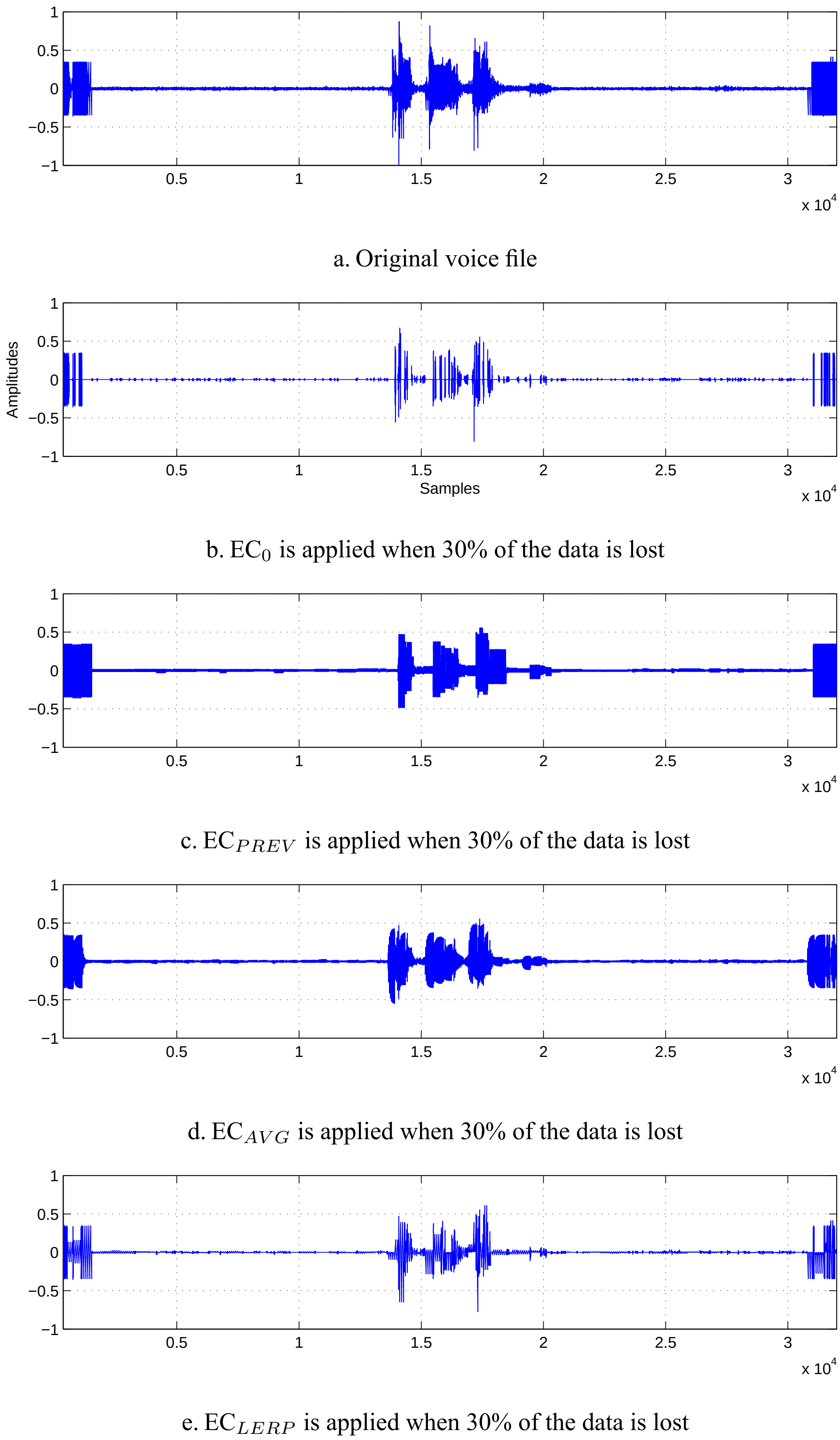


Figure 6.16. HPPR for 8-bit and 16-bit voices vs threshold values

In Figure 6.17, an overview of SERVE when the EC techniques is applied over a voice data sampled at 8KHz/8bit. Figure 6.18 demonstrates three different prioritization level when no concealment is applied on a voice data sampled at 8KHz/16bit. The effect of prioritization on reliability is evident when approximately 70 per cent of the lost voice segments are not reconstructed. When reconstruction is coupled with prioritization, the promising results can be seen in Figure 6.19. When HPPR is at a low rate, the high prioritized segments can be utilized in reconstruction and suppresses the bursty packet losses even at the last fusion point of the network.

