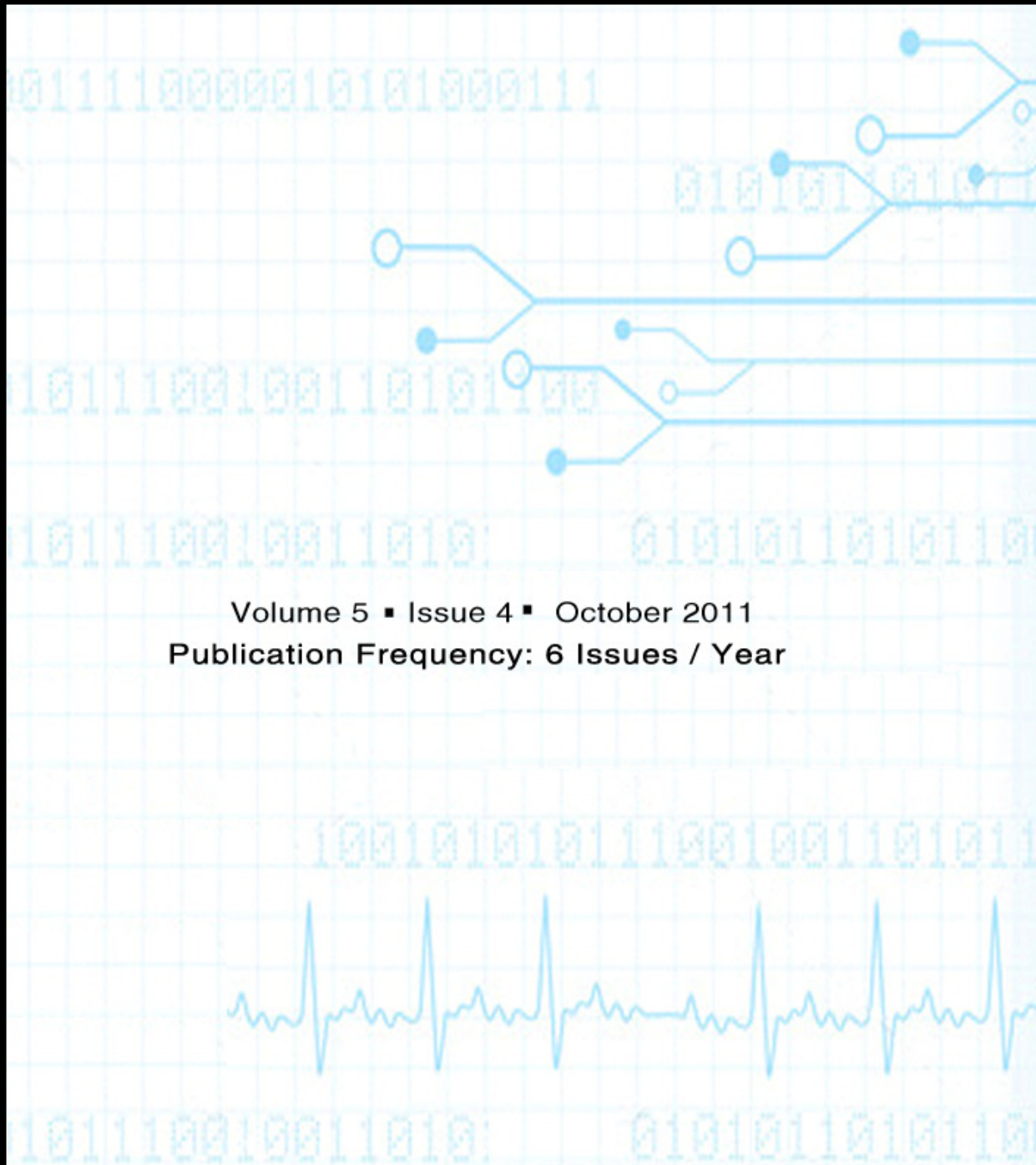Editor-in-Chief
Dr Saif alZahir

SIGNAL PROCESSING (SPIJ)

AN INTERNATIONAL JOURNAL

Volume 5 ▪ Issue 4 ▪ October 2011
Publication Frequency: 6 Issues / Year

SIGNAL PROCESSING: AN INTERNATIONAL JOURNAL (SPIJ)

VOLUME 5, ISSUE 4, 2011

EDITED BY
DR. NABEEL TAHIR

EDITORIAL PREFACE

This is third issue of volume five 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.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Starting with volume 5, 2011, SPIJ appears in more focused issues. Besides normal publications, SPIJ intend to organized special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

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)

TABLE OF CONTENTS

Volume 5, Issue 4, October 2011

Pages

Comparison Of LDM and LMS for an Application of a Speech

V.A.Mane   *vikram_mane34@yahoo.com*
Department of E&TC
Annasaheb Dange COE
Ashta, 416302, India

K.P.Paradeshi   *kiranparadeshi@rediffmail.com*
Department of E&TC
Annasaheb Dange COE
Ashta, 416302, India

S.A.Harage   *shrikant.harage@rediffmail.com*
Department of E&TC
Annasaheb Dange COE
Ashta, 416302, India

M.S.Ingavale   *mangesh_ingavale1982@rediffmail.com*
Department of E&TC
Annasaheb Dange COE
Ashta, 416302, India

Abstract

Automatic speech recognition (ASR) has moved from science-fiction fantasy to daily reality for citizens of technological societies. Some people seek it out, preferring dictating to typing, or benefiting from voice control of aids such as wheel-chairs. Others find it embedded in their Hitech gadgetry – in mobile phones and car navigation systems, or cropping up in what would have until recently been human roles such as telephone booking of cinema tickets. Wherever you may meet it, computer speech recognition is here, and it's here to stay.

Most of the automatic speech recognition (ASR) systems are based on Gaussian Mixtures model. The output of these models depends on subphone states. We often measure and transform the speech signal in another form to enhance our ability to communicate. Speech recognition is the conversion from acoustic waveform into written equivalent message information. The nature of speech recognition problem is heavily dependent upon the constraints placed on the speaker, speaking situation and message context. Various speech recognition systems are available. The system which detects the hidden conditions of speech is the best model. LMS is one of the simple algorithm used to reconstruct the speech and linear dynamic model is also used to recognize the speech in noisy atmosphere..This paper is analysis and comparison between the LDM and a simple LMS algorithm which can be used for speech recognition purpose.

Keywords : White Noise, Error Covariance Matrix, kalman Gain, LMS Cross Correlation

1. INTRODUCTION
Speech is a form of communication in everyday life. It existed since human civilizations began and even till now, speech is applied to high technological telecommunication systems. A particular field, which I personally feel, will excel   be speech signal processing in the world of telecommunications. As applications like Cellular and satellite technology are getting popular among mankind, human beings tend to demand more advance technology and are in search of improved applications. For this reason, researchers are looking closely into the four generic attributes of speech coding. They are complexity, quality, bit rate and delay. Other issues like robustness to transmission errors,

multistage encoding/decoding, and accommodation of non-voice signals such as in-band signaling and voice band modem data play an important role in coding of speech as well.

Presently Speech processing has been a growing and dynamic field for more than two decades and there is every indication that this growth will continue and even accelerate. During this growth there has been a close relationship between the developments of new algorithms and theoretical results, new filtering techniques are also of consideration to the success of speech processing.
A least mean square (LMS) adaptive filtering approach has been formulated for removing the deleterious effects of additive noise on the speech signal; unlike the classical LMS adaptive filtering scheme, the proposed method is designed to cancel out the clean true speech signal. This method takes advantage of the quasi-periodic nature of the speech signal to form an estimate of the clean speech signal at time t from the value of the signal at time t minus the estimated pitch period. For additive white noise distortion, preliminary tests indicate that the method improves the perceived speech

One of the common adaptive filtering techniques that are applied to speech is the Wiener filter. This filter is capable of estimating errors however at only very slow computations. On the other hand, the Kalman filter suppresses this disadvantage. As widely known to the world, Kalman filtering techniques are used on GPS (Global Positioning System) and INS (Inertial Navigation System). Nonetheless, they are not widely used for speech signal coding applications. According to, the reason why Kalman filter is so popular in the field of radar tracking and navigating system is that it is an optimal estimator, which provides very accurate estimation of the position of either airborne objects or shipping vessels. Due to its accurate estimation characteristic, electrical engineers are picturing the Kalman filter as a design tool for speech, whereby it can estimate and resolve errors that are contained in speech after passing through a distorted channel. Due to this motivating fact, there are many ways a Kalman filter can be tuned to suit engineering applications such as network telephony and even satellite phone conferencing. Knowing the fact that preserving information, which is contained in speech, is of extreme importance, the availability of signal filters such as the Kalman filter is of great importance.

2. EARLY APPROACHES TO SPEECH RECOGNITION
Automatic speech recognition might appear to be an almost unattainable goal. However, by concentrating on a reduced specification and by tracking the problems in a scientific and staged manner, it has been possible to make considerable progress in understanding the precise nature of the problems and in development of relevant and practical solutions [12]. However, this has not always been the case. Some of the early work, which interesting in the context of a review of different approaches to automatic speech recognition, tended to be either overambitious about the achievements that could realistically be expected to be realized or somewhat naive with regard to the real difficulties that were being tackled.

Early attempts can thus be categorized into one of two main approaches. In the fifties and sixties the main approach was based on simple principles of 'pattern matching' that in the seventies gave way to a 'knowledge engineering 'or rule based approach. Only towards the end of the seventies there was a growing awareness of the need to integrate these two approaches and move towards a clear and scientific recognition- a move that ultimately led to a maturation of ides and algorithms, which are now beginning to provide powerful exploitable solutions [10].
The following section reviews some of these early approaches to automatic speech recognition.

2.1   Pattern matching
Such systems employ two modes of operation: a 'training mode' in which example speech patterns (usually words) are stored as reference 'templates' and a recognition mode in which incoming speech patterns are compared with each reference pattern that is most similar to the input pattern determines the result. In this scheme the acoustic pattern of a speech signal typically consisted of a sequence of vectors, which had been derived from the speech waveform using some form of 'preprocessing'. For example it was common to perform a frequency analysis by means of an FFT or a filter bank in order to produce vectors that correspond to the short-time power spectrum of the signal into discrete pattern segments.

The key to success of this approach is the comparison process, and a technique called 'linear time normalization' was commonly used in order to overcome variability in the duration of spoken words. In this situation, the lengths of the patterns were 'time normalized' to a standard duration by lengthening (or shortening) the patterns the appropriate amount by using a fixed expansion (or compression) of the time scale uniformly over the entire pattern

2.2 Knowledge engineering

The knowledge-based approach popular in the early seventies was based on techniques from the field of artificial intelligence, which was the newly emerging. These techniques were applied to traditional concepts from the disciplines of phonetic and linguistics about how speech signals was organizing. The key principle was to exploit the speech knowledge through its exploit use within a rule-based framework aimed at deriving and interpretation that would be suitable for the purpose of understanding the semantic content of the signal.

2.3 Integrated approach:

This new process (popular in the late seventies) became known as 'dynamic time warping'(DTW) and it has been a highly successful technique in terms of raising performance to a level at which serious commercialization of automatic speech recognition systems could begin.

3. LMS ALGORITHM

A linear mean square (LMS) adaptive filtering approach has been formulated for removing the deleterious effects of additive noise on the speech signal; unlike the classical LMS adaptive filtering scheme, the proposed method is designed to cancel out the clean true speech signal [11]. . An adaptive LMS filter was employed to process speech in signal-to-noise ratios (S/N) varying from -8 to +12 dB. The filter configuration is commonly called noise cancellation [12][14][15].
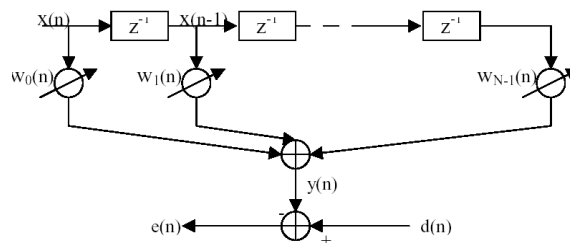
FIGURE 1: LMS model

$$Y(n) = \sum_{i=0}^{N-1} W_i(n)x(n-i)$$

$$e(n) = d(n) - y(n)$$

We assume that the signals involved are real-valued.

The LMS algorithm changes (adapts) the filter tap weights so that e(n) is minimized in the mean-square sense. When the processes x(n) & d(n) are jointly stationary, this algorithm converges to a set of tap-weights which, on average, are equal to the Wiener-Hopf solution.

The LMS algorithm is a practical scheme for realizing Wiener filters, without explicitly solving the Wiener-Hopf equation.

The conventional LMS algorithm is a stochastic implementation of the steepest descent algorithm. It simply replaces the cost function

$$\zeta = E[e2(n)] \qquad \text{Substituting } \zeta = e2(n)$$

For in the steepest descent recursion, we obtain

$$\overline{W}.(n+1) = \overline{W}(n) - \mu \overline{\nabla} e^2(n)$$

$$\dots W_{N-1}(n)$$

$$\left[\nabla = \frac{\partial}{\partial w0} \quad \frac{\partial}{\partial w1} \dots \dots \frac{\partial}{\partial w \; n-1} \right]$$

Note that the i-th element of the gradient vector   is

$$\frac{\partial e^2(n)}{\partial wi} = 2e(n)\frac{\partial e^2(n)}{\partial wi}$$

$$= -2e(n)\frac{\partial y(n)}{\partial wi}$$

$$= -2e(n)x(n-i)$$

Then

Where   (n) =[
Finally we obtain-
                    +2µe
Equation is referred to as the LMS recursion.
Summary of the LMS algorithm,
Input:
Tap -weight vector:
Input vector:
Desired output: d
Output:
Filter output: y
Tap -weight vector update:
1. Filtering: y
2. Error estimation:
3. Tap-weight vector adaption: +2µe

4. LINEAR DYNAMIC MODEL (KALMAN FILTER)

In 1960, R.E. Kalman published his famous paper describing a recursive solution to the discrete-data linear filtering problems [10]. Since that time, the Kalman filter has been the subject of extensive research and application, particularly in the area of autonomous or assisted navigation. It is only a tool – It aids mankind in solving problems, however, it does not solve any problem all by itself. This is however not a physical tool, but a mathematical one, which is made from mathematical models. In short, essentially tools for the mind. They help mental work become more efficient, just like mechanical tools, which make physical work less tedious. Additionally, it is important to understand its use and function before one can apply it effectively .It is a computer program - It uses a finite representation of the estimation problem, which is a finite number of variables; therefore this is the reason why it

is said to be "ideally suited to digital computer implementation". However, assuming that these variables are real numbers with infinite precision, some problems do happen. This is due from the distinction between finite dimension and finite information, and the distinction between "finite" and "manageable" problem sizes. On the practical side when using Kalman filtering, the above issues must be considered according to references [1; 2; 3; 4].
Mathematical analysis of Kalman filter

Following discussions from references [4; 5; 6; 7; 10]
After going through some of the introduction and advantages of using Kalman filter, we will now take a look at the process of this magnificent filter. The process commences with the addresses of a general problem of trying to estimate the state $x \in \mathfrak{R}^n$ of a discrete-time controlled process that is governed by a linear stochastic difference equation:
xk=Axk-1+Buk+wk-1   (1.1)
With measurement
zk=Hxk+vk   (1.2)
The random variables wk and vk represent the process and measurement noise respectively. We assume that they are independent of each other, and with normal probability distributions
p(w)~N(0,Q)   (1.3)
p(v)~N(0,R)   (1.4)

.
Ideally, the process noise covariance Q and measurement noise covariance R matrices are assumed to be constant, however in practice, they might change with each time step or measurement.

