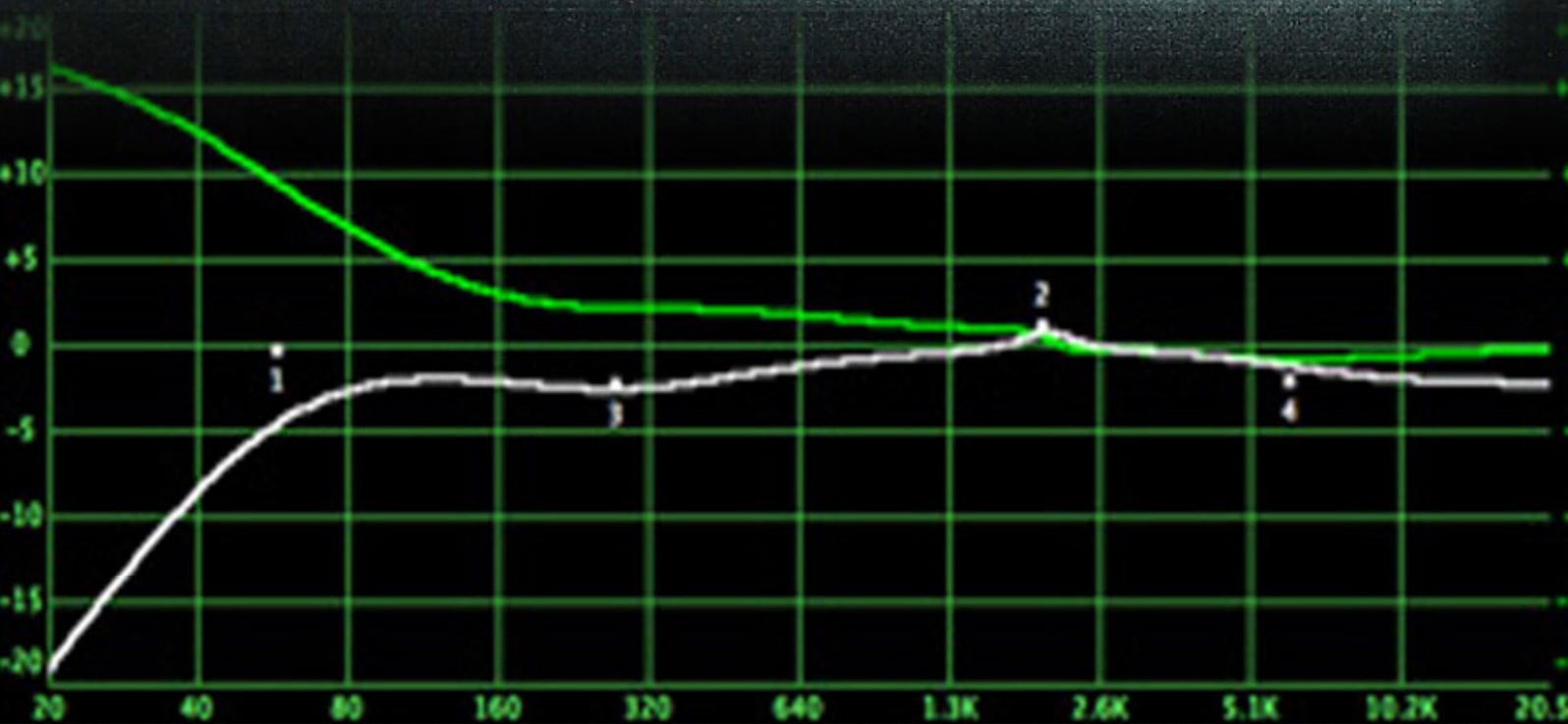


Signal Processing: An International Journal (SPIJ)

ISSN : 1985-2339

VOLUME 4, ISSUE 4

PUBLICATION FREQUENCY: 6 ISSUES PER YEAR



Signal Processing: An International Journal (SPIJ)

Volume 4, Issue 4, 2010

Edited By
Computer Science Journals
www.cscjournals.org

Editor in Chief Professor Dr. Saif alZahir

Signal Processing: An International Journal (SPIJ)

Book: 2010 Volume 4 Issue 4

Publishing Date: 30-10-2010

Proceedings

ISSN (Online): 1985-2339

This work is subjected to copyright. All rights are reserved whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication of parts thereof is permitted only under the provision of the copyright law 1965, in its current version, and permission of use must always be obtained from CSC Publishers. Violations are liable to prosecution under the copyright law.

SPIJ Journal is a part of CSC Publishers

<http://www.cscjournals.org>

© SPIJ Journal

Published in Malaysia

Typesetting: Camera-ready by author, data conversion by CSC Publishing Services – CSC Journals, Malaysia

CSC Publishers

Editorial Preface

This is fourth issue of volume four of the Signal Processing: An International Journal (SPIJ). SPIJ is an International refereed journal for publication of current research in signal processing technologies. SPIJ publishes research papers dealing primarily with the technological aspects of signal processing (analogue and digital) in new and emerging technologies. Publications of SPIJ are beneficial for researchers, academics, scholars, advanced students, practitioners, and those seeking an update on current experience, state of the art research theories and future prospects in relation to computer science in general but specific to computer security studies. Some important topics covers by SPIJ are Signal Filtering, Signal Processing Systems, Signal Processing Technology and Signal Theory etc.

This journal publishes new dissertations and state of the art research to target its readership that not only includes researchers, industrialists and scientist but also advanced students and practitioners. The aim of SPIJ is to publish research which is not only technically proficient, but contains innovation or information for our international readers. In order to position SPIJ as one of the top International journal in signal processing, a group of highly valuable and senior International scholars are serving its Editorial Board who ensures that each issue must publish qualitative research articles from International research communities relevant to signal processing fields.

SPIJ editors understand that how much it is important for authors and researchers to have their work published with a minimum delay after submission of their papers. They also strongly believe that the direct communication between the editors and authors are important for the welfare, quality and wellbeing of the Journal and its readers. Therefore, all activities from paper submission to paper publication are controlled through electronic systems that include electronic submission, editorial panel and review system that ensures rapid decision with least delays in the publication processes.

To build its international reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for SPIJ. We would like to remind you that the success of our journal depends directly on the number of quality articles submitted for review. Accordingly, we would like to request your participation by submitting quality manuscripts for review and encouraging your colleagues to submit quality manuscripts for review. One of the great benefits we can provide to our prospective authors is the mentoring nature of our review process. SPIJ provides authors with high quality, helpful reviews that are shaped to assist authors in improving their manuscripts.

Editorial Board Members

Signal Processing: An International Journal (SPIJ)

Editorial Board

Editor-in-Chief (EiC)

Dr. Saif alZahir

University of N. British Columbia (Canada)

Associate Editors (AEiCs)

Professor. Wilmar Hernandez

Universidad Politecnica de Madrid (Spain)

Dr. Tao WANG

Universite Catholique de Louvain (Belgium)

Dr. Francis F. Li

The University of Salford (United Kingdom)

Editorial Board Members (EBMs)

Dr. Thomas Yang

Embry-Riddle Aeronautical University (United States of America)

Dr. Jan Jurjens

University Dortmund (Germany)

Dr. Jyoti Singhai

Maulana Azad National institute of Technology (India)

TABLE OF CONTENTS

Volume 4, Issue 4, October 2010

Pages

- 175 - 200 Multi-Target Classification Using Acoustic Signatures in Wireless Sensor Networks: A survey
Ahmad, Ala Al-Fuqaha
- 201 - 212 Effective Preprocessing in Modeling Head-Related Impulse Responses Based on Principal Components Analysis
Hugeng , Dadang Gunawan, Wahidin Wahab
- 213 - 218 Performance Analysis of Convolution Coded WLAN Physical Layer under Different Modulation Techniques
Sanjeev Kumar, Ginni Sharma, Parveen Kumar, Anita Suman
- 219 - 227 Spectral Analysis of Sample Rate Converter
Manish Sabraj, Vipin Kakkar
- 228 - 238 A Gaussian Clustering Based Voice Activity Detector for Noisy Environments Using Spectro-Temporal Domain
Sara Valipour, Farbod Razzazi, Azim Fard
- 239 - 246 On Channel Estimation of OFDM-BPSK and -QPSK over Nakagami-m Fading Channels
Neetu Sood, Ajay K Sharma, Moin Uddin

Multi-Target Classification Using Acoustic Signatures in Wireless Sensor Networks: A survey

Ahmad Aljaafreh

*Electrical Engineering Department
Tafila Technical University
Tafila, 66110, P.O.Box 179, Jordan*

aljaafreh@ieee.org

Ala Al-Fuqaha

*Computer Science Department
Western Michigan University
Kalamazoo, MI 49008, USA*

ala.al-fuqaha@wmich.edu

Abstract

Classification of ground vehicles based on acoustic signals using wireless sensor networks is a crucial task in many applications such as battlefield surveillance, border monitoring, and traffic control. Different signal processing algorithms and techniques that are used in classification of ground moving vehicles in wireless sensor networks are surveyed in this paper. Feature extraction techniques and classifiers are discussed for single and multiple vehicles based on acoustic signals. This paper divides the corresponding literature into three main areas: feature extraction, classification techniques, and collaboration and information fusion techniques. The open research issues in these areas are also pointed out in this paper. This paper evaluates five different classifiers using two different feature extraction methods. The first one is based on the spectrum analysis and the other one is based on wavelet packet transform.

Keywords: Signal classification, feature extraction, distributed sensors, sensor fusion.

1. INTRODUCTION

Wireless sensor network (WSN) is a network of spatially distributed, densely deployed, and self organized sensor nodes, where a sensor node is a platform with sensing, computation and communication capabilities. WSN is an emerging technology because of the advances in technologies of: Micro-Electro-Mechanical Systems (MEMS), Microprocessors, wireless communication and power supply. New technologies provide cheap small accurate: sensors, processors, wireless transceivers, and long-life batteries. Sensor node is the integration of all of these technologies in a small board, like the ones in Fig. 3 part (b), it is called mote. Fig. 3 part (a) shows the basic architecture of the mote. All of the above motivate researchers and practitioners to design, deploy and implement networks of these sensor nodes in many applications. WSN has the following characteristics: concern is about the data but not about the sensor node itself, low cost, constrained power supply, static network, topology may change because of sensor node or link failure, sensor nodes are prone to destruction and failure, dense deployment, self-organization, and spatial distribution. WSN is used in many remote sensing and data aggregation applications [1],[2]. Detection, classification, and tracking are the main signal processing functions

of the wireless sensor networks [3]. WSNs increase the covered area, redundancy of the sensors, and decision makers, which improves the performance and reliability of the decision making. To understand the work, design and operation of the WSNs see Refs. [4],[5]. Refs. [4],[6] categorizes the applications and describes the implementation of the WSNs. A survey of the architecture and sensor nodes deployment in WSNs is presented in Ref. [7]. WSN is a cost efficient technology. However, it has some constraints. Limited energy, limited bandwidth, and limited computational power are the main constraints of WSNs [8]. Therefore, to implement any digital signal processing algorithm it needs to be an intelligent signal processing and decision making algorithm with the following requirements: power efficiency, robustness, and scalability. In WSNs, observed data could be processed at the sensor node itself, distributed over the network, or at the gateway node. WSNs can be utilized for distributed digital signal processing [9]-[11]. Research in classification in wireless sensor networks can be divided into two areas: hardware area (platforms, sensors), and software area (signal processing algorithms, collaboration, and networking techniques) [12]. The signal processing techniques and collaboration schemes that are used in ground vehicle classification in WSN based on acoustic signals are surveyed, as in Fig 2, in this paper. Target classification in WSN is to label or categorize a target that passing through the area that is monitored by the WSN to one of a predefined classes based on an extracted feature vector. Classification in WSNs can be considered as a process as in Fig. 4, where a feature vector is extracted from the input signal, then classified, then the information is fused to come up with the final decision. Most of the researcher are interested in improving the performance of this process through selection and design an efficient tool, as in Table 1, for one of the followings tasks :

- Feature Extraction
- Classification Techniques
- Information Fusion

The remainder of the paper is organized as follows. Section 2 presents the recent methods that are used to extract features from the vehicle acoustic signals for single and multiple targets. Section 3 discusses the classification techniques. Section 4 presents the information fusion techniques. Section 5 outlines the the open research. And finally, conclusions are discussed in section 6.

Reference	Feature Extractor	Classifier	Classes Number	Classification Rate	Fusion Method
[12]	TESPAR	ANN	2	up to 100%	-
[13] and [14]	DWT	MPP	2	98,25%	-
[15]	HLA, PSD	ANN	4	HLA: 92%, PSD: 94%	-
[16]	HLA	ANN	18	88%	running sum
[17]	HLA	MAP	6	89%	-
[18]	MFCC	GMM, HMM and ML	9	77%, 88%	-
[19]	FFT, DWT, STFT, PCA	kNN, MPP	4	85%, 88%	MRI
[20]	STFT, PCA	ANN	3	-	-
[21]	FFT, PSD, AR	kNN, ML, SVM	2	78% - 97%	-
[22]	DWT	ANN	4	73%	-
[23]	CC	HMM	9	96%	-
[24]	WPT	LDA, CART	3	-	-
[25]	CWT	ANN	6	95%	-
[26]	TVAR, PCA	ANN	6	83%-95%	-
[27]	BHM	CART	9	90%	Decision Fusion
[28]	STFT, RID	ANN, MVG	6	up to 87%	-
[29]	EE, PCA	ANN, Fuzzy	5	up to 97%	-

		Logic			
[30]	AR	ANN	4	up to 84%	-
[31]	FFT, PSD	kNN	-	-%	-
[32]	FFT, WDT	kNN	2	62%	Dempster-Shafer, MV
[33]	FFT	Template Matching	8	-%	template storing
[34]	PSD	kNN, ML	2	77%, 89%	Distributed Classification
[35]	-	kNN, ML, SVM	2	69%, 68%, 69%	MAP Bayesian, Nearest Neighbor, Majority Voting, Distance-based
[36]	Harmonic and Frequency Components	SVM	5	85%	modified Bayesian (decision level)
[37]	Harmonic set	MVG	3-5	70-80%	-
[38]	STFT, PCA	C4.5, KNN, PNN, SVM	4	60-93%	-
[39]	WPT	CART	-	-	-
[40]	MFCC	RNN	4	85%	-
[41]	PSD	KNN, ML, SVM	2	up to 97%	-
[42]	PSD, PCA	SVM	3	up to 93%	-
[43]	MFCCs	GMM	2	up to 94.1%	CART
[44]	FFT, WT	KNN, MPP, K-Means	3	95.5%	-
[45]	WPT	ML, ANN	3	up to 98%	-
[46]	PSD	ANN	4	up to 99%	-
[47]	WPT	cascaded fuzzy classifier (CFC)	3	-	Dempster-Shafer (DS)

Table 1: Recent feature extraction and classification techniques used for vehicle classification based on acoustic signals.

2. FEATURE EXTRACTION OF ACOUSTIC SIGNATURE

Feature extraction is the most significant phase of the classification process. To classify an object, a set of features of that object is extracted to label that object to one of a predefined classes. This set of features is generated from a source signal as in Fig. 1. Feature extraction can be considered as dimensionality reduction technique. In feature extraction certain transforms or techniques are used to select and generate the features that represent the characteristic of the source signal. This set of features is called a feature vector. Feature vectors could be generated in time, frequency, or time \ frequency domain.

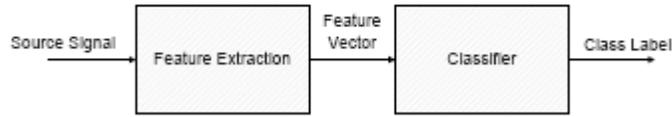


Figure 1: Classification block diagram.

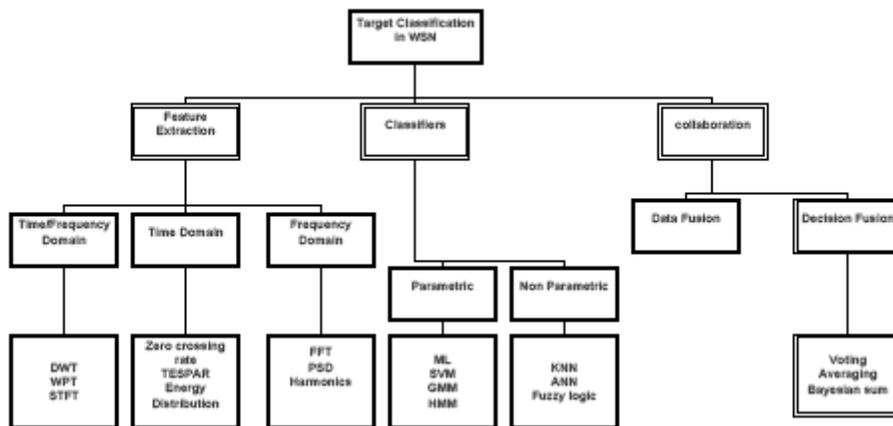


Figure 2: Taxonomy of the techniques that are used in target classification using acoustic signature in wireless sensor networks

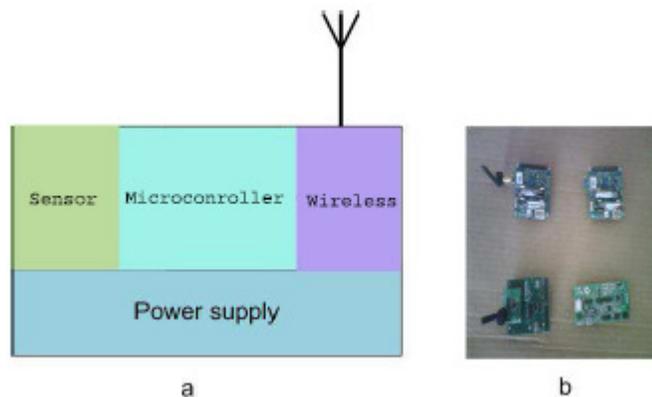


Figure 3: Wireless sensor node examples in part (b) and the common architecture of a sensor node in part (a).

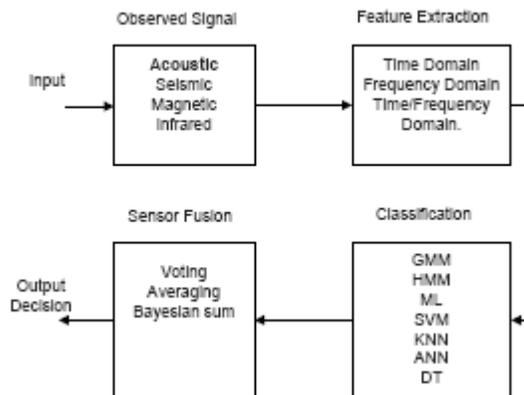


Figure 4: A summary diagram of the feature extraction, classification, and collaboration algorithms that are used in vehicle classification using WSNs.

2.1 Time Domain

The computation of feature vector in time domain is usually simple. Ref. [48] discusses two time-domain feature generation methods. The first method is based on the energy distribution of the signal, where the energy of a short time window of the source signal is used to discriminate between classes. The second method is based on counting the number of zero crossings of a signal within a time interval. The energy envelope (EE) in time domain is considered in [29]. Time Encoded Signal Processing and Recognition (TESPAR) is a method that is used in speech waveform encoding. TESPAR is used in [12] to generate features from vehicle acoustic and seismic signals. TESPAR is based on the duration and shape of the portion of the waveform that is between two zero crossings.

Principal Component Analysis (PCA) is popular statistical tools that is used for dimensional reduction. PCA is based on finding the principal eigenvectors of the covariance matrix of the set of signals. PCA is used as a feature extraction method in [19, 20, 38, 42].

2.2 Frequency Domain

Frequency based feature generation methods, like Fast Fourier Transform (FFT), are common approaches in vehicle classification [16], [20], [31], [33]-[35], [49]. In [31] Fast Fourier Transform (FFT) and Power Spectral Density (PSD) are used to extract feature vectors. Similarly in [35], the first 100 of 512 FFT coefficients are averaged by pairs to get a 50-dimensional FFT-based feature vector with resolution of 19.375 Hz and information for frequencies up to 968.75 Hz. Ref. [34] presents schemes to generate low dimension feature vectors based on PSD using an approach that selects the most common frequency bands of PSD in all the training sets for each class. Ref. [33] proposes an algorithm that uses the overall shape of the frequency spectrum to extract the feature vector of each class. Principal component eigenvectors of the covariance matrix of the zero-mean-adjusted samples of spectrum are also used to extract the sound signature as in [20]. Some vehicle acoustic signatures have a pattern of relation between the harmonics amplitude. Harmonics are the peaks of the spectral domain. The relation between the amplitude and the phase of these peaks is used to form the feature vector. Harmonic Line Association (HLA) feature vector is used in [30], where the magnitude of the second through 12th harmonic frequency components are considered as the feature vector to be used for vehicle classification. Different algorithms are used to estimate the fundamental frequency. In [36], two sets of features are extracted from the vehicle sound. The first one is based on the harmonic vector. The second one is a key frequency feature vector. In [37], the number of harmonics is modeled as a function of

the vehicle type. Looking for stable features other than the harmonics relation, Ref. [50] models the vehicle acoustic signature by a coupled harmonic signal. Cepstral coefficients (CC) are the coefficients of the inverse Fourier Transform of the log of the magnitude of the spectrum [23]. Mel-frequency cepstral coefficients (MFCC) is used in [18],[40],[43] as a feature extractor, where the feature vector is made up of few of the lowest cepstrum coefficients. Mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, where the log power spectrum on a nonlinear mel scale of frequency is transformed based on a linear cosine transform.

Two types of spectral features are explored in [21]: Non-parametric FFT-based PSD estimates, and Parametric PSD estimates using autoregressive (AR) modeling of time series. In AR model the value of any variable at any time is modeled as a function of a finite number of the past values of that variable. The number of the involved past values is called the model order. AR of the first order is a Markov chain, where the current value depends only on the previous value. A random variable X can be modeled at time t using AR of order P as follows:

$$x_t = \sum_{k=1}^p \theta_k x_{t-k} + \omega_t \tag{1}$$

where θ_k denotes the corresponding autoregressive coefficients. ω_t is a white gaussian noise with zero mean. If θ_k is varying with time then the AR process is called Time Varying Autoregressive (TVAR). TVAR is used to model the acoustic signal in [26, 51]. A filter bank is used based on the biology based hearing model (BHM) as a feature extraction system in [27].

Feature Extraction Example using the Spectrum Distribution:

The goal is to develop a scheme for extracting a low dimension feature vector, which is able to produce good classification results. The first feature extraction technique of acoustic signals in this paper is based on the low frequency band of the overall spectrum distribution. The low frequency band is utilized, because most of the vehicle's sounds come from the rotating parts, which rotate and reciprocate in a low frequency, mainly less than 600 Hz as it is clear in Fig. 6. Sounds of moving ground vehicles are recorded at the nodes at a rate of 4960 Hz as in Fig.5. After the positive detection decision, a signal of event is preprocessed as the following:

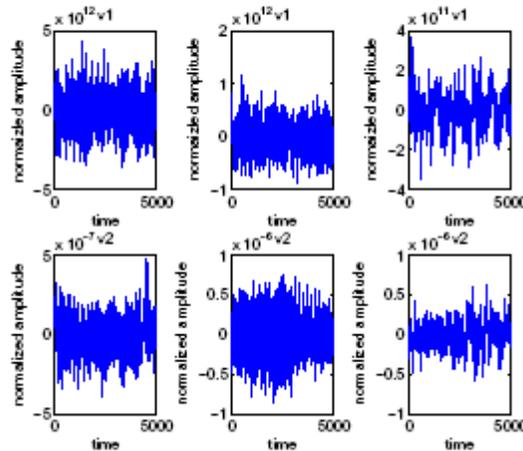


Figure 5: Time domain for three different sounds for two different vehicles v1 and v1.

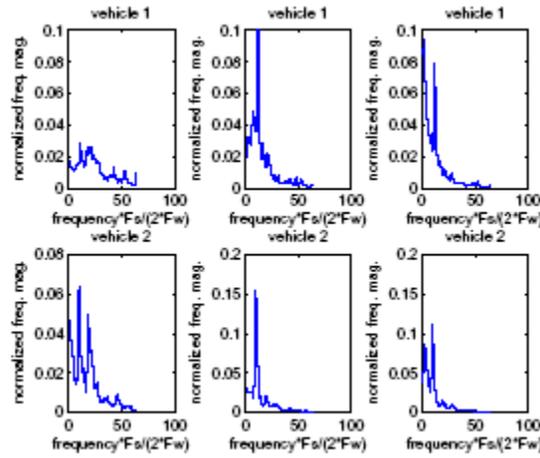


Figure 6: Frequency distribution for three different sounds for two different vehicles v1 and v1. Fs is sampling frequency =4960. Fw is FFT window size=512.

DC bias should be removed by subtracting the mean from the time series samples.

$$x_i(n) = x_i(n) - \frac{1}{N} * \sum_{n=1}^N x_i(n) \quad (2)$$

Feature vector will be the median of the magnitude of the STFT of a signal of event. It will be computed as the following: the magnitude of the spectrum is computed by FFT for a hamming window of size 512, without overlapping.

$$X_i(W) = FFT(x_i(n)) \quad (3)$$

After this, the spectrum magnitude is normalized for every frame

$$X_i(W) = \frac{X_i(W)}{\sum_{W=1}^K X_i(W)} \quad (4)$$

where K is the window size. The median of all frames is considered as the extracted feature vector.

$$X_{if}(W) = median(X_i(W)) \quad (5)$$

The mean of all frames could also be considered as the extracted feature vector.

$$X_{if}(W) = \frac{1}{Z} \sum_{i=1}^Z X_i(W) \quad (6)$$

where $z = N/k$. The first 64 points of the median of the spectrum magnitude contain up to 620 Hz. This gives a 64 dimensional vector that characterizes each vehicle sound. We compared feature extraction using the mean and the median. The median gives better results, specially for noisy environments. Fig.7 displays the acoustic spectral distribution of vehicle 1 and vehicle 2. For the unknown utterance, the same steps are done, except one frame of FFT is considered as the feature to be classified to reduce the computational cost, because this FFT computation is performed online. This can be extended to have multiple frames, but this will increase the computational cost.

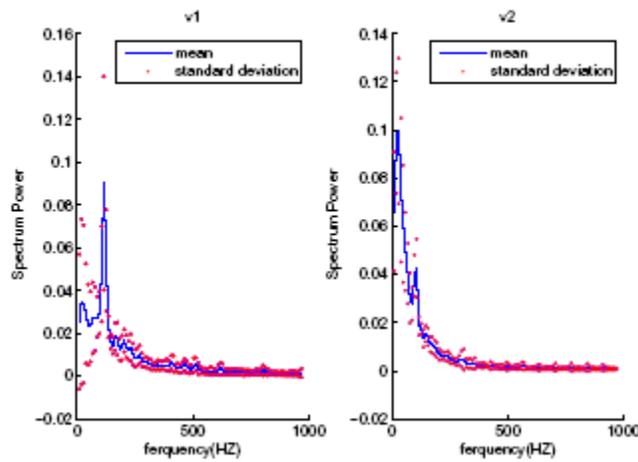


Figure 7: Acoustic spectra distribution of vehicle 1 and vehicle 2. To the left is vehicle 1

2.3 Time Frequency Domain

Short Time Fourier Transform (STFT) is used in [38] to transform the overlapped acoustic Hamming windowed frames to a feature vector. Ref. [52] proposes a probabilistic classifier that is trained on the principal components subspace of the short-time Fourier transform of the acoustic signature. Wavelet transforms provide multi-resolution time-frequency analysis [53]. Wavelet transforms (WT) is the the projection of a signal onto the wavelet. Wavelet is a series of functions

$\psi_{a,b}(t)$ derived from a base function $\psi(t)$ by translation and dilation.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-p}{a}\right) \tag{7}$$

where a is called scale parameter, b is called translation or shift parameter, and $\psi(t)$ is called wavelet base function. Wavelet Transform is called CWT when values of a and b are continuous, and it is called DWT when they are discrete [54]. Discrete Wavelet Transform (DWT) approximation coefficients y are calculated by passing the time series samples x through a low pass filter with impulse response g .

$$y(n) = x(n) * g(n) = \sum_{k=-\infty}^{\infty} x(k)g(n-k)$$

The signal is also decomposed simultaneously using a high-pass filter h . The outputs from the high-pass filter are the detail coefficients. The two filters are related to each other. DWT is exactly the same as the Wavelet Packet Transform (WPT) except that in DWT the next level is the result of one step of wavelet transform of the approximation branch and not the detail one. Wavelet packet transform can be viewed as a tree structure. The root of the tree is the time series of the vehicle sound. The next level is the result of one step of wavelet transform. Subsequent levels in the tree are obtained by applying the wavelet transform to the low and high pass filter results of the previous step's wavelet transform. The Branches of the tree are the blocks of coefficients. Each block represents a band of frequency. Feature extraction of acoustic signals is based on the energy distribution of the block coefficients of wavelet packet transform. A wavelet-based acoustic signal analysis of military vehicles is presented in [13, 55]. Discrete Wavelet Transform

(DWT) is used in [13] and [14] to extract features using statistical parameters and energy content of the wavelet coefficients. Wavelet Packet Transform (WPT) has a higher frequency resolution than the DWT [56]. WPT is also used to extract vehicle acoustic signatures by obtaining the distribution of the energies among blocks of wavelet packet coefficients like in [24],[39]. Ref. [53] has a proof that wavelet analysis methods is suitable for feature extraction of acoustic signals

Feature Extraction Using Wavelet Packet Transform:

After the positive detection decision, a one second time series is preprocessed as the following:

- The wavelet packet transform is applied for this signal then the energy of each block coefficients of the (L) level is calculated.
- This approach provides a vector of length = original time series length / 2^L . Which is considered the feature vector.