6.3.3. Discussion

In terms of implementation costs, we cannot lead any notable interpretation for EC techniques. Only the network buffer utilized in EC_0 makes a difference among others. However, this error concealment has the least effect on the voice quality, as the experimental results indicate. For voice quality assessment, a quite performance enhancement is gathered in PrSERVE when in comparison to SERVE. However, for HPPRs greater than 30 per cent, system cannot handle priority determination for 8-bit data resolutions.



a. Original voice file

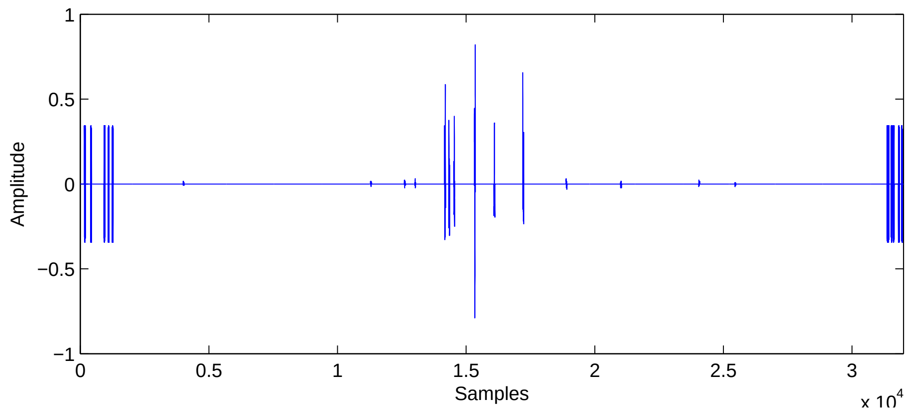
b. EC_0 is applied when 30% of the data is lost

c. EC_{PREV} is applied when 30% of the data is lost

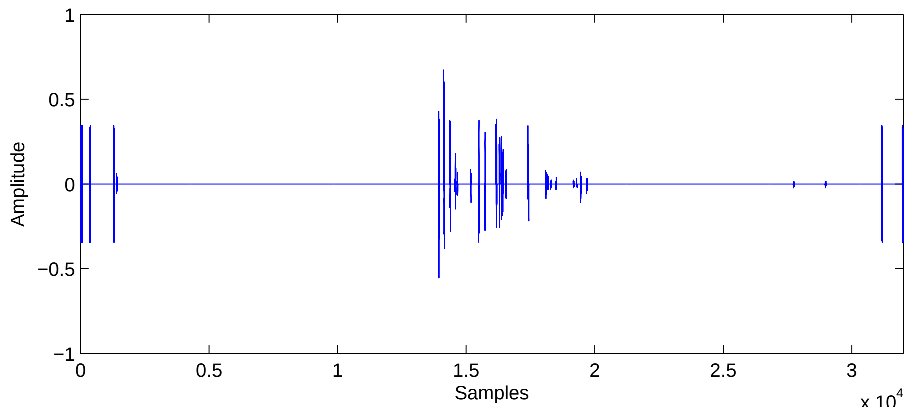
d. EC_{AVG} is applied when 30% of the data is lost

e. EC_{LERP} is applied when 30% of the data is lost

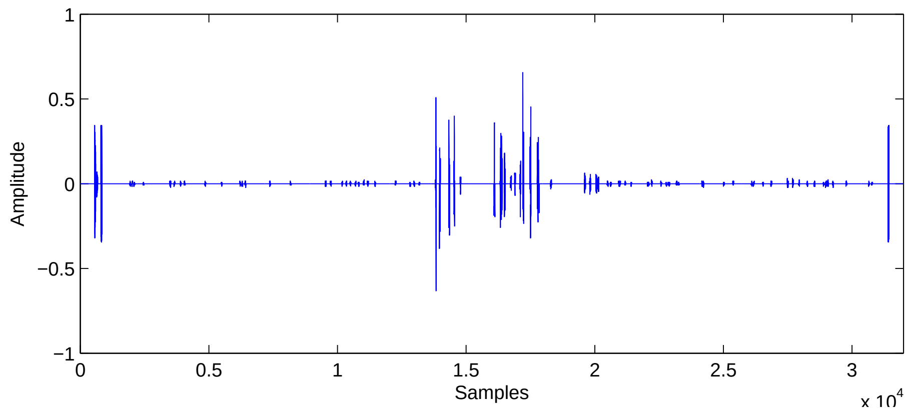
Figure 6.17. Reconstructed versions of the same voice with different EC techniques