In the absence of either a driving function or process noise, the matrix nxn A in the difference equation (1.1) relates the state at the previous time step to the state at k-1 to the current step k. In practice, A might change with each time step, however here it is assumed constant. The nxl matrix B relates the optional control input to the state x., which is a matrix in the measurement equation (1.2), which relates the state to the measurement. In practice x might change with each time step or measurement, however we assume it is constant.

4.1 Discrete Kalman Filter

This section will begin with a broad overview, covering the "high-level" operation of one form of the discrete Kalman filter. After presenting this high-level view, I will narrow the focus to the specific equations and their use in this discrete version of the filter. How does the Kalman filter works? Firstly, it estimates a process by using a form of feedback control loop whereby the filter estimates the process state at some time and then obtains feedback in the form of (noisy) measurements. As such, these equations for the Kalman filter fall into two groups: "Time Update equations" and "Measurement Update equations". The responsibilities of the time update equations are for projecting forward (in time) the current state and error covariance estimates to obtain the priori estimates for the next time step. The measurement update equations are responsible for the feedback i.e. for incorporating a new measurement into the priori estimate to obtain an improved posteriori estimate. The time update equations can also be thought of as "predictor" equations, while the measurement update equations can be thought of as "corrector" equations. By and large, this loop process of the final estimation algorithm resembles that of a predictor-corrector algorithm for solving numerical problems just like the one shown in fig below
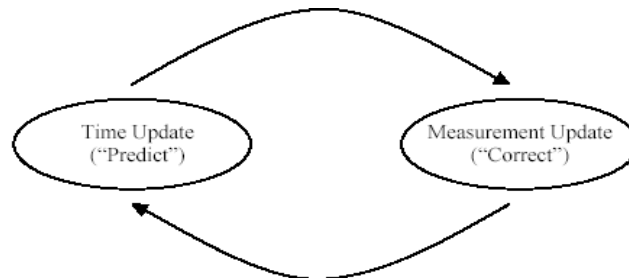
FIGURE 2: Discrete Kalman Filter Cycle

As the time update projects the current state estimate ahead in time, the measurement update adjusts the projected estimate from the time update by an actual measurement at that particular time. The specific equations for the "time" and "measurement" updates are presented below in Table1

Xk=Axk-1+Buk 1.5
Pk=Apk-1AT+Q 1.6

Table 1: Time update equations

Kk=PkHT(HPkHT+R)-1 1.7
Xk=xk+Kk(zk-Hxk) 1.8
Pk=(I-KkH)Pk 1.9

TABLE 2: Measurement equations

Once again, notice how the time update equations in Table.1 project its state, x the filter are discussed in the earlier section. By referring to Table 1, it is obvious that the first task during the measurement update is to compute the Kalman gain, By (1.7) in the table above is to actually measure the process in order to obtain, and then to generate a posteriori state estimate, by incorporating the measurement as in (1.8). Once again, notice the repeated equation of (1.7) here and (1.8) for completeness. Finally, the last step is to obtain a posteriori error covariance estimate via (1.9). Thus, after each time and measurement update pair, this loop process is repeated to project or predict the new time step priori estimates using the previous time step posteriori estimates. This recursive nature is one of the very appealing features of the Kalman filter it makes practical implementations much more feasible than (for example) an implementation of a Wiener filter which is designed to operate on all of the data directly for each estimate. Instead, the Kalman filter recursively conditions the current estimate on all of the past measurements. The high-level diagram of Fig 2 is combined with the equations from Table .1 and Table 2, in Fig.2 as shown below, which offers a much more complete and clear picture of the operation of the Kalman Filter.
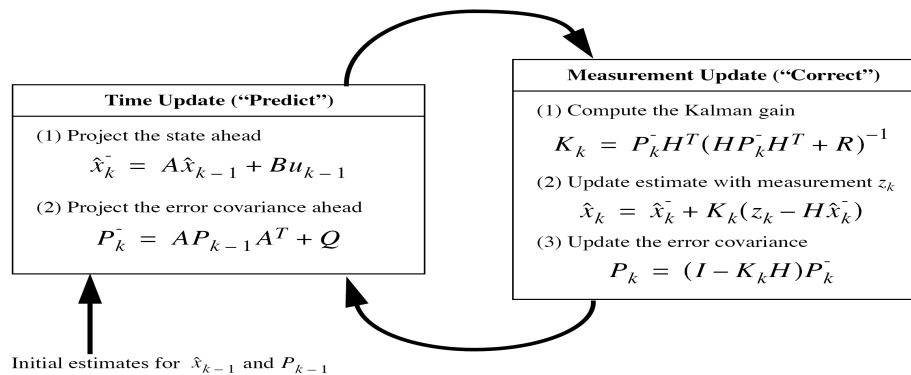

Time Update ("Predict")

(1) Project the state ahead

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1}$$

(2) Project the error covariance ahead

$$P_k^- = AP_{k-1}A^T + Q$$

Measurement Update ("Correct")

(1) Compute the Kalman gain

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1}$$

(2) Update estimate with measurement z_k

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-)$$

(3) Update the error covariance

$$P_k = (I - K_k H)P_k^-$$

Initial estimates for $\hat{x}_{k-1}$ and P_{k-1}

FIGURE 1: Complete picture of Kalman filter

The measurement noise covariance is usually measured before the operation of the filter when it comes to the actual implementation of Kalman filter. Generally, measuring the measurement noise covariance is practically possible due to the fact that the necessary requirement to measure the process noise covariance (while operating the filter), therefore it should be possible to take some off-line sample measurements in order to determine the variance of the measurement noise. As for determining of the process noise covariance, it will be generally more difficult. This is due to the reason that the process to be estimated is unable to be directly observed. Sometimes a relatively simple (poor) process model can produce acceptable results if one "injects" enough uncertainty into the process via the selection of. (Certainly, one would hope that the process measurements are reliable**).** In either case, whether or not a rational basis is chosen for the parameters, superior filter performance (statistically speaking) can be obtained by tuning the filter parameters R and Q. closing under conditions where R and Q are in fact constant, both the estimation error covariance and the Kalman gain will stabilize quickly and then remain constant (see the filter update equations in Fig 2). If this is the case, these parameters can be pre-computed by either running the filter off-line, or for example by determining the steady-state value.

5. COMPARISON OF LDM AND LMS

Optimal adjustment parameters of the adaptive filter with LMS algorithm in the practical application of suppression of additive noise in a speech signal for voice communication with the control system. By the proposed method, the optimal values of parameters of adaptive filter are calculated with guarantees the stability and convergence of the LMS algorithm [9] same as that of the LDM[16].

The proposed methods of recognition of speech give the following results on three different speeches

S1 is a noiseless speech sample and S1white is the S1 speech captured by white noise. The original speech sample, S1white and reconstructed speech samples with LDM and LMS are shown below. The cross covariance between the reconstructed and noiseless speech samples is also shown below.
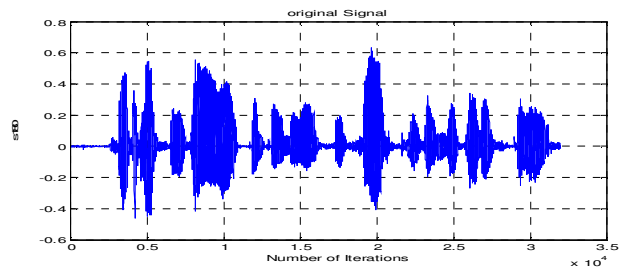
FIGURE 1a. S1 original speech sample
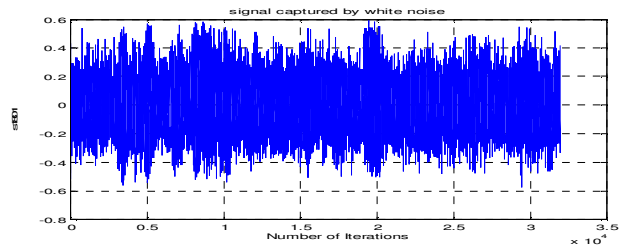
Fig 1b. S1 white speech sample captured by white noise
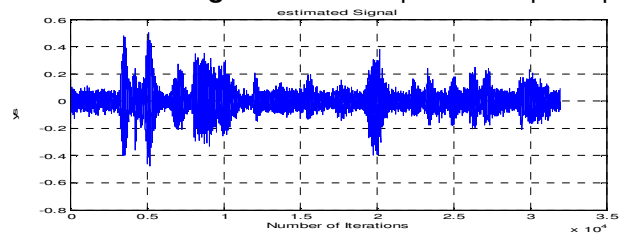
Fig 1c. Reconstructed speech by LDM

Fig 1d. Reconstructed speech by LMS

Fig 1e. Cross correlation between original and reconstructed signal by LDM

Fig 1f. Cross correlation between original and reconstructed signal by LMS

S2 is a noiseless speech sample and S2street is the S2 speech captured by street noise. The original speech sample S2, S2street and reconstructed speech samples with LDM and LMS are shown below. The cross covariance between the reconstructed and noiseless speech samples is also shown below.

Fig 2a. S2 original speech sample
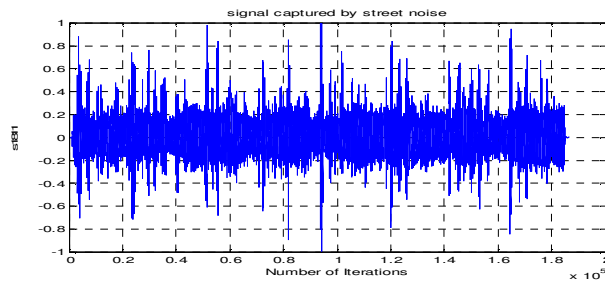
Fig 2b. S2street speech sample captured by street noise
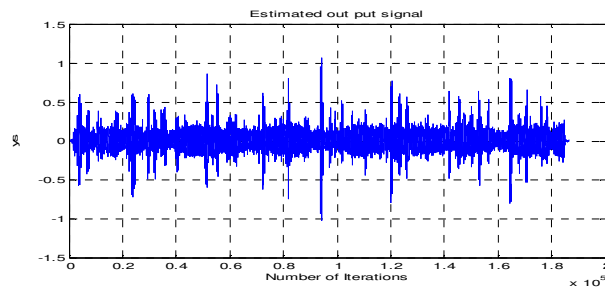
Fig 2c. Reconstructed speech by LDM
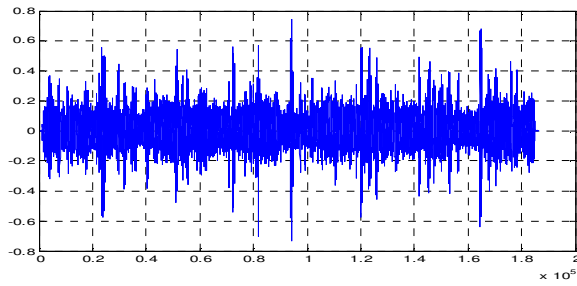
Fig 2d. Reconstructed speech by LMS
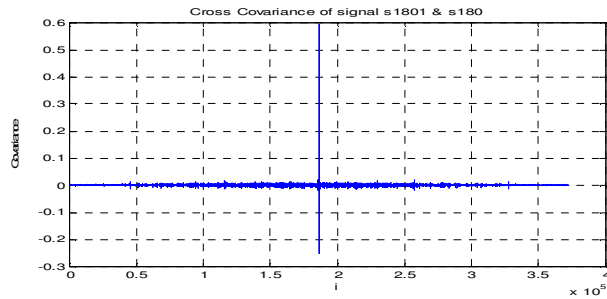
Fig 2e. Cross correlation between original and reconstructed signal by LDM
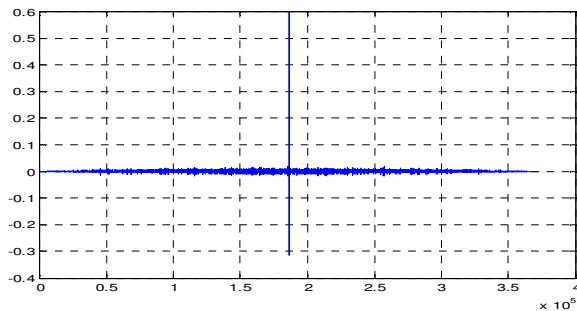
Fig 2f. Cross correlation between original and reconstructed signal by LMS

S3 is a noiseless speech sample and S3 and is the S3 speech captured by random noise (artificially generated). The original speech sample S3 and reconstructed speech samples with LDM and LMS are shown below. The cross covariance between the reconstructed and noiseless speech samples is also shown below.
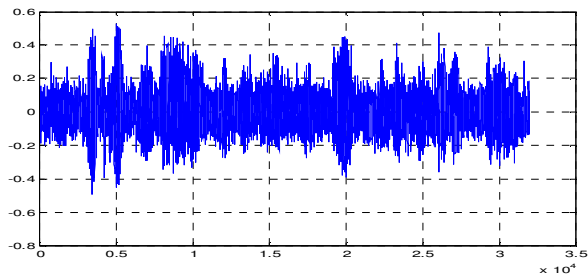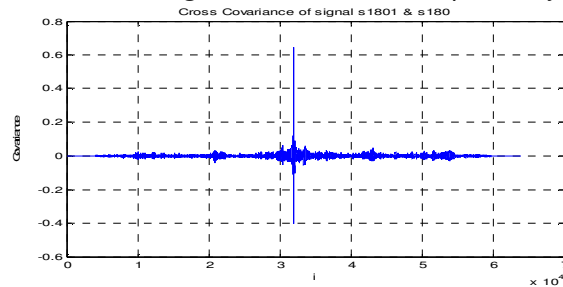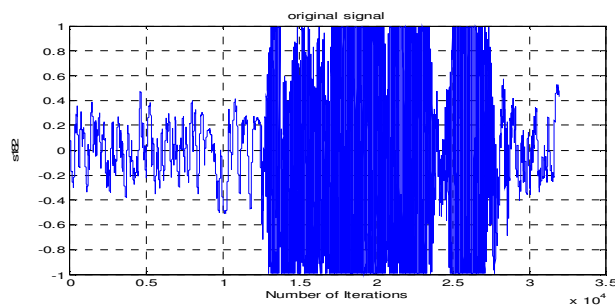