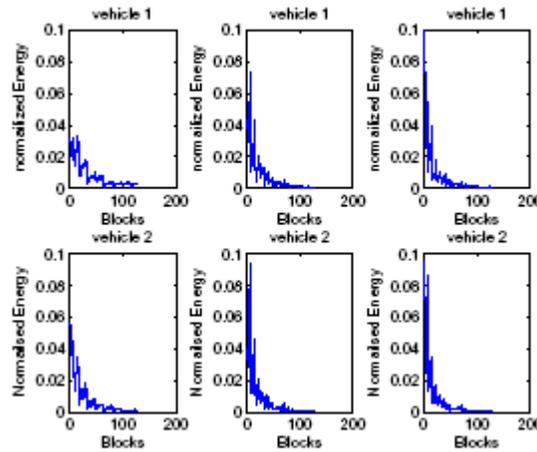


Figure 8: Wavelet block energy distribution for vehicle one in first row for three different sounds and for vehicle 2 in the second row.

Fig. 8 displays the blocks energy distribution for vehicle 1 and vehicle 2. In this paper we used classification rate as the metric for the evaluation of the feature extraction performance. But this metric depends on the classifier itself. Thus, we compare the classification rate for two classifiers as shown in Fig. 12.

2.4 Feature Extraction Performance Using Separability Measures

Separability measures provide a measure of class discriminability based on feature space partitioning. A good feature vector extractor provides close feature vectors for the same class, and far feature vectors for distinct classes. The goal is to have a feature extraction method that has high distance between distinct classes and low distance within each class. The metric, in this paper, is the separability ratio (sr), which is the ratio between the intraclass distance and the average interclass distance [57].

$$sr = \frac{D_g}{D_l} \tag{8}$$

$$D_g = \sum_{i=1}^C \frac{P_i}{n_i} \sum_{k=1}^{n_i} [(V_{ik} - m_i)(V_{ik} - m_i)^T]^{\frac{1}{2}} \tag{9}$$

D_s represents the average of the variances of distance within all classes. V_{ik} is the normalized feature vector. C is the number of classes. P_i is the probability of class i . n_i number of vectors in class i . m_i is the mean vector for class i .

$$D_s = \sum_{i=1}^C P_i [(m_i - m)(m_i - m)^T]^{-\frac{1}{2}} \quad (10)$$

D_l represents the average of the distances between all classes. m is the mean for all classes.

$$m_i = \frac{\sum_{k=1}^{n_i} V_{ik}}{n_i} \quad (11)$$

$$m = \frac{\sum_{i=1}^C \sum_{k=1}^{n_i} V_{ik}}{n_i} \quad (12)$$

The smaller the ratio is the better the separability. This means that the best feature extraction scheme is the one that decreases D_s and increases D_l .

3. CLASSIFICATION TECHNEQUES

Classifiers provide the functions or the rules that divide the feature space into regions, where each region corresponds to a certain class. Having a number of N -dimensional feature vectors for each class, a function can be deduced to partition the N feature space to number of regions, where each region represents a class. This process is called supervised learning or classification. Classifiers can be categorized to parametric or non-parametric. Some researches combine multiple different classifiers, called compound classifiers.

3.1 Parametric Classifiers

Parametric classifiers are the classifiers that can be represented in closed-form. for instance, assuming that the distribution of a certain class as a parametric form such as Gaussian. Some classifiers are based on discrimination function with a certain parametric form such as support vector machine. Below are the most parametric classifiers that have been used in vehicle classification based on acoustic signature.

3.1.1 Bayesian Classifier

Bayesian classifier is a probabilistic classifier based on using Bayes' theorem. Maximum likelihood (ML) is used to estimate the Bayesian classifier parameters. Maximum A posteriori Probability (MAP) can also be considered as a generalization of ML. Each class is assumed to be independent instances of parametric distributed random process. A naive Bayes classifier is a variation of the Bayesian classifier with assumption of an independent feature model. Bayesian classifier is used in many research papers with assumption that each class is a normal distributed random process [35],[41],[58],[59] where features vector S of each class C is assumed to be independent instances of normally distributed random process.

$$p(S | \theta_i) \approx N(\mu_i, \Sigma_i) \quad (13)$$

$i=1,2 \dots C$. μ_i and Σ_i are the mean and covariance matrix respectively. θ_i is the parameter set of i^{th} distribution, $\theta_i = \{\mu_i, \Sigma_i\}$.

$$p(S | \theta_i) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(S - \mu)^H \Sigma^{-1}(S - \mu)\right] \tag{14}$$

To represent each training set of each class as a distribution with $\hat{\Sigma}$ and $\hat{\mu}$ parameters the likelihood of $\hat{\theta}$

$$l_i(\theta) = \sum_{k=1}^n \ln P(s_k | \theta) \tag{15}$$

should be maximized by equating $\Delta_{\theta^i} l_i = 0$, then ML estimations of μ and Σ are

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^n s_k \tag{16}$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{k=1}^n (s_k - \hat{\mu})(s_k - \hat{\mu})^H \tag{17}$$

for minimum error classification the a posteriori probability should be maximized

$$p(\theta_i | S) = \frac{p(S | \theta_i)p(\theta_i)}{\sum_{j=1}^C p(S | \theta_j)p(\theta_j)} \tag{18}$$

$h_i(x)$ denote the logarithmic version of $p(\theta_i | S)$

$$h_i(x) = \ln p(S | \theta_i) + \ln p(\theta_i) - G \tag{19}$$

where G is constant can be ignored in the optimization

$$h_i(x) = -\frac{1}{2}(S - \mu)^H \Sigma_i^{-1}(S - \mu) - D \tag{20}$$

where D is constant then any vehicle feature vector x is classified to class i according to the discriminant functions $h_i(x)$ if $h_i(x) > h_j(x)$ for all $j \neq i$. Linear discriminant Analysis (LDA) assumes that the class covariances Σ_i are identical. LDA is used as a linear classifier in [24] ML is the optimum classifier but it needs large number of training set. Training ML classifier with small number of training set will not give an invertible covariance matrix. This makes it hard to compute the discriminant functions.

3.1.2 Support Vector Machine (SVM)

SVM is widely used as a learning algorithm for classifications and regressions. SVM classify data x_i by class label $y_i \in \{+1, -1\}$ given a set of examples $\{x_i, y_i\}$ by finding a hyperplane $wx + b$, $x \in R^n$ which separate the data point x_i of each class. as in Fig. 9.

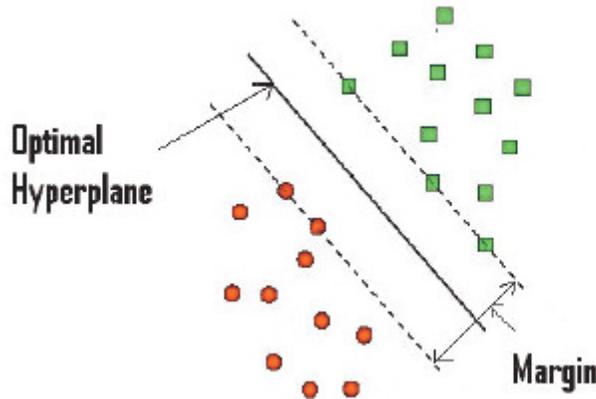


Figure 9: An Example of Two Classes Problem. Squares and Circles Represent Class 1 and Class 2 Respectively.

$$g(x) = \text{sign}(wx + b) \tag{21}$$

where w is the weight vector, b is the bias. SVM choose the hyperplane that maximize the distance between the hyperplane and the closest points in each feature space region which are called support vectors. So the unique optimal hyperplane is the plane that maximize this distance

$$|wx_i + b| / PwP \tag{22}$$

This is equivalent to the following optimization problem

$$\min_{w,b} \frac{PwP^2}{2}, \text{ s.t. } y_i(w^T x_i + b) \geq 1 \tag{23}$$

For the cases that nonlinear separable, a kernel function maps the input vectors to a higher dimension space in which a linear hyperplane can be used to separate inputs. So the classification decision function becomes:

$$\text{sign}\left(\sum_{i \in SVs} \alpha_i^0 y_i K(p, p_i) + b\right) \tag{24}$$

where SVs are the support vectors. α_i^0 and b are a lagrangian expression parameters.

$K(p, p_i)$ is the kernel function. It is required to represent data as a vector of a real number to use SVM to classify moving ground vehicles. Performance of SVM classifier for vehicle acoustic signature classification for both feature extraction methods is also evaluated. It is found that SVM is a good classifier for stochastic signal in WSN [12], [42], [60]–[63].

3.1.3 Gaussian Mixture Model (GMM)

Due to the constraints in WSN resources, parametric models such as Gaussian mixture model is preferred to non-parametric models [43]. Modeling of acoustic signal in WSN using a parametric model, like GMM requires little resources, and has a good pattern matching performance [43]. GMM is a statistical method that is used for classification and clustering. GMM is a linear combination of M-Gaussian pdfs. Let x be a N-dimensional feature vector, then the distribution of x is as follows:

$$f_m(x) = \sum_{i=1}^m \alpha_i \phi(x; \theta_i) \tag{25}$$

where $\sum_{i=1}^m \alpha_i = 1$, $\alpha_i \geq 0: i \in 1, \dots, m$

α_i is the mixing weight, $\phi(x; \theta_i)$ is the Gaussian mixture component. Component i has N-variate Gaussian density function with weight α_i , mean vector μ_i , and covariance matrix Σ_i .

Expectation maximization (EM) is one of the common algorithm that is used to obtain the GMM parameters $\Phi_i = (\alpha_i, \mu_i, \Sigma_i)$ from the training set. The GMM generated from the training set will be used in vehicle classification as in Fig. 10.

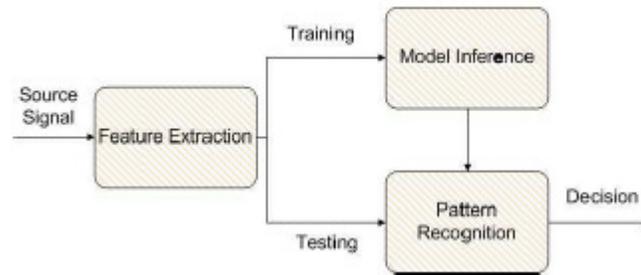


Figure 10: Main block diagram of pattern recognition

Any vehicle feature vector x is classified to class C_i if it maximizes $p(C_i | x) = p(x | C_i) \cdot p(C_i)$. If all the classes are assumed to occur with the same probability, then the concern is to maximize $p(x | C_i)$ for every possible class. GMM is used as a classifier in WSN based on the features that are extracted from the vehicle sounds in [43]. Ref. [23] concludes that the GMM, as a parametric classifier, outperforms the KNN and SVM classifiers, and it also concludes that GMM needs relatively less resources.

3.1.4 Hidden Markov Model (HMM)

Acoustic signals could be modeled as HMM. HMM has a specific discrete number of unobserved states; each state has a transition probability to any other state and an initial probability. Each state may be considered as representing a certain sound of the vehicle [23]. Ref. [23] models the cepstral coefficients that are obtained from the time domain signal as HMM, where the pdfs of the states are assumed to be Gaussian with non-zero means and with a diagonal covariance matrix. Modeling the vehicle sounds as HMM is based on the assumption that the acoustic signal of the vehicle is consisting of a sequence of a discrete number of sounds, where the statics of each sound of these sounds is described by a separate state. The parameters of the HMM are: the state transition probability to any other state, the initial probability for each state, and the

observation pdf parameters for each state. Estimation of the maximum likelihood parameters of the HMM given a data set of the vehicle sounds can be done by a special case of the Expectation-maximization algorithm called the Baum-Welch algorithm; it is also known as the forward-backward algorithm. HMM implementation for vehicle classification is based on the estimation of the sequence of states, given a sequence of observations. Some known algorithms are used for that such as the Viterbi algorithm. GMM is static pattern model, while HMM is a sequential pattern model.

3.1.5 Minimum Distance Approach

Minimum Distance Approach (MPP) is a simple parametric classifier that is based on the minimum distance between the feature vector and the average of the class distribution. MPP assume the distribution of the training set of each class as Gaussian distribution. MPP measure the distance between the average of each class distribution and the feature vector of the test data, then it corresponds the test data to the class that has the minimum distance. MPP is used as a classifier in Refs [13],[19],[44] .

3.2 Non-parametric Classifiers

In this kind of classifiers no assumptions are made about the probability density function for each class.

3.2.1 KNN classifier

KNN is a simple and accurate method for classifying objects based on the majority of the closest training examples in the feature space. KNN is rarely used in wireless sensor networks because it needs large memory and high computation. In our experiments we set K to be three. So the KNN classifier finds the closest three neighbors out of all the training set. After that, the KNN classifier counts how many of these three is in class one and how many is in class two then based on the majority classify the tested one. KNN is implemented in many literatures as a benchmark to evaluate other classifiers [21],[31],[34],[35],[44]

3.2.2 Artificial Neural Network (ANN)

Artificial neural networks are a learning intelligent tools used to solve problem that hard to be modeled analytically. A key feature of neural networks is the ability to learn from examples of input/ output pairs in a supervised fashion. In [40], rough neural network is used to recognize type of vehicles in WSN. The network has 25 input neurons, corresponding to the 25-dimensional feature vector, 25 hidden layer neurons, and 4 output neurons. Neural network classifier and the maximum likelihood classifier are compared concerning their advantages and disadvantages in [45]. Artificial neural networks were used as a technique to classify and identify vehicles based on acoustic signals in [46], where a two-layer back propagation neural network is trained by the output of a self organized maps neural network.

3.2.3 Fuzzy Logic Rule-Based classifier

Fuzzy Logic Rule-Based classifier (FLRBC) maps the input to the output based on some rules [64]. Fuzzy logic inference is a simple approach to solving a problem rather than attempting to model it mathematically. Empirically, the fuzzy logic inference depends on human's experience more than the technical understanding of the problem. As in Fig. 11, fuzzy logic inference consists of three stages:

1. Fuzzification: map any input to a degree of membership in one or more membership functions, the input variable is evaluated in term of the linguistic condition.
2. Fuzzy inference: fuzzy inference is the calculation of the fuzzy output.
3. Defuzzification: defuzzification is to convert the fuzzy output to a crisp output.



Figure 11: The stages of fuzzy logic inference design.

Cascaded fuzzy classifier is proposed in [47] to classify ground moving vehicles locally at sensor nodes in WSN.

3.2.4 Decision Tree

Decision tree is a nonlinear classifier that depend on a multistage decision system, where the classes are sequentially rejected until reach the accepted class. This kind of classifier split the feature space into unique regions, where each region represents a class [65]. In Refs. [38],[43],[66] decision tree is utilized as a classifier. C4.5 algorithm is used to generate a decision tree in [38]. Decision tree is sometimes describes as classification tree or regression tree. Classification And Regression Tree (CART) analysis is used to refer to both of classification and regression.

3.3 Classifiers Evaluation and Comparison

Classification evaluation is to measure how accurate the classifier is. There are three main classification metrics that are used to evaluate the performance of the classifier namely: Correct classification rate, detection probability, and false alarm probability. Correct classification rate is the ratio of the correctly classified events from all the samples for all classes. Detection probability for a certain class is the ratio of the correctly classified events from the samples of that class. False alarm probability for a certain class is the ratio of the wrongly classified events for that class from all samples of other classes. Classifiers might have a high accuracy if tested with the same training data. Thus, a classification validation is needed. There are two common methods that are used for classification validation. The first is called hold-out or split-sample method, where a single part of the data is set a side for testing. The second method is the cross validation method. k-fold cross-validation divides the data into k subsets. The classifier is trained k times, each time leaving out one of the subsets from training to be used for testing the classifier. If k equals the sample size, the method is called leave-one-out. Ref [45] compared the recognition rate and the robustness of two classifiers, neural network classifier and maximum likelihood classifier. Neural fuzzy techniques for classification of vehicle acoustic signal is used in Ref. [29]. In [38], Four different classifiers decision tree (C4.5), KNN, probabilistic neural network (PNN) and SVM are compared. The classification results indicate the performance of SVM outperforms C4.5, KNN, and PNN. SVM, ML, and KNN are used in [35] to evaluate their data set. In this paper, five different classifiers are compared as in Fig. 12.

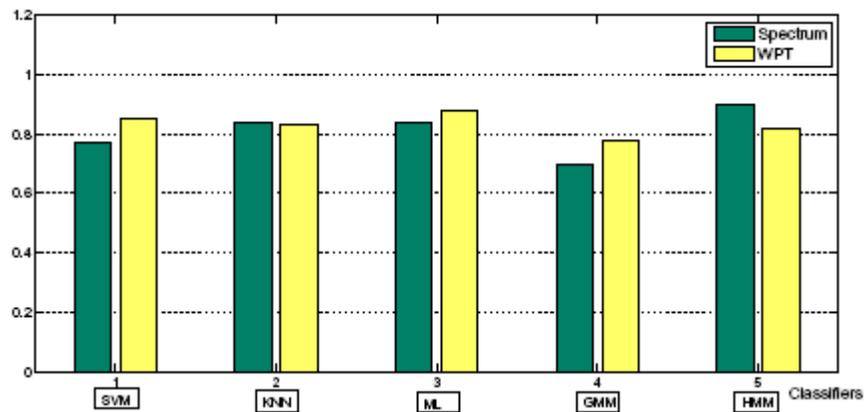


Figure 12: A comparison of the correct classification rates for different classifiers for two different feature extraction methods. The first one is based on the spectrum analysis and the other one is based on wavelet packet transform.

3.4 Single Target Classification

Single vehicle classification is to label or classify a vehicle to one of the predefined classes. Classification is based on the feature vector, which is generated from the observed continuous stochastic signal. This stochastic signal could be acoustic, seismic, electromagnetic, or any other kind of signals. The feature vector will be the input of the classifier. The classifier could be any kind of the classifiers that are mentioned in the classifiers section. Classifiers could be parametric or non parametric. The parameters of the parametric one are estimated from a training set. Non parametric classifiers are also trained by a training set. Single vehicle classification techniques can be used for multiple vehicle classification after sources separation or with the assumption that the sensor will not observe two or more vehicles at the same time. This assumption is not realistic, especially in battlefield scenarios.

3.5 Multiple Targets Classification

Many researchers assume that multiple targets could be separated in time and space. However, in many of the surveillance applications this assumption is not realistic, where multiple targets can exist in the sensing range of one sensor at the same time as in Fig. 13. This makes the sensor observe a combination of signals. The combination of multiple signals can be modeled as linear, nonlinear, or convoluting combination of the single target signals. Ref. [67] exploits the classifier that is trained in single target classification to classify a convoy of vehicles. Most of the literature models the multiple targets classification problem as a Blind Source Separation (BSS) problem. BSS problem has been tackled in the literature in many ways, such as neural network [68]-[70] and Independent Component Analysis (ICA). ICA is frequently used to separate or extract the source signals [71]-[75]. In [76] the source extraction problem in wireless sensor network is studied in two different sensor network models. Fast fixed-point ICA algorithm is used for source separation [51] presents a statistical method based on particle filtering for the multiple vehicle acoustic signals separation problem in wireless sensor networks. In [77], a recognition system links BSS algorithm with an acoustic signature recognizer based on missing feature theory (MFT). The result of the comparison between FastICA, Robust SOBI, and their work shows that both of former algorithms are better for mixtures of two signals and more. Refs. [78],[79] discuss problem of source estimation in sensor network for multiple targets detection, then a distributed source estimation algorithm is developed. These solutions have some drawbacks that make it hard to be implemented in WSNs. It is evident that the manager nodes need to perform source separation and source number estimation. These tasks are computationally intensive when executed on the individual sensor nodes. The manager node does the following: estimation of the number of sources, separation or extraction of sources, classification of sources. [80] presents a system that is able to recover speech signal in the presence of additive non-stationary noise. This done

through a combination of the classification and mask estimation. Ref. [37] uses a multi-variate Gaussian classifier for classifying individual acoustic targets after beamforming the received signal in the direction of the targets. We direct the reader for more information in beamforming to [81]. Classification of multiple targets without the separation of the sources based on multiple hypothesis testing is an efficient way of classification [82]. A distributed classifiers based on modeling each target as a zero mean stationary Gaussian random process and the same for the mixed signals is proposed in Ref. [83].

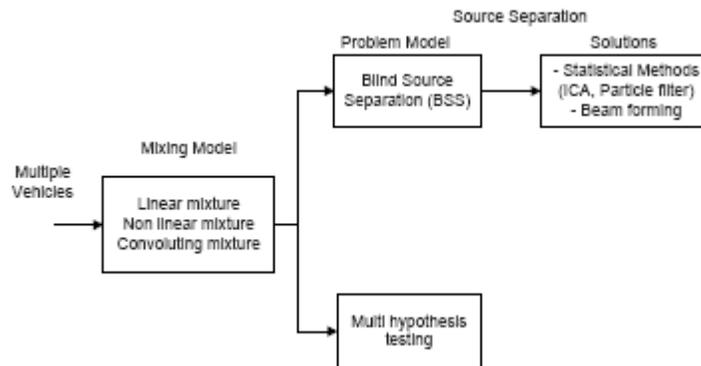


Figure 13: A summary diagram of the techniques that are used in multiple vehicle classification using WSNs.

4. COLLABORATIVE CLASSIFICATION

Data fusion, information fusion, data aggregation, multisensor integration and sensor fusion are terms that have been used interchangeably in the literature to refer to the idea and theory of fusion. In WSNs data fusion is needed for the following reasons: sensor failure, sensor and technology imprecision, limitations in spatial and temporal coverage [84]. Information fusion can be classified according to the relationships between the sources as: redundant, cooperative, or complementary. It also can be classified according to the abstraction level of the processed data as: low-level (raw data), medium-level, high-level (decision fusion), or multi-level fusion.

The main objective of collaborative classification is to extract the most beneficial information from the collected data. Because every target has its own signature according to the type of generated signal. Deployment of different kinds of sensors, in the same sensor node or in different sensor nodes, increases the performance of collaborative signal processing. This stems from the fact that every sensor type has a different interference and a measurements noise. Efficient and reliable decision making needs data fusion and collaborative signal processing. Distributed classification algorithms fuse signal or decisions from multiple sensor nodes, then classify the targets based on a priori statistical information [85],[86]. Collaboration could be across the sensor nodes, or within a sensor node only when it includes multiple modalities of data. Ref. [67] shows the improvement in classification error because of the collaboration. WSNs have two kinds of collaboration signal processing models: data fusion and decision fusion. For more information in data and decision fusion see [87],[88]. Ref. [89] analyzes both models. Exploiting the collaboration among the sensor nodes enhances even the sensing coverage for the network [90]. Decision fusion has less accuracy than data fusion. However, data fusion has more computation and communication overhead than the data fusion. In decision fusion, sensor nodes do some processing then send the decision to the manager node as in Fig. 15, where these decisions could be hard or soft decisions [91]-[93]. Manager node fuses all the decisions and come up with the best decision. Rules of decision fusion could be based on: voting, averaging, Bayesian, Dempster-Shafer (DS) sum or other rules as in [94]. An example of decision fusion is the tracking system that is proposed in [95], where detection and classification are performed in the sensor

node while tracking is performed in the sink node. Data and decision fusion are increasingly implemented in sensor network because of hardware improvement, advances in processing methods, and advances in sensor technology [96]. Fig 14 shows some of the data fusion applications. Data and decision fusions techniques answer the question of how to combine the data and decisions from the sensor nodes in a network to obtain a final decision with optimal resource consumption. Sensor nodes make the measurements, then send raw measurements, processed measurements, or local posterior to the fusion center. Fusion architecture can be hierarchical or distributed. In hierarchical fusion, the data or decision is fused from the sensor node to the higher level fusion center. While in distributed fusion architecture, the data or decision is broadcasted to all other fusion centers.

There are various scenarios of data and decision fusion for single and multiple sensors within the sensor node or cross over the network. Ref. [97] has a survey that focuses on the decision fusion. Ref. [98] studies a distributed decision fusion, where the local decision makers send the ranking of the hypotheses, rather than just the most likely hypothesis. A consensus algorithm which weighs the measurements of the neighboring nodes is presented in [99].

Data fusion from seismic and acoustic signal improves the classification accuracy [67]. In Ref. [43], a decision tree generated by the classification and regression tree algorithm is used to fuse the information from heterogeneous sensors. Multi-resolution integration (MRI) algorithm is used in [19] as a data fusion algorithm to handle the uncertainty of sensor output. MRI algorithm is based on building an overlap function from the outputs of the sensors, where this function is resolved at finer scales.

5. OPEN REASEARCH

For single node processing, researcher use mathematical or statistical tools to extract the features that can represent a unique signature for each vehicle or class. Then another tools are used to classify any new utterance to one of the predefined classes. The main goals of this kind of research is either to increase the performance of the classification or decrease the complexity. The relation between performance and complexity is a trade-off relation. Thus the research is open in both areas. Optimal classifier is the classifier that increase the performance and decrease the complexity. Thus, it will be so beneficial to have a standard metrics that have both measures and have a public data base where researcher can evaluate their algorithms based on these metrics using a public data base. Single node processing versus collaborative processing is a hot area of research. Collaborative processing answers the question of how to combine the data and decisions from the sensor nodes in a network to obtain a final decision with optimal resource consumption. Collaborative processing is an open area for research. Fusion modeling, where the signal to noise ratio for both acoustic and communication channel are considered, is critical to find the optimal number of decision makers. Utilization of time and spatial correlation across different nodes is crucial in collaborative processing. All of the above is related to the signal processing techniques that are used for feature extraction, classification, and data and decision fusion. However this is related to the communication protocols that are used in WSN.

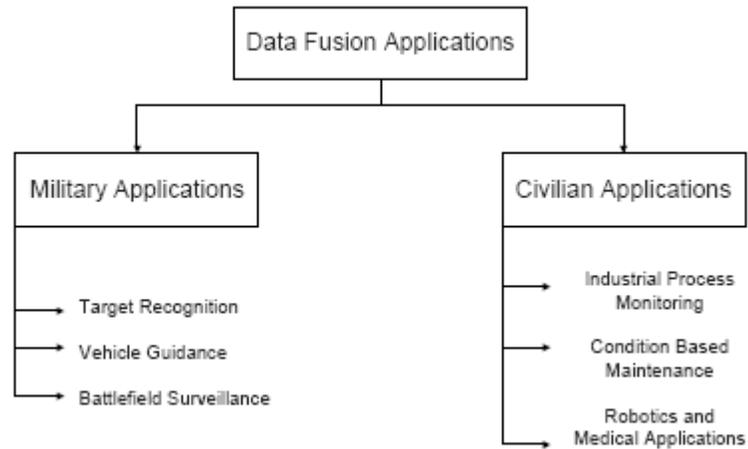


Figure 14: Data fusion applications.

6. CONCLUSIONS

The recent research related to classification process of ground vehicles in wireless sensor network, based on acoustic signals, is reviewed in this paper. Classification process involves three main components: feature extraction, classifier design and selection, and information fusion. Feature extraction methods are classified based on the extraction domain to time, frequency, and time/frequency. Classifiers are also categorized into parametric and non parametric classifiers. In wireless sensor networks decision fusion is preferred rather than data fusion because of the power constraints. This paper proposed two feature extraction methods. One is based on wavelet packet transform and the other is based on spectrum distribution. both methods give almost the same separability and classification rate.