a. HPPR=10% at fusion 5

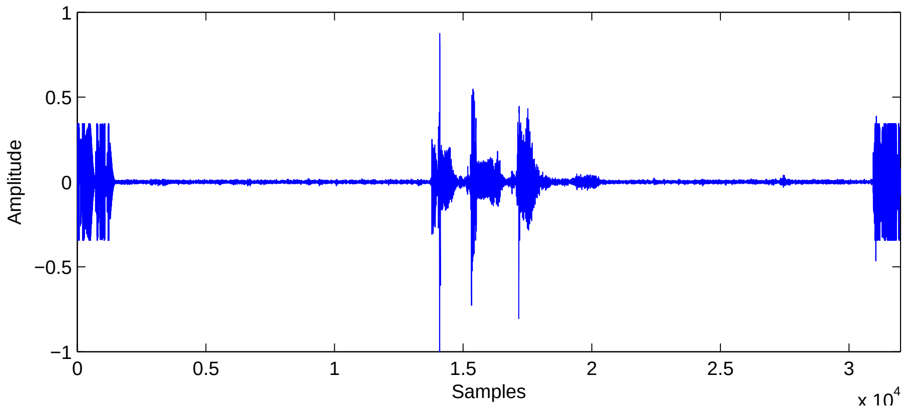


b. HPPR=30% at fusion 5

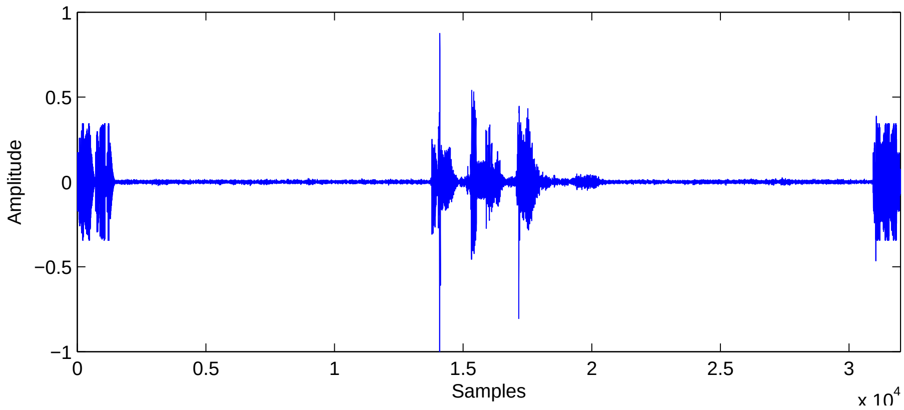


c. HPPR=50% at fusion 5

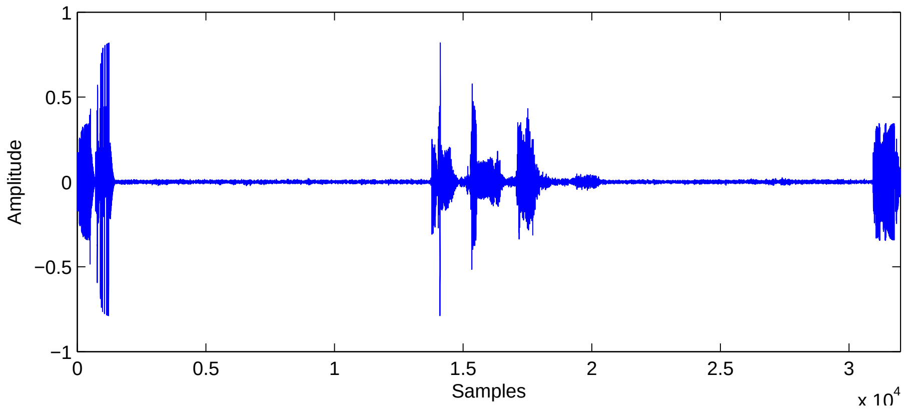
Figure 6.18. Effect of HPPR when NC is applied on the same voice



a. HPPR=5% at fusion 1



b. HPPR=5% at fusion 3



c. HPPR=5% at fusion 5

Figure 6.19. Effect of reconstruction over bursty losses when HPPR is low

7. CONCLUSION AND FUTURE WORK

The capability and the accuracy of the existing applications in WSNs are highly improved with multimedia utilization along new data handling approaches. The number of applications related with audio/voice transmission in WMSNs is promisingly increasing. Given the requirements and challenges in the literature, this thesis mainly focuses on voice coding and transmission in WMSNs. With the consideration of intrinsic system necessities and multi-dimensional characteristics of the voice data, several proposals are presented with an extensive performance assessments.