Fig 3a. S3 original speech sample
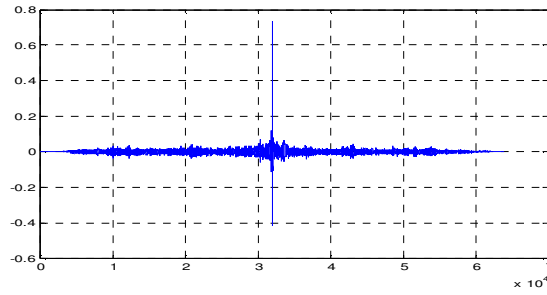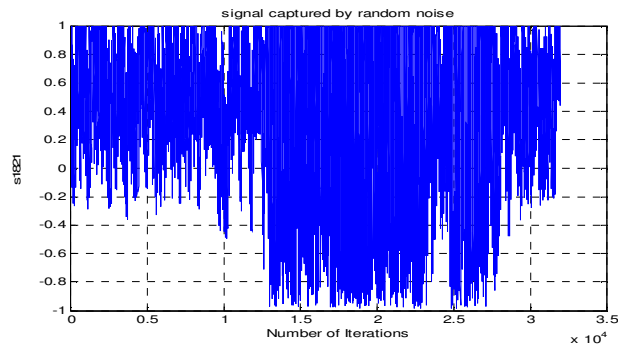
Fig 3b. S3street speech sample captured by random noise
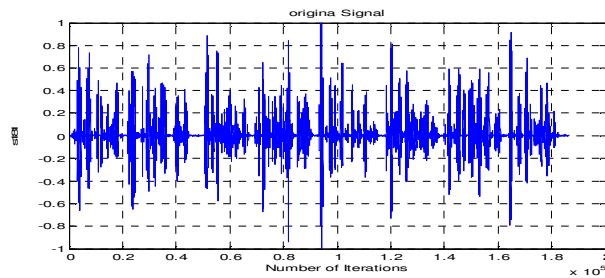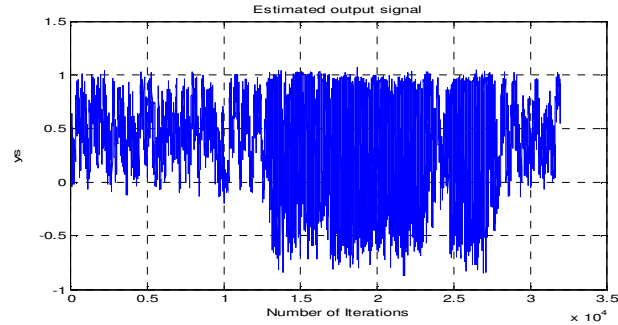

Fig 3c. Reconstructed speech by LDM
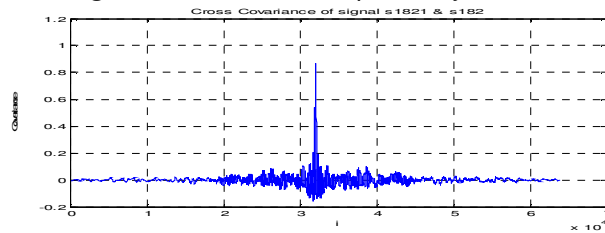
Fig 3 Cross correlation between original and reconstructed signal by LDM
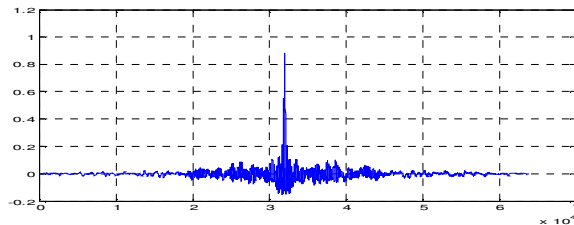
Fig 3e. Cross correlation between original and reconstructed signal by LMS

Speech	LDM	LMS
S1	0.65	0.736
S2	0.595	0.5996
S3	0.870	0.8836

TABLE 3: comparison of LDM and LMS in terms of cross correlation

6. CONCLUSION

After the working over the two models it is concluded that we can use the LDM for speech but LMS algorithm is also one of the methods which is simple and efficient algorithm. After comparing the results we found that for Speech sample S1 captured by White noise LMS algorithm is having better results than LDM. For Speech sample S2 captured by Street noise both algorithms are

having approximately same results. For Speech sample S3 captured by Random noise again LMS algorithm is having better results than LDM. Overall we found LMS Algorithm is giving promising results for the above speech samples considered for experimentation.

7. REFERENCES

Book Chapters

[1] M.S. Grewal and A.P. Andrews, Kalman Filtering Theory and Practice Using MATLAB 2nd edition, John Wiley & Sons, Canada, 2001, pp 1-12

Book

[2] An Introduction to the Kalman FilterGreg Welch and Gary Bishop UNC-Chapel Hill, TR 95-041, July 24, 2006

Dissertations and Theses

[3] J. Frankel,"Linear Dynamic Models for automatic speech recognition", Ph.D. dissertation, The center for Speech Technology Research, University of Edinburgh, UK, 2003.

Article in a Journal

[4] Joe Frankel and Simon King, speech Recognition using Linear Dynamic Models", Member, IEEE Member,IEE Manuscript received September 2004. This work is supported by EPSRC grant GR/S21281/01 Joe Frankel and Simon King are both with the Center for speech Technology Research, University of Edinburgh.

[5] Discriminative training for large vocabulary speech recognition using minimum classification errors" by Eric McDermott member IEEE,TimonthyJ Hazen,Member IEEE, Jonathan Le Roux,Atsushi Nakamura,Member,IEEE,and Shigeru Katagiri,Fellow,IEEE IEEE transaction on Audio speech and language processing vol 15 No1 2007

[6] Kalman, R. E. 1960. "A New Approach to Linear Filtering and Prediction Problems," Transaction of the ASME--Journal of Basic Engineering, March 1960.

[7] Kalman-Filtering Speech Enhancement Method Based on a Voiced-Unvoiced Speech Model, Zenton Goh, Kah-Chye Tan, Senior Member, IEEE, and B. T. G. Tan IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 7, NO. 5, SEPTEMBER 1999

[8] The stability of variable step-size LMS algorithms Gelfand,S.B.; Yongbin Wei; Krogmeier, J.V.; Sch. of Electr. & Comput. Eng., Purdue Univ., West Lafayette, IN Signal Processing, IEEE Transactions on Issue Date Dec 1999.

[9] Application of optimal settings of the LMS adaptive filter for speech signal processing Computer Science and Information Technology (IMCSIT), Proceedings of the 2010.

[10] C. R. Watkins, "Practical Kalman Filtering in Signal Coding", New Techniques in Signal Coding, ANU, Dec 1994.

Articles from Conference Proceedings (published)

[11] LMS Adaptive filtering for enhancing the quality of noisy speech Sambur, M. ITT Defense Communications Division, Nutley, N. J. Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '78.

[12] Application of the LMS adaptive filter to improve speech communication in the presence of noise Chabries, D. Christiansen, R. Brey, R. Robinette, M. Brigham Young University, Provo, UT, USA This paper appears in: Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82.

[13]   Speech enhancement using a Kalman-based normalized LMS algorithm Mahmoodzadeh, A. Abutalebi, H.R. Agahi, H. Electr. Eng. Dept., Yazd Univ., and Yazd This paper appears in: Telecommunications, 2008. IST 2008. International Symposium on Issue Date 27-28 Aug. 2008

[14]   Reduction of nonstationary acoustic noise in speech using LMS adaptive noise cancelling Pulsipher, D. Boll, S. Rushforth, C.Timothy, L. Sandia Laboratories, Livermore, CA This paper appears in: Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79.

[15]   Adaptive noise canceling for speech signals Sambur, M.
ITT Defense Communication Division, Nutley, NJ This paper appears in: Acoustics, Speech and Signal Processing, IEEE Transactions on Issue Date : Oct 1978

[16]   Paper in IEEE explore entitled "Comparison of LDM and HMM for an Application of a Speech" by Mane, V.A., Patil, A.B.,   Paradeshi, K.P., Dept. of E&TC, Annasaheb Dange COE, Ashta, India in International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom), 2010 Issue Date: 16-17 Oct. 2010 On page(s): 431-436 Location: Kottayam Print ISBN: 978-1-4244-8093-7 References Cited: 13 INSPEC Accession Number: 11679354 Digital Object Identifier: 10.1109/ARTCom.2010.65 Date of Current Version: 03 December 2010

A Low Power Digital Phase Locked Loop With ROM-Free Numerically Controlled Oscillator

M. Saber   *mohsaber@tsubaki.csce.kyushu-u.ac.jp*
Department of Informatics
Kyushu University
744 Motooka, Nishi-ku, Fukuoka-shi,89-0395,Japan

Y. Jitsumatsu   *jitsumatsu@inf.kyushu-u.ac.jp*
Department of Informatics
Kyushu University
744 Motooka, Nishi-ku, Fukuoka-shi,89-0395,Japan

M. T. A. Khan   *tahir@apu-u.ac.jp*
Ritsumeikan Asia Pacific University, College of Asia Pacific Studies
1-1 Jumonjibaru, Beppu, Oita, 874-8577, Japan

Abstract

This paper analyzes and designs a second order digital phase-locked loop (DPLL), and presents low power architecture for DPLL. The proposed architecture reduces the high power consumption of conventional DPLL, which results from using a read only memory (ROM) in implementation of the numerically controlled oscillator (NCO). The proposed DPLL utilizes a new design for NCO, in which no ROM is used. DPLL is designed and implemented using FPGA, consumes 237 mw, which means more than 25% saving in power consumption, and works at faster clock frequency compared to traditional architecture.

Keywords: Digital Phase locked loop (DPLL), Field Programmable Gate Array (FPGA), Software Defined Radio (SFDR), Read Only Memory (ROM), Spurious Free Dynamic Range (SFDR).

1. INTRODUCTION

Software Defined Radios (SDRs) are leading the integration of digital signal processing (DSP) and radio frequency (RF) capabilities. This integration allows software to control communications parameters such as the frequency range, filtering, modulation type, data rates, and frequency hopping schemes. SDR technology can be seen in wireless devices used for different applications in military, civil applications, and commercial network. Compared to conventional RF transceiver technologies, the advantage of SDR is its flexibility. SDR provides the ability to reconfigure system performance and functions on the fly [1].

In order to take advantage of such digital processing, analog signals must be converted to and from the digital domain. This is done using analog-to-digital (ADC) and digital-to-analog (DAC) converters. To take full advantage of digital processing, SDRs keep the signal in digital domain as much as possible, digitizing and reconstructing as close as possible to the antenna. Despite an ADC or DAC connected directly to an antenna is a required end goal, there are issues with selectivity and sensitivity that need an analog front [2].

Phase-locked loop (PLL) is one of the most important building blocks necessary for modern digital communications, which is used as a frequency synthesizer in RF circuits, or to recover time and carrier in the baseband digital signal processing. A complete understanding of the concept of PLL includes many study areas such as RF circuits, digital signal processing, discrete time control systems, and communication theory [3]. Traditional PLL consists of three parts; phase frequency detector (PFD), loop filter, and voltage controlled oscillator (VCO).

The traditional analog PLL faces many design problems such as voltage supply noise, temperature noise, and large area consumed by loop filter components like resistors and capacitors. On the other hand DPLL, formed of all digital components, provides a high immunity to supply voltage noise and temperature variation. Moreover, DPLL can be designed by using hardware description language (HDL) with any standard cell library. Thus, the time for redesign and check for errors is reduced. Therefore, DPLL provides a good solution to analog PLL design problems. Unfortunately, DPLL has a critical disadvantage, i.e., high power consumption resulting from the numerically-controlled oscillator (NCO) [4].

The high power consumption of NCO is the result of using ROM, which contains the sampled amplitudes of a sinusoidal waveform. As accuracy of the generated signal increases, the size of ROM increases, which causes high power consumption and reduces the speed of the circuit. We propose a DPLL architecture in which the traditional NCO is replaced by a circuit which generates a cosine waveform using a piecewise-linear approximation.

In section 2, PLL operation is explained. The traditional NCO is described in section 3. Section 4 illustrates a modified NCO which can solve the problems of traditional NCO. In section 5 mathematical model of DPLL in both Z-domain and S-domain is illustrated. In section 6 simulation results. In section 7 hardware implementation of modified NCO and modified DPLL is presented and in the end some conclusions are given.

2. PHASE LOCKED LOOP

PLL is an important component in many types of communication systems. It works in two different manners; to synchronize a carrier in frequency and phase or to operate as a synthesizer. The block diagram of DPLL is shown in Fig. 1. It consists of three main blocks, phase/frequency detector (PD), loop filter and NCO.
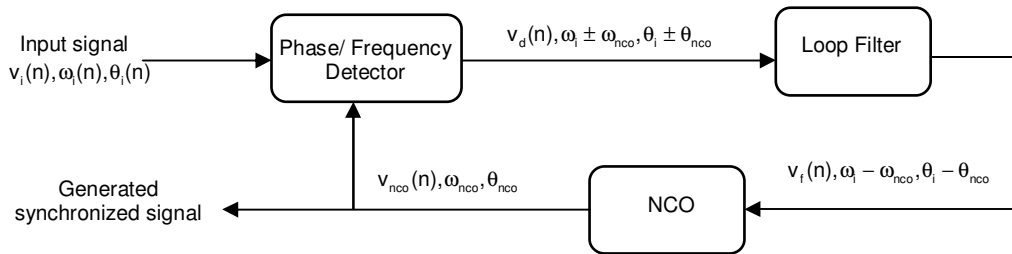
FIGURE 1: Digital phase locked loop in discrete time domain.

The operation of DPLL is as follows: without input signal applied to the system, NCO generates a signal with a center frequency (f_c), which is called the free running frequency. The input signal applied to the system is

$$v_i(n) = A_i \sin(\omega_i\, n + \theta_i), \tag{1}$$

where A_i is the amplitude, ω_i is the angular frequency, and θ_i is the phase of the input signal. Feedback loop mechanism of PLL will force NCO to generate a sinusoidal signal $v_{nco}(n)$

$$v_{nco}(n) = A_o \sin(\omega_{nco}\, n + \theta_{nco}), \tag{2}$$

where A_o is the amplitude, ω_{nco} is the angular frequency and θ_{nco} is the phase of the signal generated by NCO. θ_{nco} is given by

$$\theta_{nco}(n) = k_v \sum_{i=-\infty}^{N} v_f(i), \tag{3}$$

where k_v is the NCO gain constant and $v_f(n)$ is the filter output. If k_m denotes the phase detector (multiplier) gain, then output of the phase detector is

$$v_d(n) = \frac{k_m A_i A_o}{2} \sin(\omega_i\, n + \theta_i) \cos(\omega_{nco}\, n + \theta_{nco})$$

$$= \frac{k_m A_i A_o}{2} [\sin((\omega_i + \omega_{nco})\, n + \theta_i + \theta_{nco}) + \sin((\omega_i - \omega_{nco})\, n + \theta_i - \theta_{nco})], \tag{4}$$

The first term in (4) corresponds to high frequency component, and the second term corresponds to the phase difference between $v_i(n)$ and $v_{nco}(n)$. Loop filter will remove the first term in (4). If $\omega_i = \omega_{nco}$, then phase difference can be obtained as

$$v_f(n) = k_d[\sin(\theta_i - \theta_{nco})], \tag{5}$$

where $k_d = \dfrac{k_m A_i A_o}{2}$. If $(\theta_i - \theta_{nco}) \quad 1$, then $V_f(n)$ is approximated by

$$v_f(n) \approx \frac{k_d A_i A_o}{2}(\theta_i - \theta_{nco}). \tag{6}$$

This difference voltage is applied to the NCO. Thus, the control voltage $v_f(n)$ forces the NCO output frequency to change up or down to reduce the frequency difference between ω_{nco} and ω_i. The equation of the generated frequency of NCO is

$$\omega_{nco}(n) = \omega_c + v_f(n), \tag{7}$$

where ω_c is the center frequency of NCO. If the input frequency ω_i is close to ω_{nco}, the feedback manner of PLL causes NCO to synchronize or lock with the incoming signal. Once it is locked, the generated signal of NCO will synchronize the input signal in phase and frequency.