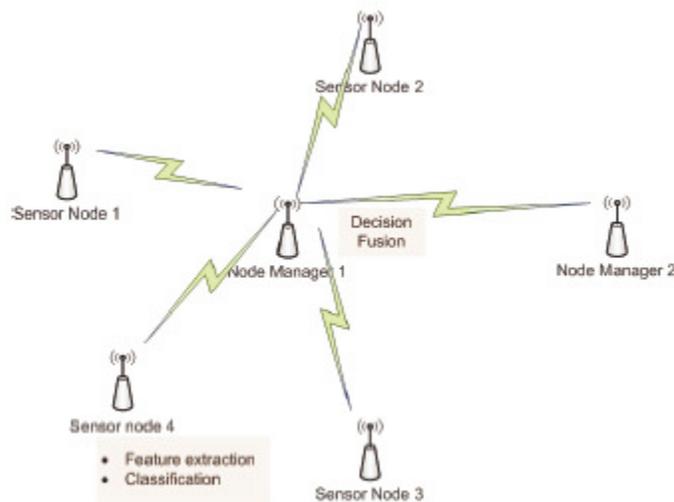


Figure 15: One cluster of a wireless sensor network.

3. REFERENCES

- [1] K. Romer, F. Mattern. "The design space of wireless sensor networks". *IEEE Wireless Communications*, 11(6)54–61, 2004
- [2] T. Arampatzis, J. Lygeros, S. Manesis. "A survey of applications of wireless sensors and wireless sensor networks". In *Intelligent Control, Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation*, 2005
- [3] R. Tan, G. Xing, J. Wang, and H. C. So. "Exploiting reactive collaborative target detection in wireless sensor networks". *IEEE Transactions on Mobile Computing*, 99(1): 5555.
- [4] L. B. Saad and B. Tourancheau. "Multiple mobile sinks positioning in wireless sensor networks for buildings". In *Sensor Technologies and Applications, 2009. SENSORCOMM '09. Third International Conference*, 2009
- [5] R. Zurawski. "Keynote: Wireless sensor network in industrial automation". In *Embedded Software and Systems, ICESS '09. International Conference*, 2009
- [6] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci. "Wireless sensor networks: a survey". *Computer Networks*, 38(4):393–422, 2002. [Online]. Available at: <http://www.sciencedirect.com/science/article/B6VRG-44W46D4-1/2/f18cba34a1b0407e24e97fa7918cdfdc>
- [7] P. Gajbhiye and A. Mahajan. "A survey of architecture and node deployment in wireless sensor network". In *Applications of Digital Information and Web Technologies, ICADIW, First International Conference*, 2008
- [8] T. He, S. Krishnamurthy, J. A. Stankovic, T. Abdelzaher, L. Luo, R. Stoleru, T. Yan, L. Gu, J. Hui, and B. Krogh. "Energy-efficient surveillance system using wireless sensor networks". In *MobiSys '04: Proceedings of the 2nd international conference on Mobile systems, applications, and services*. New York, NY, USA: ACM, 2004
- [9] T. Canli, M. Terwilliger, A. Gupta, and A. Khokhar. "Power - time optimal algorithm for computing fft over sensor networks". In *SenSys '04: Proceedings of the 2nd international conference on Embedded networked sensor systems*. New York, NY, USA: ACM, 2004
- [10] C. F. Chiasserini and R. R. Rao. "On the concept of distributed digital signal processing in wireless sensor networks". In *Proceedings of MILCOM*, 2002
- [11] D. Estrin, L. Girod, G. Pottie, M. Srivastava "Instrumenting the world with wireless sensor networks". In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, 4: 2033–6, 2001
- [12] G. P. Mazarakis and J. N. Avaritsiotis. "Vehicle classification in sensor networks using time-domain signal processing and neural networks". *Microprocess. Microsyst.*, 31(6) 381–392, 2007
- [13] H. Choe, R. Karlsen, G. Gerhart, and T. Meitzler. "Wavelet-based ground vehicle recognition using acoustic signals". In *Proceedings of the SPIE - The International Society for Optical Engineering, Conference Paper, Wavelet Applications III, Orlando, FL, USA, 2762: 434–45, SPIE*, 1996

- [14] A. Khandoker, D. Lai, R. Begg, M. Palaniswami. "Wavelet-based feature extraction for support vector machines for screening balance impairments in the elderly". *Neural Systems and Rehabilitation Engineering*, IEEE Transactions on, 15(4):587–597, 2007
- [15] M. C. Wellman, N. Srour and D. B. Hillis. "Feature extraction and fusion of acoustic and seismic sensors for target identification". In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, G. Yonas, Ed., 3081:139–145, 1997
- [16] G. Succi, T. Pedersen, R. Gampert and G. Prado. "Acoustic target tracking and target identification-recent results". In *Proceedings of the SPIE - The International Society for Optical Engineering*, 3713:10–21, 1999
- [17] M. E. Hohil, J. R. Heberley, J. Chang, A. Rotolo. "Vehicle counting and classification algorithms for unattended ground sensors". E. M. Carapezza, Ed., SPIE, 5090(1):99–110, 2003. Available: <http://link.aip.org/link/?PSI/5090/99/1>.
- [18] M. E. Munich. "Bayesian subspace methods for acoustic signature recognition of vehicles". In *Proc.EUSIPCO*, 2004
- [19] X. Wang H. Qi. "Acoustic target classification using distributed sensor arrays". In *Proc. IEEE ICASSP*, 4:4186–4189, 2002
- [20] H. Wu, M. Siegel, P. Khosla. "Vehicle sound signature recognition by frequency vector principal component analysis". *Instrumentation and Measurement*, IEEE Transactions on, 48(5):1005–1009, 1999
- [21] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed. "Detection, classification and tracking of targets in distributed sensor networks". In *IEEE Signal Processing Magazine*, 2002
- [22] H. Maciejewski, J. Mazurkiewicz, K. Skowron, and T. Walkowiak. "Neural networks for vehicle recognition". In *Proc. 6th Conference on Microelectronics for Neural Networks, Evolutionary and Fuzzy System*, 1997
- [23] W. J. Roberts, H. W. Sabrin, and Y. Ephraim. "Ground vehicle classification using hidden markov models". In *Atlantic coast technologies Inc.*, Silver Spring MD, 2001
- [24] A. Averbuch, E. Hulata, V. Zheludev, and I. Kozlov. "A wavelet packet algorithm for classification and detection of moving vehicles". *Multidimensional Syst. Signal Process*, 12(1):9–31, 2001
- [25] J. E. Lopez, H. H. Chen, and J. Saulnier. "Target identification using wavelet-based feature extraction and neural network classifiers". in *CYTEL SYSTEMS INC HUDSON MA*, 1999
- [26] K. B. Eom. "Analysis of acoustic signatures from moving vehicles using time-varying autoregressive models". *Multidimensional Syst. Signal Process*, 10(4):357–378, 1999
- [27] L. Liu. "Ground vehicle acoustic signal processing based on biological hearing models". Master's thesis, University of Maryland, 1999
- [28] N. B. Thammakhoune and S. W. Lang. "Long range acoustic classification". Sanders a Lockheed Martin Company, Tech. Rep., 1999
- [29] S. Somkiat "Neural fuzzy techniques in vehicle acoustic signal classification". Ph.D.

dissertation, chair-Vanlandingham, Hugh F. 1997

[30] A. Y. Nooralahiyan, H. R. Kirby, D. McKeown. "Vehicle classification by acoustic signature". *Mathematical and computer modeling*, 27:9–11, 1998

[31] M. Baljeet, N. Ioanis, H. Janelle. "Distributed classification of acoustic targets in wireless audio-sensor networks". *Comput. Netw.*, 52(13):2582–2593, 2008

[32] L. Chun-Ting, H. Hong, F. Tao, L. De-Ren, S. Xiao. "Classification fusion in wireless sensor networks". *ACTA AUTOMATICA SINICA*, 32(6):948–955, 2006

[33] S. S. Yang, Y. G. Kim¹, H. Choi. "Vehicle identification using discrete spectrums in wireless sensor networks". *Journal Of Networks*, 3(4):51–63, 2008

[34] B. Malhotra, I. Nikolaidis, and M. Nascimento. "Distributed and efficient classifiers for wireless audio-sensor networks". In *Networked Sensing Systems*, 2008. INSS 2008. 5th International Conference on, 2008

[35] M. F. Duarte, Y. H. Hu. "Vehicle classification in distributed sensor networks". *Journal of Parallel and Distributed Computing*, 64(7):826–838, 2004, computing and Communication in Distributed Sensor Networks. Available at: <http://www.sciencedirect.com/science/article/B6WKJ-4CXD0JJ-1/2/64f671263463155e2afd7bf778c3a7dd>

[36] B. Guo, M. Nixon, and T. Damarla. "Acoustic information fusion for ground vehicle classification". In *Information Fusion*, 2008 11th International Conference, 2008

[37] R. D. T., P. Tien, and L. Douglas. "An algorithm for classifying multiple targets using acoustic signatures", 2004

[38] H. Xiao¹, Q. Yuan¹, X. Liu¹, and Y. Wen. "Advanced Intelligent Computing Theories and Application, with Aspects of Theoretical and Methodological Issue." Springer Berlin / Heidelberg, 2007

[39] A. Amir, Z. V. A., R. Neta, S. Alon "Wavelet-based acoustic detection of moving vehicles". *Multidimensional Syst. Signal Process.*, 20(1):55–80, 2009

[40] H. Qi, X. Tao, and L. H. Tao. "Vehicle classification in wireless sensor networks based on rough neural network". In *ACST'06: Proceedings of the 2nd IASTED international conference on Advances in computer science and technology*. Anaheim, CA, USA: ACTA Press, 2006

[41] D. Li, K. Wong, Y. H. Hu, and A. Sayeed. "Detection, classification, and tracking of targets". *Signal Processing Magazine, IEEE*, 19(2):17–29, 2002

[42] Q. Xiao-xuan, J. Jian-wei, H. Xiao-wei and Y. Zhong-hu. "An approach of passive vehicle type recognition by acoustic signal based on svm". 2009

[43] Y. Kim, S. Jeong, D. Kim, T. S. Lopez. "An efficient scheme of target classification and information fusion in wireless sensor networks". *Personal Ubiquitous Comput.*, 13(7): 499–508, 2009

[44] Y. Sun and H. Qi. "Dynamic target classification in wireless sensor networks". In *Pattern Recognition, ICPR, 19th International Conference*, 2008

- [45] W. Duan, M. He, Y. Chang and Yan Feng. "Acoustic objective recognition in wireless sensor networks". In 2009 4th IEEE Conference on Industrial Electronics and Applications, 2009
- [46] R. Mgya, S. Zein-Sabatto, A. Shirkhodaie, and W. Chen. "Vehicle identifications using acoustic sensing". In SoutheastCon, 2007 Proceedings. IEEE, 2007
- [47] P. Qiang, W. Jianming, C. Hongbing, L. Na, and L. Haitao. "Improved ds acoustic-seismic modality fusion for ground-moving target classification in wireless sensor networks". Pattern Recogn. Lett., 28(16):2419–2426, 2007
- [48] S. Erb. "Classification of vehicles based on acoustic features". Master's thesis, Begutachter: Univ.-Prof. Dipl.-Ing. Dr. Bernhard Rinner, 2007
- [49] A. Aljaafreh and L. Dong. "An evaluation of feature extraction methods for vehicle classification based on acoustic signals". In Networking, Sensing and Control (ICNSC), 2010 International Conference, 2010
- [50] D. Lake. "Harmonic phase coupling for battlefield acoustic target identification". In Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on, 4:2049–2052, 1998
- [51] Y. Kaia, H. Qia, W. Jianminga, L. Haitao. "Multiple vehicle signals separation based on particle filtering in wireless sensor network". Journal of Systems Engineering and Electronics, 19(3):440–446, 2008
- [52] M. E. Munich. "Bayesian subspace methods for acoustic signature recognition of vehicles". In 12th European Signal Processing Conf), 2004
- [53] H.-l. Wang, W. Yang, W.-d. Zhang, and Y. Jun. "Feature extraction of acoustic signal based on wavelet analysis". In ICESSYMPPOSIA '08: Proceedings of the 2008 International Conference on Embedded Software and Systems Symposia. Washington, DC, USA: IEEE Computer Society, 2008
- [54] X. Shao, L. Sun. "An application of the continuous wavelet transform to resolution of multicomponent overlapping analytical signals". Analytical Letters, 34(2):267–280, 2001 [Online]. Available at: <http://dx.doi.org/10.1081/AL-100001578>
- [55] H. C. Choe. "Signature detection, recognition, and classification using wavelets in signal and image processing". Ph.D. dissertation, Texas A&M University, Department of Electrical Engineering, 1997
- [56] R. Karlsen, T. Meitzler, G. Gerhart, D. Gorsich, and H. Choe. "Comparative study of wavelet methods for ground vehicle signature analysis". In Proceedings of the SPIE - The International Society for Optical Engineering, 2762: 314–24, 1996
- [57] Wang, Lipo, Fu, and Xiuju. "Data Mining with Computational Intelligence". Springer Berlin Heidelberg, 2005
- [58] M. Wichli and T. Braun. "Event classification and filtering of false alarms in wireless sensor networks". Parallel and Distributed Processing with Applications, International Symposium on, 0:757–764, 2008
- [59] D. Janakiram, V. A. M. Reddy and A. P. Kumar. "Outlier detection in wireless sensor networks using bayesian belief networks". In Communication System Software and

Middleware, First International Conference, 2006

[60] D. S. Kim, M. A. Azim, and J. S. Park. "Privacy preserving support vector machines in wireless sensor networks". In ARES '08: Proceedings of the 2008 Third International Conference on Availability, Reliability and Security Washington, DC, USA: IEEE Computer Society, 2008

[61] X. Wang, D.-w. Bi, L. Ding, and S. Wang. "Agent collaborative target localization and classification in wireless sensor networks." *Sensors*, 7(8):1359–1386, 2007. [Online]. Available at: <http://www.mdpi.com/1424-8220/7/8/1359>

[62] D. Tran and T. Nguyen. "Support vector classification strategies for localization in sensor networks". In Communications and Electronics, 2006. ICCE '06. First International Conference, 2006

[63] L. Yip, K. Comanor, J. C. Chen, R. E. Hudson, K. Yao, and L. Vandenberghe. "Array processing for target doa, localization, and classification based on aml and svm algorithms in sensor networks". In AML and SVM Algorithms in Sensor Networks, Proc. of the 2nd International Workshop on Information Processing in Sensor Networks (IPSN03, 2003

[64] H. Wu and J. M. Mendel. "Classification of battlefield ground vehicles using acoustic features and fuzzy logic rule-based classifiers". *Fuzzy Systems, IEEE Transactions* 15(1):56–72, 2007

[65] S. Thoerodoridis and K. Koutroumbas. "Pattern Recognition". Elsevier Inc., 2009

[66] B. Lantow. "Applying distributed classification algorithms to wireless sensor networks a brief view into the application of the sprint algorithm family". In Networking, ICN, Seventh International Conference, 2008

[67] C. Meesookho, S. Narayanan, and C. Raghavendra. "Collaborative classification applications in sensor networks". In Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002

[68] A. Prieto, C. G. Puntonet, B. Prieto, and M. Rodríguez-A´lvarez. "A competitive neural network for blind separation of sources based on geometric properties". In IWANN '97: Proceedings of the International Work-Conference on Artificial and Natural Neural Networks. London, UK: Springer-Verlag, 1997

[69] Y.-W. Chen, X.-Y. Zeng, and Z. Nakao. "Blind separation based on an evolutionary neural network". *Pattern Recognition, International Conference on*, (2): p. 2973, 2000

[70] M. Solazzi and A. Uncini. "Spline neural networks for blind separation of post-nonlinear-linear mixtures". *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 51(4):817–829, 2004

[71] J.-M. Ye, X.-L. Zhu, and X.-D. Zhang. "Adaptive blind separation with an unknown number of sources". *Neural Comput.*, 16(8):1641–1660, 2004

[72] T.-Y. Sun, C.-C. Liu, S.-J. Tsai, and S.-T. Hsieh. "Blind source separation with dynamic source number using adaptive neural algorithm". *Expert Syst. Appl.*, 36(5):8855–8861, 2009

[73] W. Penny, S. Roberts, and R. Everson. "Hidden markov independent components for biosignal analysis". In *Advances in Medical Signal and Information Processing*, 2000. First International Conference on (IEE Conf. Publ. No. 476), 2000

- [74] X. Wang, H. Qi, and H. Du. "Distributed source number estimation for multiple target detection in sensor networks". In *Statistical Signal Processing, 2003 IEEE Workshop*, 2003
- [75] Y. Kai, H. Qi, W. Jianming, and L. Haitao. "Multiple vehicle signals separation based on particle filtering in wireless sensor network". *Journal of Systems Engineering and Electronics*, 19(3):440-446, 2008. [Online]. Available at :<http://www.sciencedirect.com/science/article/B82XK-4SVTN2V-5/2/379989fc284a479b6102967e30b0769a>
- [76] H. Chen, C. K. Tse, and J. Feng "Source extraction in bandwidth constrained wireless sensor networks". *IEEE TRANSACTION ON CIRCUITS AND SYSTEMS-II:EXPRESS BRIEFS*, 55(9):947–951, 2008
- [77] H. Qi, X. Tao, and L. H. Tao. "Multiple target recognition based on blind source separation and missing feature theory". In *Computational Advances in Multi-Sensor Adaptive Processing 1st IEEE International Workshop on Volume*, 2005
- [78] F. Silva, J. Heidemann, R. Govindan, and D. Estrin. "Frontiers in Distributed Sensor Networks". CRC Press, Inc., 2003
- [79] X. Wang, H. Qi, and H. Du. "Distributed source number estimation for multiple target detection in sensor networks". In *Statistical Signal Processing, 2003 IEEE Workshop*, 2003
- [80] R. J. Weiss and D. P. W. Ellis. "Estimating single-channel source separation masks: Relevance vector machine classifiers vs. pitch-based masking". In *Proceedings of the ISCA Tutorial and Research Workshop on Statistical and Perceptual Audition (SAPA)*, 2006
- [81] R. Braunling, R. M. Jensen, M. A. Gallo. "Acoustic target detection, tracking, classification, and location in a multiple-target environment". G. Yonas, Ed., *SPIE*, 3081(1): 57–66, 1997 [Online]. Available at: <http://link.aip.org/link/?PSI/3081/57/1>
- [82] E. Drakopoulos, J. J. Chao, C. C. Lee. "A two-level distributed multiple hypothesis decision system". *37(3):380–384*, 1992
- [83] J. H. Kotecha, V. Ramachandranand, A. M. Sayeed. "Distributed multitarget classification in wireless sensor networks". *23(4):703–824*, 2005
- [84] E. F. Nakamura, A. A. F. Loureiro, A. C. Frery. "Information fusion for wireless sensor networks: Methods, models, and classifications". *ACM Comput. Surv.*, 39(3):9, 2007
- [85] R. Brooks, P. Ramanathan, A. Sayeed. "Distributed target classification and tracking in sensor networks". *Proceedings of the IEEE*, 91(8):1163–1171, 2003
- [86] A. Aljaafreh and L. Dong. "Hidden markov model based classification approach for multiple dynamic vehicles in wireless sensor networks". In *Networking, Sensing and Control (ICNSC), 2010 International Conference*, 2010
- [87] J. Llinas and D. Hall. "An introduction to multi-sensor data fusion". In *Circuits and Systems, ISCAS '98. Proceedings of the 1998 IEEE International Symposium on* 6:1998
- [88] I. Liggins, M.E., C.-Y. Chong, I. Kadar, M. Alford, V. Vannicola, S. Thomopoulos. "Distributed fusion architectures and algorithms for target tracking". *Proceedings of the IEEE*, 85(1):95–107, 1997

- [89] Z. H. Kamal, M. A. Salahuddin, A. K. Gupta, M. Terwilliger, V. Bhuse, and B. Beckmann. "Analytical analysis of data and decision fusion in sensor networks". In *ESA/VLSI*, 2004
- [90] G. Xing, R. Tan, B. Liu, J. Wang, X. Jia, and C.-W. Yi. "Data fusion improves the coverage of wireless sensor networks". In *MobiCom '09: Proceedings of the 15th annual international conference on Mobile computing and networking*. New York, NY, USA: ACM, 2009
- [91] D. J. Miller, Y. Zhang, G. Kesidis. "Decision aggregation in distributed classification by a transductive extension of maximum entropy/improved iterative scaling." *EURASIP Journal on Advances in Signal Processing*, 21: 2008
- [92] A. D'Costa, V. Ramachandran, A. Sayeed. "Distributed classification of gaussian space-time sources in wireless sensor networks". *Selected Areas in Communications, IEEE Journal on*, 22(6):1026–1036, 2004
- [93] A. Aljaafreh and L. Dong. "Cooperative detection of moving targets in wireless sensor network based on fuzzy dynamic weighted majority voting decision fusion". In *Networking, Sensing and Control (ICNSC), International Conference*, 2010
- [94] J. Kittler, M. Hatef, R. P. Duin, and J. Matas. "On combining classifiers". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3): 226–239, 1998
- [95] X. Wang and S. Wang. "Collaborative signal processing for target tracking in distributed wireless sensor networks". *J. Parallel Distrib. Comput.*, 67(5):501–515, 2007
- [96] D. Hall, J. Llinas. "An introduction to multisensor data fusion". *Proceedings of the IEEE*, 85(1)6–23,1997
- [97] A. Sinha, H. Chen, D. Danu, T. Kirubarajan, and M. Farooq. "Estimation and decision fusion: A survey." *Neurocomputing*, 71(13-15):2650– 2656, 2008, *artificial Neural Networks (ICANN 2006) / Engineering of Intelligent Systems (ICEIS 2006)*. [Online]. Available at: <http://www.sciencedirect.com/science/article/B6V10-4SFS0KH-8/2/f59b9b186c4bbbfe157bf08a07f72c4f>.
- [98] E. Drakopoulos, J. Chao, and C. Lee. "A two-level distributed multiple hypothesis decision system". *Automatic Control, IEEE Transactions on*, 37(3):380–384,1992
- [99] A. Speranzon, C. Fischione, and K. Johansson. "Distributed and collaborative estimation over wireless sensor networks". In *Decision and Control, 45th IEEE Conference*, 2006

Effective Preprocessing in Modeling Head-Related Impulse Responses Based on Principal Components Analysis

Hugeng

*Department of Electrical Engineering
University of Indonesia
Depok, 16424, Indonesia*

hugeng@ui.ac.id

Wahidin Wahab

*Department of Electrical Engineering
University of Indonesia
Depok, 16424, Indonesia*

wahidin.wahab@ui.ac.id

Dadang Gunawan

*Department of Electrical Engineering
University of Indonesia
Depok, 16424, Indonesia*

guna@eng.ui.ac.id

Abstract

It was found in previous works in modeling head-related impulse responses (HRIRs) using principal components analysis (PCA), both in frequency and time domain, that different sets of measured HRIRs were used, which were obtained from measurements by various institutions involving different kinds of subjects of human being, anesthetized live cat and acoustic manikin. Groups of researchers also applied different number of basis functions resulted from PCA, i.e. 4 – 10 basis functions. Then, the performance of the models was tested using different parameters, i.e. spectral distortion score and mean square error (MSE). Since there were varied factors mentioned above, a fair comparison among these models is difficult to achieve. Using PCA, we modeled the original HRIRs, minimum-phase HRIRs, direct-pulse HRIRs, normalized HRIRs in the time domain. However, in frequency domain, the models of magnitude head-related transfer functions (HRTFs), log-magnitude HRTFs, standardized log-magnitude HRTFs were performed. We performed a comprehensive comparison of various types of preprocessing of the previous data types in modeling HRIRs based on PCA using ten basis functions, CIPIC HRTF Database, and MSE. Our results showed that models of magnitude HRTFs had overall smallest average MSE. On the other hand, the best models in time domain were achieved from minimum-phase HRIRs.

Keywords: HRIR Model, HRTF Model, Principal Components Analysis.

1. INTRODUCTION

Human being can employ two ears to distinguish directions of various sound sources even by visually impaired people. There are three primary cues in perceiving the direction of a sound source, i.e. interaural time difference (ITD), interaural level difference (ILD), and spectral modification caused by pinnae, head, and torso. These primary sound cues are encrypted in binaural Head-Related Transfer Functions (HRTFs). HRTF is defined as the transfer function of the acoustic filter of human auditory system, in frequency domain, from a sound source to the entrance of ear canal. The counterpart of HRTF in time domain is known as head-related impulse response (HRIR). By convolving binaural HRIRs with a sound signal, a listener can localize the direction of the sound. Spatial sound synthesis has been widely utilized in various fields because of the great supports of digital signal processing in headphone system and multichannel speaker system. To implement binaural HRTFs in the creation of virtual auditory space, control of ITD, ILD, and spectral modification is the most essential part to give information of sound direction. On the horizontal plane, ITD and ILD play the main role in perceiving the direction of the sound source [1]. Although these two cues are remaining almost the same on the median plane, perception of sound direction is much affected by spectral modification mainly due to reflection and diffraction of pinnae folds [2].

HRIRs are resulted by reflections and diffractions of sound wave by human body, therefore, HRIRs vary significantly from subject to subject [3-4]. Poor vertical effect and front-back reversal were observed when non-individualized HRIRs were used to generate spatial sound image [3]. Application of individual HRIRs will be the most appropriate approach to create precise localization cues. However, measuring individual HRTFs of a listener is a time consuming process and economically expensive. Recently researchers have studied many other approaches to overcome the individual variability of HRIRs without directly measuring individual HRIRs. One effective and simple approach is modeling the HRIRs which includes several parameters that can be adapted by anthropometric measurements of a listener to produce his / her individualized HRIRs.

Several authors had proposed functional models of HRTFs [5-7]. They sought a mathematical equation that represented the HRTF as a function of frequency and direction, such that the models could provide explicit mathematical relationship between the HRTF and source location. Functional approaches would reduce the storage requirement and represent the HRTF at an arbitrary direction. However, attempts to describe the HRTF in a simple mathematical equation have been of only limited success. Following the approach as in [6], Algazi et al. proposed a structural model of HRTF, which attempted to relate functional HRTF to anthropometric measurements [8].

In this research, we concerned on the low-dimensional and orthogonal models for a set of HRIRs which have been generated by using the Karhunen-Loeve expansion or Principal Components Analysis (PCA). Kistler and Wightman [4] proposed a model based on PCA and minimum-phase reconstruction. The set of HRIRs was preprocessed to result in a set of logarithms of HRTF magnitudes. Then, PCA was employed to the logarithms of the HRTF magnitudes after the removal of direction-independent and subject-dependent spectral features. Middlebrooks and Green [9] and Hu et al. [10-11], also applied the same preprocessing to the HRIRs data as in [4]. Another research was the spatial feature extraction and regularization model for the HRTFs proposed by Chen et al. [12]. They preprocessed the set of HRIRs to become a set of complex-valued HRTFs. Using PCA, the models of HRTFs were expressed as weighted combinations of a set of complex-valued eigen transfer functions. The sample weights were determined by projecting all measured complex-valued HRTFs onto the eigen transfer functions. A functional representation for weights was attained by applying a thin plate generalized spline smoothing model to regularize the sample weights. This approach maintained the phase of the spectral components and model accuracy at a whole upper 3/4 sphere but it dealt with large amounts of complex-valued computation in matrix-vector products and two-dimensional splines.

Another group of researchers, Inoue et al. [13], attempted to estimate HRTFs using anthropometric measurements. They modeled the magnitude HRTFs using PCA and the weights were estimated by a regression model using anthropometric measurements. The estimated HRTF of the main response was reconstructed as a minimum-phase response. Then, ITDs were modeled using multiple regression analysis from anthropometric measurements. The complete models of HRTFs for both ears were reconstructed by combining the magnitudes model and ITDs model. They evaluated the estimation performance using a spectral distortion (SD) score. Our previous work [14] has similar approach as [13] in preprocessing the set of HRIRs. However, we evaluated the model performance using mean square error (MSE). Xu et al. [15] also proposed the modeling of HRIRs in the frequency domain. They preprocessed the database of HRIRs into a set of standardized log-magnitude HRTFs. Because of the variations of HRTFs, it is necessary to standardize them before performing PCA. The log-magnitude HRTFs were standardized by subtracting the mean and dividing by the standard deviation. This preprocessing was different from that of [4] only at the dividing by standard deviation. They developed a method to individualize HRTFs using anthropometric measurements. They compared the existing [4],[13] and their improved methods by a paired-sample *t*-test. The performance of the HRTFs models was also measured using SD score. Using PCA, Sodnik and Tomazic [16] modeled the linear amplitude spectra of HRTF, called magnitude HRTF by other researchers, thus excluding the phase spectrum and ITD. They concentrated on the azimuth perception on the horizontal plane. The performance of the HRTF models was measured by subjective localization tests from 15 subjects.