With a bottom-up approach, a set of sequenced encoding schemes, NAIVE and SERVE, are presented in which robust and lightweight transmission environment requirements, reliable and straightforward data handling techniques are taken into consideration, respectively. An all-embracing priority-based and error-resilient encoding scheme, PrSERVE, is proposed for an activity monitoring scenario. In this scheme, the results of preferential data preservation are analyzed with regard to the performance outcomes of the subset schemes.

The results of NAIVE show that a simple buffering management accommodated on the sensor nodes can yield in a robust data transmission. In this scheme, several voice properties and in-network qualifications are analyzed and evaluated. One step forward, the robustness of the network is coupled with data reliability during transmission in SERVE. By the implementation of simple reconstruction strategies, the effects of error concealment techniques are assessed. The results indicate that the need for a different voice quality improvement technique is a must since the reconstruction schemes presented do not excessively contribute on voice quality when the packet loss rates dramatically increases. So, we presented a multi-level priority-based encoding scheme which can be supported with a disjoint multi-path in order to preserve the preferential information inside the data. By utilizing a two-level prioritization on the data segments, a significant improvement is obtained in voice quality. A general voice transmission framework is presented to embrace the engaged encoding schemes. This framework addresses several problems faced in multimedia transmission such as, delay, delay-jitter, additional bandwidth, congestions caused by streaming data, and QoS support. The transmission mechanism included in it, which is based on the bursty and regulated transmission of voice data segments with prioritization via multiple paths, makes it possible to address the mentioned problems. Another advantage of this framework is scalability in terms of available memory and bandwidth.

Aside from the encoding and transmission schemes, the multi-dimensional properties of the voice data is assessed in so far as the WSN capabilities permit. The common conventional features of the voice data—sampling frequency and bit depth, are evaluated in a large set of controlled experiments. By this means, an optimal data extent is offered in order to be utilized in the framework. Moreover, as the network property, several packet widths holding voice data segments inside are investigated for an efficient data transfer.

Data being dispatched in the lossy multi-hop sensor network environment is qualified with objective evaluation measures, a modified version of R-factor and SNR. With a simple approach, we have included sampling frequency and overall packet loss rates in this formulations in order to perceptually assess the voice intelligibility.

In pursuit of bringing these studies into action, a real wireless voice transmission testbed is established to determine data requirements during voice capturing and to examine the effect of real erratic channels on voice quality. Thousands of voice transmissions are realized. In the testbed, an orthogonal backchannel mechanism is employed for gathering the data loss patterns occurred in each hop. By projecting the patterns to different group of voice data, the accuracy of the results is justified. The test results are also stored in a text-based database to make use of for future experiments. The variety of the outcomes are supported with a large set of simulation tests in order to disclose quality gradients of the streaming data. The basic characteristics of voice data commonly used in the literature—sampling frequency and bit depth are essayed with random sampling and sensitivity analysis approaches. The characteristics of the network are analyzed with a vast number of time-varying tests. Implementation costs of the reliable data transmission approaches and error correction methods are evaluated on a real implementation.

The result of the extensive tests revealed that the EC algorithms are robust to packet losses in WSN. Thus, it removes the delay and additional bandwidth requirements introduced by the reliability and error correction mechanisms i.e. ARQ and FEC. Another advantage of the algorithm is that it does not require additional processes and capabilities at the intermediate nodes. However, bursty losses cannot be handled by these methods. For this reason, hybrid utilization of EC with prioritization is also promising. Since, conceptually, both of them are related with different aspects of reconstruction of voice data. The result of their synergy is studied in detail by using previously stored real testbed results. The results indicate that a remarkable performance gain is attained by employing PrSERVE scheme. Our results show the effectiveness of the architecture designed for a priority-based transmission in a WMSN.

On the back of these contributions, following objectives are accomplished in this thesis. A lightweight coding scheme, which significantly decreases the energy consumption of the network by minimizing the required bandwidth, is devised. The need for computationally more capable nodes at the source is eliminated. A simple database is created including thousands of real voice transmission loss patterns which can be used in future works, instead of unrealistic simulations. Additionally, a voice transmission framework considering the shortcomings and needs of WMSN is devised.

In a limited number of works in the literature dealing with audio/voice utilization in WMSNs, the proposed approaches for error concealment and prioritization are investigated in an acute angle. Most of the studies on multimedia streaming in sensor networks consider video and still-image content forms and by-pass audio/voice utilization. With its computational simplicity, PrSERVE scheme can be merged with still-image streaming and pave the way for simplistic video encoding schemes in WMSNs. On the other hand, there are several audio/voice quality assessment tools in the sensor network field. By improving the considerations of our R-Factor utilization, the perceptual evaluation of the streams may be yield in more precise objectiveness. As a result, a comprehensive architecture for any application domain in WMSNs can be implemented.