3. TRADITIONAL NCO

Voltage Controlled Oscillator (VCO), which is used in analog PLL generates a sinusoidal waveform whose frequency depends on the input voltage. NCO, which is used in DPLL, generates a digital (sampled) sinusoidal waveform with a fundamental frequency determined by the digital input value (n-bits). As shown in Fig. 2, NCO consists of ROM, and accumulator. The output signal of the accumulator is used as address to the ROM. The input signal to the accumulator consists of the sum of an offset (ω_c) corresponding to the free running frequency, and v_f which is the output of the loop filter [5]. The general equation of generated frequency from NCO is

$$f_{nco} = (\frac{v_f}{2^j} + \omega_c) \times f_{clk}. \tag{8}$$

where f_{nco} is the generated frequency, ω_c is the center frequency, v_f is an integer value and lies in the range $(-2^{j-1} \leq v_f \leq 2^{j-1})$, j is number of bits or width of the accumulator, which is 16 bits, and f_{clk} is the clock frequency.

The operation of NCO is as follows: first assuming that the system clock frequency is 50MHz, j=16 and $\omega_c = 1310$, the free running frequency is 1 MHz. Then, as shown in Fig. 3 there are 50 sampling points in one cycle of 1 MHz sinusoidal waveform. NCO generates exactly one cycle of sinusoidal waveform when the input value (v_f) is equal to zero. Since the offset value is 1310, every clock cycle the accumulator accumulates the offset value. Then in 50 cycles the accumulated value will increase by one. The accumulator output will address this value to the ROM and extract the cosine amplitudes values stored in it.

When the input value is greater than zero, the accumulation speed becomes higher. Thus in less than 50 cycles of clock frequency the accumulator increases by 1, this will generate a higher frequency than 1MHz. When the input value is less than 0, a frequency lower than 1 MHz is generated. The problem with using a ROM is that, its size increases to achieve a high spectral purity of the generated waveform. This leads to high power consumption and slow operation of the system.
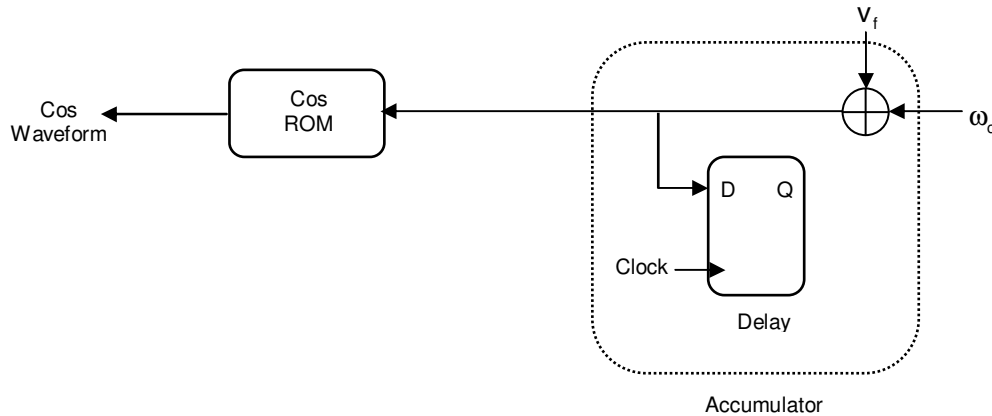
FIGURE 2: Numerically controlled oscillator structure.
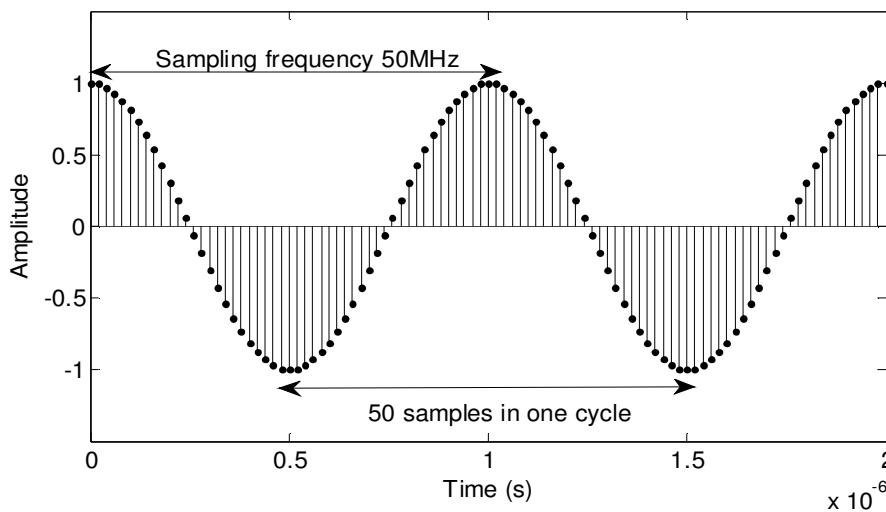
FIGURE 3: Output waveform of NCO.

3.1 Previous Work

NCO which generates sine or cosine output as shown in Fig. 2 differs mostly in the implementation of ROM block. This block is the slowest and consumes high power. The problem of ROM is that, its size grows exponentially with the width of the phase accumulator. Since one normally desires a large number of bits to achieve fine frequency tuning and high spectral purity, several techniques have been invented to limit the ROM size while maintaining suitable performance.

One technique uses the quarter wave symmetry of sine function to reduce the number of saved samples by 4, in which ROM saves only the amplitudes of first quarter and through additional hardware the other quarters are generated [6]. Truncating accumulator output (remove number of most significant bits (MSBs)) is a common method to reduce the size of ROM but this method introduces spurious harmonics [7].

Different angular decomposition techniques proposed to reduce the ROM size consist of splitting the ROM into a number of smaller ROMs, each ROM is addressed by a portion of truncated accumulator output. Generated samples of each ROM are added to form a complete sinusoidal waveform. In order to introduce more reduction in the ROM size, many techniques have been proposed to make an initial approximation of the sine amplitude from the value of the phase angle, and to use the ROM or a combination of ROMs to store correction values [8:11]. Although these methods reduce the power consumption but they still use ROM which causes a residual of high power consumption.

Many other techniques have been proposed using piecewise continuous polynomials to approximate the first quadrant of the sine function. One of them is based on a Taylor-series expansion [12], a simplified 4th degree polynomial [13] and 4th degree Chebyshev polynomials [14]. The drawbacks of the above techniques are that they require additional hardware to make extra computations which increase the complexity of the circuit. The additional hardware consumes power consumption which supposed to reduce.

4. MODIFIED NCO

4.1 Proposed Architecture

In proposed architecture no ROM is used, to provide fast switching, and less power consumption. Instead of using a ROM a piecewise linear approximation is used, that is representing the first quarter of the cosine waveform as linear lines, each line fits a linear equation with slope and bias. Depending on the symmetry of the cosine waveform (have 4 quarters), it can easily deduce the other 3 quarters of the cosine waveform from only the first quarter. The first quarter of the cosine function is divided into eight piecewise linear segments of equal length of the form:

$$\cos(t) \approx a_i t + b_i, \quad \frac{i}{16}\pi \le t < \frac{i+1}{16}\pi, \quad i=0,1,.....7, \tag{9}$$

where a_i is the segment slope and is limited to 4 bits, and b_i is the constant or bias limited to 8 bits. Slopes and biases are chosen using the minimum mean square error (MMSE) criterion, that minimizes the integrated mean square error between the ideal $\cos(t)$ and the approximated cosine function $p(t)$.

$$mmse = \int_{t=0}^{\pi/2} [\cos(t) - p(t)]^2 dt. \tag{10}$$

Fig. 4 shows a comparison between ideal and approximated cosine waveforms. It seems to be the same except the top and bottom of the waveform, that is because of the linear segments.

M. Saber, Y. Jitsumatsu & M. T. A. Khan
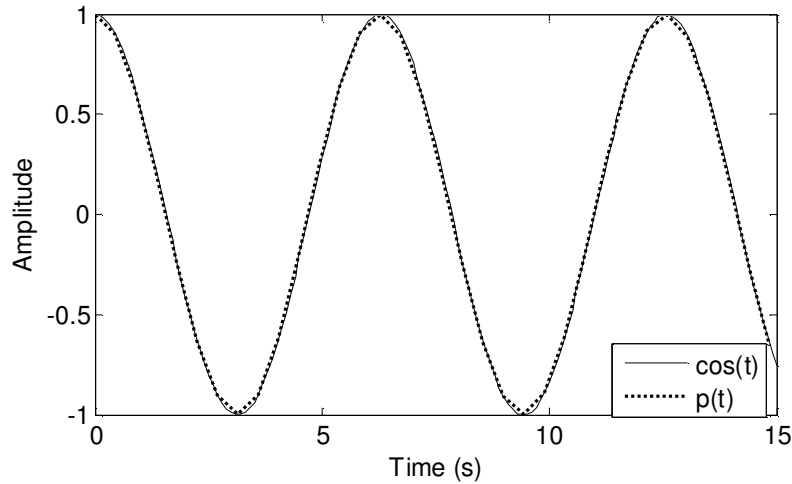
FIGURE 4: Approximated and Ideal cosine waveforms.

The modified NCO consists of two main components and two negation units. Fig. 5 shows the block diagram of each component and the corresponding waveform. Accumulator receives the input signal $v_f(n)$ which represents the phase difference between θ_i and θ_{nco}. The accumulator works as a circular counter. A complete rotation of the accumulator represents one cycle of the output waveform. The accumulator receives a signal with eight bits-length, and the width of the accumulator is j=16 bits, so truncation is done to the output signal of the accumulator to be X signal with L=10 bits' length. The first two most significant (MSBs) bits of the accumulator are used to control the operation of NCO. 2nd MSB controls the sign of signal X before performing the piecewise linear calculation. This negative sign is needed to substitute in the linear function to generate all quarters of the cosine waveform. Second negation is done at the output stage to correct position of second and third quarters. This negation is controlled using XOR function between 1st, and 2nd MSB.
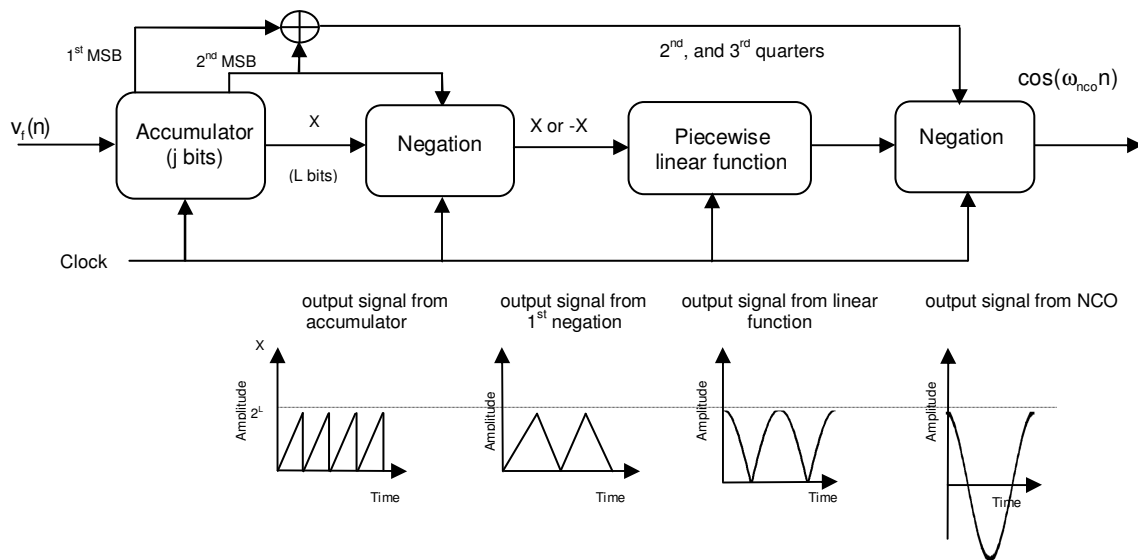
FIGURE 5: Structure of modified NCO.

4.2 Spurious Free Dynamic Range (SFDR)

SFDR is defined as the ratio between the RMS value of the fundamental frequency (maximum signal component) and the RMS value of the next largest noise or harmonic distortion component, (which is referred to as a "spurious" or a "spur") at its output. SFDR is usually measured in dBc (i.e. with respect to the carrier frequency amplitude) or in dBFS (i.e. with respect to the ADC's full-scale range). Depending on the test condition, SFDR is observed within a pre-defined frequency window or from DC up to Nyquist's frequency of the converter (ADC or DAC). Fig. 6 shows how SFDR is measured [15]. Since the modified NCO depends on linear approximation to generate digital samples of cosine waveform, the spectrum of the generated waveform contains spurs at all the spectrum frequencies, and SFDR is used to measure the spectral purity of the generated frequencies.
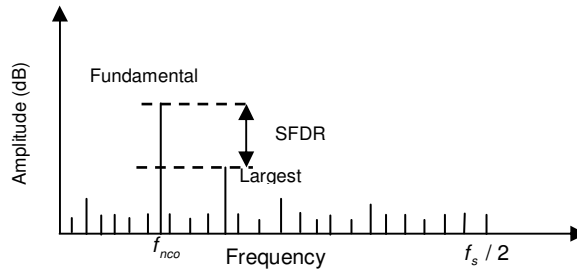
FIGURE 6: SFDR measure.

5. DPLL MATHEMATICAL MODEL

A mathematical model for DPLL is built in z-domain, and s-domain to study the ability of the system to maintain phase tracking when exited by phase steps, frequency steps, or other excitation signals. Fig. 7 and Fig. 8 shows mathematical model of the system in both Z-domain and S-domain respectively.
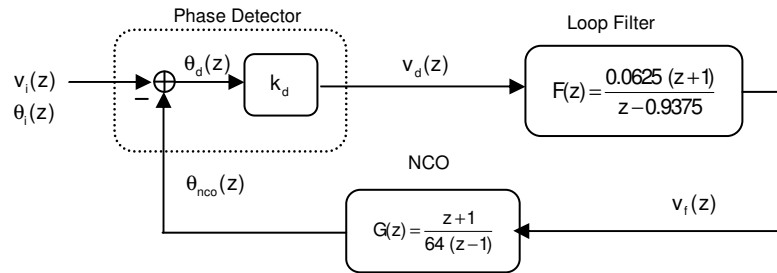
FIGURE 7: DPLL in Z-domain.
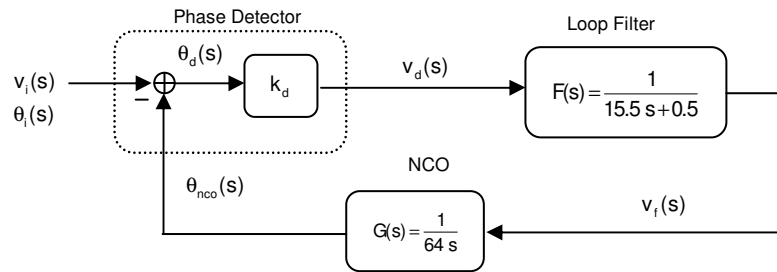
FIGURE 8: DPLL in S-domain.