Similar work of modeling HRTF using PCA in the time domain was performed by Wu et al. [17]. Each HRIR of the set of all HRIRs used in their model was normalized by its level gain. They defined the square root of the energy of HRIR as its level gain. The normalized HRIRs, having same energy and onset time, would have different spectral characteristics. The variances of HRIRs are decreased by these normalizing procedures. The object for measuring HRIRs was an anesthetized live cat. They resulted in that the modeled HRIRs were nearly identical to the measured HRIRs. Other works were proposed in [18-19]. Before PCA, they preprocessed the median HRIRs to remove the initial time delay and to extract the early response that lasted for certain time interval since the arrival of the direct pulse. Shin and Park [18] extracted only the response of pinna with length of about 0.23 ms (10 samples with sampling frequency, f_s , of 44,100 Hz). In spite of this, Hwang and Park [19] included also the response of head and torso. They employed 1.5 ms of HRIR (67 samples with same f_s). They individualized the resulted HRIR model subjectively and then tested the individualization method by subjective listening tests. We also proposed an individualization method for HRIR model using PCA based on multiple regression analysis [20]. We preprocessed the measured HRIRs from CIPIC HRTF Database into minimum-phase HRIRs using cepstrum analysis. We tested our individualization method both objectively and subjectively. The performance of our HRIRs models was tested objectively using MSE as used in [4],[17].

The above mentioned previous works in modeling HRIRs, both in frequency and time domain, used different sets of HRIRs, which were obtained from measurements by various institutions using subjects of human being, anesthetized live cat and acoustic manikin. In modeling HRIRs using PCA, they also used different number of basis functions, 4-10 basis functions. Also the performance of the models was tested using different parameters, i.e. SD score and MSE. Due to the varied subjects, sets of HRIRs, number of basis functions used in PCA and performance parameters used, a fair comparison among these models is difficult to obtain.

2. MODELING HRIRs BASED ON PCA

This paper presents a comprehensive comparison of various preprocessings of a set of HRIRs in modeling HRIRs based on PCA using ten basis functions and CIPIC HRTF Database [21]. The goal of our work is to obtain an effective preprocessing in modeling HRIRs based on PCA by analyzing this comparison.

Types of Preprocessing Used

In this research, we used the following data types in modeling HRIRs, in the time domain, i.e. original HRIRs (directly from database, without preprocessing), direct-pulse HRIRs or HRIRs with initial time delay removed [18-19], minimum-phase HRIRs [20], and normalized HRIRs [17]. In spite, in the frequency domain, the data types used in modeling HRIRs were magnitude HRTFs [13-14],[16], log-magnitude HRTFs [4],[9-11], and standardized log-magnitude HRTFs [15]. We defined the above terms of data types to avoid ambiguity of possibly different terms used among groups of researchers. Each preprocessing involved in achieving related data type is explained in more detail in the given references.

As proposed in [19], direct-pulse HRIRs were obtained before PCA by removing the initial time delay and extracting the early samples that lasted for 1.5 ms or 67 samples with $f_s = 44,100$ Hz since the arrival of direct pulse. The initial time delay of HRIR indicates the propagation time of sound from sound source to eardrum. If it is required later, it can be re-inserted afterwards. The response of 1.5 ms includes the effects of pinna, head and torso. As explained in [14], the minimum-phase HRIR can be obtained through the calculation of real cepstrum of its original HRIR, which has arbitrary phase. It can be said that the minimum-phase HRIR is the removed initial time delay version of HRIR, similar with the direct-pulse HRIR. Original HRIR and its correspond minimum-phase HRIR have the same magnitude spectrum in the frequency domain. We took only 67 first samples out of 200 samples in the minimum-phase HRIR. Thus, in both types of preprocessing, the size of dataset to be analyzed in PCA was reduced without loss of meaningful information by the preprocessing.

Due to the head shadow effect, each HRIR has different energy which equals to the sum-square of samples of that HRIR. The square root of the energy was defined in [17] as the level gain of the HRIR. The normalized HRIRs were obtained by normalizing all original HRIRs by their correspond level gains. The normalized HRIRs, having the same energy and onset time, would have different spectral characteristics. The normalizing procedures decrease the variances of HRIRs. We took only 120 samples of the original HRIRs and the normalized HRIRs for modeling in PCA because these responses already include responses of pinna, head, and torso, on other side, the level of rest samples is very small and not significant.

Complex HRTFs were attained by implementing 256-points fast Fourier transform (FFT) to all original HRIRs from the database used. The magnitude HRTFs are then the absolute values of these complex HRTFs. Only 128 first frequency components of a magnitude HRTF, $|HRTF|$, were taken into analysis because of the symmetry property of a magnitude spectrum. Then, we defined a log-magnitude HRTF as twenty times the base-10 logarithm of its magnitude HRTF ($20 \log_{10} |HRTF|$). Finally, a standardized log-magnitude HRTF was obtained by standardizing or normalizing a log-magnitude HRTF with its standard deviation.

Principal Components Analysis

We explain below only the modeling of magnitude HRTFs using PCA with the same number of basis functions, subjects, and sound directions used as in the modeling of other data. We took also only 128 first frequency components of log-magnitude HRTF and standardized log-magnitude HRTF for analysis in PCA because of the same reason as before. The variable N below is equal 120 in modeling original HRIRs and normalized HRIRs and is replaced by 67 when modeling minimum-phase HRIRs and direct-pulse HRIRs. In modeling other data than magnitude HRTFs, the data of magnitude HRTFs is simply replaced by the desired data.

The entire magnitude HRTFs were computed from left-ear and right-ear HRIRs of 45 subjects from all sound sources with 1250 directions in sphere. There are 25 azimuths and 50 elevations of sound sound directions for each ear of a subject, so that a total of 112,500 magnitude HRTFs were produced. A matrix composed of DTFs is needed by PCA. The original data matrix, \mathbf{H} ($N \times M$), is composed of magnitudes of HRTFs, in which, each column vector, \mathbf{h}_i ($i=1,2,\dots,M$), represents a magnitude HRTF of an ear of a subject in a direction in sphere. The number of magnitude HRTFs

of each subject is 2500 (2 ears x 1250 directions). Hence, the size of \mathbf{H} is 128 x 112,500 (N=128, M=112,500). The empirical mean vector ($\boldsymbol{\mu}$: Nx1) of all magnitude HRTFs is given by,

$$\boldsymbol{\mu} = (1/M) \sum_{i=1}^M \mathbf{h}_i. \quad (1)$$

The DTFs matrix, \mathbf{D} , is the mean-subtracted matrix and is given by,

$$\mathbf{D} = \mathbf{H} - \boldsymbol{\mu} \cdot \mathbf{y}, \quad (2)$$

where \mathbf{y} is a 1xM row vector of all 1's. The next step is to compute a covariance matrix, \mathbf{S} , that is given by

$$\mathbf{S} = \mathbf{D} \cdot \mathbf{D}^* / (M-1) \quad (3)$$

where * indicates the conjugate transpose operator. The basis functions or principal components (PCs), \mathbf{v}_i ($i=1,2,\dots,q$), are the q eigenvectors of the covariance matrix, \mathbf{S} , corresponding to q largest eigenvalues. If $q = N$, then the DTFs can be fully reconstructed by a linear combination of the N PCs. However, q is set smaller than N because the goal of PCA is to reduce the dimension of dataset. An estimate of the original dataset is obtained here by only 10 PCs, which account for more than 90% (exactly 94.30%) variance in the original data \mathbf{D} . By using only 10 PCs to model magnitude HRTFs, we expected to obtain satisfactory good results. The PCs matrix, $\mathbf{V} = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_N]$, that consisted of complete set of PCs can be obtained by solving the following eigen equation,

$$\mathbf{S} \mathbf{V} = \boldsymbol{\Lambda} \mathbf{V} \quad (4)$$

where $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_{128}\}$, is a diagonal matrix formed by 128 eigen values, where each eigen value, λ_i , represents sample variance of DTFs that was projected onto i-th eigen vektor or PC, \mathbf{v}_i .

Then, the weights of PCs (PCWs), \mathbf{W} (10x112,500), that correspond to all DTFs, \mathbf{D} , can be obtained as,

$$\mathbf{W} = \mathbf{V}^* \cdot \mathbf{D}, \quad (5)$$

where PCs matrix now was reduced to $\mathbf{V} = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_{10}]$. PCWs represent the contribution of each PC to a DTF. They contain both the spatial features and the inter-individual difference of DTF. Thus, the matrix consisted of models of magnitude HRTFs, $\hat{\mathbf{H}}$, is given by,

$$\hat{\mathbf{H}} = \mathbf{V} \cdot \mathbf{W} + \boldsymbol{\mu} \cdot \mathbf{y}. \quad (6)$$

The performance of the models of magnitude HRTFs, resulting from PCA, was evaluated by comparing the mean-square error of the disparity between the approximated magnitude HRTFs and the original magnitude HRTFs calculated from database, to the mean-square error of the original magnitude HRTFs in percentage, which is stated as

$$e_j(\theta, \phi) = 100 \% \times \frac{\|\mathbf{h}_j(\theta, \phi) - \hat{\mathbf{h}}_j(\theta, \phi)\|^2}{\|\mathbf{h}_j(\theta, \phi)\|^2} \quad (7)$$

where $\mathbf{h}_j(\theta, \phi)$ is the j-th original magnitude HRTF in the direction with azimuth, θ , and elevation, ϕ , $\hat{\mathbf{h}}_j(\theta, \phi)$ is the corresponding approximated or model of magnitude HRTF, $\mathbf{h}_j(\theta, \phi)$. As the error increases, the performance of the model of magnitude HRTF deteriorates. On the contrary better localization results will be achieved with small error, $e_j(\theta, \phi)$, which is called MSE by some researchers [4],[17].

3. EXPERIMENTS' RESULTS AND ANALYSES

Analysis of Percentage Cumulative Variance and Average MSE

Table 1 shows the percentage cumulative variance of DTFs or direct impulse responses (DIRs) of each data type in the database explained by PC-1 (\mathbf{v}_1) until PC-2, PC-5, PC-10, and PC-13 (\mathbf{v}_2 , \mathbf{v}_5 , \mathbf{v}_{10} , \mathbf{v}_{13}) respectively. Percentage cumulative variance is obtained from the percentage of cumulative sum of first largest eigen values that correspond to first PC until a particular PC, compared to the total cumulative sum of all PCs. As we can see, some cumulative variances of PCs from several data types do not exceed 90% for first 10 PCs, especially worse from original HRIRs and normalized HRIRs. This fact influences the performance of models of those data types. The average MSE across sound directions and subjects of the models of original HRIRs was only 48.5% using 10 PCs, similarly that of normalized HRIRs was only 32.0%. These bad results due to the time delay and many details included in those data types. It is quite difficult for PCA to estimate the time delay and details in those HRIRs.

By using PCA with 10 PCs, the best result for modeling HRIRs in time domain was achieved from minimum-phase HRIRs. The cumulative variance and average MSE were 90.40% and 7.26% respectively. In the frequency domain, we attained best result from magnitude HRTFs with 94.30% of cumulative variance and only 3.30% of average MSE. Much smaller average MSEs are observed from models in frequency domain since they have smoother spectra than impulse responses in time domain.

By using the same setup for PCA, the same database, and the same performance parameter as in [18], our models of minimum-phase HRIRs in median plane outperformed the models of direct-pulse HRIRs in median plane as proposed by Hwang and Park [18]. Using 10 PCs, their models had average MSE of about 6.67% [18] compared to the average MSE of our models of 5.31%. We had also performed the individualization of magnitude HRTFs in horizontal plane using multiple linear regression (MLR) [14]. Comparable work was done by Hu et al. [10]. However, they performed the individualization of log-magnitude HRTFs in the horizontal plane. As we explained in [14], by comparing the performance shown in [10], our individualization of magnitude HRTFs was much better than that of their individualization of log-magnitude HRTFs. We believed that the selection of preprocessing used in our work supported the better results.

The application of more PCs would reduce the modeling error between each data type of HRTF in database and its model, however, it costed more computing time and larger memory space. The PCs-matrix, \mathbf{V} , that at first has 128x128 elements was reduced into a matrix of only 128x10 elements in the cases of HRTFs modeling since we used only the first 10 PCs out of all 128 PCs. Thus, we needed only 10 PCWs to perform the model. One can see obviously the advantage of PCA in reducing significantly the memory space needed.

Data Type	Cumulative Variance (%)				Average MSE (%)			
	2PCs	5PCs	10PCs	13PCs	2PCs	5PCs	10PCs	13PCs
Original HRIRs	27.98	50.90	70.89	79.16	83.41	66.25	48.50	41.49
Direct pulse HRIRs	47.02	73.07	89.15	92.91	42.02	20.97	8.91	6.24
Minimum phase HRIRs	56.98	77.89	90.40	93.57	28.97	18.44	7.26	4.82
Normalized HRIRs	25.97	46.41	67.26	75.18	72.37	52.38	32.00	24.26
Magnitude HRTFs	72.61	86.94	94.30	96.13	11.35	6.95	3.30	2.35
Log-magnitude HRTFs	76.46	85.91	92.38	94.24	17.81	10.17	5.17	3.70
Standardized log-magnitude HRTFs	73.20	81.48	88.69	91.16	18.61	12.46	7.43	5.71

TABLE 1: Percentage Cumulative Variance Explained and Average MSE across Directions and Subjects.

Because of lack of space in this paper, the following results' analyses were performed only for models of minimum-phase HRIRs and of magnitude HRTFs using only 10 PCs and the corresponding MSEs in the sphere from Subject-003 of CIPIC HRTF Database.

Analysis of Basis Functions of PCA in Modeling Minimum Phase HRIRs

The modeling method, PCA, of minimum-phase HRIRs was performed as explained in the following. At the beginning, all original HRIRs in the database (with total number of 112,500 HRIRs) were converted into a set of minimum-phase HRIRs using the method explained in [19]. Mean of these HRIRs was then calculated. This mean was subtracted from each minimum-phase HRIR to give corresponding minimum-phase direct impulse response (DIR). All minimum-phase DIRs were applied to compute samples covariance matrix, \mathbf{S} , as shown in Eq. 3. By solving the eigen equation (Eq. 4) that involving \mathbf{S} , we obtained matrix of basis functions, \mathbf{V} , that consisted of 67 complete basis functions, i.e. $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_{67}]$. A perfect modeling of minimum-phase DIRs was achieved by employing complete basis functions. But we used only 10 basis functions (PCs) which explained 90.4% variance of all minimum-phase DIRs.

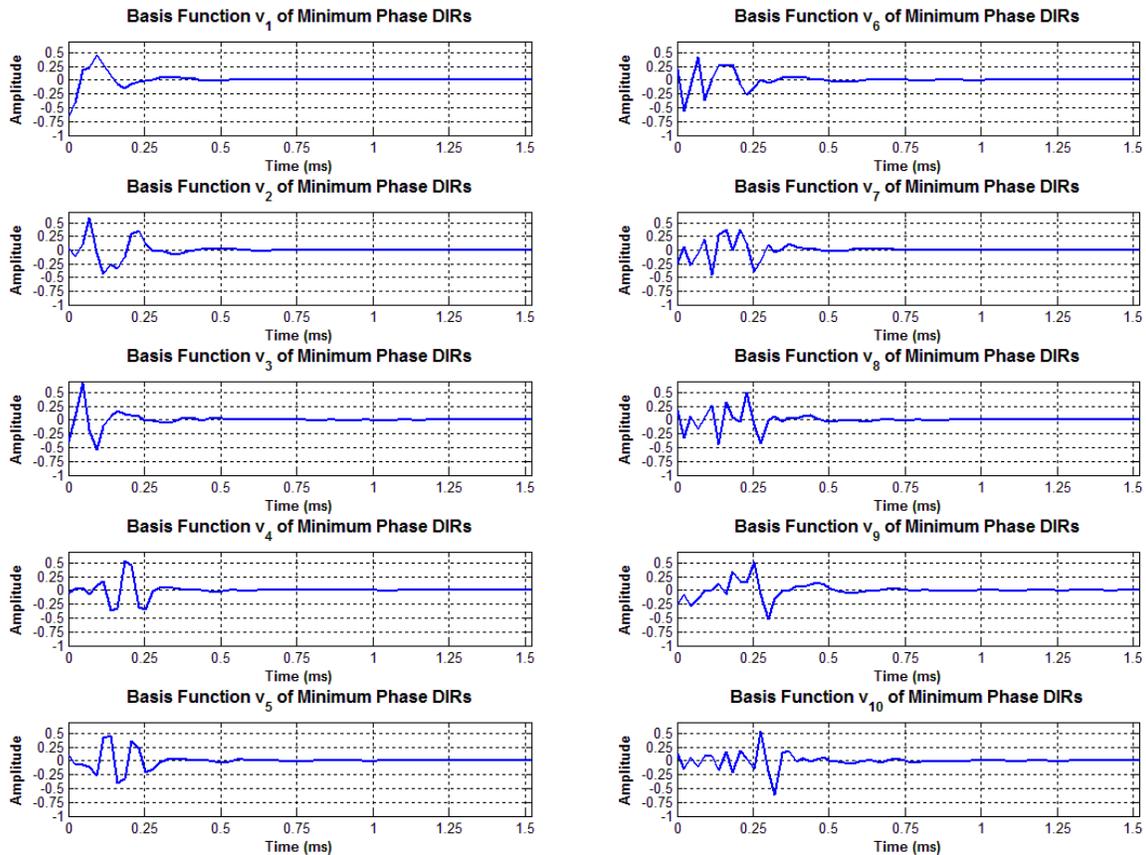


FIGURE 1: The First Ten Basis Functions Extracted from PCA in Modeling Minimum Phase HRIRs

Fig. 1 shows basis functions \mathbf{v}_1 to \mathbf{v}_{10} that were applied in modeling minimum-phase DIRs. Basis functions \mathbf{v}_1 to \mathbf{v}_{10} correspond to 10 consecutive largest eigen values, from the largest to the smallest one ($\lambda_1 > \lambda_2 > \dots > \lambda_{67}$ where λ_i is the i -th eigen value resulted from PCA). As shown in the figure, \mathbf{v}_1 has the simplest form. When the eigen value become smaller, the correspond basis function has a form with more details. Basis functions \mathbf{v}_1 to \mathbf{v}_5 have non zero amplitudes from 0.3 ms, but \mathbf{v}_6 to \mathbf{v}_{10} have non zero amplitudes from 0.6 ms. The intervals of amplitude levels were in general similar for these basis functions.

Analysis of MSEs of Models of Minimum Phase HRIRs in Sphere from Subject-003

From the observation of experiments' results, there are similar errors that occur in the models of both ears in the particular regions of interest. Hence, we discuss here only the errors of left ear models of minimum-phase HRIRs. Fig. 2 shows MSEs of the models of minimum-phase HRIRs

from left ear of Subject-003 in sphere of all sound directions. As seen, small MSEs occur in center regions around the head, i.e. region near median plane with azimuth 0° and plane above head with elevation 90° of sound directions. Left ear average MSE across directions on the median plane is 4.23% and that of the plane above head is 2.03%. Across directions on horizontal plane (elevation 0°), the left ear average MSE is 5.15%. Directions in the ipsilateral side provided smallest MSEs. Ipsilateral side of left ear is region with azimuth $-90^{\circ} < \theta < 0^{\circ}$. Left ear average MSE across directions with azimuth -80° is 2.65%.

Less fine left ear models of minimum-phase HRIRs occur at the regions below subject, i.e. planes with elevation -45° (front below) and 230.6° (rear below), that have average MSEs 13.18% and 12.62% respectively. Worst left ear modeling occurs for contralateral plane with azimuth 65° , that has average MSE of 15.33° . However, average MSE on the contralateral plane with azimuth 80° is only 6.86%. Across all directions in sphere, left ear average MSE was 5.07% and that of right ear was 5.39%.

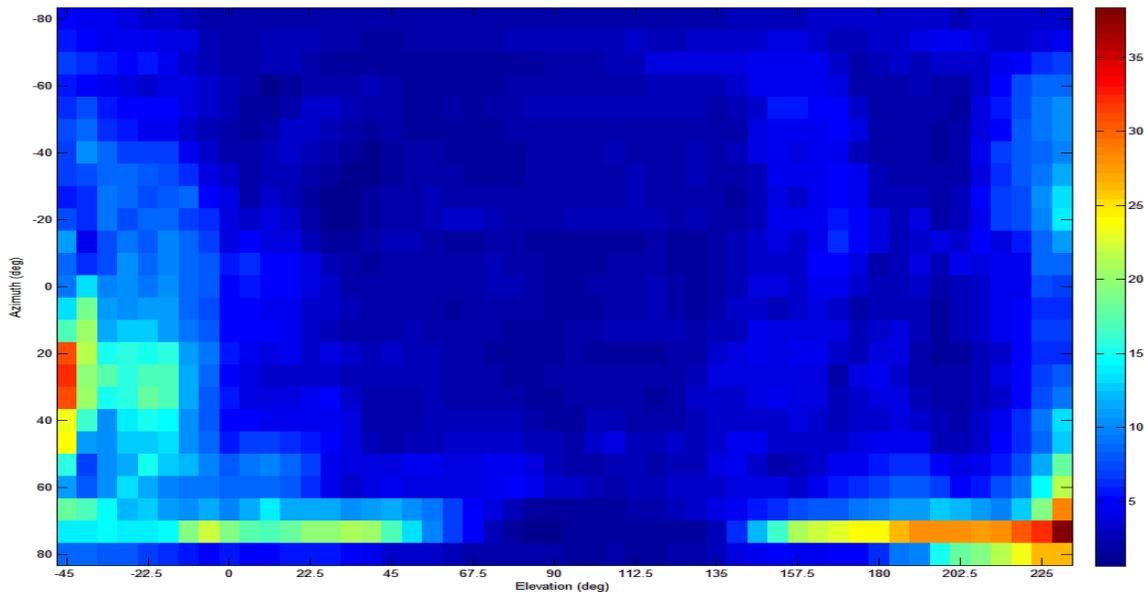


FIGURE 2: Percentage Mean Square Errors among Original and Models of Minimum Phase HRIRs

Analysis of Basis Functions of PCA in Modeling Minimum Phase HRIRs

It is shown from Fig. 3 that all of the first 10 basis functions can be said to be constant near zero at frequencies below 2 – 3 kHz because there are no dependencies between sound directions and variations of magnitude HRTFs in this frequency region. The sum of linear combination of these ten basis functions will be near zero or in other words that it is independent from the sound directions. Above ± 3 kHz, all basis functions possess non-zero magnitudes. Higher frequency variations in these basis functions (except in first basis function) represent higher frequency peaks and notches that depend on sound directions of all magnitude HRTFs. As happened in modeling minimum-phase HRIRs, first basis functions in this modeling has also the simplest form of magnitude spectrum. One can said this basis function as the amplification in the region from about 3 kHz to 22.05 kHz. More detailed magnitude spectrum is found at the basis function which corresponds to smaller eigen value. In general, the magnitudes of the spectrum can be seen to be in similar levels interval.

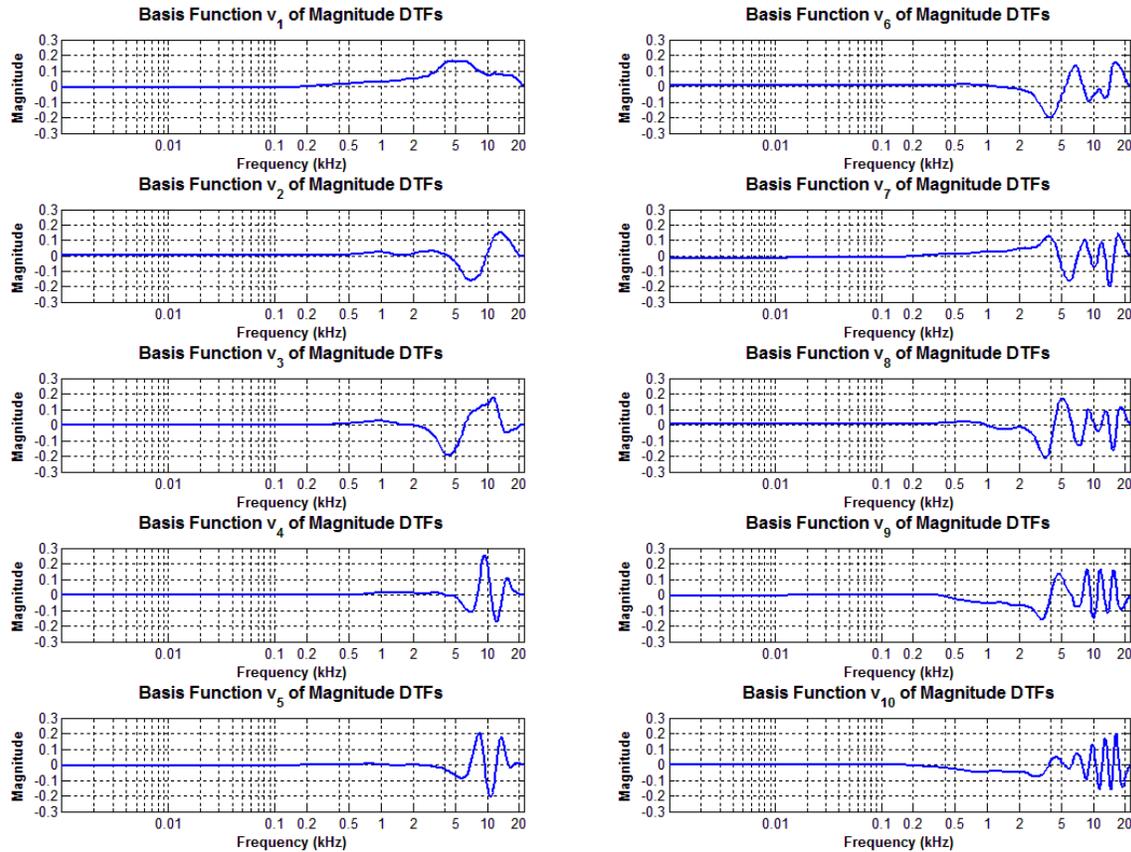


FIGURE 3: The First Ten Basis Functions Extracted from PCA in Modeling Magnitude HRTFs

Analysis of MSEs of Models of Magnitude HRTFs in Sphere from Subject-003

We can see in Fig. 4, a very good performance of models of magnitude HRTFs from Subject-003 across all directions in sphere with MSE about 5%. Fig. 4 shows MSEs of the models of magnitude HRTFs from left ear of Subject-003 of sound sources in sphere. Region with smallest MSEs is the region in the center of sphere, i.e. region around median plane. Average MSE across directions on median plane is 1.95%, while that on horizontal plane is 2.39%. The average MSE of directions on nearest ipsilateral plane with azimuth -80° is 1.47%, while that of directions on farthest contralateral plane with azimuth 80° is 2.62%. Smallest average MSE was achieved from the directions on the plane above head with elevation 90° , i.e. 1.15%. On front below plane with elevation -45° and rear below plane with elevation 230.6° , the average MSEs are worse, i.e. 6.30% and 5.05% respectively. The overall left ear average MSE in sphere is 2.44 and that of right ear is 2.53%.

For models of both minimum-phase HRIRs and magnitude HRTFs, fine performance of models in the region near the hearing ear causes from the fact that the HRIRs has generally larger energies compared to those of directions far from the hearing ear. One advantage of minimum-phase HRIRs is that they have removed time delays. On the other hand, magnitude HRTFs have quite smooth spectrum. PCA produced a set of PCs that could approximate well respective data types.