REFERENCES

1. Akyildiz, I., W. Su, Y. Sankarasubramaniam and E. Cayirci, “Wireless sensor networks: a survey”, *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.

2. Yick, J., B. Mukherjee and D. Ghosal, “Wireless sensor network survey”, *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

3. Akyildiz, I., T. Melodia and K. Chowdhury, “A survey on wireless multimedia sensor networks”, *Computer Networks*, vol. 51, no. 4, pp. 921–960, 2007.

4. Gurses, E. and O. Akan, “Multimedia communication in wireless sensor networks”, in *Annales des Télécommunications*, vol. 60, p. 872, Presses Polytechniques Romandes, 2005.

5. Culurciello, E. and A. Andreou, “CMOS image sensors for sensor networks”, *Analog Integrated Circuits and Signal Processing*, vol. 49, no. 1, pp. 39–51, 2006.

6. Mahdi, a. and D. Picovici, “Advances in voice quality measurement in modern telecommunications”, *Digital Signal Processing*, vol. 19, no. 1, pp. 79–103, 2009.

7. Wang, H., D. Peng, W. Wang, H. Sharif and H.-H. Chen, “Energy-Aware Adaptive Watermarking for Real-Time Image Delivery in Wireless Sensor Networks”, *2008 IEEE International Conference on Communications*, , no. 1, pp. 1479–1483, 2008.

8. Azimi-Sadjadi, M. R., G. Kiss, B. Feher, S. Srinivasan and a. Ledeczi, “Acoustic source localization with high performance sensor nodes”, *Proceedings of SPIE*, vol. 6562, pp. 65620Y–65620Y–10, 2007.

9. Berisha, V. and a. Spanias, “Real-Time Implementation of a Distributed Voice Activity Detector”, *Fourth IEEE Workshop on Sensor Array and Multichannel Processing, 2006.*, pp. 659–662, 2006.

10. Facchinetti, T., M. Ghibaudi, E. Goldoni and A. Savioli, “Real-time voice streaming over IEEE 802.15. 4”, in *Computers and Communications (ISCC), 2010 IEEE Symposium on*, pp. 985–990, IEEE, 2010.

11. Alesii, R., F. Graziosi, L. Pomante and C. Rinaldi, “Exploiting WSN for Audio Surveillance Applications : the VoWSN Approach”, pp. 520–524, 2008.

12. Petracca, M., 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*, pp. 19–24, 2009.

13. Bele, I. V., “Dimensionality in voice quality.”, *Journal of voice : official journal of the Voice Foundation*, vol. 21, no. 3, pp. 257–72, 2007.

14. Xu, J., K. Li, Y. Shen, G. Min and W. Qu, “Adaptive Energy-Efficient Packet Transmission for Voice Delivering in Wireless Sensor Networks”, *2009 Sixth IFIP International Conference on Network and Parallel Computing*, pp. 86–92, 2009.

15. Wang, H., D. Peng, W. Wang, H. Sharif, J. Wegiel, D. Nguyen, R. Bowne and C. Backhaus, “Artificial Immune System based image pattern recognition in energy efficient Wireless Multimedia Sensor Networks”, in *MILCOM 2008 - 2008 IEEE Military Communications Conference*, pp. 1–7, IEEE, 2008, ISBN 978-1-4244-2676-8.

16. Mangharam, R., A. Rowe, R. Rajkumar and R. Suzuki, “Voice over Sensor Networks”, *2006 27th IEEE International Real-Time Systems Symposium (RTSS’06)*, pp. 291–302, 2006.

17. Palafox, L. E. and J. A. Garcia-Macias, “Wireless Sensor Networks for Voice Capture in Ubiquitous Home Environments”, *2009 4th International Symposium on Wireless Pervasive Computing*, pp. 1–5, 2009.

18. Carvalho, L., E. Mota, R. Aguiar, a.F. Lima, J. de Souza and a. Barreto, “An E-Model Implementation for Speech Quality Evaluation in VoIP Systems”, *10th IEEE Symposium*

on Computers and Communications (ISCC'05), pp. 933–938, 2005.

19. Li, X., D. Tao, X. Gao and W. Lu, "A natural image quality evaluation metric", *Signal Processing*, vol. 89, no. 4, pp. 548–555, 2009.