The phase transfer function of the system in Z-domain is

$$\frac{\theta_{nco}(z)}{\theta_i(z)} = \frac{k_d \ F(z) \ G(z)}{1 + K_d \ F(z) \ G(z)} = \frac{1 + z^{-1} + z^{-2}}{1025 - 1982 \ z^{-1} + 961 \ z^{-2}} \ . \tag{11}$$

To get the step response of the system a relation between $v_f(z)$ and $v_i(z)$ is needed. Assuming the input signal is a unit step of frequency at constant phase

$$\frac{v_f(z)}{v_i(z)} = \frac{F(z)}{1 + F(z) \ G(z)} = \frac{64 \ (1 - z^{-2})}{1025 - 1982 \ z^{-1} + 961 \ z^{-2}} \ . \tag{12}$$

Using bilinear transformation, the previous equations are obtained in S-domain

$$\frac{\theta_{nco}(s)}{\theta_i(s)} = \frac{k_d \ F(s) \ G(s)}{1 + k_d \ F(s) \ G(s)} = \frac{1}{992 \ s^2 + 32 \ s + 1} \tag{13}$$

$$\frac{v_f(s)}{v_i(s)} = \frac{F(s)}{1 + G(s) \ F(s)} = \frac{64 \ s}{992 \ s^2 + 32 \ s + 1} \tag{14}$$

In the test for stability, DPLL is subjected to a test signal representing a unit step of frequency at constant phase using (14) with $f_s = 50$ MHz [16-17]. As shown in Fig. 9, the system is stable with overshoots at the transient state.
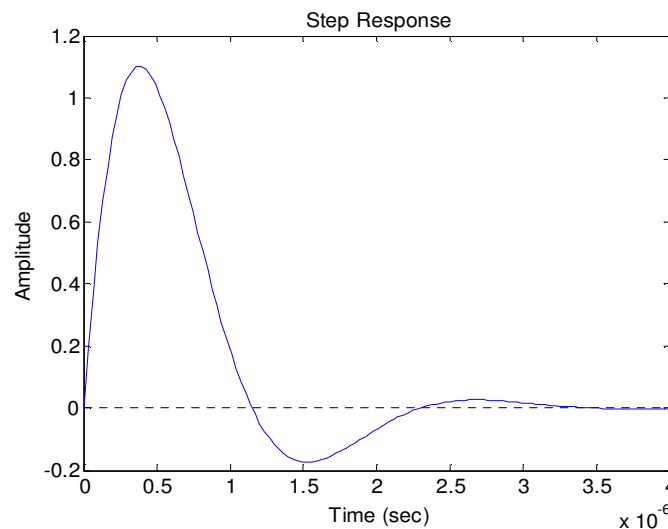
FIGURE 9: DPLL in S-domain.

6. SIMULATION RESULTS

6.1 SFDR of Modified NCO
To measure the SFDR a discrete Fourier transform (DFT) is done for a long repetition period of the generated signal from modified NCO. Difference between the amplitude of the fundamental output frequency and the amplitude of the largest spurs in the dynamic range is noted. Fig. 10 shows the output spectrum for input word of value 1317 representing $v_f(n)$, at a clock frequency

of 50 MHz and an accumulator width j=16. The fundamental frequency is approximately 2 MHz with -30.057 dB, and the spurious appears at 14.46 MHz with -89.925 dB, so SFDR=59.868 dBc.
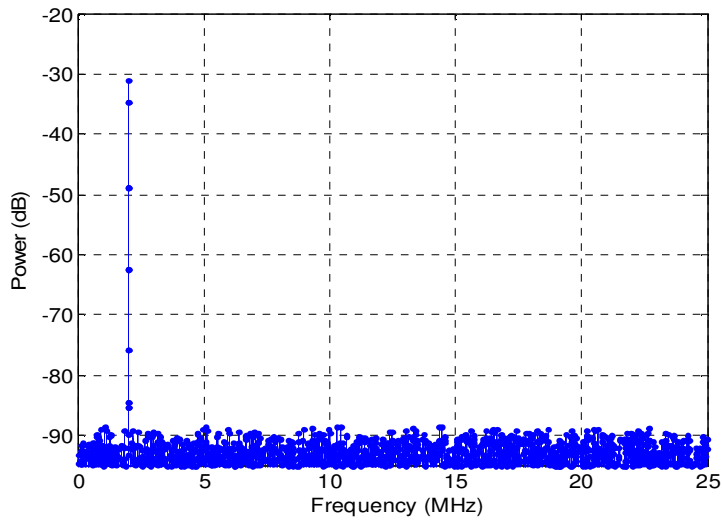
FIGURE 10: SFDR for fundamental frequency of 2 MHz.

6.2 DPLL Synchronization

In this section, we investigate the performances of proposed DPLL's using computer simulations. The proposed DPLL has the following parameters:

$f_s = 10$ MHz, $k_m = 1$, $k_v = 1024$, $A_i = A_o = 1$, $\omega_{nco} = 1$ MHz, and $\theta_{nco} = 0$.

Two types of simulations are done. In the first one, DPLL receives a signal with phase difference ($\omega_i = 1$ MHz, $\theta_i = \pi / 4$), DPLL response is shown in Fig. 11. In the second case input signal has both phase difference and frequency difference ($\omega_i = 1.01$ MHz, $\theta_i = \pi / 4$), DPLL response is shown in Fig.12.
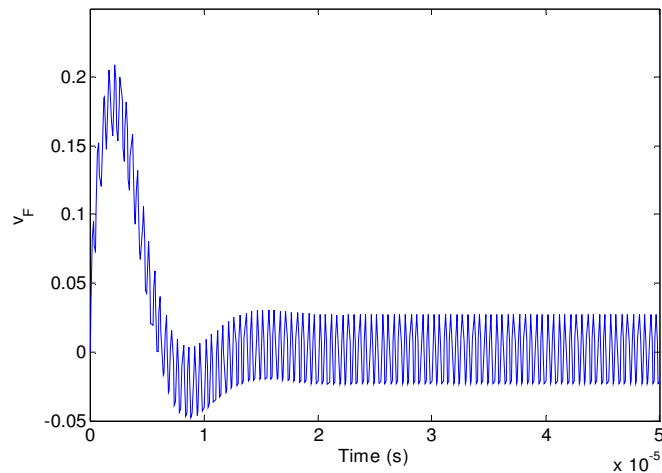
FIGURE 11: DPLL response in case of phase difference.
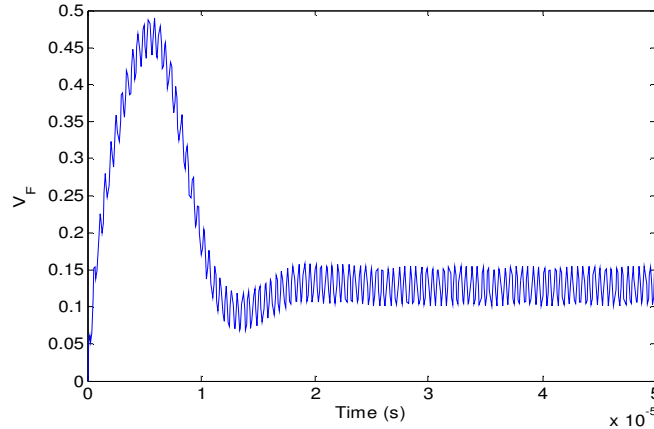
FIGURE 12: DPLL response in case of phase and frequency difference.

6.3 Proposed DPLL vs. Traditional DPLL

The main objective of this simulation is to compare the performance of the proposed DPLL with traditional DPLL; to be sure that replacing ROM with linear approximation did not affect the operation of DPLL. In this simulation both architectures have the same parameters.

$f_s = 10$ MHz, $k_m = 1$, $k_v = 1024$, $A_i = A_o = 1$, $\omega_{nco} = 1$ MHz, and $\theta_{nco} = 0$.

An input signal with $\omega_i = 1.02$ MHz and $\theta_i = \pi/2$. is applied to both architectures. Both responses are shown in Fig.13 which indicates that the performance of the proposed DPLL is not affected by the modified NCO. i.e. the ability of locking phase or frequency of the input signal is not affected. This means the proposed DPLL saves power consumption compared to traditional DPLL without affecting the performance of DPLL.
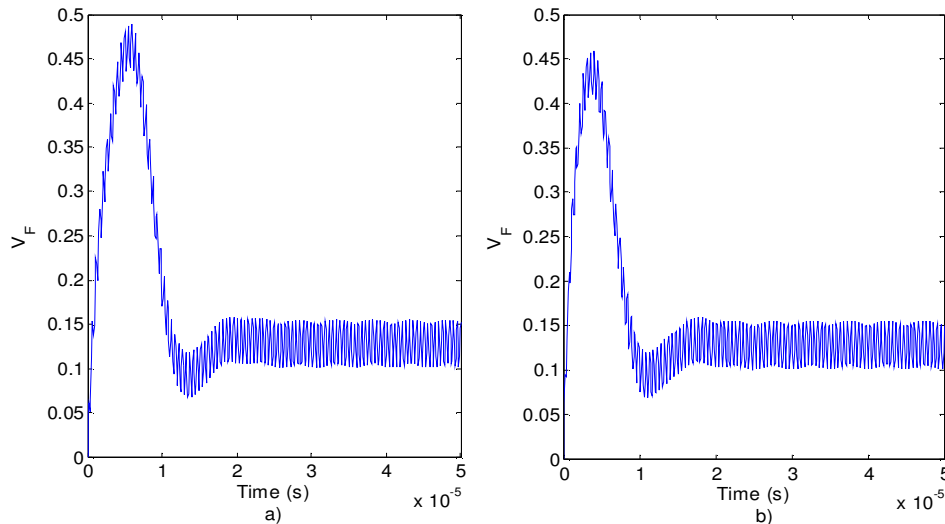
FIGURE 13: DPLL response in case of phase and frequency difference
a) Response of traditional DPLL.   b) Response of proposed DPLL.

7. HARDWARE IMPLEMENTATION

Hardware implementation of modified NCO, and modified DPLL is done using VHDL code using Xilinx system generator Simulink tool [18:20]. The architecture of modified NCO is shown in Fig. 14.
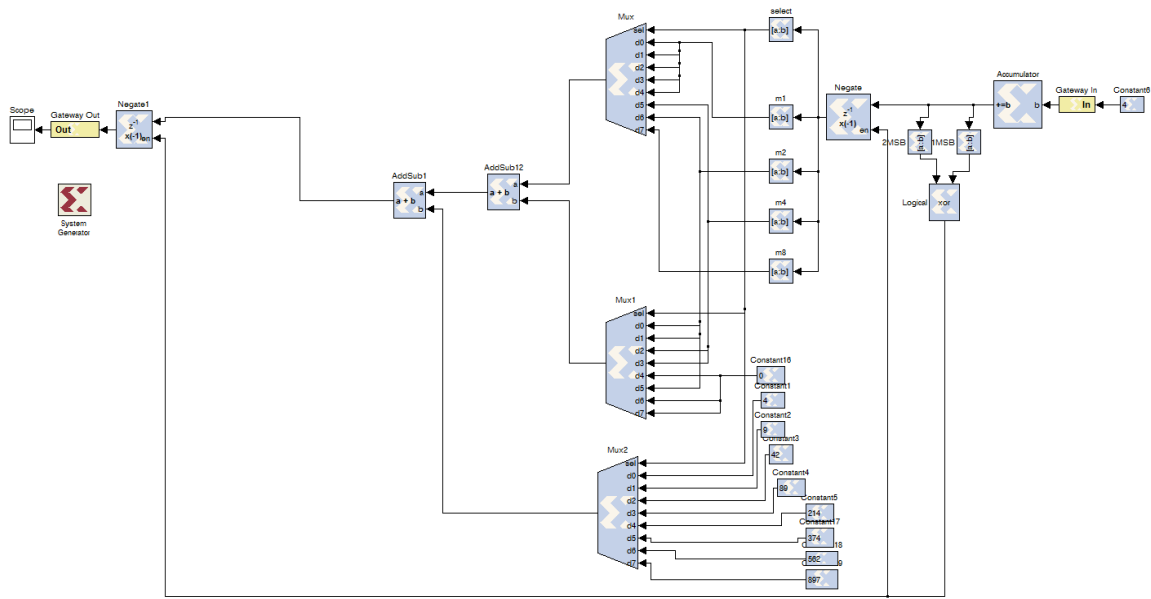
FIGURE 14: Modified NCO model

Implementation of linear segments requires slopes and constants. The slopes are chosen using MMSE as mentioned before and the slopes' accuracy is limited to a fraction four bits. m1 represents the full truncated output from the accumulator, m2 is half m1, m4 is half m2 (quarter m1) and m8 is half m4 (eighth m1). The first three MSBs generated from the accumulator are used to control three multiplexers. The first two multiplexers are forming the slope value, and the third multiplexer form the constant value. According to the selected signal, the linear equations are chosen through the multiplexers to form the complete linear equation.

The architecture of modified DPLL is shown in Fig. 15; the architecture uses the modified NCO instead of traditional NCO. The simulation is done at clock frequency 50 MHz. All signals are binary signals with different widths. The input signal is a binary signal of 8 bits width representing a sinusoidal signal at frequency 1 MHz. Fig.16 shows the simulation waveforms as an analog signal, the input signal (input1) of frequency 1 MHz is multiplied by the modified NCO signal (input2), and the output signal is passed through the digital filter. The final output shows that digital implementation agrees with the simulation waveform.
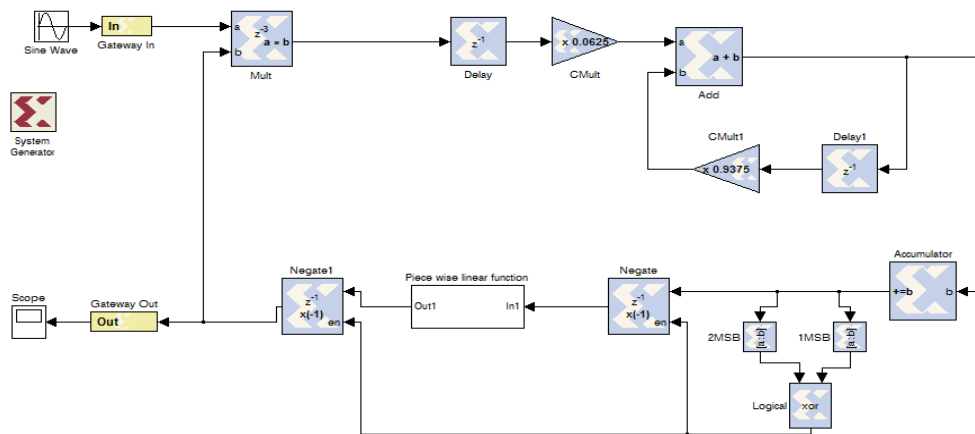
FIGURE 15: Modified DPLL model.