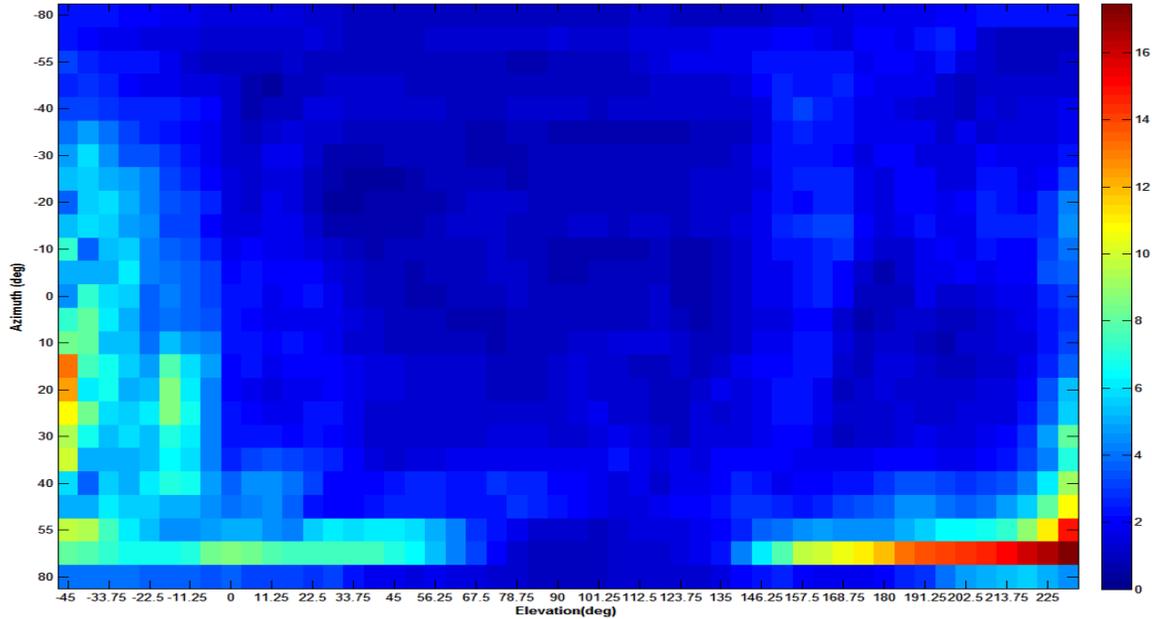


FIGURE 4: Percentage Mean Square Errors among Original and Models of Magnitude HRTFs

4. CONCLUSION

We proposed, using PCA, the modeling of magnitude HRTFs in frequency domain and the modeling of minimum-phase HRIRs in time domain. Using PCA with 10 basis functions from CIPIC HRTF Database, we compared the performances of models of 7 data types, i.e. original HRIRs, minimum-phase HRIRs, direct-pulse HRIRs, and normalized HRIRs, in time domain; also magnitude HRTFs, log-magnitude HRTFs, and standardized log-magnitude HRTFs, in frequency domain. Magnitude HRTFs showed the best performance with smallest average MSE across all HRIRs in database. On the other hand, the best models in time domain were achieved from minimum-phase HRTFs.

5. ACKNOWLEDGMENT

The authors wish to thank all staffs of CIPIC Interface Laboratory of California University at Davis, USA for providing the CIPIC HRTF Database. This work was supported by a grant from Electrical Engineering Department, Industrial Technology Faculty, Trisakti University, Jakarta, Indonesia.

6. REFERENCES

1. D. R. Begault. "3-D sound for virtual reality and multimedia", Moffett Field, CA: Ames Research Center, 2000
2. J. Hebrank and D. Wright. "Spectral cues used in the localization of sound sources on the median plane". J. Acoust. Soc. Amer., 56:1829-1834, 1974
3. E. M. Wenzel, M. Arruda, D. J. Kistler and F. L. Wightman. "Localization using nonindividualized head-related transfer functions". J. Acoust. Soc. Amer., 94:111-123, 1993
4. D. J. Kistler and F. L. Wightman. "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction". J. Acoust. Soc. Amer., 91(3):1637-1647, 1992

5. D. W. Batteau. "*The role of the pinna in human localization*". In Proceedings of Royal Society, London, 168(series B):158-180, 1967
6. K. Genuit. „*A description of the human outer ear transfer function by elements of communication theory*“. In Proceedings of 12th Int. Congress on Acoustics. Toronto, ADSTR, B6-8, 1986
7. J. Chen, B. D. Van Veen and K. E. Hecox. "*External ear transfer function modeling: A beamforming approach*". J. Acoust. Soc. Amer., 92:1933-1945, 1992
8. V. R. Algazi, R. O. Duda, R. P. Morrison and D. M. Thompson. "*Structural composition and decomposition of HRTF*". In Proceedings of IEEE WASPAA2001. New Paltz, NY, pp. 103-106, 2001
9. J. C. Middlebrooks and D. M. Green, "*Observations on a principal components analysis of head-related transfer functions*". J. Acoust. Soc. Amer., 92(1):597-599, 1992
10. H. Hu, J. Zhou, H. Ma and Z. Wu. "*Head related transfer function personalization based on multiple regression analysis*". In Proceedings of IEEE Int. Conf. on Computational Intelligence and Security, Guangzhou, China, vol. 2, pp.1829 – 1832, 2006
11. H. Hu, L. Chen and Z. Wu. "*The estimation of personalized HRTFs in individual VAS*". In Proceedings of IEEE 4th Int. Conf. on Natural Computation, Jinan, China, pp. 203-207, 2008
12. J. Chen, B. D. Van Veen and K. E. Hecox. "*A spatial feature extraction and regularization model for the head-related transfer function*". J. Acoust. Soc. Amer., 97(1):439-452, 1995
13. N. Inoue, T. Kimura, T. Nishino, K. Itou and K. Takeda. "*Evaluation of HRTFs estimated using physical features*". Acoust. Sci. & Tech., 26(5):453-455, 2005
14. Hugeng, W. Wahab and D. Gunawan. "*Improved method for individualization of HRTFs on horizontal plane using reduced number of anthropometric measurements*". J. of Telecommunications, 2(2):31-41, May 2010
15. S. Xu, Z. Li and G. Salvendy. "*Improved method to individualize head-related transfer function using anthropometric measurements*". Acoust. Sci. & Tech., 29(6):388-390, 2008
16. J. Sodnik and S. Tomazic. "*Directional information in head related transfer functions*". In Proceedings of TENCON 2004 – 2004 IEEE Region 10 Conf., 2:100-103. Chiang Mai, Thailand, 2004
17. Z. Wu, F. H. Y. Chan, F. K. Lam and J. C. K. Chan. "*A time domain binaural model based on spatial feature extraction for the head-related transfer function*". J. Acoust. Soc. Amer., 102(4):2211-2218, 1997
18. K. H. Shin and Y. Park. "*Enhanced vertical perception through head-related impulse response customization based on pinna response tuning in the median plane*". IEICE Trans. Fundamentals, E91-A(1):345-356, Jan. 2008
19. S. Hwang and Y. Park. "*HRIR customization in the median plane via principal components analysis*". In Proceedings of AES 31st Int. Conf. New Directions in High Resolution Audio. London, UK, June 2007

20. Hugeng, W. Wahab and D. Gunawan. "*Enhanced individualization of head-related impulse response model in horizontal plane based on multiple regression analysis*". In Proceedings of IEEE 2010 2nd Int. Conf. on Computer Engineering and Applications (ICCEA 2010), 2:226-230. Bali Island, Indonesia, March 2010
21. V. R. Algazi, R. O. Duda, D. P. Thompson and C. Avendano, "*The CIPIC HRTF Database*". In Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'01). New Paltz, NY, USA, Oct. 2001

Performance Analysis of Convolution Coded WLAN Physical Layer under Different Modulation Techniques

Ginni Sharma

Lecturer/Electronics & Comm. Engg./ Global Institutes of Technology & Mgmt., Gurgaon, Haryana, India

erginnisharma@gmail.com

Sanjeev Kumar

Asst. Professor/ Electronics & Comm. Engg./ Amritsar College of Engg. & Technology, Amritsar, 143001, India

sanjeev_be@indiatimes.com

Anita Suman

Lecturer/Electronics & Comm. Engg./ Beant College of Engg. & Technology, Gurdaspur, 143521, India

Parveen Kumar

Asst. Professor/Electronics & Comm. Engg./ Beant College of Engg. & Technology, Gurdaspur, 143521, India

parveen.klair@gmail.com

Abstract

WLAN plays an important role as a complement to the existing or planned cellular networks which can offer high speed voice, video and data service up to the customer end. The aim of this paper is to analysis the performance of coded WLAN system for different digital modulation schemes (BPSK, 16-PSK, QPSK, 4-QAM and 16-QAM) under AWGN channel. The performance of convolution encoder WLAN system is in terms of graph between BER and SNR. We also verify the system performance with different code rates ($1/2$, $1/3$, $2/3$ and $3/4$) and different constraint length.

Keywords: OFDM, BER, SNR, WLAN, AWGN, Constraint length (K) and Code rate (r).

1. INTRODUCTION

The Wireless Local Area Network (WLAN) technology is defined by the IEEE 802.11 family of specifications. The standard defines a medium access control (MAC) sub-layer and three physical (PHY) layers. The goal of the IEEE 802.11 protocol is to describe a wireless LAN that delivers services commonly found in wired networks, e.g., throughput, reliable data delivery, and continuous network connections. Orthogonal Frequency Division Multiplexing (OFDM) is a very attractive technique to achieve the high-bit-rate data transmission and is used in WLAN standard. The OFDM system divides the wide signal bandwidth into many narrowband sub channels that are transmitted in parallel. The subcarriers are orthogonal to each other means that they are

mathematical independent. In 1960, Chang [1] postulated the principle of transmitting messages simultaneously through a linear band-limited channel without ICI and ISI. The Saltzberg [2] in 1967 analyzed the performance of such a system. The major contribution to the OFDM technique is given by Weinstein and Ebert [3] which demonstrated the use of the discrete Fourier transform (DFT) to perform the baseband modulation and demodulation. Peled and Ruiz [4] suggested the filling of guard space with the cyclic extension of the OFDM symbol which solves the problem of orthogonality over dispersive channel.

A convolutional coding is done by combining the fixed number of input bits. The input bits are stored in the fixed length shift register and they are combined with the help of mod-2 adders. This operation is equivalent to binary convolution and hence it is called the convolutional coding. Figure 1 shows that for every input message bit two encoded bits V1 and V2 are transmitted one after the other. K is the constraint length of the encoder as it is defined as the number of shifts over which input message bit can influence the encoder output. The k is the number of message bits and n is the number of encoded output bits. Therefore the code rate, r, of the encoder is $r = k/n$.

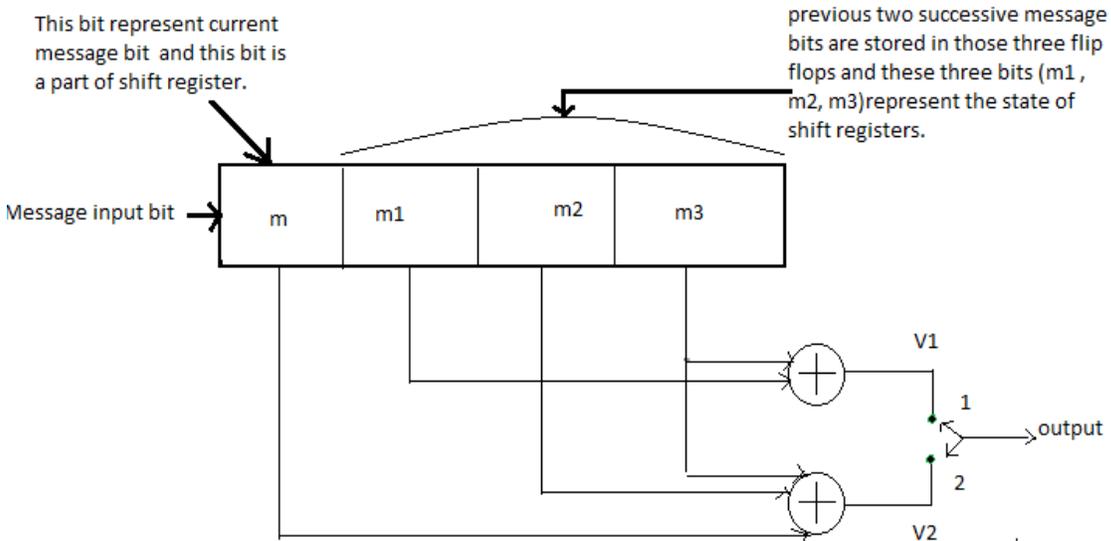


FIGURE 1: Convolutional encoder with $K=4$, $k=1$ and $n=2$

Commonly k and n parameters range from 1 to 8, K from 2 to 10 and code rate from $1/8$ to $7/8$, except for deep space application where code rate is as low as $1/100$ or even longer to be employed.

2. PHYSICAL LAYER STRUCTURE OF WLAN

The complete channel encoding setup at transmitting side and decoding setup at receiving side of the WLAN physical layer is shown in figure 2. In this setup, the input binary data stream is ensured against errors with convolution codes and interleaved. The convolutionally encoded bits are interleaved further prior to convert into each of the either four complex modulation symbols in BPSK, QPSK, 16-PSK, 4-QAM, 16-QAM modulation. The symbols which are digitally modulated transmitted in parallel on subcarriers through implementation as an Inverse Fast Fourier Transform (IFFT). To mitigate the effects of inter-symbol interference (ISI), each block of IFFT coefficients is typically presented by a cyclic prefix [5, 6, 7].

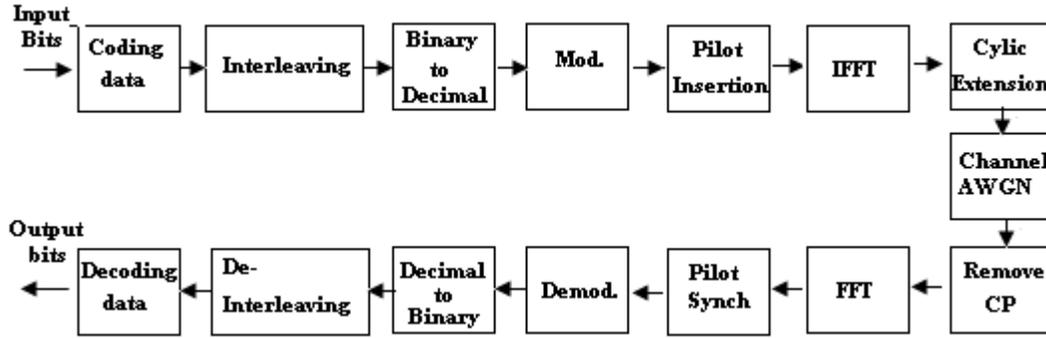


FIGURE 2: Block diagram showing WLAN Physical Layer transceiver

At the receiving side, a reverse process (including deinterleaving and decoding) is performed to obtain the original data bits. The degradation of OFDM performance due to frequency offset or/and phase noise is much more severe in comparison with single carrier modulation [8, 9]. A few techniques to reduce the frequency and phase error of OFDM may be found in [10, 11]. The methods to extend IEEE 802.11 to incorporate adaptive antennas, thereby enhancing security is given by Carey, J.M. [12].

3. SIMULATION RESULTS

The WLAN system using different modulation schemes in the presence of AWGN channel was simulated using Matlab. The different digital modulation schemes used for the simulation are BPSK, QPSK, 16-PSK, 4-QAM and 16-QAM. We are using Convolution encoder with different code rates (1/2, 2/3, 1/3, 3/4) and with different constraint length. Figure 3 shows the graph of BER vs SNR for BPSK modulation with different code rates and constraint length (K). Whereas Figure 4 and Figure 5 shows the performance of system for QPSK and 16-PSK modulation.

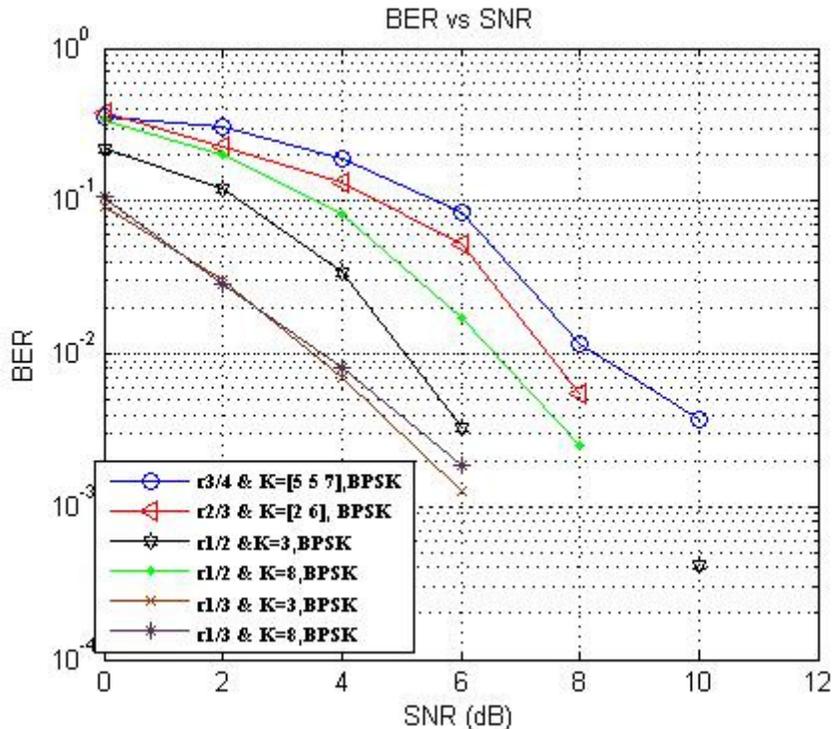


FIGURE 3: BER Vs SNR of BPSK in AWGN channel

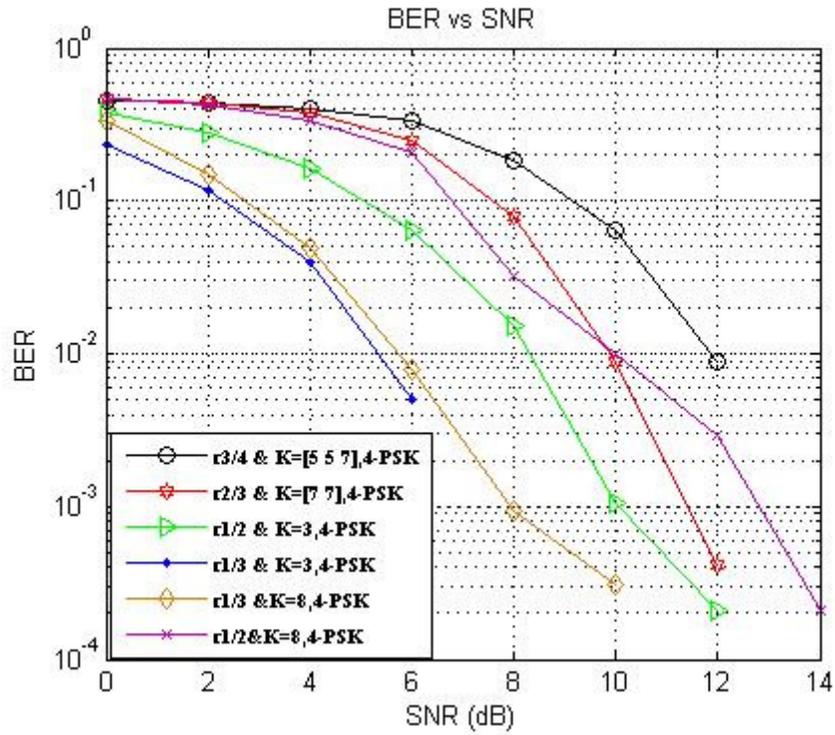


FIGURE 4: BER Vs SNR of QPSK in AWGN channel

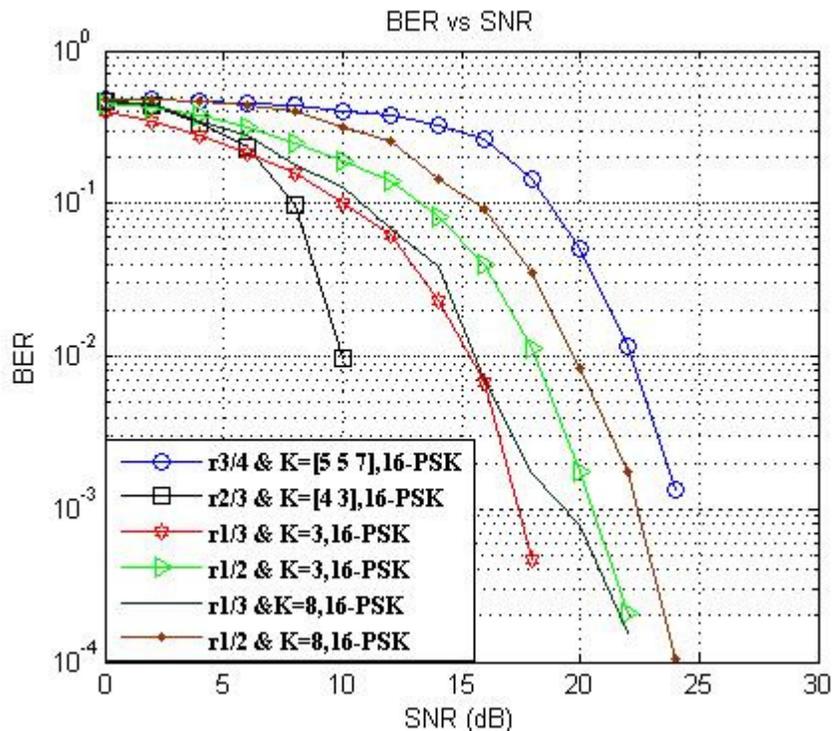


FIGURE 5: BER Vs SNR of 16-PSK in AWGN channel

4. CONCLUSION

A performance analysis of WLAN system adopting convolutional encoding with block interleaver has been carried out. The BER curves were used to compare the performance of different modulation techniques. Performance results highlight the impact of modulation scheme and show that the implementation of an rated convolutional code under different modulation. It is concluded from plots that Convolutional encoder with rate equals to $1/3$ perform better in 4-PSK (QPSK) and BPSK as compared to other code rate. However the convolution encoder with code rate equals to $2/3$ gives better result compared to other for 16-PSK modulation. It is also concluded that BPSK modulation with rate equals to $1/3$ gives better result as compared to other modulation i.e QPSK and 16-PSK.

5. REFERENCES

- [1] Chang, R. W. "Synthesis of Band-Limited Orthogonal Signals for Multichannel Data Transmission". Bell Systems Technical Journal, 45:1775–1796, 1960
- [2] Saltzberg, B. R. "Performance of an Efficient Parallel Data Transmission System". IEEE Trans. on Communications, COM-15(6):805–811, 1967
- [3] Weinstein, S. B., P. M. Ebert. "Data Transmission of Frequency Division Multiplexing Using the Discrete Frequency Transform". IEEE Trans. on Communications, COM-19(5): 623–634, 1971
- [4] Peled, A., and A. Ruiz, "Frequency Domain Data Transmission Using Reduced Computational Complexity Algorithms". Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP '80), Denver, CO, 1980

- [5] W. Hneiti, Dr. N. Ajlouni, A. Arab. "*Performance Enhancement of Wireless Local Area Networks*".
- [6] Md. A. Islam, R. U. Mondal and Md. Z. Hasan. "*Performance Evaluation of Wimax Physical Layer under Adaptive Modulation Techniques and Communication Channels*".
- [7] J. Chuang, N. Sollenberger. "*Beyond 3G: Wideband wireless data access based on OFDM and dynamic packet assignment*". IEEE Communications Mag., 38:78–87, 2000
- [8] A. R. S. Bahai and B. R. Saltzberg. "*Multicarrier digital communications, theory and applications of OFDM*". Kluwer Academic Publishers, pp. 192 (2002)
- [9] S. Glisic. "*Advanced Wireless Communications, 4G Technology*". John Wiley & Sons Ltd: Chichester, 2004
- [10] J. Armstrong. "*Analysis of New and Existing Methods of Reducing Intercarrier Interference Due to Carrier Frequency Offset in OFDM*". IEEE Transactions on Comm., 47:365–369, 1999
- [11] J. Yia-Yun, C. Jian, Q. Ying, T. Yang-ling. "*A Method to Eliminate Effects of Phase Noise in OFDM Synchronization System*". Journal of Shanghai University, 3(3):214–217, 1999
- [12] Carey, J.M. Grunwald, D. "*Enhancing WLAN security with smart antennas: a physical layer response for information assurance*". IEEE Vehicular Technology Conference, 2004. VTC2004-Fall. 2004

Spectral Analysis of Sample Rate Converter

Manish Sabraj

*School of Electronics & Communication Engg
Shri Mata Vaishno Devi University
Jammu and Kashmir, 182320, India*

manishsabraj@yahoo.co.in

Vipin Kakkar

*School of Electronics & Communication Engg
Shri Mata Vaishno Devi University
Jammu and Kashmir, 182320, India*

vipan.kakkar@smvdu.ac.in

Abstract

The aim of digital sample rate conversion is to bring a digital audio signal from one sample frequency to another. The distortion of the audio signal introduced by the sample rate converter should be as low as possible. The generation of the output samples from the input samples may be performed by the application of various methods. In this paper, a new technique of digital sample-rate converter is proposed. We perform the spectral analysis of proposed digital sample rate converter.

Keywords: Sample Rate Converter, Spectral Analysis, Upsample-Downsample Filter, Sigma-Delta Modulator, Frequency Detector.

1. INTRODUCTION

Sample Rate Conversion (SRC) is a process by which the audio sample rate gets changed without affecting the pitch of the audio [1]. This process is necessary in different situations: Digital Audio Workstation (DAW) users often record and edit at a high sample rate, and then down-sample the audio to get it onto various media. This sample rate conversion can either be done by the DAW during or after the bounce, or in a separate application after bouncing. In another scenario, sample rate conversion is necessary when audio material recorded for a specific media (e.g. CD) gets transferred to a different media (e.g. DVD, DAT or Digital Video). For example, a DVD audio project requires sample rate conversion from 96 kHz to 44.1kHz in order to be transferred to CD, and a CD audio project requires conversion from 44.1 kHz to 48 kHz to be transferred to Digital Video format.

It is very important for the sample rate conversion to be as transparent as possible. Ideally, when converting from an original into a new sample rate, we would like the converted signal fidelity to be as high as if we had directly sampled it from the original analog signal. This degree of perfect transparency is possible only in theory, since we would need a computer with infinite memory and infinite processing power to achieve it. In practice, however, a very high degree of transparency can be achieved with a high-quality sample rate converter.

In this paper, analysis results are shown for a new method of digital sample rate converter [2]. The operation principle of the new method of sample rate conversion is very simple. An input sample is directly transferred to the output, while per unit of time, a certain amount of these samples is omitted or repeated, depending on the difference in input and output sample

frequencies. The omission, acceptance or repetition of a sample is called ‘validation’. In order to get the simplest hardware implementation, the choice has been made to use only the take-over operation and the repetition operation in the current system solution. This means that the output sampling frequency of the sample rate converter is always larger than the input sample frequency.

The process of repeating samples inevitably introduces errors. The resulting output samples will have correct values, but as a result of the validation operation, they are placed on the output time grid with a variable time delay with respect to the input time grid. As a consequence, the output sequence should be viewed as the input sequence, having the correct signal amplitude, which is sampled at wrong time moments. The effect is the same as sampling the input signal by a jittered clock [3]. As a result, it can be stated that the time error mechanism introduced by the validation algorithm is time jitter.

If all input samples would be transferred to the output grid without the repetition or omission of a certain amount of them, then the output signal would be just a delayed version of the input signal, exhibiting the same shape. It is the repetition and omission (in the current system setup only the repetition) of input samples that give rise to a variation in time delay for each individual output sample. This variation in individual time delays introduces phase errors. As a result of this, the shape of the output signal will be distorted [4].

The time errors introduced by the conversion process can be reduced considerably by applying upsampling and downsampling techniques. The input sample rate of the converter will be higher so that the conversion errors are smaller, resulting in smaller time jitter. These techniques do not suffice when we want to achieve the very high analog audio performance required for professional applications [5]. By using a sigma-delta modulator (noise shaper) as control source for the conversion process, the time errors will be shaped to the higher frequency region. As a result, the audio quality (in the baseband) of the signal will be preserved, provided that enough bandwidth is created by upsampling of the input signal. The high frequency (out of base band) phase modulation terms can be filtered by a decimation filter or an analog low-pass filter which is directly placed after the sample-rate converter [6]. Figure 1 shows the block diagram of the complete sample-rate converter.