20. Chen, B., "Audio Recognition with Distributed Wireless Sensor Networks by", *Computer*, 2010.

21. Chiasserini, C.-F. and E. Magli, "Energy consumption and image quality in wireless video-surveillance networks", in *Personal, Indoor and Mobile Radio Communications, 2002. The 13th IEEE International Symposium on*, vol. 5, pp. 2357 – 2361 vol.5, 2002.

22. Zhu, Q. F. and Y. Wang, *Error Control and Concealment for Video Communication*, p. 163, Marcel Dekker, Inc., New York, 1999, ISBN 9780824719289.

23. Akyildiz, I., T. Melodia and K. Chowdhury, "Wireless Multimedia Sensor Networks: Applications and Testbeds", *Proceedings of the IEEE*, vol. 96, no. 10, pp. 1588–1605, 2008.

24. Boluk, P. S., S. Baydere and A. E. Harmanci, "Robust Image Transmission Over Wireless Sensor Networks", *Mob. Netw. Appl.*, vol. 16, pp. 149–170, 2011.

25. Irgan, K., C. Unsalan and S. Baydere, "Priority Encoding of Image Data in Wireless Multimedia Sensor Networks for Border Suveillance", in *Proceedings of the 25th International Symposium on Computer and Information Sciences ISCIS*, Springer, LNEE, 2010.

26. Boluk, P., S. Baydere and A. Harmanci, "Perceptual quality-based image communication service framework for wireless sensor networks", *Wireless Communications and Mobile Computing*, 2011.

27. Polastre, J., R. Szewczyk and D. Culler, "Telos: enabling ultra-low power wireless research", in *IPSN 2005. Fourth International Symposium on Information Processing*

in Sensor Networks, 2005., vol. 00, pp. 364–369, IEEE, 2005.

28. Turkes, O. 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)*, pp. 317 –322, IEEE, 2011.

29. Misra, S., M. Reisslein and G. Xue, “A survey of multimedia streaming in wireless sensor networks”, *IEEE Communications Surveys & Tutorials*, vol. 10, no. 4, pp. 18–39, 2008.

30. Shu, L., Y. Zhang, Z. Zhou, M. Hauswirth, Z. Yu and G. Hynes, “Transmitting and Gathering Streaming Data in Wireless Multimedia Sensor Networks Within Expected Network Lifetime”, *Mobile Networks and Applications*, pp. 306–322, 2008.

31. Rong-lin, H., Y. Jian-rong, G. Xia-jun, G. Xiang-ping and C. Li-Qing, “The Research and Design on TDD Voice WSN”, in *Multimedia Technology (ICMT), 2010 International Conference on*, pp. 1–4, IEEE, 2010.

32. Jurdak, R., A. G. Ruzzelli, G. M. P. O. Hare and C. V. Lopes, “Reliable Symbol Synchronization in Software-Driven Acoustic Sensor Networks”, *Communications Society*, , no. 1, pp. 823–827, 2007.

33. Berisha, V., H. Kwon and A. Spanias, “Real-Time Acoustic Monitoring Using Wireless Sensor Motes beaior related Information sreid ggndertand ang acousticvue tigo speriovllan vss albciier”, *Analysis*, pp. 847–850, 2006.

34. Bele, I. V., “Reliability in perceptual analysis of voice quality.”, *Journal of voice : official journal of the Voice Foundation*, vol. 19, no. 4, pp. 555–73, 2005.

35. Zhang, J., G. Zhou, S. H. Son and J. A. Stankovic, “Ears on the Ground : An Acoustic Streaming Service in Wireless Sensor Networks”, *Digital Signal Processing*, pp. 3–4.

36. Pazarloglou, A., S. George, R. Stoleru and R. Gutierrez-Osuna, “Demo abstract: Signal reconstruction with subnyquist sampling using wireless sensor networks”, in *Information Processing in Sensor Networks, 2009. IPSN 2009. International Conference on*, pp. 423–424, IEEE, 2009.

37. Mangharam, R., A. Rowe, R. Rajkumar and R. Suzuki, “Voice over Sensor Networks”, *2006 27th IEEE International Real-Time Systems Symposium (RTSS’06)*, pp. 291–302, 2006.

38. Pazarloglou, A., R. Stoleru and R. Gutierrez-Osuna, “High-Resolution Speech Signal Reconstruction in Wireless Sensor Networks”, *2009 6th IEEE Consumer Communications and Networking Conference*, vol. 1, no. Figure 2, pp. 1–5, 2009.

39. Brunelli, D. and L. Teodorani, “Improving audio streaming over multi-hop ZigBee networks”, *2008 IEEE Symposium on Computers and Communications*, pp. 31–36, 2008.