M. Saber, Y. Jitsumatsu & M. T. A. Khan
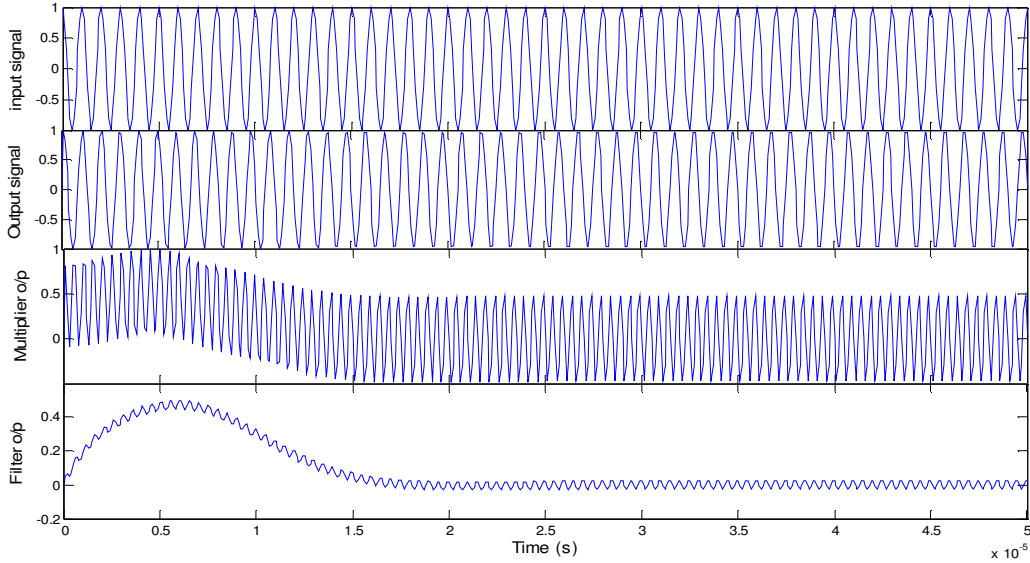
FIGURE 16: VHDL simulations of DPLL.

To recognize how much the modified NCO reduces the power consumption, logic elements and operation with faster frequency. A comparison between traditional NCO, which uses ROM block and modified NCO, is done by implementing both architectures on the same FPGA device (Xilinx-Spartan-3A DSP Xc3d3400a-5fg676). This comparison gives an idea of how much could be the improvements in power consumption, reduction in the occupied number of logic elements and faster frequency. As illustrated in Table1, the modified NCO reduces about 40% of total logic elements used in traditional NCO, and did not use memory bits, which leads to save the power consumption by about 25% and operation at a faster frequency about 1.8 times the speed of traditional NCO. Comparison is also done with the traditional DPLL (which uses a traditional NCO) and modified DPLL (which uses modified NCO). Table 2 shows the result of comparison; it is clear that the modified DPLL consumed less power, occupied less area and worked faster than the traditional DPLL, with no degradation in system operation such as locking range.

	Traditional NCO	Modified NCO
Slices	108	64
Flip Flops	21	17
Block RAMs	60	0
Look up table (LUT)	210	116
IOBs	24	24
Maximum Frequency	151.461 MHz	284.738 MHZ
Power consumption	0.264 Watt	0.197 Watt

TABLE 1: Implementation results comparison of NCO.

	Traditional DPLL	Modified DPLL
Slices	162	64
Flip Flops	37	17
Block RAMs	60	0
Look up table (LUT)	299	116
IOBs	24	24
Maximum Frequency	101.241 MHz	205.279 MHZ
Power consumption	0.314 Watt	0.237 Watt

TABLE 2: Implementation results comparison of DPLL.

8. CONCLUSION

Second order DPLL architecture has been described, analyzed and implemented to be suitable for any application. The problem of high power consumption of DPLL has been solved by replacing the traditional NCO (the main component in DPLL) with a modified ROM. The traditional NCO uses ROM, which results in high power consumption as well as slower operation. The proposed architecture reduces power consumption, area consumption and works at a higher frequency than the traditional one.

9. ACKNOWLEDGEMENT

10. REFERENCES

[1] M. Dillinger. K. Madani, N. Alonistioti. Software Defined radio: Architectures, systems, and functions. John Willey & Sons Inc., 2003.

[2] T. J. Rouphael. RF and Digital Signal Processing For Software-Defined Radio: A Multi standard Multi-Mode Approch. John Willy & Sons Inc., 2008

[3] R. E. Best. Phase-Locked Loops: Design, Simulation, and Application. 6th ed, McGraw-Hill, 2007.

[4] S. Goldman. Phase Locked-Loop Engineering Hand Book of Integrated Circuit. Artech House Publishers, 2007

[5] B. Goldberg. Digital Frequency Synthesis Demystified: DDS and Fractional-N PLLs. Newnes ,1999.

[6] V.F. Kroupa, Ed. Direct Digital Frequency Synthesizers. IEEE Press,1999.

[7] V.F. Kroupa, V. Cizek, J. Stursa, H. Svandova. "Spurious signals in direct digital frequency synthesizers due to the phase truncation." IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 47, no. 5, pp. 1166-1172, September 2000.

[8] H.T. Nicholas III, H. Samueli and B. Kim. "The optimization of direct digital frequency synthesizer performance in the presence of finite word length effects," in Proc. of the 42nd Annual Frequency Control Symposium, 1988, pp. 357-363.

[9] A. Yamagishi, M. Ishikawa, T. Tsukahara, and S. Date. "A 2-V, 2-GHz low-power direct digital frequency synthesizer chipset for wireless communication." IEEE Journal of Solid-State Circuits, vol. 33, pp.210-217, February 1998.

[10] A. M. Sodagar, G. R. Lahiji, "Mapping from phase to sine-amplitude in direct digital frequency synthesizers using parabolic approximation." in IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing, vol. 47, pp. 1452-1457, December 2000.

[11] J.M.P. Langlois, D. Al-Khalili. "ROM size reduction with low processing cost for direct digital frequency synthesis," in Proc. of the IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, August 2001, pp. 287-290.

[12] L.A. Weaver, R.J. Kerr. "High resolution phase to sine amplitude conversion." U.S. patent 4 905 177, Feb. 27,1990.

[13] A.M. Sodagar, G.R. Lahiji. "A novel architecture for ROM-less sine-output direct digital frequency synthesizers by using the 2nd-order parabolic approximation," in Proc. of the 2000 IEEE/IEA International Frequency Control Symposium and Exhibition, 7-9 June 2000, pp. 284-289.

[14] K.I. Palomaki, J. Niitylahti. "Direct digital frequency synthesizer architecture based on Chebyshev approximation," in Proc. of the 34th Asilomar Conference on Signals, Systems and Computers, Oct. 29th – Nov. 1st., 2000, pp. 1639-1643.

[15] J. Rudy. CMOS Integrated Analog-to-Digital and Digital-to-Analog Converters. Springer, 2003.

[16] J. G. Proakis, G. Dimitri, Manolakis. Digital Signal Processing. Prentice Hall,1996.

[17] Naresh K. Sinha. Linear Systems. John Wiley & Sons Inc.,1991.

[18] Xilinx Inc. system generator for DSP user guide. Xilinx, 2009.

[19] W.Y. Yang. Matlab/Simulink for Digital Communication. A-Jin, 2009.

[20] P. Chu. FPGA Prototyping by VHDL Examples: Xilinx Spartan-3 Version. Wiley-Interscience, 2008.

A Non Parametric Estimation Based Underwater Target Classifier

Binesh T   *bineshtbt@gmail.com*
Department of Electronics,
Cochin University of Science and Technology,
Cochin 682 022, India

Supriya M. H.   *supriya@cusat.ac.in*
Department of Electronics,
Cochin University of Science and Technology,
Cochin 682 022, India

P.R. Saseendran Pillai   *prspillai@cusat.ac.in*
Department of Electronics,
Cochin University of Science and Technology,
Cochin 682 022, India

Abstract

Underwater noise sources constitute a prominent class of input signal in most underwater signal processing systems. The problem of identification of noise sources in the ocean is of great importance because of its numerous practical applications. In this paper, a methodology is presented for the detection and identification of underwater targets and noise sources based on non parametric indicators. The proposed system utilizes Cepstral coefficient analysis and the Kruskal-Wallis H statistic along with other statistical indicators like F-test statistic for the effective detection and classification of noise sources in the ocean. Simulation results for typical underwater noise data and the set of identified underwater targets are also presented in this paper.

Keywords: Cepstral Coefficients, Linear Prediction Coefficients, Forward Backward Algorithm, Kruskal-Wallis H Statistic, F-test Statistic, Median, Sum of Ranks.

1. INTRODUCTION

Underwater acoustic propagation depends on a variety of factors associated with the channel in addition to the characteristic properties of the generating source. Studies on noise data waveforms generated by man made underwater targets and marine species are significant as they will unveil the general characteristics of the noise generating mechanisms. The composite ambient noise containing the noise waveforms from the targets, received by the hydrophone array systems are processed for extracting the target specific features. Though quite a large number of techniques have been evolved for the extraction of source specific features for the task of identification and classification, none of them are capable of providing the complete set of functional clues. Of these, many of the techniques are complex and some of them lead to ambiguities in the decision making process. Since classification of noise sources using certain traditional techniques yields low accuracy rates, many improved approaches based on non-parametric and parametric modeling have been mentioned in open literature [1]. Some of the modern approaches for the extraction of spectral profiles give more emphasis to spectral resolutions and increased signal detection capabilities while others rely on the extraction and utilization of acceptable features of underwater signal sources .The proper identification and classification of underwater man-made and biological noise sources can utilize the cepstral feature extraction and non parametric statistical approaches which do not rely on any

assumption that the data are drawn from a given probability distribution and includes non parametric statistical models, inference and statistical tests. The Kruskal-Wallis H test is a non parametric test and the H statistic can be efficiently employed in different statistical situations[2]. The underwater noise signal is sampled, processed and cepstral features are extracted and hence the sample set of transition probability values of the system model is estimated. The H statistic, F-test statistic, Median value and Sum of Ranks are estimated for the sample sets of various underwater signals, the transition probability values and a reference signal, which were found to be occupying non overlapping value ranges and can be utilized in the system design for the identification of underwater signal sources.

2. PRINCIPLES

Cepstral coefficients are widely used as features for a variety of recognition and classification applications. In a cepstral transformation, the convolution of two signals $x_1[n]$ and $x_2[n]$ becomes equivalent to X_c, which is the sum of the cepstra of the two signals.

$$X_c = \hat{x}_1[n] + \hat{x}_2[n] \tag{1}$$

Defined otherwise, P discrete cepstrum coefficients[3], c_p where $p = 0,.....P-1$ define an amplitude envelope $|H(\omega)|$ equals $\exp(c_0 + 2\sum_p c_p \cos(p\omega))$ with p varying from 1 to P-1.
The Inverse Fourier Transform of the log amplitude gives the cepstral coefficients. The discrete cepstrum coefficients can be described by a set , at frequencies ω_k with amplitudes X_k with k= 1,....P. This can be expressed mathematically:

$$X(\omega) = \sum_{k=1}^{P} (X_k \partial(\omega - \omega_k)) \tag{2}$$

where $\partial(\omega)$ denotes the Dirac delta distribution. The calculation of c_p can be done by minimizing the square difference of $|H(\omega)|$ and $|X(\omega)|$.

Non parametrical analysis provides effective methods for target detection and classification of underwater targets. Such a strategy may also be incorporated into a hierarchical classification framework, where a target is first assigned to a class and later with additional information, it may be identified as a particular target within that class. In order to train a statistical model for each class, many methods can be used, which may consist of several training states. The system can be trained on the target data associated with their respective classes. Statistical non parametric tests can be considered as an alternative for comparisons of data of which the distribution is not Gaussian[4]. The exact distribution of H-statistic in the Kruskal-Wallis test is conventionally fitted to a Chi-squared approximation. In state based models, the sequence of tokens generated by it may give some information about the sequence of states. Even though the states possess different attributes, for many practical applications there will be often some physical significance associated to the set of states and their transition probabilities. The proposed procedure can utilize a codebook to estimate the required parameters. In a codebook, a large number of observational vectors of the training data is clustered into a certain number of observational vector clusters using K- means iterative procedure. Based on this clustered observational vectors, estimates of the parameters are generated during system modulation.

2.1 LPC Analysis
Linear Prediction Coefficients(LPC) Analysis is used to calculate the Cepstral coefficients. LPC is a powerful modeling technique used for signal analysis. LPC encodes a signal by finding a set of weights on earlier signal values that can predict the next signal value. Linear prediction coefficients can be transformed to cepstral coefficients which is a more robust set of parameters. In matrix form,

Ra = **r** (3)

Where **r** is the autocorrelation vector, **a** is the LPC vector and **R** is the Toeplitz matrix of **r**. The solution is:

a = R⁻¹r (4)

2.2 Cepstral Coefficients and Clustering

The p Cepstral coefficients c_m, for m=0,1…p-1 derived from the set of LPC coefficients using the LPC to Cepstral coefficient recursion[5].

K-means is one of the learning algorithms that solve the clustering problem . It is an algorithm to cluster *n* objects based on attributes into *K* partitions, where *K < n*. It attempts to find the centers of natural clusters in the data. It assumes that the object attributes form a vector space. The main idea is to define *K* centroids, one for each cluster. The result it tries to achieve is to minimize the total intra-cluster variance, or, the squared error function [6]

$$V = \sum_{i=1}^{K}(x_j - \mu_i)^2$$ (5)

where there are *K* clusters S_i, *i = 1, 2, ….K*, and μ_i is the centroid or mean point of all the points x_j which will form the elements of S_i and considered in the above computation.

2.3 Forward-Backward Algorithm

The Forward-Backward Algorithm is an algorithm for computing the probability of a particular observation sequence. Let the forward probability $\alpha_j(t)$ for some model M with N states be defined as $\alpha_j(t)=P(o_1,…..,o_t),x(t)=j|M$). That is, $\alpha_j(t)$ is the joint probability of observing the first t vectors and being in state j at time t.

This recursion is based on the fact that the probability of being in state *j* at time *t* and having observation o_t can be found by adding the forward probabilities for all possible previous states *i* weighted by the transition probability a_{ij} . Also,

$$\alpha_N(T) = \sum_{i=2}^{N-1}(\alpha_i(T)a_{iN})$$ (6)

and *P(O|M)* equals $\alpha_N(T)$.

The backward probability $\beta_j(t)$ is defined as:
$$\beta_j(t) = P(o_{t+1},….,o_T \mid x(t) = j, M)$$ (7)

The forward probability is a joint probability and the backward probability is a conditional probability. Also, $\alpha_j(t)$ $\beta_j(t)= P(O,x(t)=j|M)$. Hence the probability of state occupation becomes $S_j(t)= P(x(t)=j|O,M)$ which in turn equals P(O, x(t)=j|M) ÷P(O|M). Let P(O|M) be denoted by P_o. Then

$$S_j(t) = \frac{1}{P_o}\alpha_j(t)\beta_j(t)$$ (8)

2.4 H-Statistic

Statistical indicators measure the significance of the difference between the performance of different systems and can be used to grade the systems if the performance difference is significant. Kruskal-wallis H-test is a non parametric test[7] of hypothesis whose test statistic can be effectively utilized in underwater signal classification. The H-statistic is given by:

$$H = \frac{12}{N(N+1)}\sum_{j=1}^{G}\frac{R_j^2}{N_j} - 3(N+1)$$ (9)

where G is the total number of samples, N_j, $j= 1,...G$, is the size of sample j, R_j, $j = 1,...G$, is the rank of the sample j. Let (R_j^2/N_j) of the different sample sets be termed as C which forms an intermediate parameter in H estimation and

$$N = \sum_{j=1}^{G} N_j \qquad (10)$$

2.5 F-Statistic

A F-test is a statistical test which is usually applied when comparing statistical models and is used to assess if the expected values of a quantitative variable within several pre-defined groups have difference among each other. The test statistic in an F-test is the ratio of two scaled sums of squares following Chi-squared distribution, indicating different sources of variability. The F-test statistic is given as the ratio of 'Between-Group variability'(BG) to 'Within-Group variability'(WG). The two terms can be defined mathematically as follows:

$$BG = \sum_{i} n_i (y_{iav} - Y_{av}) / (N_g - 1) \qquad (11)$$

where y_{iav} denotes the sample mean in the ith sample group, n_i is the number of observations in the ith group and Y_{av} denotes the overall mean of the data. Also

$$WG = \sum_{ij} (Y_{ij} - y_{iav})^2 / (N_0 - N_g) \qquad (12)$$

where Y_{ij} is the jth observation in the ith out of N_g groups and N_0 is the overall sample size.