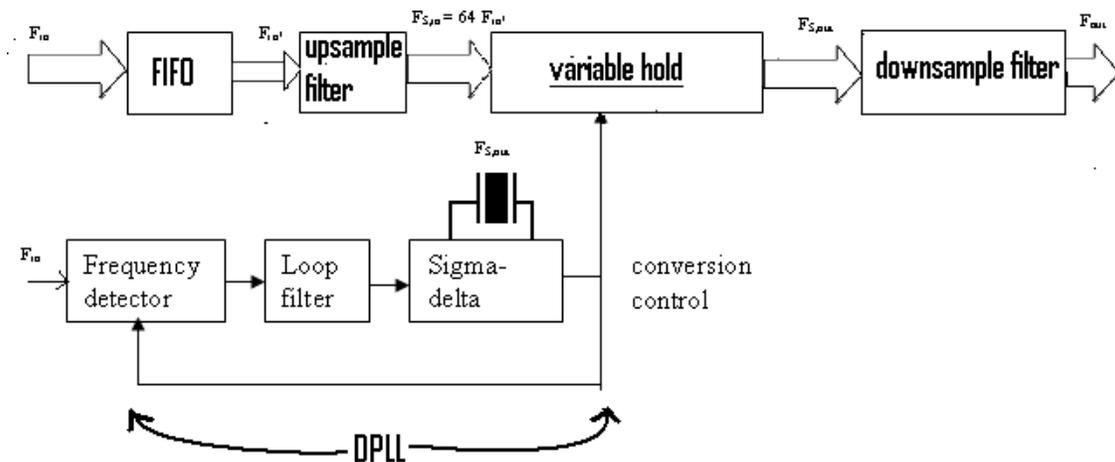


FIGURE 1: Block diagram of the sample-rate converter.

As has already been mentioned, only the input sample take over operation will be employed here in order to get the simplest hardware. This means that the input sample frequency of the converter must be always be smaller than the output sample frequency. With this restriction imposed, it is assured that all input samples are used in the output sequence, none of them being

omitted. The extra output samples per unit of time are inserted in the output sequence by repetition of their previous output samples.

For practical problems in which only a finite number of samples of the analog signal are available, say $x(nT)$, $n = 0, 1, \dots, N - 1$, the band-limited analog signal and transform often are still modeled by the sampling representations but with the unobserved samples set equal to zero, i.e., $x(nT) = 0$ for $n < 0$ or $n > N - 1$. With this model the Fourier transform in the band $Mod(\omega) < \pi / T$ is represented by the discrete Fourier transform (DFT). In either case, the problem of resampling with a different sample interval is in principle solved, because one can reconstruct the original analog signal, or an acceptable model of it, and then resample at will (Crochiere and Rabiner [1], [7], Pridham and Mucci [8], [9], Shaefer and Rabiner [10]).

In spite of the availability and utility of the Fourier and sampling theorem representations it is sometimes preferable to employ a simpler interpolation scheme than that involving the sinc kernel in order to reduce the computational load. In such cases it is important to consider the approximation error and its influence on the ability of the new sample set adequately to represent the original signal.

2. THEORETICAL GUIDELINES FOR SPECTRAL ANALYSIS

In this part, the properties of the proposed sample-rate converter in the frequency domain will be investigated. It is observed that the first order approximation of the amplitude error is accurate enough, even for the worst case situation. The continuous-time description of the first-order model is:

$$y(t) = x(t) - x(t) \cdot \Delta t(t)$$

(1)

Figure 2 gives a block diagram of this first-order model.

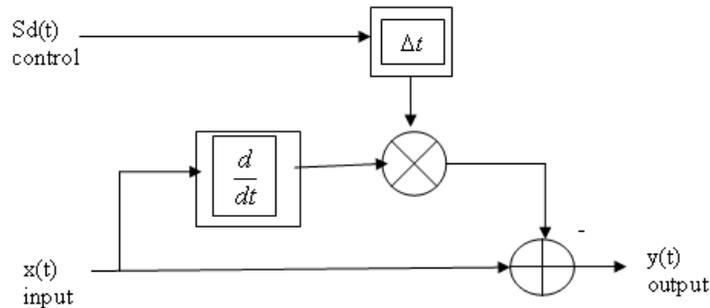


FIGURE 2: Block diagram of the first-order model.

When we want to convert (1) to discrete-time, we have to keep in mind that the output samples have a different sample time than the input samples. We will therefore enter two discrete-time variables; $k.T_{S,out}$ for the output samples and time delays, and $l.T_{S,in}$ for the input samples (shortly denoted by k and l). The discrete-time model now becomes:

$$y(k) = x(l) - x(l) \cdot \Delta t(k)$$

(2)

Normally the derivative of a discrete-time signal is determined by the amplitude difference of the current sample and its previous sample, divided by the sample time. In our case it is more correct to use the difference between the next input sample and the present input sample, because the position in time of the present output sample is between the time moments of those two input samples. For the time derivative of the input signal $x(l)$ we obtain:

$$\dot{x}(l) = \frac{x(l+1) - x(l)}{T_{S,in}} = F_{S,in} \cdot [x(l+1) - x(l)] \quad (3)$$

Substituting this into (2), we get

$$y(k) = x(l) - F_{S,in} \cdot [x(l+1) - x(l)] \cdot \Delta t(k) \quad (4)$$

In order to become known with the spectral density of the output signal $y(k)$, we must firstly determine the correlation function:

$$R_{yy}(n) = E\{y(k) \cdot y(k+n)\} \quad (5)$$

This correlation function describes the correlation between the present output sample (k_0) and the (k_0+n) -th output sample and is therefore dependent on the output sample time $T_{S,out}$. Note that n must be an integer.

The problem arises that for a time step of n samples ($=n \cdot T_{S,out}$) in the output signal we must know the corresponding time step in the input signal. Assume that this time step is equal to $m \cdot T_{S,in}$, that is, $y(k+n)$ corresponds to $x(l+m)$. The relation between m and n then becomes:

$$n \cdot T_{S,out} = m \cdot T_{S,in} \Rightarrow m = \frac{T_{S,out}}{T_{S,in}} \cdot n \Rightarrow m = \frac{F_{S,in}}{F_{S,out}} \cdot n \quad (6)$$

The conversion factor for the sample-rate conversion process is not necessarily a rational number, which implies that m is not necessarily an integer. For the calculation of the discrete-time correlation function we need both n and m , as we have two discrete-time variables. The problem is that the discrete-time input signal $x(l+m)$ is not defined when m is not an integer. We must therefore conclude that the correlation function $R_{yy}(n)$ of the discrete-time output signal can not be solved analytically.

Consider the discrete-time description of the first-order model (2). For the calculation of an output sample on time moment k (somewhere between l and $l+1$) the discrete-time derivative of the input signal $x(l)$ on time moment k is needed. This derivative is determined using the two adjacent input samples (4). Suppose $x(t)$ is the continuous-time signal constructed out of the input samples $x(l)$ using *linear interpolation*. The continuous-time derivative of this input signal is in fact similar to the discrete-time derivative given by (3). In fact we deduce our discrete-time analysis from the continuous-time analysis. In order to find out the spectral properties of the sample-rate converter, it is therefore allowed that we use the *continuous-time* description given by (1).

3. SIMULATION AND PERFORMANCE ANALYSIS

In this part, the properties of the sample-rate converter will be demonstrated and compared to the theoretical results. Firstly, the time domain simulations will be shown for a certain conversion factor. Next, the Fourier transformations of these time domain signals will give the frequency domain presentation of the sample-rate conversion process.

3.1 Time Domain Simulations

In this subsection the results of a time domain simulation are presented. The results have been obtained by simulating the sample-rate converter using a third-order sigma-delta modulator. The input signal consists of a single sinewave with a frequency F_{in} of 20kHz and an amplitude A of 1[Volt]. The output sampling frequency equals $128F_s=5.6448\text{MHz}$ while the input sampling frequency is chosen to be $30.13579F_s=1.328988\text{MHz}$. The latter is chosen much smaller than $128F_s$ so that the distortion in the output signal due to much repetitions will be clearly visible. The DC input of the sigma-delta modulator can be calculated as 0.52912828125. The average number of repetitions will be 3.25. Figure 3 shows a plot of the three most important signals involved in the conversion process. The upper trace is the sigma-delta control signal, the signal in the middle is the output signal of the sample-rate converter and the lower trace shows the corresponding time error for each output sample.

It can be seen that the output signal of the converter is fairly distorted due to the large number of repetition samples. The plot style of the corresponding time error signal is staircase, because with this plot style the stepwise behaviour of this signal is illustrated best.

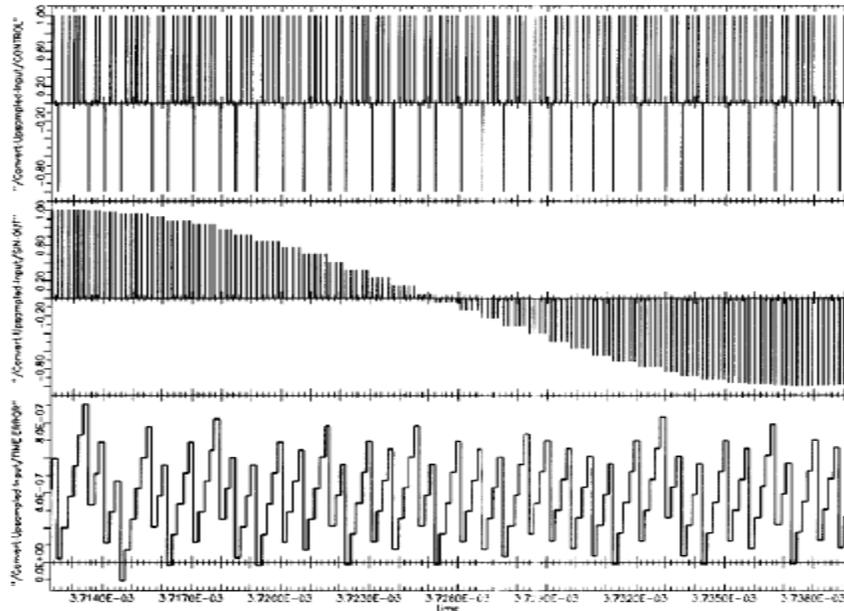


FIGURE 3: Time domain signals of the sample-rate conversion process: sigma-delta control, output sinewave, time error signal. $F_{s,in}=30.13579F_s$, $F_{s,out}=128F_s$.

3.2 Frequency Domain Results

The frequency spectra of the signals in figure 3 are obtained by taking the Fourier transform of these signals. Figure 4 shows the spectra of the sigma-delta output signal and the time error signal, while figure 5 shows the frequency spectrum of the output signal. The plots have a logarithmic frequency-axis.

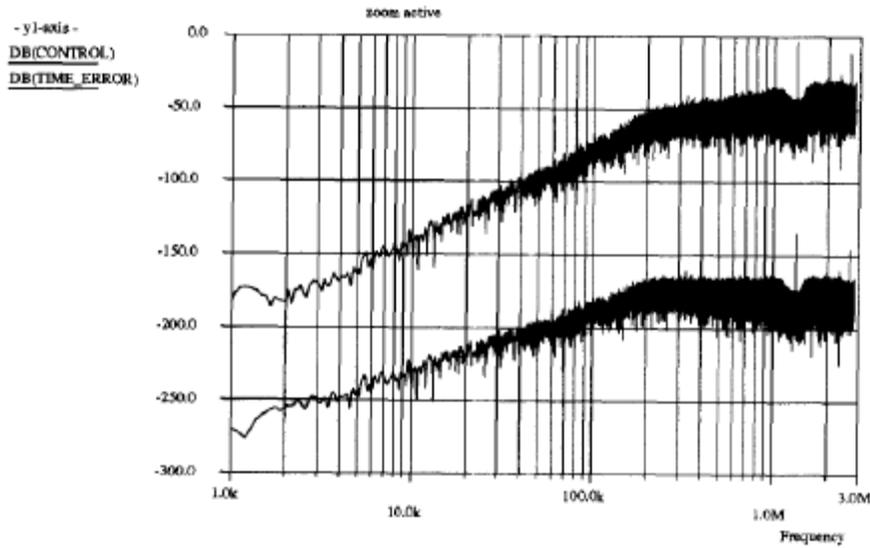


FIGURE 4: Sigma-delta spectrum (higher trace) and time error spectrum (lower trace) for $F_{S,in}=30.13579F_s$, $F_{S,out}=128F_s$ and $F_{in}=20\text{kHz}$.

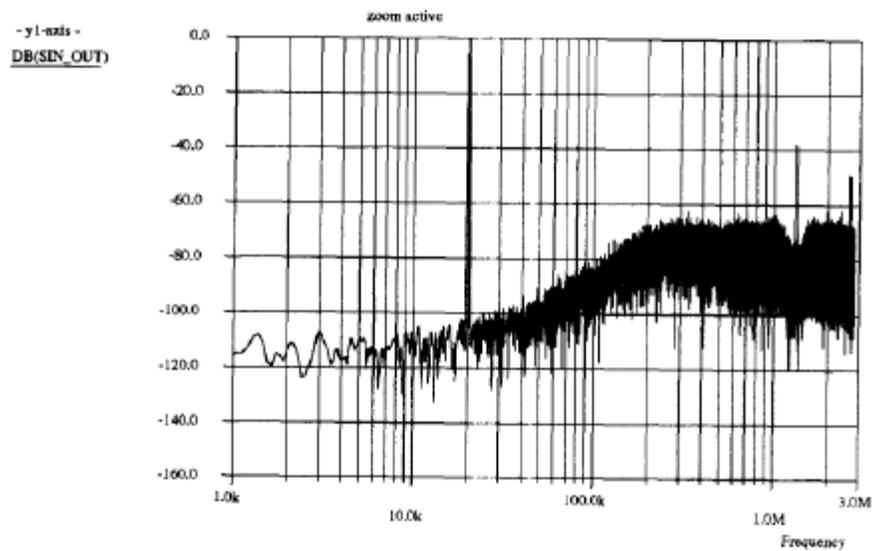


FIGURE 5: Frequency spectrum of the output signal for $F_{S,in}=30.13579F_s$, $F_{S,out}=128F_s$ and $F_{in}=20\text{kHz}$.

In figures 6 and 7 a zoom-in is given of the figures 4 and 5 respectively, around the “first” spectral peak in the output spectrum of the sigma-delta modulator. The plots have a linear frequency-axis.

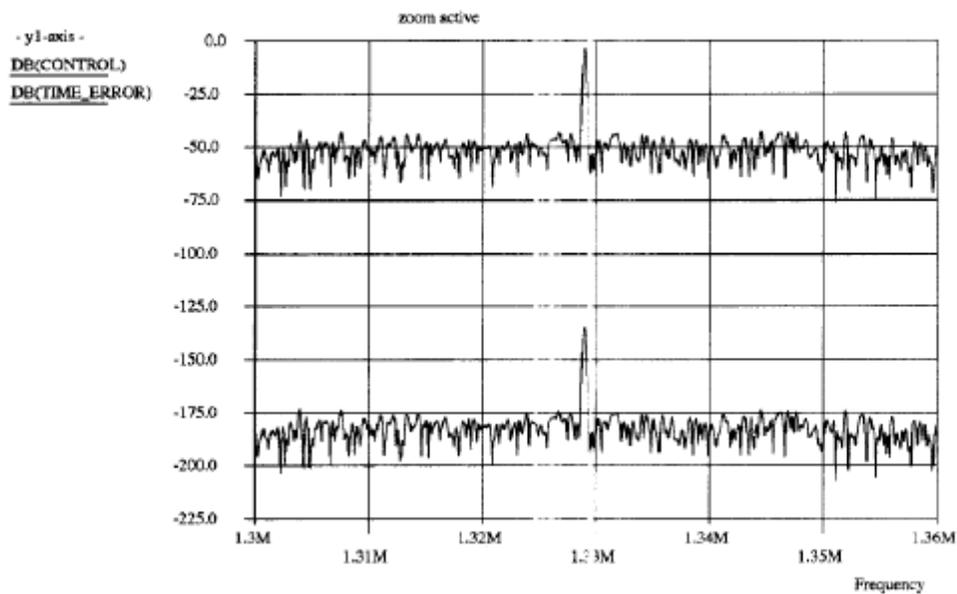


FIGURE 6: A zoom-in of figure 4

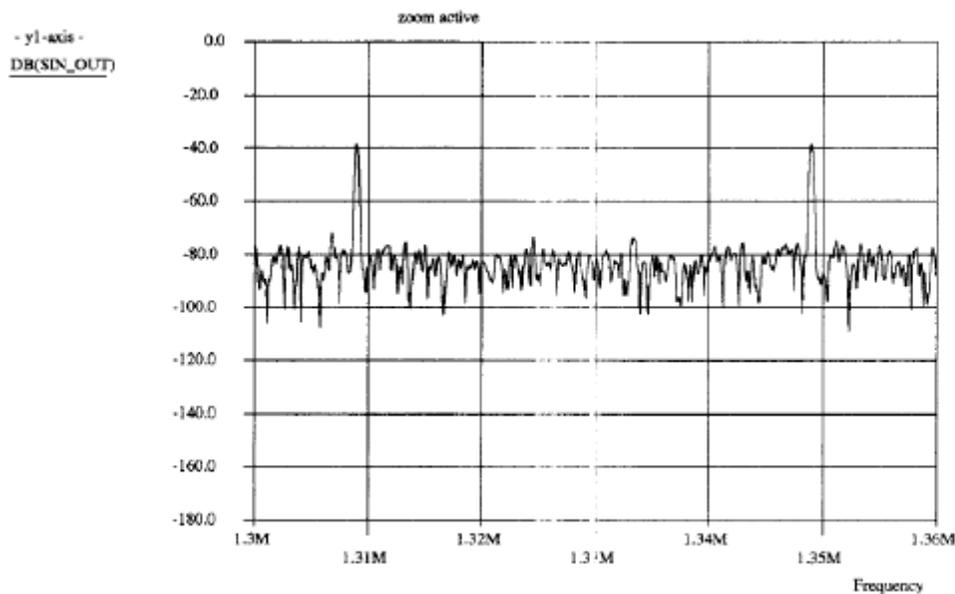


FIGURE 7: A zoom-in of figure 5.

From the figures 4 to 7, we can make the following observations:

- The slope in the sigma-delta spectrum is +60 dB per decade, which corresponds to a third order sigma-delta modulator, while the time error spectrum shows a slope of +40 dB per decade. The additional roll-off in the time error spectrum amounts 20 dB in comparison with the sigma delta spectrum, as was expected. The shaping of the time error is indeed one order lower than the shaping of quantization noise to the high frequency region, second order in this case.

- The time error spectrum indeed contains the same spectral components as the sigma-delta spectrum (figures 4 and 6), which was expected. The sigma-delta spectrum always contains a spectral peak at the input sampling frequency $F_{S,in}$, which is for this case 1.328988 MHz and at multiples of this frequency. These peaks can also be observed in the time error spectrum (having a different amplitude).
- The spectrum of the output signal of the sample-rate converter (figure 5) contains spectral peaks at $F_{S,in}-F_{in}$ and $F_{S,in}+F_{in}$ with $F_{in}=20$ kHz (figure 7). This corresponds to the theoretical considerations. The spectral peaks of the sigma-delta modulator are frequency shifted over a frequency f_{in} to the right and to the left.
- For frequencies between 20 kHz and 200 kHz, the output spectrum shows a slope of about +40 dB per decade (figure 5). For frequencies above 200 kHz, the spectrum is flat. In the audio base band (0-20kHz), the spectrum is also fairly flat. It should be noticed that the negative frequencies must be taken into account when we look at frequency-shifted spectra [7].

4. CONCLUSION

It is concluded that the frequency domain results obtained in this paper correspond to the theoretical dependencies. It appears that due to the frequency-shift of the time error spectrum, the spectrum of the output signal of the sample-rate converter is fairly flat in the audio base band. The addition of the two frequency-shifted spectra causes more quantization noise to fall into this base band. As a result, the signal-to-noise ratio of the output signal will be smaller: the performance of the output sinewave is degraded.

5. REFERENCES

- [1] R. Crochiere and L. Rabiner. “*Mullinale Digital Signal Processing*”. Englewood Cliffs, NJ: Prentice Hall, 1983
- [2] Therssen, D. “*A new Principle for sample-rate conversion of digital signals*”. Leuven, Belgium, 1990
- [3] Baggen, C.P.M.J. “*An Information Theoretic Approach to Timing Jitter*”. Doctoral Dissertation, University of California, San Diego. 1993
- [4] R. Fitzgerald, W. Anderson. “*Spectral distortion in sampling rate conversion by zero-order polynomial interpolation*”. *IEEE Trans. Signal Processing*, 40:1576–1578, 1992
- [5] H.G. Gökler, G. Evangelista, and A. Groth. “*Minimal Polyphase Implementation of Fractional Sample Rate Conversion*”. *Signal Processing*, 81(4):673-691, 2001
- [6] Groth, H.G. Gökler. “*Efficient Minimum Group Delay Block Processing Approach to Fractional Sample Rate Conversion*”. ISCAS '01, Sydney, Australia, II:189-192, 2001
- [7] R. E. Crochiere, L. R. Rabiner. “*Interpolation and decimation of digital signals-A tutorial review*”. *Proc. IEEE*, 69(3):300-331, 81
- [8] R. G. Pridham, R. A. Mucci. “*A novel approach to digital beamforming*”. *J. Acousr. Soc. Amer.*, 63(2):425-434, 1978
- [9] R. G. Pridham, R. A. Mucci. “*Digital interpolation beamforming for low-pass and bandpass signals*”. *Proc. IEEE*, 67(6):904-919, 1979

- [10] R. W. Schafer, L. R. Rabiner. "*A digital signal processing approach to interpolation*". Proc. IEEE, 61:692-702, 1973
- [11] Groth, and H.G. Gökler. "Polyphase Implementation of Unrestricted Fractional Sample Rate Conversion," Internal Report, Available at: <http://www.nt.ruhr-uni-bochum.de/lehrstuhl/mitarbeiter/alex.html>, 2000

A Gaussian Clustering Based Voice Activity Detector for Noisy Environments Using Spectro-Temporal Domain

Sara Valipour

*Faculty of Engineering Islamic Azad
University Ghaemshahr Branch
Ghaemshahr, Iran*

Valipour_Sara@yahoo.com

Farbod Razzazi

*Faculty of Engineering Islamic Azad University,
Science and Research Branch
Tehran, Iran*

razzazi @ srbiau.ac.ir

Azim Fard

*Communications Regulatory Authority
Tehran, Iran*

azimfard@cra.ir

Nafiseh Esfandian

*Faculty of Engineering Islamic Azad
University Ghaemshahr Branch
Ghaemshahr, Iran*

Na_Esfandian@yahoo.coo

Abstract

In this paper, a voice activity detector is proposed on the basis of Gaussian modeling of noise in the spectro-temporal space. Spectro-temporal space is obtained from auditory cortical processing. The auditory model that offers a multi-dimensional picture of the sound includes two stages: the initial stage is a model of inner ear and the second stage is the auditory central cortical modeling in the brain. In this paper, the speech noise in this picture has been modeled by a 3-D mono Gaussian cluster. At the start of suggested VAD process, the noise is modeled by a Gaussian shaped cluster. The average noise behavior is obtained in different spectrotemporal space in various points for each frame. In the stage of separation of speech from noise, the criterion is the difference between the average noise behavior and the speech signal amplitude in spectrotemporal domain. This was measured for each frame and was used as the criterion of classification. Using Noisex92, this method is tested in different noise models such as White, exhibition, Street, Office and Train noises. The results are compared to both auditory model and multifeature method. It is observed that the performance of this method in low signal-to-noise ratios (SNRs) conditions is better than other current methods.

Keywords: Voice activity detector, Spectro-temporal Domain, Gaussian modeling, Auditory model.

1. INTRODUCTION

In general, sound signal is composed of two parts, speech and non-speech. The latter is either silence or background noise. Accordingly, detection of speech signal from non-speech signal, known as voice activity detection (VAD), is one of the most important issues in the speech processing systems. In particular, the complexities increases in low SNRs where there are challenging in VAD design. One of the applications of VAD is in speech enhancement systems [1], where VAD is used to estimate noise characteristics from the silence parts of the signal. Robust speech recognition [2], speech coding [3] and echo cancellation are among the other applications of VAD.

The first, but of course the most, usual VAD algorithm has been presented in [4]. There are other VAD algorithms as well. In [5], a VAD has presented on the basis of MFCC features and SVM, as MFCC coefficients provide good information of the speech signal. Sohn in [6] has used a Gaussian statistical model for VAD. As another work, has obtained a VAD based on Taylor series [7]. Chang has performed a class from VAD algorithm using different statistical models. Moreover, he has combined Gaussian model, complex laplacian and gamma probability density equations to analytically characterize statistical properties of noise and speech parts [8]. Another VAD has been obtained based on the generalized Gamma distribution [9], where a distribution of noise spectra and noisy speech spectra has been obtained based on inactive speech intervals. In all these algorithms, the results are not promising in low signal to noise ratios (SNR) and VAD performance in low SNRs has remained as a challenging issue.

In this research, our proposed VAD algorithm is based on spectro-temporal representation of the speech. The idea is based on neuro-physiological and psycho acoustical investigations in various stages of auditory system. This model consists of two main stages. The first one is the auditory system which represents the acoustic signal as an auditory spectrogram. The second stage, which is the central cortical stage, is the stage of analyzing the spectrogram by using a set of 2D filters. The new successful achievements in the spectro-temporal studies reveal a significant improvement of the performance for enhancement systems [1], Speech Recognition [10] and also robust pitch extraction [11].

In this work, a VAD algorithm is proposed on the basis of noise Gaussian model in the spectro-temporal domain. Evaluating the efficiency, it is shown that the spectro-temporal domain is a suitable space for this separation. The rest of the paper is organized as follows. In section 2, the auditory model and spectro-temporal model are briefly reviewed. In section 3, the proposed VAD method is presented and analyzed in the spectro-temporal domain. In section 4, the method is evaluated and compared to other methods. Finally, the paper is briefly concluded in Section 5.

2. SPECTRO-TEMPORAL MODEL

Auditory Model

The auditory model has been obtained on the basis of neuroscience and biology researches. They are achieved based on two different sections of the auditory systems, mammals and, in particular, humans [12]. The model has two main parts [13-14]. In the first part of this model, the acoustic signal is represented by an auditory spectrogram [14]. While in the next part, the auditory spectrogram is analyzed using a set of 2D filters [15].

calculation of auditory spectrogram

In first part of auditory model, the auditory spectrogram of input signal is calculated by passing various through stages. The stages are shown in figure 1 [12-15].

As shown in figure 1, the input signal, enters a hair cell stage after passing through a filter bank which makes a 2D representation of the input signal. This part consists of three stages: a high-pass time domain filter, a nonlinear compression stage, and a low-pass time domain filter. The output of this stage is applied to a lateral inhibitory network which is in fact a first-order frequency domain derivative, followed by a half-wave rectifier and a low pass time domain integrator. At the

final stage, the auditory spectrogram of the signal is calculated. The analytical characterization of sequential stages for the first section of the auditory model is given as follow [1].

$$y_{coch} = s(t) *_{\tau} h(t, x) \tag{1}$$

$$y_{AN}(t, x) = g((\partial y_{coch}(t, x) / \partial t) *_{\tau} w(t)) \tag{2}$$

$$y_{LIN}(t, x) = Max((\partial y_{AN}(t, x) / \partial x), 0) \tag{3}$$

$$y_{final}(t, x) = y_{LIN}(t, x) *_{\tau} \mu(t, x) \tag{4}$$

In the above relations, the operator $*_{\tau}$ shows the convolution in time domain.