40. Gurses, E. and O. Akan, “Multimedia communication in wireless sensor networks”, in *Annales des télécommunications*, vol. 60, p. 872, PRESSES POLYTECHNIQUES ROMANDES, 2005.

41. Rajendran, V., K. Obraczka and J. J. Garcia-Luna-Aceves, “Energy-efficient collision-free medium access control for wireless sensor networks”, *Proceedings of the first international conference on Embedded networked sensor systems - SenSys ’03*, p. 181, 2003.

42. Van Hoesel, L. and P. Havinga, “A lightweight medium access protocol (LMAC) for wireless sensor networks”, in *Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004. Proceedings of the 2004*, vol. 12, pp. 205–208, Citeseer, 2004, ISBN 4907764219.

43. Abeysekera, S. S., K. P. Padhi and S. Member, “An Investigation of Window Effects on the Frequency Estimation Using the Phase Vocoder”, *Language*, vol. 14, no. 4, pp.

1432–1439, 2006.

44. Körber, H.-j., H. Wattar and G. Scholl, “Modular Wireless Real-Time Sensor / Actuator Network for Factory Automation Applications”, vol. 3, no. 2, pp. 111–119, 2007.

45. Zhan, Y., S. Miura, J. Nishimura and T. Kuroda, “Human Activity Recognition from Environmental Background Sounds for Wireless Sensor Networks”, *2007 IEEE International Conference on Networking, Sensing and Control*, pp. 307–312, 2007.

46. Sheybani, E. and G. Javidi, “Dimensionality Reduction and Noise Removal in Wireless Sensor Network Datasets”, *2009 Second International Conference on Computer and Electrical Engineering*, pp. 674–677, 2009.

47. Member, S., “Performance of Transport Protocols for Multimedia Communications in Wireless Sensor Networks ÃĆÂÍ”, *October*, vol. 11, no. 10, pp. 826–828, 2007.

48. Lu, S., V. Bharghavan and R. Srikant, “Fair scheduling in wireless packet networks”, *IEEE/ACM Transactions on Networking (TON)*, vol. 7, no. 4, pp. 473–489, 1999.

49. Biswas, R., K. Chowdhury and D. Agrawal, “Data-centric attribute allocation and retrieval (DCAAR) scheme for wireless sensor networks”, *IEEE International Conference on Mobile Adhoc and Sensor Systems Conference, 2005.*, pp. 422–429, 2005.

50. Perillo, M., “Sensor management policies to provide application QoS”, *Ad Hoc Networks*, vol. 1, no. 2-3, pp. 235–246, 2003.

51. Boulis, A. and M. Srivastava, “Node-level energy management for sensor networks in the presence of multiple applications”, *Wirel. Netw.*, vol. 10, no. 6, pp. 737–746, 2004.

52. Dasgupta, K. and M. Kukreja, “Topology-aware placement and role assignment for energy-efficient information gathering in sensor networks”, *Proceedings of the Eighth IEEE Symposium on Computers and Communications ISCC*, pp. 341–348, 2003.

53. Floréen, P., P. Kaski, J. Kohonen and P. Orponen, "Exact and approximate balanced data gathering in energy-constrained sensor networks", *Theoretical computer science*, vol. 344, no. 1, pp. 30–46, 2005.

54. Sadagopan, N. and B. Krishnamachari, "Maximizing data extraction in energy-limited sensor networks", in *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, vol. 3, pp. 1717–1727, IEEE, 2004.

55. Ordónez, F. and B. Krishnamachari, "Optimal information extraction in energy-limited wireless sensor networks", *Selected Areas in Communications, IEEE Journal on*, vol. 22, no. 6, pp. 1121–1129, 2004.

56. Politis, I., M. Tsagkaropoulos, T. Dagiuklas and S. Kotsopoulos, "Power Efficient Video Multipath Transmission over Wireless Multimedia Sensor Networks", *Mobile Networks and Applications*, pp. 274–284, 2008.

57. Maimour, M., C. Pham and J. Amelot, "Load repartition for congestion control in multimedia wireless sensor networks with multipath routing", in *Wireless Pervasive Computing, 2008. ISWPC 2008. 3rd International Symposium on*, pp. 11 –15, 2008.

58. Wu, H. and a. Abouzeid, "Error resilient image transport in wireless sensor networks", *Computer Networks*, vol. 50, no. 15, pp. 2873–2887, 2006.

59. Felemban, E. and E. Ekici, "MMSPEED: multipath Multi-SPEED protocol for QoS guarantee of reliability and. Timeliness in wireless sensor networks", *IEEE Transactions on Mobile Computing*, vol. 5, no. 6, pp. 738–754, 2006.