2.6 Median (M) and Sum of Ranks (R)

The statistical estimate Median (M) is an important characteristic of signals from any underwater source. It is a measure of the skewness of the sampled signal distribution and also an indicator of the amplitude variations in the sample set of the particular signal. The Median of the signal can be estimated as that amplitude value in the sample set from which there occurs equal numbers of positive and negative amplitude deviations. The M parameter, along with H and F values helps in the classification of a particular signal. The other statistical estimate used along side H, F and M parameters in the proposed system is the Sum of Ranks (R). It gives a measure of the relative gradation of signal amplitude variations of the signal, taking into consideration, the sample location indices in the sample set of the underwater signal. The R parameter can be estimated for a sample set of by reordering the samples in the increasing order of amplitudes and replacing the original samples with their respective ranks, in the distribution. A minimum rank of unity can be assigned to a sample. For equal valued samples, average of the corresponding rank can be assigned. The sum of all the individual sample ranks will give the parameter R, which forms an important property, when utilized along with other parameters of the system. For the underwater signals with closely related H and F parameters, the R parameter can be helpful for identification in association with the M parameter.

3. METHODOLOGY

The methodology consists of various stages and the different steps involved in the extraction of feature vectors are furnished below.

3.1 Cepstral Coefficient Extraction
3.1.1 Sampling and Frame Conversion

The noise data waveforms emanating from the underwater target of interest have been sampled and recorded as a wave file data, which is sampled to be converted to frames of N_s samples, with adjacent frames being separated by m_d samples[5]. Denoting the sampled signal by s[n], the lth frame of data by x_l[n], and there are L frames, then

$$x_l[n] = s[m_d l + n] \qquad (13)$$

Where $n = 0, 1, ..., N_s -1$, and $l = 0, 1, ...L-1$.

3.1.2 Windowing

Each individual frame is windowed to minimize the signal discontinuities at the boundaries of each frame. If the window is defined as *w[n]*, then the windowed signal x_w is

$$x_w = x_l[n]w[n] \tag{14}$$

where $0 < n < N_s\text{-}1$.

Hamming window is used as a typical window for the autocorrelation method of LPC.

A frame based analysis of the noise data waveform has been performed to generate the sample vector, which can be used to estimate the statistics needed for target classification. The sampled signal is partitioned into frames of N_s samples, and consecutive frames are spaced m_d samples apart. Each frame is multiplied by a N_s-sample Hamming window, and LP analysis is performed[8]. The Linear Prediction Coefficients are then converted to the required number of Cepstral coefficients, which are weighted by a raised sine window.

3.2   Vector Quantization

The next step in the system is a clustering process which can be used to generate a code book which in turn is utilized in the estimation of transition probability vector. The K-means algorithm has been used to fix the centroids of a cluster model. The extracted cepstral coefficients of the underwater signal source are being utilized as the data in this vector quantization process of unique cluster identification. A matrix is defined, which represents the data which is being clustered, in a concatenation of K clusters, with each row corresponding to a vector. The cluster centroids are generated as a vector with the cluster identity. The sum of square error function is used in the algorithm, and a log of the error values after each iteration can be returned in a variable. The maximum number of iterations can also be specified.

3.3   Transition Probability Vector Generation

A Vector of transition probabilities can be generated from the vector quantized output, for the estimation of the Decision Statistics. The algorithm for the generation of the transition probability vector is as follows:

START:
    Segregate the data into Frames.
    Windowing the Frames using Hamming Window.
    Generation of Linear Prediction Coefficients.
    LPC to Cepstral Coefficient conversion.
    Vector Quantization and code book generation.
    Set N_{it} = maximum iterations

 LABEL 1:
   While (count <= N_{it})
  {
    Compute the forward probability $\alpha_j(t)$ for all states j at times t.
    Compute the backward probability $\beta_j(t)$.
    If (P(O|M)<= value of previous iteration)
     {
      go to LABEL 2
     }
    Estimate Transition Probability $S_j(t)$.
    count = count + 1.
  }

LABEL 2:
  Generate a single column vector by concatenating individual columns of the estimated transition probability matrix.
END

3.4 Decision Statistics Estimation

The H and F statistics are estimated as illustrated in Fig 1 with the three sample set consisting of the previously generated transition probability vector, a down sampled version of the original underwater signal and a predefined reference sample vector. A correction for ties can be made by dividing the H-statistic value by a Correction Factor(CF) defined as follows:

$$CF = 1 - \frac{1}{(N^3 - N)} \sum_{i=1}^{g} (t_i^3 - t_i)$$

(15)

where g is the number of groupings of different tied ranks, and t_i is the number of tied values within group i that are tied at a particular value. This correction usually makes only negligibly small change in the value of test statistic unless there are large numbers of ties. Additional statistical parameters like Median and Sum of Ranks can also be estimated along with, for the underwater signal being processed.
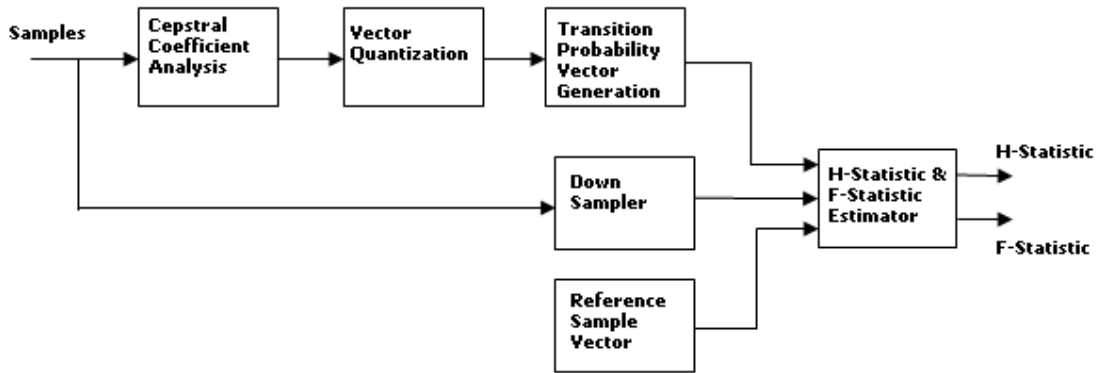
FIGURE 1: Estimation of Decision Statistics

4. IMPLEMENTATION

The sampled underwater noise source is divided into frames of 400 samples (N_s). Consecutive frames are spaced 19 samples apart. Each frame is multiplied by an N_s-sample Hamming window. Because of lower side lobe levels, Hamming window is a good choice for comparatively accurate signal processing systems. Each windowed set of samples is auto correlated to give a set of coefficients. Then linear prediction coefficient analysis is done on the autocorrelation vector to estimate the LP coefficients and using recursion method, linear prediction coefficients are converted to cepstral coefficients. They are then weighted by a raised sine window function. By applying K-means algorithm, K centroids are defined, one for each cluster. Random selection of K vectors is done. K=16 is selected in the algorithm. The next step is to take each vector and associate it to the nearest centroid. At this   point, readjusting the centroids is done based on the new assignment. The algorithm minimizes the squared error   function mentioned earlier. Thus, vector quantization is carried out and unique clusters are defined for the particular underwater noise waveform.

4.1 Sample Sets Under Consideration

Using Forward-Backward re-estimation algorithm, the transition probabilities for the twenty states of the system model are estimated leading to the generation of the transition probability vector

which is considered as the first sample set. A vector of down sampled values of the underwater noise source with a down sampling factor of 0.5 forms the second sample set while a reference sample set of 1000 samples with sample values of 0.5 for the first 500 samples and 0.25 for the next 500 samples as depicted in Fig 2, forms the third sample set.

The Kruskal-Wallis H-statistic is estimated with the correction factor to obtain the Chi-squared statistic approximation. The F-statistic approximation is also estimated for the system. The Median(M) of the underwater signal and Sum of Ranks(R), taking into consideration, the three vectors, of the same underwater signal are also evaluated. The estimated values for the four parameters of different underwater noise sources possess divergent statistical properties which can be utilized in the effective identification and classification of the unknown underwater signal source under consideration.

5. RESULTS AND DISCUSSIONS

The system has been validated using simulation studies and the estimated H-statistic as well as F-statistic approximations, median values(M) and sum of ranks(R) of different underwater signal sources have been tabulated in Table 1.

Underwater signal source	Estimated H-Statistic Approximation value	Estimated F-Statistic Approximation value	Estimated Median value(M)	Estimated Sum of Ranks value(R)
Shors	2090	3465	-0.0025	833927
Toadfish	1798	2322	0.001975	1002781
Beluga	2044	3242	-0.00158	908441
Bagre	2420	5706	0.03316	1445904
Outboard	1951	2791	0.00355	971748
Damsel	2115	3616	0.0012571	827679
Sculpin	1172	933	0.21805	1414338
Atlantic croaker	1987	3023	-0.0004	862176
Spiny	2450	6076	-0.005633	631137
BlueGrunt	2097	3570	0.0003167	860600
Dolphin	2146	3455	-0.00108	863228
01m	1172	940	0.0772	1313128
Barjack	2021	3050	0.00228	892434
Bow1	2168	3939	-0.0049167	782094
Boat	1494	1451	0.0024	1136117
Chord	2160	3783	0.000625	778549
3Blade	1837	2372	-0.004733	988073
Torpedo	2563	9757	-0.007817	540386
Rockhind	2075	3394	0.0013125	864103
Snap1	2117	3632	-0.000483	823856
Scad	1990	2893	0.0006667	869278
Finwhale	2134	3875	-0.000453	793392
Seal1	2051	3187	0.0241	1040226
Garib	1896	2635	-0.049514	969721
Grunt	1955	3259	0.00235	888618
Ocean Wave	2054	3558	-0.006425	844440
Minke	2130	3476	0.0001	823722
Hump	2156	3838	-0.010267	786830
Seatrout	2051	3251	0.01018	934365
Silverperch	2064	3193	0.0031	855612
Cavitate	1877	2559	-0.007275	1004192
Sklaxon	2141	3744	-0.00995	807558
Submarine	1644	1843	-0.040775	1012841
Badgear	2060	3453	-0.000217	852301
Seacat	1731	2580	-0.003825	985634
Searobin	1844	2394	-0.002425	962476

TABLE 1: Underwater signal sources and their estimated values of H-statistic, F-statistic, Median and Sum of Ranks.

The Reference Sample Set of the type depicted in Fig 2, having a statistical variance of 0.0156 has been considered in the proposed technique. Also, the Coefficient of Variation (CV) which is defined as the ratio of the Standard Deviation to modulus of Mean, for this reference sample set is seen to be 0.124.
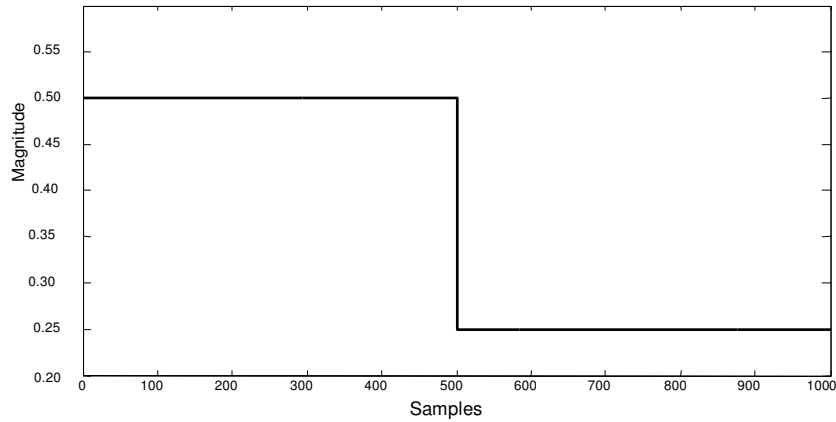
FIGURE 2: Plot of Reference Sample Set values used in the system.

The (H, F, M, R) components form the recognition parameter for a given underwater signal source. The plots of the loglikelihood in transition probability estimation for the underwater noises of Toad Fish and Submarine are depicted in Fig 3 (a) and (b). The unknown underwater signal is processed and the extracted H,F,M,R components are assigned to known underwater signal categories by judiciously matching the component parameters. The signals listed out in Table 1 have been tested with the system, utilizing the (H,F,M,R) components and correct recognition has been obtained except for the Searobin and 3Blade underwater signals. The system possesses a tolerance specification of ±1% for the parameters used in this technique.

(a)   (b)
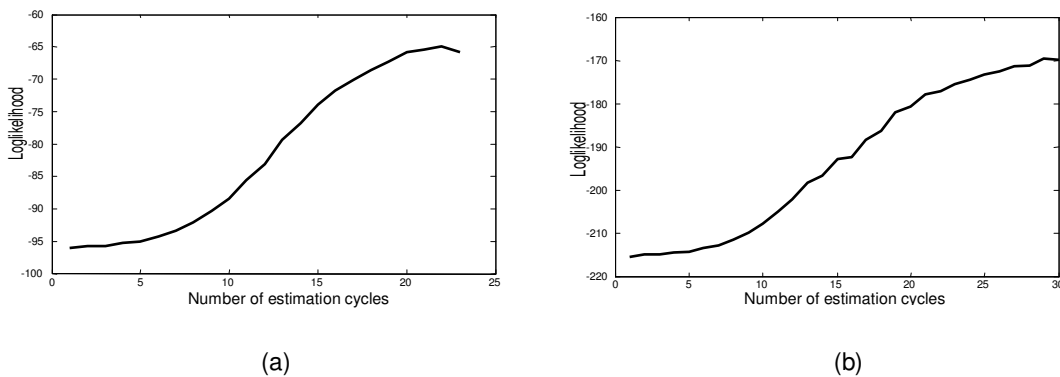
FIGURE 3: Plots of loglikelihood in Transition probability estimation for (a) Toad Fish (b) Submarine.

The proposed system is optimized for the classification of underwater noise sources in the ocean. Non-parametric estimators and the featured statistical indicators possess increased robustness essential for the efficient classification capability of a system. State Transition Probability estimation has been utilized in the design of Hidden Markov Model based speech recognition systems [1][9]. In this underwater target classifying system, the transition probabilities form a significant sample set in the estimation of recognition parameters of a particular signal. The simulated results, using the four components, show high recognition capability of the system for underwater signals. The increased computational complexity of the system is offset by the

improved classification efficiency, while upholding the inherent advantages of non-parametric classifiers.