The central auditory section

In this section, the auditory spectrogram is analyzed to extract the spectro-temporal features[16]. The signal is applied through a bank of 2-D filters. The contents of spectro-temporal modulation of the auditory spectrogram are determined using selected filter banks, centered along tonotopic axis [17]. The spectrotemporal impulse response of these filters is called the spectro-temporal response field (STRF). STRFs are 2-D Gabor wavelets.

$$STRF_{+} = \Re\{H_{rate}(t; \omega, \theta) \cdot H_{scale}^{*}(f; \Omega, \phi)\} \tag{5}$$

$$STRF_{-} = \Re\{H_{rate}^{*}(t; \omega, \theta) \cdot H_{scale}(f; \Omega, \phi)\} \tag{6}$$

where \Re is the real part, $*$ is complex conjugate, ω is speed and Ω is scale. θ and ϕ are a phase of asymmetry along time and frequency domain respectively. In addition, H_{rate} and H_{scale} may be analytically extracted from h_{rate} and h_{scale} [1].

$$H_{rate}(t; \omega, \theta) = h_{rate}(t; \omega, \theta) + j\hat{h}_{rate}(t; \omega, \theta) \tag{7}$$

$$H_{scale}(f; \Omega, \phi) = h_{scale}(f; \Omega, \phi) + j\hat{h}_{scale}(f; \Omega, \phi) \tag{8}$$

where $\hat{}$ shows the Hilbert transformation. h_{rate} and h_{scale} are respectively the temporal and spectral impulse responses [1].

$$h_{rate}(t; \omega, \theta) = h_r(t; \omega) \cos\theta + \hat{h}_r(t; \omega) \sin\theta \tag{9}$$

$$h_{scale}(f; \Omega, \phi) = h_s(f; \Omega) \cos\phi + \hat{h}_s(f; \Omega) \sin\phi \tag{10}$$

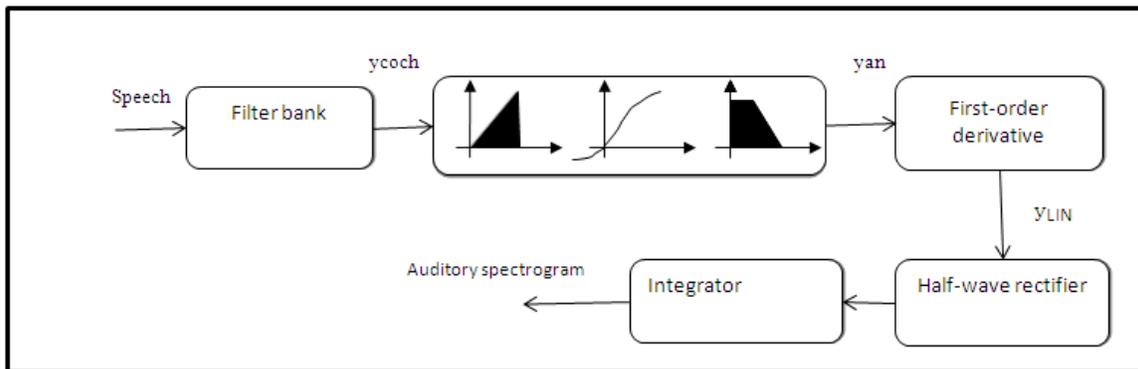


FIGURE 1: Different stages of first part of the auditory model

The impulse responses are obtained as hereunder for various frequencies and times [1]

$$h_r(t; \omega) = \omega h_2(\omega t) \tag{11}$$

$$h_s(f; \Omega) = \Omega h_2(\Omega f) \tag{12}$$

The auditory spectrum, after passing through STRFs is transformed into a 4-D cortical picture. These four dimensions are frequency, time, speed and scale [1].

$$r_+(t, f, \omega, \Omega, \theta, \phi) = y(t, f) *_{t,f} STRF_+(t, f, \omega, \Omega, \theta, \phi) \quad (13)$$

$$r_-(t, f, \omega, \Omega, \theta, \phi) = y(t, f) *_{t,f} STRF_-(t, f, \omega, \Omega, \theta, \phi) \quad (14)$$

where $*_{t,f}$ is the 2D convolution with respect to time and frequency. r_+ and r_- are respectively the spectro-temporal responses of downward (+) and upward (-) STRFs. The wavelet transformation is obtained from the filters h_{rate} and h_{scale} as below:

$$h_{TW}(t; \omega) = h_r(t; \omega) + j\tilde{h}_r(t; \omega) \quad (15)$$

$$h_{ZW}(f; \Omega) = h_s(f; \Omega) + j\tilde{h}_s(f; \Omega) \quad (16)$$

The complex response of downward and upward selective filters is as follows:

$$Z_+(t, f; \Omega, \omega) = y(t, f) *_{t,f} [h_{TW}^*(t; \omega) h_{ZW}(f; \Omega)] \quad (17)$$

$$Z_-(t, f; \Omega, \omega) = y(t, f) *_{t,f} [h_{TW}(t; \omega) h_{ZW}^*(f; \Omega)] \quad (18)$$

Finally, for each speech frame, two 3-D complex valued matrices are obtained for downward and upward filters respectively.

3. VAD METHOD IN SPECTRO-TEMPORAL DOMAIN

In the proposed VAD method, a Gaussian model is applied to model the noise cluster shape in spectro-temporal domain. In this method, it is attempted to estimate the cluster shape of 3-D spectro-temporal representation of noise (silent) using a Gaussian function. The concept originates from the fact that the shape of the noise cluster, created by large amplitude points in spectro-temporal domain is similar to a 3-D Gaussian hyper-surface. Therefore, the parameters of this function should be corresponded to the average of spectro-temporal representation of noise frames. The block diagram of such the noise modeling is shown in figure 2.

As shown in figure 2, the cortical picture of input noise is calculated for each frame with three dimensions of frequency, speed and scale. It is divided into two separate downward and upward representations. In the proposed method, in order to model noise samples in spectro-temporal space, we calculate the parameters of Gaussian model for downward and upward magnitude, separately.

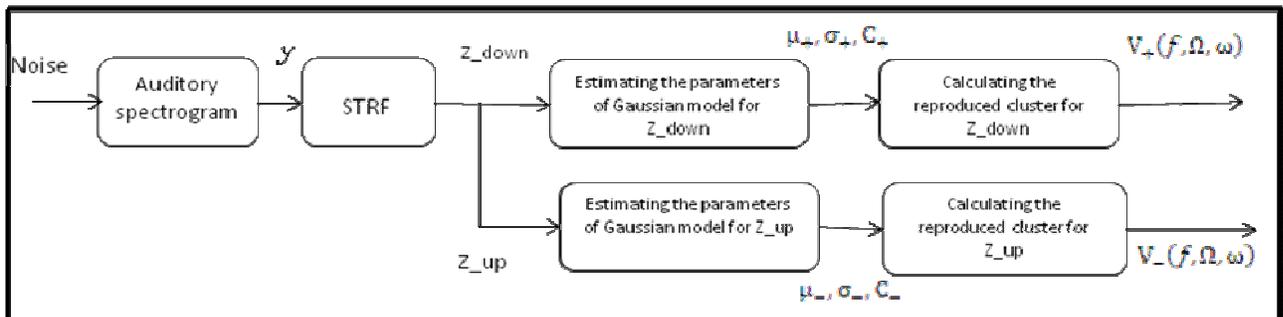


FIGURE 2: Noise Gaussian modeling in the spectro-temporal domain

The parameters include mean, covariance and gain respectively:

$$\mu_{\pm} = \frac{\sum_{t,f,\Omega,\omega} |Z_{\pm}(t, f, \Omega, \omega)| \begin{pmatrix} f \\ \Omega \\ \omega \end{pmatrix}}{\sum_{t,f,\Omega,\omega} |Z_{\pm}(t, f, \Omega, \omega)|} \quad (19)$$

$$\sigma_{\pm} = \frac{\sum_{t,f,\Omega,\omega} |Z_{\pm}(t, f, \Omega, \omega)| \begin{pmatrix} f \\ \Omega \\ \omega \end{pmatrix} - \mu_{\pm} \begin{pmatrix} f \\ \Omega \\ \omega \end{pmatrix}}{\sum_{t,f,\Omega,\omega} |Z_{\pm}(t, f, \Omega, \omega)|} \quad (20)$$

$$C_{\pm} = \frac{\sum_{t,f,\Omega,\omega} |Z_{\pm}(t, f, \Omega, \omega)|}{T} \quad (21)$$

After estimating the parameters of the Gaussian model for the noise, the reproduced cluster demonstrates the average behavior of the noise in sampling points of the spectro-temporal space for each frame. The reproduced cluster may be formulated as:

$$V_{\pm}(f, \Omega, \omega) = \frac{C_{\pm}}{(2\pi)^{3/2} |\sigma_{\pm}|} \cdot e^{-\frac{1}{2} \begin{pmatrix} f \\ \Omega \\ \omega \end{pmatrix} - \mu_{\pm} \begin{pmatrix} f \\ \Omega \\ \omega \end{pmatrix}} \cdot \sigma_{\pm}^{-2} \begin{pmatrix} f \\ \Omega \\ \omega \end{pmatrix} - \mu_{\pm} \begin{pmatrix} f \\ \Omega \\ \omega \end{pmatrix}} \quad (22)$$

The distance of each frame of input signals with this pattern represents the similarity measure of the frame behavior to the noise. Therefore, a distance measure is proposed to calculate the similarity of the modeled cluster and the input frame. For each frame, after spectro-temporal representation, the magnitude of downward and upward representation is calculated:

$$P_{t\pm}(f, \Omega, \omega) = |Z_{\pm}(t, f, \Omega, \omega)| \quad (23)$$

Our tests have shown that the phase of cortical space is not an acceptable criterion for determining speech and noise sections. Therefore, only the magnitude section of this signal has been used. The distance measure of the input frame and the modeled cluster is proposed as:

$$D_{\pm} = \frac{1}{N} \cdot \frac{\sum_{f,\Omega,\omega} (V_{\pm}(f, \Omega, \omega) \cdot |P_{t\pm}(f, \Omega, \omega) - V_{\pm}(f, \Omega, \omega)|)}{\sum_{f,\Omega,\omega} V_{\pm}(f, \Omega, \omega)} \quad (24)$$

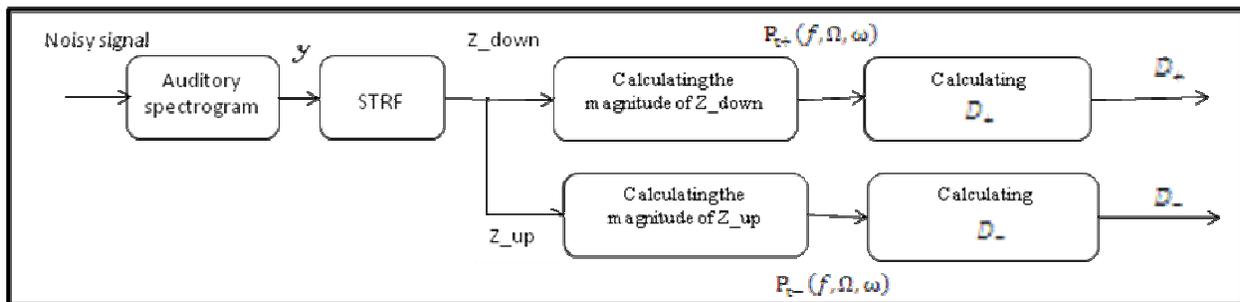


FIGURE 3: Decision making procedure for a frame in the proposed VAD system

In fact, the above relation is the weighted mean of two 3-D hyper-surfaces resulting from present frame and the average statistical behavior of the noise. The weight of this averaging has been determined in such a way that full-energy points would be more effective in this averaging. In figure 3, the block diagram shows the decision making procedure for a frame in the proposed VAD system.

Our VAD method is based on thresholding the resulted difference with an empirically set threshold.

$$VAD = \begin{cases} 0 & D_{\pm} \leq Th \\ 1 & D_{\pm} > Th \end{cases} \quad (25)$$

Determining the suitable threshold has been performed by testing in various noisy conditions which optimization results are given in section 4.

4. TESTS AND RESULTS

Evaluation framework

In the conducted tests, the speech signals are sampled in 16 KHz sampling frequency, 16 bit resolution. The length of each frame was assumed to be 4 ms. To build noisy signals; we took noises from NOISEX92 database [18] and added them to the clean signal. NOISEX database includes airport, babble, car, exhibition, office, restaurant, train, subway, street and white noises. Exhibition noise as representative of the human uproar, street noise representing open space, office noise representing office environment, Train noise representing industrial environment and white noise as the worst noise were selected. In addition, the noise was added to the clean signal in different SNRs with amounts -15, -10, -5, 0, 5, 10, 15, 20, 30, 40 db.

The proposed VAD system accuracy was measured by PS2S and PN2N probabilities. The measured parameters are defined as:

$$P_{S2S} = \frac{N_{S2S}}{N_S} \cdot 100 \quad (26)$$

$$P_{N2N} = \frac{N_{N2N}}{N_N} \cdot 100 \quad (27)$$

In the above relation, P_{S2S} is the probability of correct classification of speech frames in percents and P_{N2N} is the probability of correct classification of silence frames in percents. also, N_S is the total number of speech frames, N_N is the total number of silence frames, N_{S2S} is the number of correctly classified speech frames and N_{N2N} is number of correctly classified silence frames.

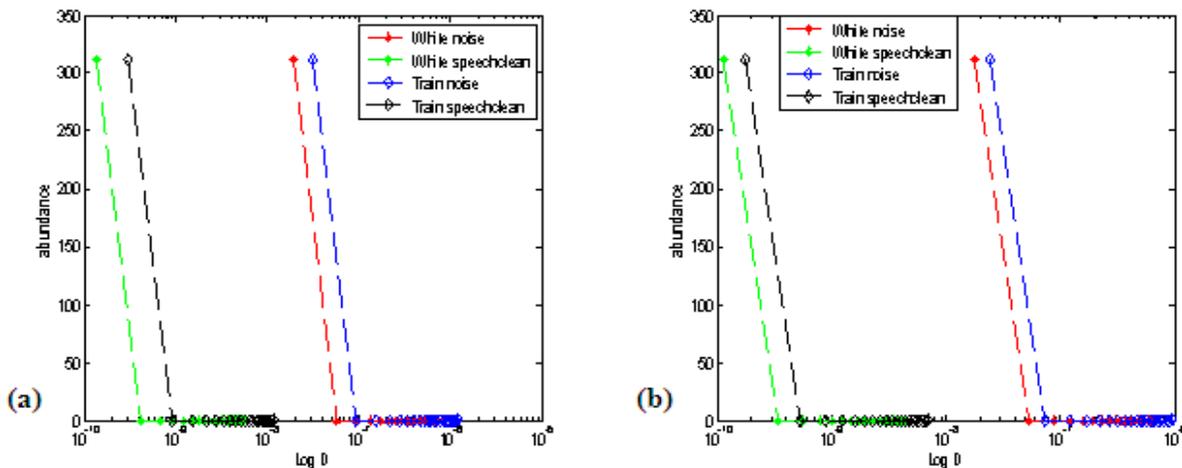


FIGURE 4: The histogram of noise behavior and speech. (a): The histogram of upward magnitude of noise behavior and speech. (b): The histogram of downward magnitude of noise behavior and speech

Histogram test

In figure (4-a) and (4-b), the histogram of upward magnitude of noise behavior and speech, and downward magnitude of noise behavior and speech have been shown in white and train noises respectively. The aim of this test is to show that Gaussian model is suitable for the noise. As seen in figures (4-a) and (4-b), we have shown that noise and speech have completely a separate behavior on the D axis.

Effect of different noises on proposed VAD

In the next experiment, the behavior of proposed VAD system was studied in different noises environments and various SNRs. In tables 1 and 2, the trend of changes in speech and non speech signal classification rates is given for various SNRs and in 5 different types of noise. As it is seen in table 1, street noise had a better behavior comparing to other noises. The system is well behaved in white, street and train noises in zero SNR. In addition, exhibition noise had worse behavior comparing to other 4 noises. In table 2, it may be observed that the classification rate of non-speech signals is equal to 100 percent for all noises and all SNRs.

Also, the figures (5-a) and (5-b) show the effect of various noises in various SNRs on a correct percentage of downward and upward magnitude of speech signal respectively. As it can be observed in both figures, a correct percentage of upward and downward magnitude of speech signal is in the Exhibition noise in zero SNR is around 59 and 69 percent and in SNR -5, is around 0 and 3 percents respectively. This may be explained by the fact that exhibition noise is human uproar and it provided the worst behavior comparing to other noises. As it may be observed in the figure, Street noise provides a better behavior comparing to other noises. Actually, this noise is produced by cars and is typically independent on the speech signal. Therefore, it is easily separable from speech signal in spectro-temporal domain.

SNR	z	PS2S White	PS2S Exhibition	PS2S Street	PS2S Train	PS2S Office
-15	Z_up	24.88	0	51.17	0	28.64
	Z_down	5.16	0	58.68	0	32.86
-10	Z_up	53.52	0	56.81	51.17	38.03
	Z_down	69.01	0	73.71	3.28	38.02
-5	Z_up	59.15	0	78.87	57.74	52.11
	Z_down	92.96	3.28	91.55	82.63	64.79
0	Z_up	92.96	54.93	93.90	91.08	58.69
	Z_down	95.77	62.44	94.37	93.90	91.55
5	Z_up	97.18	68.07	96.24	94.8	92.02
	Z_down	97.18	91.55	95.77	95.31	94.37
10	Z_up	100	92.49	99.53	96.71	95.30
	Z_down	99.53	94.37	97.65	97.18	96.24
15	Z_up	100	95.77	100	100	97.65
	Z_down	100	96.24	99.06	98.59	98.12
20	Z_up	100	98.12	100	100	100
	Z_down	100	98.12	100	100	100
30	Z_up	100	100	100	100	100
	Z_down	100	100	100	100	100
40	Z_up	100	100	100	100	100
	Z_down	100	100	100	100	100

TABLE 1: Speech signal diagnosis correctness percentage in different noises and various SNRs.

SNR	z	PN2N White	PN2N Exhibition	PN2N Street	PN2N Train	PN2N Office
-15	Z_up	100	100	100	100	100
	Z_down	100	100	100	100	100
-10	Z_up	100	100	100	100	100
	Z_down	100	100	100	100	100
-5	Z_up	100	100	100	100	100
	Z_down	100	100	100	100	100
0	Z_up	100	100	100	100	100
	Z_down	100	100	100	100	100
5	Z_up	100	100	100	100	100
	Z_down	100	100	100	100	100
10	Z_up	100	100	100	100	100
	Z_down	100	100	100	100	100
15	Z_up	100	100	100	100	100
	Z_down	100	100	100	100	100
20	Z_up	100	100	100	100	100
	Z_down	100	100	100	100	100
30	Z_up	100	100	100	100	100
	Z_down	100	100	100	100	100
40	Z_up	100	100	100	100	100
	Z_down	100	100	100	100	100

TABLE 2: Non- Speech signal diagnosis correctness percentage in different noises and various SNRs.

The figures (6-a) and (6-b) show the effect of various noises in different SNRs on a correctly classified non-speech signals using downward and upward magnitude for all 5 noises and all tested SNRs is equal to 100 percents.

Comparison of proposed VAD system behavior with other methods

In this section, the proposed VAD was compared to auditory model [19] and multifeature method [20]. The results of the three systems were shown in figures (7-a) and (7-b). In fact, the obtained results have been reported in white noise on other systems, therefore the systems are compared in these situations.

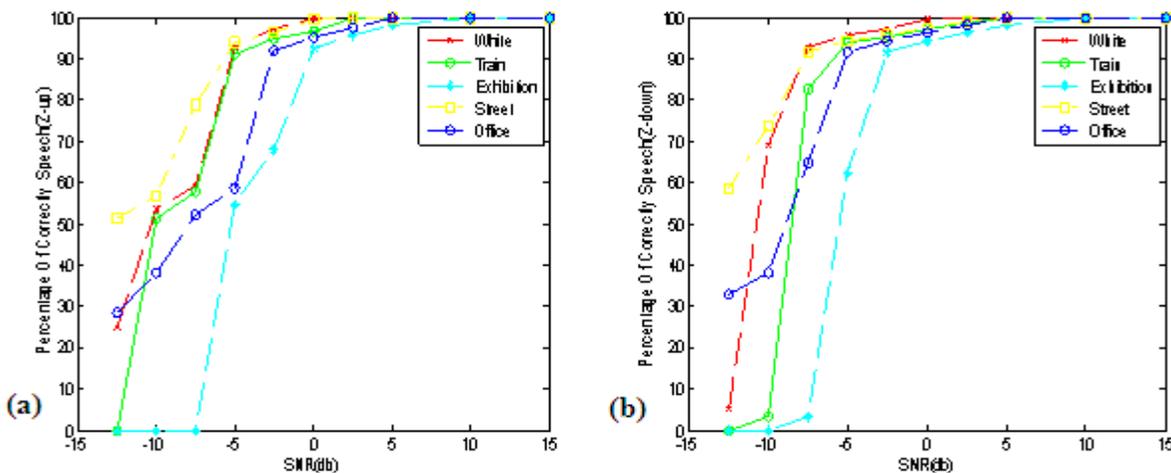


FIGURE 5: Effect of various noises in various SNRs on a correct percentage of magnitude of speech signal. (a): Effect of various noises in various SNRs on a correct percentage of upward magnitude of speech signal. (b): Effect of various noises in various SNRs on a correct percentage of downward magnitude of speech signal.

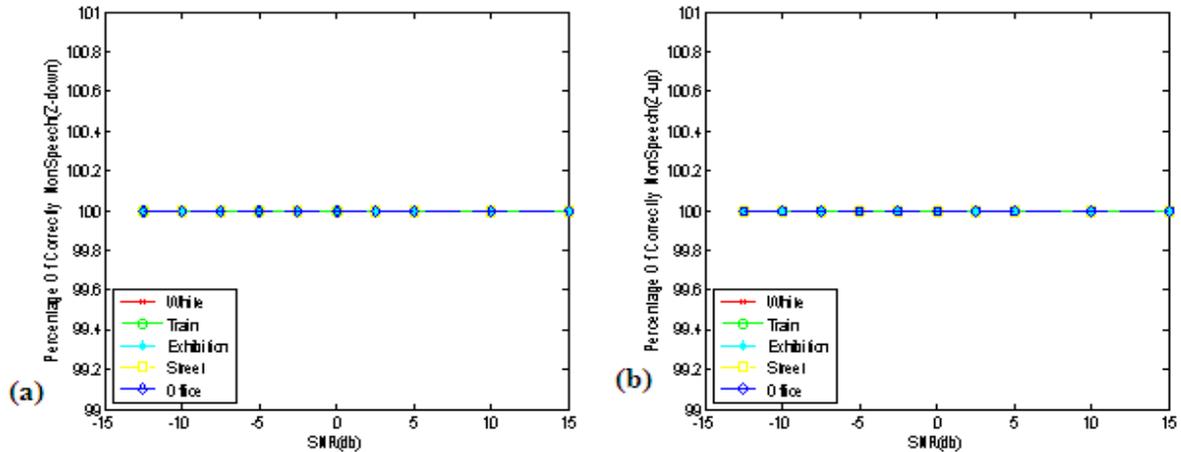


FIGURE 6: Effect of various noises in various SNRs on a correct percentage of magnitude of non-speech signal. (a): Effect of various noises in various SNRs on a correct percentage of upward magnitude of non-speech signal. (b): Effect of various noises in various SNRs on a correct percentage of downward magnitude of non-speech signal.

In figure (7-a) it can be observed that the proposed method had a much better behavior comparing to multi-feature method. However in comparison with auditory model, it is observed that the proposed method performance was better in low SNRs. In addition, figure (7-b) which is the effect of white noise on the correctness percentage of non-speech signal, comparing three systems show that the proposed VAD had a good behavior.

Behavior of correctness change with change in threshold

It is worthy to note that with a very subtle change in threshold, the rate of speech and non-speech signal classification is reduced. In figure (8-a) and (8-b) the classification rate variations in z-up and z-down versus threshold has been analyzed respectively. As seen in the figures (8-a) and (8-b), by increasing the threshold PS2N decreases and PN2S increases.

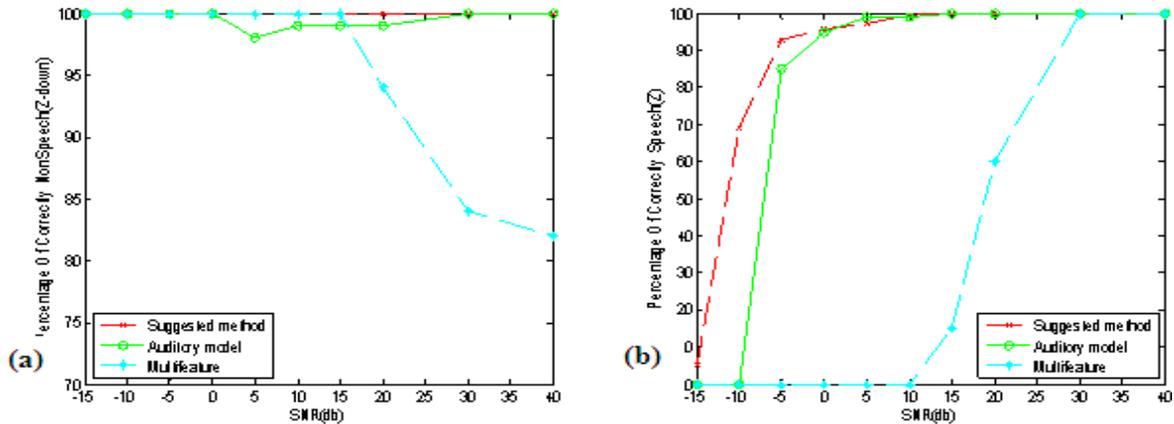


FIGURE 7: Effect of white noise on the correctness percentage for suggested method, auditory model [17] and multifeature method [18]. (a): Effect of white noise on the correctness percentage of speech signal. (b): Effect of white noise on the correctness percentage of non-speech signal

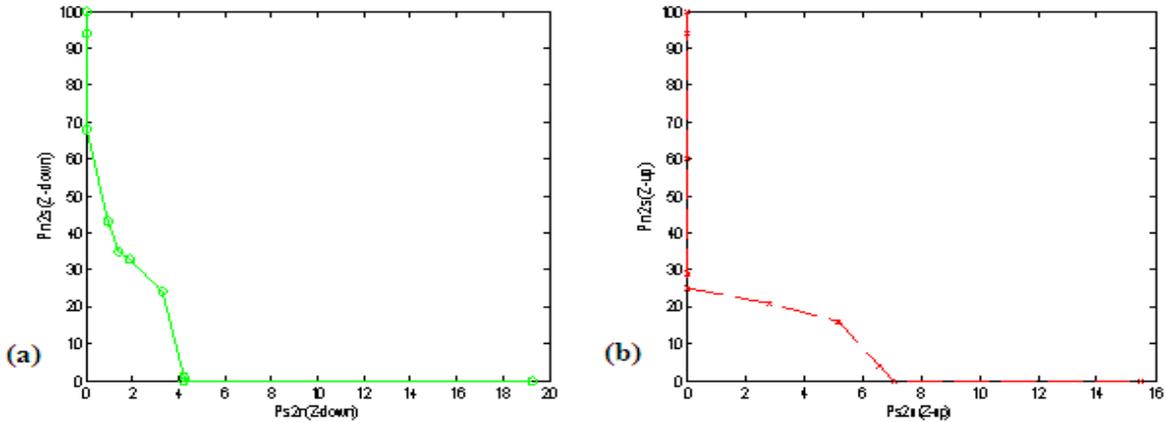


FIGURE 8: The classification rate variations in z-up (a) and z-down (b) versus threshold

5. CONCLUSION & FUTURE WORK

In this paper, a new VAD algorithm was presented on the basis of Gaussian modeling in spectro-temporal domain. The extracted features of this model in 4-D has been used in the proposed VAD. To provide the Gaussian modeling, the noise effectively passes through this space. Then a distance measurement was proposed. and applied to distinguish between noise and speech frames. Finally, the distance was compared to a given threshold for speech-silence classification. In our method, miss-classification rates were used for evaluation purposes. To provide a comparison, it was observed that the proposed method demonstrates better behavior in low SNRs. The proposed VAD algorithm can be applied to speech enhancement systems in spectro-temporal domain.