60. Albanese, A., J. Blomer, J. Edmonds, M. Luby and M. Sudan, "Priority encoding transmission", *Information Theory, IEEE Transactions on*, vol. 42, no. 6, pp. 1737 –1744, 1996.

61. Zhang, L., M. Hauswirth, L. Shu, Z. Zhou, V. Reynolds and G. Han, "Multi-priority Multi-path Selection for Video Streaming in Wireless Multimedia Sensor Networks",

in *Proceedings of the 5th international conference on Ubiquitous Intelligence and Computing*, pp. 439–452, Springer-Verlag, Berlin, Heidelberg, 2008.

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

63. Bhute, V. 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*, pp. 185–188, IEEE, 2008.

64. Tatlas, N., A. Floros and T. Zarouchas, "An Error Concealment Technique for Wireless Digital Audio Delivery", *Networks and Digital*, pp. 181–184, 2006.

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

66. Zorzi, M., "Performance of FEC and ARQ error control in bursty channels underdelay constraints", in *48th IEEE Vehicular Technology Conference, 1998. VTC 98*, vol. 2, 1998.

67. Sarisaray, P., 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*, pp. 73–86, 2007.

68. Shannon, C. E., "The mathematical theory of communication. 1963.", *M.D. computing : computers in medical practice*, vol. 14, no. 4, pp. 306–17, 1948.

69. Danskin, J. and G. Davis, "Fast lossy Internet image transmission", *of the Third ACM Conference on*, pp. 321–332, 1995.

70. Munkong, R., "Auditory perception and cognition", *IEEE Signal Processing Magazine*,

vol. 25, no. 3, pp. 98–117, 2008.

71. International Telecommunications Union, *ITU-T Recommendation G.107*, http://www.itu.int/itudocr/itu-t/aap/sg12aap/history/g107/g107ww9.doc [retrieved 25 April 2011].

72. Ries, M., R. Puglia, T. Tebaldi, O. Nemethova and M. Rupp, "Audiovisual quality estimation for mobile streaming services", in *Wireless Communication Systems, 2005. 2nd International Symposium on*, pp. 173–177, IEEE, 2005.

73. Shao, Y. and C.-H. Chang, "A generalized time-frequency subtraction method for robust speech enhancement based on wavelet filter banks modeling of human auditory system.", *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society*, vol. 37, no. 4, pp. 877–89, 2007.

74. Dutta, P., M. Grimmer, A. Arora, S. Bibyk and D. Culler, "Design of a wireless sensor network platform for detecting rare, random, and ephemeral events", in *Proceedings of the 4th international symposium on Information processing in sensor networks*, p. 70, IEEE Press, 2005.

75. Lorincz, K., 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*, vol. 3, no. 4, pp. 16–23, 2004.

76. Turkes, O. and S. Baydere, "Voice Quality Improvement with Error Concealment in Audio Sensor Networks", *Proceedings of 10th IFIP International Conference on Wired/Wireless Internet Communications (WWIC), Springer Lecture Notes in Computer Science (LNCS)*, 2012.

77. Yaghmaee, M. H. and D. A. Adjeroh, "Priority-based rate control for service differentiation and congestion control in wireless multimedia sensor networks", *Computer Networks*, vol. 53, no. 11, pp. 1798–1811, 2009.

78. Texas Instruments, *Texas Instruments MSP430x4xx Family User's Guide*, http://www.ti.com/lit/ug/slau056j/slau056j.pdf [retrieved 23 April 2011].

79. Telosb Crossbow, *TelosB Data Sheet*, http://www.willow.co.uk/TelosBDatasheet.pdf [retrieved 23 October 2010].

80. Levis, P., 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*, http://www.cs.berkeley.edu/ culler/AIIT/papers/TinyOS/levis06tinyos.pdf [retrieved 23 July 2011].

81. Gök, H., "Voice over Wireless Sensor Networks", *B.Sc. Report, Yeditepe University*, 2011.

82. Egea-Lopez, E., J. Vales-Alonso, a. Martinez-Sala, P. Pavon-Mario and J. Garcia-Haro, "Simulation scalability issues in wireless sensor networks", *IEEE Communications Magazine*, vol. 44, no. 7, pp. 64–73, 2006.

83. Gozalvez, J., M. Sepulcre and R. Bauza, "Impact of the radio channel modelling on the performance of VANET communication protocols", *Telecommunication Systems*, 2010.

84. Arğ, N., "Voice Coding and Transmission with Error Concealment in Wireless Multimedia Sensor Networks", *B.Sc. Report, Yeditepe University*, 2012.