6. REFERENCES

1. N. Mesgarani, S. A Shamma, "Speech enhancement based on filtering the spectrotemporal modulations", IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), Philadelphia, March 2005.
2. N. R. Garner, P. A. Barrett, D. M. Howard, and A. M. Tyrrell, "Robust noise detection for speech detection and enhancement", Electron. Lett., Vol. 33, no. 4, pp. 270-271, Feb. 1997.
3. J.Sohn, N. S. Kim, and W.Sung, "A statistical model-based voice activity detection", IEEE Signal Process. Lett., Vol. 6, no. 1, pp. 1-3, Jan 1999.
4. L.F. Lamel, L. R. Rabiner, A. E. Rosenberg, and J. G. Wilpon, "An improved endpoint detector for isolated word recognition", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 29, pp. 777-758, 1981.
5. T. Kinnunen, E. Chernenko, M.Tuononen, P. Fränti, and H.Li, "Voice activity detection using MFCC features and support vector machine", Int. Conf. on Speech and Computer (SPECOM07), Moscow, Russia, Vol. 2, 556-561, Oct 2007.
6. J.Sohn, W.Sung, "A voice activity detector employing soft decision based noise spectrum adaptation", IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), pp. 365-368, 1998.
7. Ángel de la Torre, Javier Ramírez, Carmen Benítez, Jose C.Segura, Luz García, Antonio J.Rubio, "Noise robust model-based voice activity detection", INTERSPEECH2006, pp. 1954-1957, Pittsburgh, 2006.
8. J. -H. Chang, N. S. Kim, and S. K. Mitra, "Voice activity detection based on multiple statistical models", IEEE Trans. Signal Processing, Vol. 56, no. 6, pp. 1965-1976, June, 2006.

9. J.W.Shin, J. -H. Chang, H. S. Yun, and N. S. Kim, "Voice Activity detection based on generalized gamma distribution", IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), Vol. 1, pp. 781-784, March 2005.
10. B. Meyer and M. Kleinschmidt, "Robust speech recognition based on localized spectro-temporal features", in Proceedings of the Elektronische Sprach-und Signalverarbeitung (ESSV), Karlsruhe, 2003.
11. C.Shahnaz, W.-P.Zhu and M.O.Ahmad, "Aspectro-temporal algorithm for pitch frequency estimation from noisy observations", in Proc. 2008 IEEE ISCAS, pp. 1704-1707, May 18-21, 2008, Seattle, USA.
12. T. Chi, P. Ru, and S. A. Shamma, "Multiresolution spectrotemporal analysis of complex sounds", Journal of the Acoustical Society of America, Vol. 118, no. 2, pp. 887-906, 2005.
13. N. Kowalski, D. A. Depireux, and S. Shamma, "Analysis of dynamic spectra in ferret primary auditory cortex I. Characteristics of signal-unit response to moving ripple spectra", J.Neurophysiology, Vol. 76, no. 5, pp.3503-3523, 1996.
14. K.Wang and S. A. Shamma, "Spectral shape analysis in the central system", IEEE Trans. Speech Process. , Vol. 3, no. 5, pp. 382-395, Sep. 1995.
15. K. Wang and S. A. Shamma, " Self-normalization and noise-robustness in early auditory representations", IEEE Trans. Speech and Audio Proc, pp: 421-435, 1994.
16. S. A. Shamma, "Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve", J. Acoust. Soc. Am., pp:1622-1632, 1985
17. S. Shamma, "Methods of neuronal modeling", in Spatial and Temporal Processing in the Auditory System, pp. 411-460, MIT press, Cambridge, Mass, USA, 2nd edition, 1998.
18. A. Varga, H. J. M. Steeneken, M. Tomlinson, and D. Jones, "The NOISEX-92 study the effect of additive noise on automatic speech recognition ", Documentation included in the NOISEX-92 CD-ROMs, 1992.
19. N. Mesgarani, S. Shamma, and M. Slaney, "Speech discrimination based on multiscale spectro-temporal modulations", IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP '04), Vol. 1, pp. 601-604, Montreal, Canada, May 2004.
20. E. Scheirer, and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator", in Int. Conf. Acoustic, Speech and Signal Processing, Vol. 2, Munich, Germany, 1997, p. 1331.

On Channel Estimation of OFDM-BPSK and -QPSK over Nakagami-m Fading Channels

Neetu Sood

*Department of Electronics and Communication Engineering
National Institute of Technology, Jalandhar, India*

soodn@nitj.ac.in

Ajay K Sharma

*Department of Computer Science Engineering
National Institute of Technology, Jalandhar, India*

sharmaajayk@nitj.ac.in

Moin Uddin

National Institute of Technology, Jalandhar, India

prof_moin@yahoo.com

Abstract

This paper evaluates the performance of OFDM - BPSK & -QPSK based system with and without channel estimation over Nakagami-m fading channels. Nakagami-m variants are generated by decomposition of Nakagami random variable into orthogonal random variables with Gaussian distribution envelopes. Performance of OFDM system in Nakagami channel has been reported here. The results yield the optimum value of m based on BER and SNR. Using this optimum value of m , Channel estimation over flat fading has been reported here. It has been depicted clearly from simulated graphs that channel estimation has further reduced the BER. However, threshold value of m has played a vital role during channel estimation.

Keywords-OFDM, Fading distribution, Nakagami-m channel, Rayleigh fading, Channel estimation, Trained Symbol.

1. INTRODUCTION

OFDM technique is a multi-carrier transmission technique, which is being recognized as an excellent method for high-speed bi-directional wireless data communication. The prime idea is that all queuing data in buffer are uniformly allocated on small sub-carriers. OFDM efficiently squeezes multiple modulated carriers tightly together reducing the required bandwidth but keeping the modulated signals orthogonal so that they do not interfere with each other. OFDM that is highly efficient technique shows favorable properties such as robustness to channel fading and inter symbol interference (ISI) and is more immune to noise. OFDM system is capable of mitigating a frequency selective fading channel to a set of parallel flat fading channels, which need relatively simple processes for channel equalization.

Rayleigh and Rician fading channels have already been deployed and studied in depth for OFDM systems. OFDM Rayleigh channel simulator for OFDM has been reported in [1]. Modification to existing model of simulator was proposed by considering the correlation between the sub channels of an OFDM system resulting into reduced computational complexity. BER performance in frequency selective Rician fading channel is studied in [2]. Estimation of OFDM system in Rayleigh faded channel is provided by many techniques, in [3], used the pilot symbol along with

the previously known channel coefficients for fast Rayleigh faded channels. In [4], Timing phase estimator for OFDM system in Rayleigh faded environment is proposed with low complexity. Nakagami-m fading distribution is another useful and important model to characterize the fading channel [5]. Kang et al. [6] modeled the OFDM- BPSK system with frequency selective fading channel. The work was further enhanced by Zheng et al. [7] by presenting asymptotic BER performance of OFDM system in frequency selective Nakagami-m channel. In [8], accurate error performance of OFDM systems was provided on basis of number of channel taps in Nakagami-m fading environment.

OFDM systems have gained an equivalent attention with flat fading environment. In [9], present the method of Channel estimation and Carrier frequency offset to design an OFDM receiver in flat fading environment. However, BER performance of OFDM system in flat fading channel using DBPSK modulation technique is studied by Lijun et al. [10]. Since, the frequency selective model of Nakagami-m channel is already presented in literature, so our motivation behind this paper is to study the performance of OFDM system using flat fading channel of Nakagami-m distribution and further to improve the BER by applying channel estimation.

This paper is organized as follows: In section 2, OFDM system model is described. In section 3, the mathematical model to generate the Nakagami-m fading channel is explained along with the OFDM transmitting signal. In section 4, channel estimation technique is discussed. The analysis of simulated results of performance of OFDM system without estimation is done in section 5, while results with estimation have been presented in section 6. Finally section 7, concludes the paper.

2. MODEL DESCRIPTION

A Complex base band OFDM signal with N subcarriers, is expressed as [11]:

$$s(t) = \sum_{k=0}^{N-1} D_k e^{j2\pi k f_o t} \quad 0 \leq t \leq T \quad (1)$$

For each OFDM symbol, the modulated data sequences are denoted by $D(0), D(1), \dots, D(N-1)$. Here, f_o denote the sub-carriers spacing and is set to $f_o = 1/T$ the condition of orthogonality. After IFFT, the time-domain OFDM signal can be expressed as [11]:

$$S(n) = \frac{1}{N} \sum_{k=0}^{N-1} D_k e^{\frac{j2\pi k f_o n}{N}} \quad (2)$$

After IFFT, the modulated signal is up-converted to carrier frequency f_c and then the following signal is produced and transmitted through channel [11]:

$$x(t) = \text{Re} \left\{ \sum_{k=0}^{N-1} D_k e^{j2\pi k (f_o + f_c) t} \right\} \quad 0 \leq t \leq T \quad (3)$$

$x(t)$ represents the final OFDM signal in which sub-carriers shall undergo a flat fading channel.

3. CHANNEL MODEL

In this paper, the sub-channel spacing ($f_o = 1/T$) is chosen so that the produced parallel fading sub-channels have flat fading characteristics. So, we have chosen Nakagami-m flat fading

channel with additive white Gaussian noise. In flat fading environment, the base-band signal at the input of receiver $y(t)$ is as described as follows [11]:

$$y(t) = x(t) * r(t) + n(t) \tag{4}$$

where, $x(t)$ denotes the base-band transmitted signal, $r(t)$ is the Nakagami-m distributed channel envelope and $n(t)$ is the additive white Gaussian noise with zero mean.

Nakagami- m fading distribution function is given by [5]

$$p_R(r) = \frac{2m^m r^{2m-1}}{\Gamma(m)\Omega^m} \exp\left(-\frac{mr^2}{\Omega}\right), r \geq 0 \tag{5}$$

Where, $\Gamma(\cdot)$ is the Gamma function, $\Omega = \overline{r^2}$ is the average power, m is fading parameter and r is Nakagami distribution envelope.

Since, Nakagami distribution encompasses Scattered, reflected and direct components of the original transmitted signal [12], it can be generated using the envelope of the both random signal processes $r_{nlos}(t)$ for non line- of- sight envelope i.e. Rician and $r_{los}(t)$ for line-of-sight i.e. Rayleigh as per the following expression[12]

$$r(t) = |r_{nlos}(t)| \exp(1 - m) + |r_{los}(t)| \cdot (1 - \exp(1 - m)) \tag{6}$$

So, this value of $r(t)$ is used as envelope of Nakagami-m distributed channel.

4. CHANNEL ESTIMATION

Channel estimation in frequency selective has different approach then compared with flat fading environment. A comparative study using Minimum Mean Squared Error (MMSE) and Least square (LS) estimator in frequency selective fading environment has been presented in [13]. The channel estimation based on comb type pilot arrangement is studied using different algorithms by bahai et. al.[14]. Semi-analytical method to evaluate BER of a quadrature phase shift keying (QPSK)-OFDM system in Nakagami, $m < 1$ fading and additive noise where pilot-assisted linear channel estimation and channel equalization is described in [15]. A novel channel estimation scheme for OFDMA uplink packet transmissions over doubly selective channels was suggested in [16]. The proposed method uses irregular sampling techniques in order to allow flexible resource allocation and pilot arrangement. In flat fading environment, estimation of the channel using trained sequence of data has been studied and implemented in [17]. Channel phase was estimated during each coherence time. Then pilot data of some required percentage of data length (referred as training percentage in simulation) is inserted into the source data. It is used to estimate the random phase shift of the fading channel and train the decision to adjust the received signal with phase recover. So, finally phase estimation using training symbol is implemented in flat fading environment.

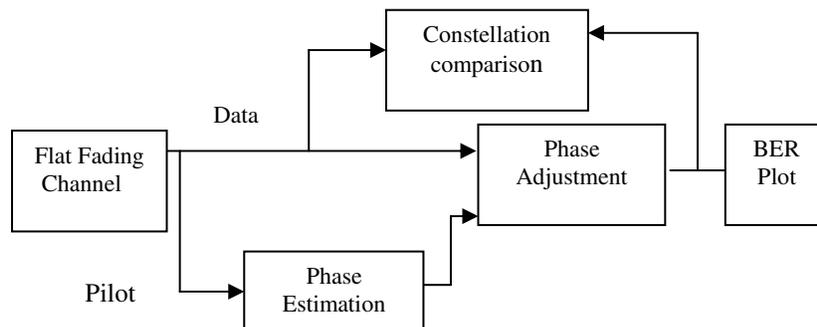


FIGURE 1: Simulation flow chart for Channel estimation [16].

In this paper, after simulating the OFDM system in Nakagami-m faded environment, OFDM system has been simulated with channel estimation scheme to obtain the improved results. The results obtained showed the significant variation in BER for with and without estimation curves.

5. RESULTS WITHOUT ESTIMATION

To analyze the performance of OFDM-BPSK and -QPSK systems over Nakagami-m fading channel, we consider the total number of sub-carriers 400, the IFFT /FFT length is chosen to be 1024 by using Guard interval of length 256.

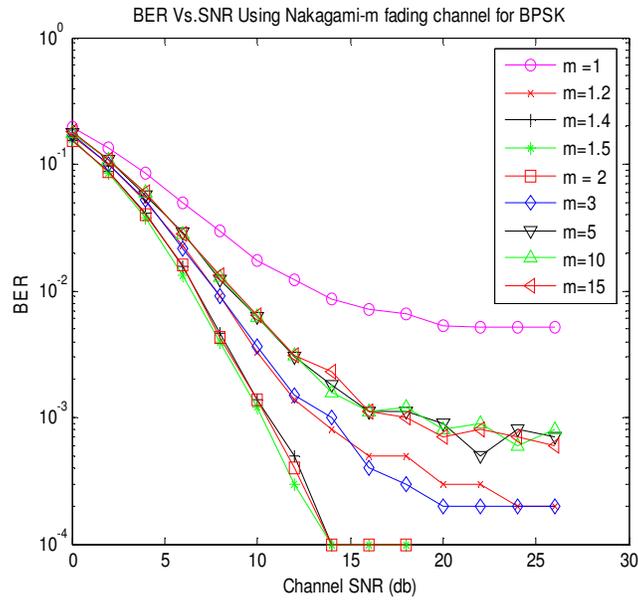


FIGURE 2: BER Vs. SNR for OFDM-BPSK system

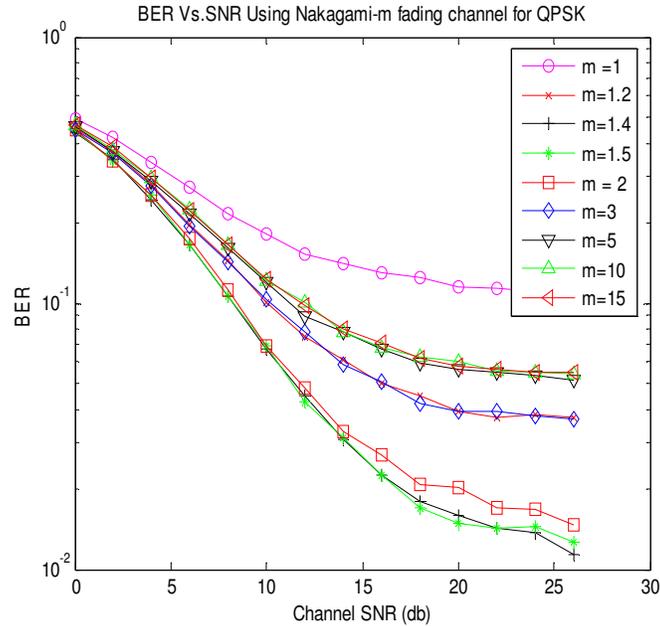


FIGURE 3: BER Vs. SNR for OFDM-QPSK System

In this section, we have presented the simulation results using MATLAB™ by implementing two modulation formats for OFDM to get threshold value of fading parameter m . Figure 2 indicates the BER Versus SNR for OFDM-BPSK with different values of fading parameter m . It is well known, that at $m = 1$, Nakagami- m fading corresponds to Rayleigh fading. So, the results for the same have been achieved through simulations. When value of m is increased, the BER starts reducing and value of 10^{-4} is reported at $m = 1.4, 1.5$ and 2 . Further, if we increase m , no reduction in BER has been reported rather it starts increasing. So threshold value of m is achieved to be 1.4 , to estimate the fading channel. This interesting fact about Nakagami- m channel has also been reported by Zheng et al. [6]. We have further analyzed OFDM system using QPSK shown in Figure 3. Results obtained for OFDM-QPSK systems are similar in nature to that of OFDM-BPSK system. BER starts decreasing with increasing value of m . It is very well depicted from graph the threshold value of $m = 1.4$ is achieved with BER value of 10^{-2} . If value of m is further increased, BER starts increasing. Here, the value of $m = 1$ gives the Rayleigh fading curve for OFDM systems.

Finally, BER performance of OFDM system in Nakagami channel degrades if we increase m beyond the certain threshold value [18].

6. RESULTS WITH ESTIMATION

Trained symbols are added to source signal as discussed in section 4. The percentage of such symbol may be varied depending upon the system response to the trained sequence. We have analyzed the results for various percentage values of trained sequence. In this paper, improved results with channel estimation and OFDM implementation has been reported with threshold value of $m = 1.4$ and varying value of training sequence over the range from 10% to 50%. Results for OFDM-BPSK and -QPSK have been indicated in Fig. 4 and 5 respectively.

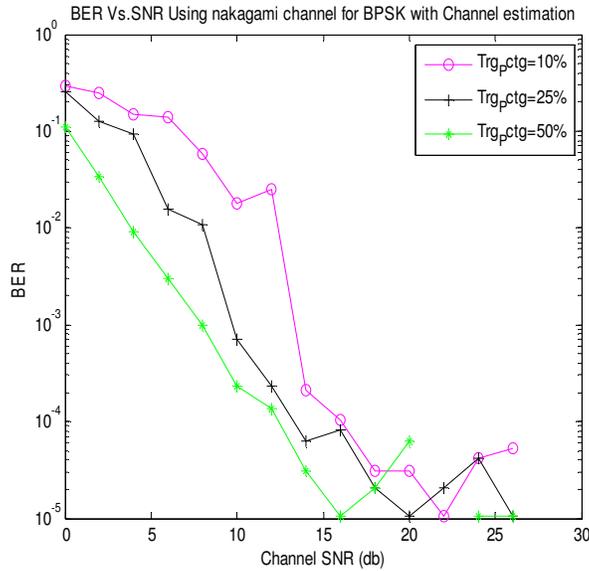


FIGURE 4: BER Vs. SNR for OFDM-BPSK System

In depth analysis of these graphs show that by increasing the amount of trained symbols BER decreases. In Figure 4, BER value at SNR of 10db has decreased from 0.017885 to 0.00022 for training percentage of 10 to 50. The final value of BER in Figure 2 has been reported to be 10^{-4} , whereas with estimation it has been reported to be 10^{-5} , Hence, there is a significant improvement over BER in the modified implementation.

However, results obtained in Figure 5 are indicating curves as compared to those obtained in Figure 4 but the value of BER for OFDM- QPSK is higher than –BPSK system. Final value of BER in Figure 4 & Figure 5 has been reported to be 10^{-2} . Channel estimation has same effect on this system, BER at SNR of 10db has decreased from 0.06587 to 0.03912 because of increased value of training symbol.

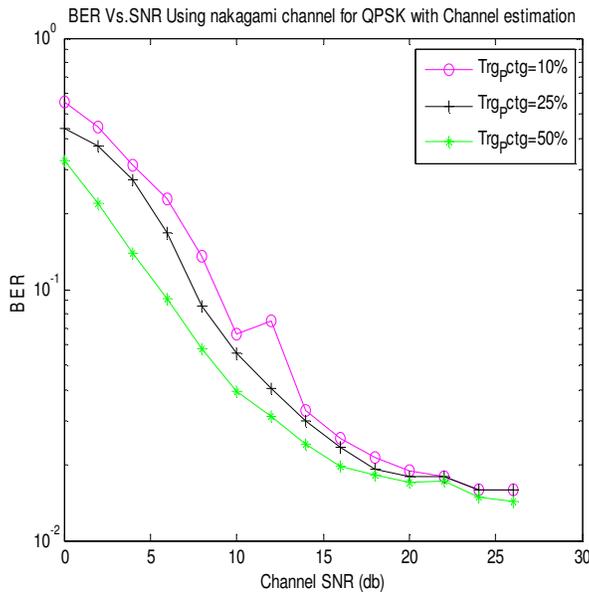


FIGURE 5: BER Vs. SNR for OFDM-QPSK System

7. CONCLUSIONS

In this paper, we have evaluated the performance of OFDM system using BPSK and QPSK with OFDM using Nakagami-m fading channel. Further the enhanced system performance has been implemented by phase estimation of channel. Threshold value of m has played a significant role in channel estimation as it provided the minimum value of BER.

8. REFERENCES

- [1] Moradi, S.Gazor, "An OFDM Rayleigh Fading Channel Simulator," In Proceedings of Vehicular Technology Conference, 2006. VTC-2006 IEEE 64th, 25-28 Sept. 2006
- [2] Jun Lu, Thiang Tjhung, Fumiyuki Adachi and Cheng Li Huang, "BER performance of OFDM-MDPSK system in Frequency –Selective Rician Fading with Diversity Reception," IEEE Trans. On Vehicular Technology, vol. 49, no. 4, pp. 1216-1225, July 2000.
- [3] Jin Goog Kim, Tae Joon and Jong Tae Lim,, "Channel estimation for OFDM over Fast Rayleigh Fading Channels," In Proceedings of world Academy of science and technology, vol. 21, pp. 455-458, Jan. 2007.
- [4] Young Jae Ryu and Dong Seog Han, "Timing phase estimator overcoming Rayleigh Fading For OFDM systems," IEEE Transaction on Consumer Electronics, vol.47, no.3, pp.370-377, Aug 2001.
- [5] M. Nakagami, "The m-distribution—A general formula of intensity distribution of rapid fading," in Statistical Methods in Radio Wave Propagation, W. C. Hoffman, Ed. Elmsford, NY: Pergamon, 1960.
- [6] Zhejin Kang, Kung Yao, Flavio Lorenzelli, "Nakagami-m Fading Modeling in the Frequency Domain for OFDM system analysis," IEEE Communication letters, vol. 7, no.10, pp. 484-486, Oct.2003.
- [7] Zheng du, Julian Cheng and Norman c. Beaulieu, "Asymptotic BER performance of OFDM in Frequency Selective Nakagami-m Channels," In Proceeding of IEEE Conference on Vehicular Technology , vol. 1, pp. 612-615, Sept. 2004.
- [8] Zheng du, Julian Cheng and Norman c. Beaulieu, "Accurate Error Rate Performance Analysis of OFDM on Frequency Selective Nakagami-m Fading Channels," IEEE Trans.on communications.vol. 54, no. 2, pp. 319-328, Feb. 2006.
- [9] Zhangyong Ma and Young-il- Kim, "A Novel OFDM receiver in Flat Fading Channel," In Proceeding of IEEE Conference on advanced communication technology, ICACT, Vol.. 2, pp. 1052-54, 2005.
- [10] Song Lijun, Tang Youxi, Li Shaoqian and Huang Shunji, " BER Performance of Frequency Domain Differential Demodulation OFDM in Flat Fading Channel," Journal of Electronic Science and Technology of China, Vol. 1, no. 1, Dec. 2003.
- [11] Ruhallah Ali Hemmati, Paeiz Azmi, "Clipping distortion mitigation in OFDM systems over fading channels by using DFT-based method" Elsevier Computers and Electrical Engineering, vol.31, issue 7, pp 431-443, Oct. 2005.
- [12] Jyoteesh Malhotra, Ajay K. Sharma, R.S Kaler, "Investigation on First Order Performance Metrics in the Nakagami-m Fading Channel," In the proceeding of conference of Design Technique for Modern Electronic Devices, VLSI and Communication Systems, 14th-15th May, 2007.
- [13] Jan-Jaap van de Beek, Ove Edfors, Magnus Sandell, Sarah Kate Wilson and Per Ola B. rjesson, "On Channel Estimation In OFDM Systems," In Proceedings of Vehicular Technology Conference (VTC 95), vol. 2, pp. 815-819, Chicago, USA, September 1995.

- [14] Sinem Coleri , Mustafa Ergen , Anuj Puri , Ahmad Bahai, "A study of channel estimation in OFDM systems," In Proceedings of Vehicular Technology Conference(VTC'02), Vancouver, Canada, September, 2002.
- [15] V. Subotic, S. Primak, "BER Analysis of Equalized OFDM Systems in Nakagami, $m < 1$ Fading," in Wireless Personal Communications, An International Journal, vol. 40, Issue 3, Feb. 2007, pp. 281 – 290.
- [16] P. Fertl and G. Matz, "Multi-user channel estimation in OFDMA uplink systems based on irregular sampling and reduced pilot overhead," In Proceeding of IEEE ICASSP 2007.
- [17] Zhifeng Chen, "Performance Analysis of Channel Estimation and Adaptive Equalization in Slow Fading Channel," University of Florida, <http://users.ecel.ufl.edu/~zhifeng>.
- [18] Neetu Sood, Ajay K Sharma, Moin Uddin, " BER Performance of OFDM-BPSK and - QPSK over Nakagami-m Fading Channels," In Proceeding of IEEE-IACC'2010, Patiala, India, Feb. 2010.

CALL FOR PAPERS

Journal: Signal Processing: An International Journal (SPIJ)

Volume: 4 **Issue:** 5

ISSN: 1985-2339

URL: <http://www.cscjournals.org/csc/description.php?JCode=SPIJ>

About SPIJ

The International Journal of Signal Processing (SPIJ) lays emphasis on all aspects of the theory and practice of signal processing (analogue and digital) in new and emerging technologies. It features original research work, review articles, and accounts of practical developments. It is intended for a rapid dissemination of knowledge and experience to engineers and scientists working in the research, development, practical application or design and analysis of signal processing, algorithms and architecture performance analysis (including measurement, modeling, and simulation) of signal processing systems.

As SPIJ is directed as much at the practicing engineer as at the academic researcher, we encourage practicing electronic, electrical, mechanical, systems, sensor, instrumentation, chemical engineers, researchers in advanced control systems and signal processing, applied mathematicians, computer scientists among others, to express their views and ideas on the current trends, challenges, implementation problems and state of the art technologies.

To build its International reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for SPIJ.

SPIJ List of Topics

The realm of International Journal of Signal Processing (SPIJ) extends, but not limited, to the following:

- Biomedical Signal Processing
- Communication Signal Processing
- Detection and Estimation
- Earth Resources Signal Processing
- Industrial Applications
- Optical Signal Processing
- Radar Signal Processing
- Acoustic and Vibration Signal Processing
- Data Processing
- Digital Signal Processing
- Geophysical and Astrophysical Signal Processing
- Multi-dimensional Signal Processing
- Pattern Recognition
- Remote Sensing

- Signal Filtering
- Signal Processing Technology
- Software Developments
- Spectral Analysis
- Stochastic Processes
- Signal Processing Systems
- Signal Theory
- Sonar Signal Processing
- Speech Processing

IMPORTANT DATES

Volume: 5

Issue: 1

Paper Submission: October 31, 2010

Author Notification: November 30, 2010

Issue Publication: December 2010

CALL FOR EDITORS/REVIEWERS

CSC Journals is in process of appointing Editorial Board Members for ***Signal Processing: An International Journal (SPIJ)***. CSC Journals would like to invite interested candidates to join **SPIJ** network of professionals/researchers for the positions of Editor-in-Chief, Associate Editor-in-Chief, Editorial Board Members and Reviewers.

The invitation encourages interested professionals to contribute into CSC research network by joining as a part of editorial board members and reviewers for scientific peer-reviewed journals. All journals use an online, electronic submission process. The Editor is responsible for the timely and substantive output of the journal, including the solicitation of manuscripts, supervision of the peer review process and the final selection of articles for publication. Responsibilities also include implementing the journal's editorial policies, maintaining high professional standards for published content, ensuring the integrity of the journal, guiding manuscripts through the review process, overseeing revisions, and planning special issues along with the editorial team.

A complete list of journals can be found at <http://www.cscjournals.org/csc/byjournal.php>. Interested candidates may apply for the following positions through <http://www.cscjournals.org/csc/login.php>.

Please remember that it is through the effort of volunteers such as yourself that CSC Journals continues to grow and flourish. Your help with reviewing the issues written by prospective authors would be very much appreciated.

Feel free to contact us at coordinator@cscjournals.org if you have any queries.

Contact Information

Computer Science Journals Sdn Bhd

M-3-19, Plaza Damas Sri Hartamas
50480, Kuala Lumpur MALAYSIA

Phone: +603 6207 1607

+603 2782 6991

Fax: +603 6207 1697

BRANCH OFFICE 1

Suite 5.04 Level 5, 365 Little Collins Street,
MELBOURNE 3000, Victoria, AUSTRALIA

Fax: +613 8677 1132

EMAIL SUPPORT

Head CSC Press: coordinator@cscjournals.org

CSC Press: cscpress@cscjournals.org

Info: info@cscjournals.org

COMPUTER SCIENCE JOURNALS SDN BHD
M-3-19, PLAZA DAMAS
SRI HARTAMAS
50480, KUALA LUMPUR
MALAYSIA