6. CONCLUSIONS

The proposed system makes use of statistical indicators along with non-parametric estimations like the cepstral coefficients for the identification and classification of underwater targets utilizing the target emanations. Using simulation studies, the H-statistic as well as F-statistic approximations along with the Median and Sum of Ranks parameters for different underwater signal sources have been estimated and are utilized for the identification of the unknown noise sources in the ocean. The system can also be augmented with other features and can be effectively used for the identification and classification of noise sources in the ocean, with improved success rates.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1]   Lawrence R. Rabiner, *"A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition"*, Proceedings of the IEEE, 77(2):257-273, 1989

[2]   W. H. Kruskal and W. A. Wallis, "*Use of ranks in one-criterion variance analysis*," Journal of American Statistics Association, 47 : 583-621, Dec.1952

[3]   Schwarz and X. Rodet *"Spectral Envelope estimation and representation for sound analysis- synthesis"*, In Proceedings of International Computer Music Conference, ICMC 99, Beijing, 1999

[4]   Dirk K. de vries and Yves Chandon, *"On the false positive rate of Statistical equipment comparisons based on the Kruskal-Wallis H-statistice"*, IEEE Transactions on Semi conductor manufacturing, 20(3), 2007

[5]   Lawrence Rabiner and Biing-Hwang Juang , "*Fundamentals of Speech Recognition"*, NJ: PTR Prentice Hall, pp. 112-117 (1993)

[6]   Donghu Li, Azimi Sadjadi, M. R and Robinson, M, *"Comparison of different Classification Algorithms for underwater target discrimination"*, IEEE Transactions on Neural Networks, 15(1), 2004

[7]   M. Hollander & D.A. Wolfe, "*Non parametric Statistical methods"*, New York, Wiley, (1973).

[8]   J. R Deller, J. G. Proakis and F. H. L Hansen, *"Discrete time processing of speech signals"*, IEEE Press, p. 71, (2000)

[9]   L. R. Rabiner and B. H. Juang, "*An Introduction to Hidden Markov Models"*, IEEE ASSP Magazine, 3(1): pp. 4-16, 1986

Mourad Talbi, Akram Aouinet, Lotfi Salhi & Adnane Cherif

New Method of R-Wave Detection by Continuous Wavelet Transform

Mourad Talbi   *mouradtalbi196@yahoo.fr*
Faculty of Sciences of Tunis/ Laboratory of
Signal Processing/ PHISICS DEPARTEMENT
University of Tunisia-Manar
TUNIS, 1060, TUNISIA

Akram Aouinet   *akramkeliba@yahoo.fr*
Faculty of Sciences of Tunis/ Laboratory of
Signal Processing/ PHISICS DEPARTEMENT
University of Tunisia-Manar
TUNIS, 1060, TUNISIA

Lotfi Salhi   *lotfi.salhi@laposte.net*
Faculty of Sciences of Tunis/ Laboratory of
Signal Processing/ PHISICS DEPARTEMENT
University of Tunisia-Manar
TUNIS, 1060, TUNISIA

Adnane Cherif   *adnane.cher@fst.rnu.tn*
Faculty of Sciences of Tunis/ Laboratory of
Signal Processing/ PHISICS DEPARTEMENT
University of Tunisia-Manar
TUNIS, 1060, TUNISIA

Abstract

In this paper we have employed a new method of R-waves detection in electrocardiogram (ECG) signals. This method is based on the application of the discretised Continuous Wavelet Transform (CWT) used for the Bionic Wavelet Transform (BWT). The mother wavelet associated to this transform is the Morlet wavelet. For evaluating the proposed method, we have compared it to others methods that are based on Wavelet Transform (WT). In this evaluation, the used ECG signals are taken from MIT-BIH database. The obtained results show that the proposed method outperforms some conventional techniques used in our evaluation.

Keywords: Continuous Wavelet Transform, Electrocardiogram, Hard Thresholding, R-wave Detection.

1. INTRODUCTION

The electric currents in the heart have been measured and recorded for more than a hundred years, but the term electrocardiogram (ECG) was introduced by Willem Einthoven in 1893 at a meeting of the Dutch Medical Society. The electrocardiogram is considered to be the backbone of cardiology, and can be recorded fairly easily with surface electrodes on the surface of the limbs or chest. The ECG records the electrical activity, this typical tracing consists of a series of repetitive waves namely P, Q, R, S and T. The P wave represents left and right atrial depolarization, ventricular contractions (both right and left) show as a series of 3 waves, Q-R-S know as the QRS complex, the last common wave in an ECG is the T wave, this reflects the period of ventricular repolarization. A cardiologist can look at a patient's electrocardiogram and determine the presence of disturbances in the intervals, amplitudes and areas of these waves. QRS complex is the most prominent feature in electrocardiogram because of its specific shape; therefore it is taken as a reference in ECG feature extraction. R wave detectors are very useful tools in

Mourad Talbi, Akram Aouinet, Lotfi Salhi & Adnane Cherif

analyzing ECG features thus form the basis of ECG feature extraction [1]. The development of accurate and quick methods for automatic ECG feature extraction is of major importance, especially for the analysis of long recordings (Holters and ambulatory systems). In fact, beat detection is necessary to determine the heart rate, and several related arrhythmias such as Tachycardia, Bradycardia and Heart Rate Variation [2]. All methods used by scientists are to help cardiologists to gain time to interpret results and improve the diagnostic.

In this paper, we proposed a technique using discretized continuous wavelet transform (CWT), 'Morlet' mother wavelet has been selected for detection of R-wave. The method described is robust, does not require any pre-processing stage, simple to implement and the selection of detail signal C4 has been justified. Finally, the ECG signals used in the experiments are obtained from MIT-BIH database [3].

2. MATERIAL

2.1 Continuous Wavelet Transform (CWT)
Morlet first introduced the idea of wavelets as a family of function constructed from translations and wavelets of a single function called the 'mother wavelet'. The wavelet analysis has been introduced as a windowing technique with variable-sized regions. Wavelet decomposition introduces the notion of scale as an alternative to frequency and maps a signal into time-scale plan. The wavelet analysis is the decomposition of a signal into sine waves of different frequencies [4]. Mathematically, the continuous wavelet transform of a function $x(t)$ is defined as the integral transform of $x(t)$ with a family of wavelet functions, $\Psi_{a,b}(t)$:

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \cdot \Psi\left(\frac{t-b}{a}\right) dt, a \in IR_+^*, b \in IR \tag{1}$$

The function $\Psi(t)$ is commonly called the mother wavelet and the family of function $\Psi_{a,b}(t)$ is called daughter wavelets. The daughter wavelets are derived from scaling and shifting the mother wavelet. The scale factor a represents the scaling of the function $\psi(t)$, and the shift factor b represents the temporal translation of the function. It is important to know that determination of CWT scale parameter and mother wavelets are very significant in ECG feature extraction [4].

In this work, we have used the discretized CWT employing the Morlet wavelet. This discretized CWT is used for the Bionic Wavelet Transform (BWT) introduced by Yao et al [5].

2.2 Wavelet Selection
The selection of the analyzing function in wavelet transforms, which is called the mother wavelet, has a significant effect on the result of analysis and should be selected carefully based on the nature of the signal [6]. But there is no universal method suggested to select a practical wavelet. They are several wavelet families like Biorthogonal, Coiflets, Daubechies, Morlet, Symlets etc. In this study, 'Morlet' mother wavelet has been selected for feature extraction. The analysis shows that extracted features from ECG signal by using the Morlet mother wavelet would be simple to compute, easy to understand, and the results are very good. Figure 1 shows the real and imaginary parts of the complex Morlet mother wavelet.

2.3 Data Base
The data available from MIT-BIH Arrhythmia Database [3] is the standard used by many researchers. The MIT-BIH database contains many data sets of electrocardiogram signals, mostly abnormal or unhealthy electrocardiograms, but it also contains normal electrocardiograms that can be used as a reference base [7]. This contains two lead ECG signals of 48 patients. The selected Arrhythmias are Premature Atrial Beat (PAB), Premature Ventricular Beat (PVB), Right Bundle Branch Block (RBBB), and Left Bundle Branch Block (LBBB).

Mourad Talbi, Akram Aouinet, Lotfi Salhi & Adnane Cherif
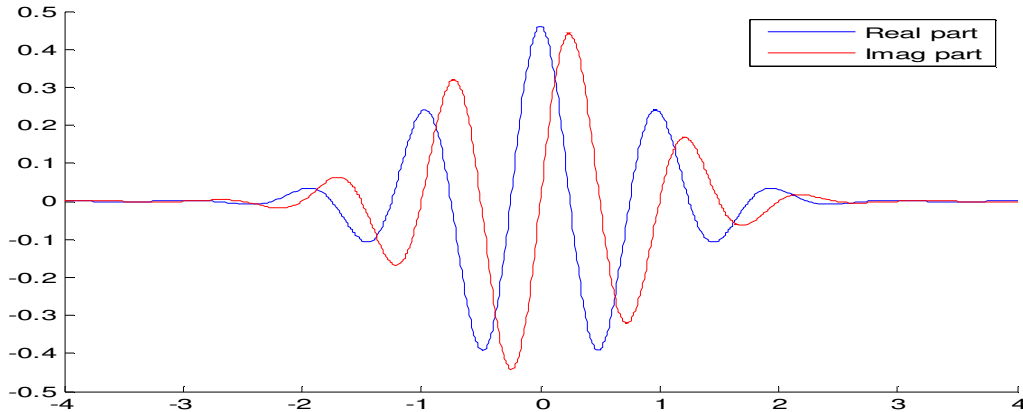
FIGURE 1: The Morlet Wavelet (its real part and imaginary part).

3. MATERIAL

The ECG signals taken from MIT-BIH arrhythmia database are converted in to Matlab format (.mat files). The ECG signal is sampled at 360 Hz with a resolution of 11 bits. In this section, we have developed and evaluated a robust method R-Wave detection based on Continuous Wavelet Transform. This technique is summarized by the following steps:

Step1: we decompose the ECG signal into 8 scales by using the modified discretized continuous wavelet transform MMycwt which is used by BWT.
Step2: we chose the best wavelet coefficient to perform the detection of R-wave: this selection is based on the research work of Awadhesh Pachauri et al [1].
Step3: we apply hard thresholding to that coefficient by using the appropriate threshold.
Step4: we detect R-waves by using the step3: the positions of the R-waves are those having amplitudes that are greater than the value of the selected threshold.

The Figure 2 gives in details the different steps of the proposed technique and they will be detailed in the next paragraph.
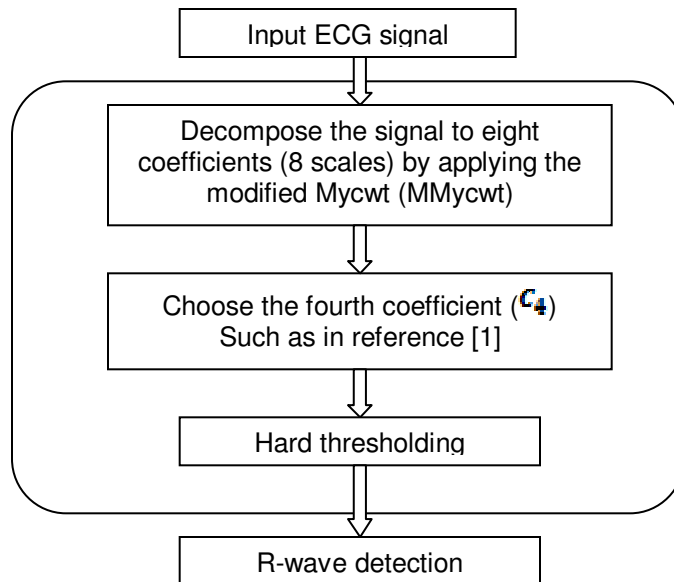
FIGURE 2: The different steps of the proposed technique of the R-waves detection.

3.1 MMycwt

For an ECG signal, the most important feature is the frequency range in which its main components occur [8]. Despite the existence of some other components like VLPs, we are interested in this paper in P, Q, R, S and T waves such as in the reference [8]. In references [9, 10], the value of (the initial center frequency of the mother wavelet) is equal to 15165.4Hz. As the scale increases, the center frequency goes smaller and smaller in the following way:

$$f_m = f_0/q^m, q > 1, m = 1, 2, \ldots \qquad (2)$$

We don't need such high frequency for ECG signals. Omid et al [8] have optimized the value of by running the program for different values of and then minimizing the gradient of error variance by comparing the results-numerically and morphologically with each other. It has been found that if belongs to the range of 360 to 500Hz there would be no much distortion on the analyzed ECG signals [8]. In their work, Omid et al [8] have chosen 400Hz as the value of f_0. Hence, in our work, we have chosen $f_0 = 250 Hz$ in order to obtain the MMycwt. In this paper, we have chosen the value 1.1623 as that of q such as in the reference [9, 10].

Every ECG signals under test are decomposed up to 8 levels. The maximum number of decomposition level depends upon total number of samples present in the signal.

$$n = 2^N \qquad (3)$$

where N is the total number of levels of decomposition and n is the total number of samples in the ECG signal.

3.3. Selection of Detail Coefficient (C_4)

According to the reference [1], it was shown by simulation that the wavelet coefficient in level four, owns the highest coefficient of cross correlation with the original signal therefore we have chosen in this work, this coefficient to detect R-peaks.

3.4. Thresholding

After applying the CWT to the input ECG signal, the fourth wavelet coefficient we apply the hard thresholding to fourth wavelet coefficient, C_4 and the threshold is selected to be:

$$Thr = \alpha \times \max(C_4) \qquad (4)$$

$$\text{If } C_4(i) \le Thr$$
$$C_4(i) = 0$$

where α is a positive parameter belonging to the range of 0.3 to 0.9.

4. RESULTS AND VALIDATION

The algorithm has been tested on MIT-BIH arrhythmia databases in which every recording is of 30 minutes duration, 10 records were tested for R peaks to evaluate our algorithm. In our evaluation of the proposed technique, we have calculated the Sensitivity, the Positive predictivity and the Error which:

- Sensitivity:
$$S_e = \frac{TP}{TP+FN} \qquad (5)$$

- Positive productivity:
$$P^+ = \frac{TP}{TP+FP} \qquad (6)$$

Mourad Talbi, Akram Aouinet, Lotfi Salhi & Adnane Cherif

• Error :
$$\%error = \frac{FP+FN}{Total\,beats}$$
(7)

Table1 shows that our method achieves very good detection performance. This algorithm attains sensitivity of 99.96% and a positive predictivity of 99.84% without the need to apply any pretreatment to the original signal.

Tape (N°)	Total N° beats	FP beats	FN beats	P^+ (%)	S_e (%)
100	2273	0	0	100	100
101	1865	0	2	100	99.89
102	2187	13	4	99.40	99.81
103	2084	0	0	100	100
104	2230	21	0	99.06	100
105	2572	0	0	100	100
106	2027	0	0	100	100
107	2137	0	1	100	99.95
111	2124	0	0	100	100
112	2539	0	0	100	100

TABLE 1: Performance of the proposed classification model for test data.

Table2 shows that the proposed method outperforms some conventional techniques used in our evaluation such as the techniques of Arzeno et al. [11], Mahmoodabadi et al. [2] and Hubin and Jiankiang [12]. The technique of Rym Besrour et al. [13] gives the best result in term of %error and the proposed technique comes in the second place. The latter gives the best result in term of Se % and the technique of Jasko [14] is the best in term of P+ %.

| QRS detector	S_e %	P^+ %	% error
Arzeno et al.[11]	99.29	99.24	1.47
	99.57	99.59	0.84
	98.07	99.18	2.75
Huabin and Jiankiang [12]	99.68	99.59	0.73
Josko [14]	99.86	99.91	0.23
Mahmoodabadi et al.[2]	99.18	98	2.82
Rym Besrour et al [13]	99.92	99.88	0.19
Martinez et al. [15]	99.80	99.86	0.34
This work	**99.96**	**99.84**	**0.2**

TABLE 2: R wave's detection results on MIT-BIH database.

The positions of the R peaks are detected and marked by the symbol '*' on the original signal. Figure 3 illustrates some examples of R-wave detection using the proposed technique.

(a)
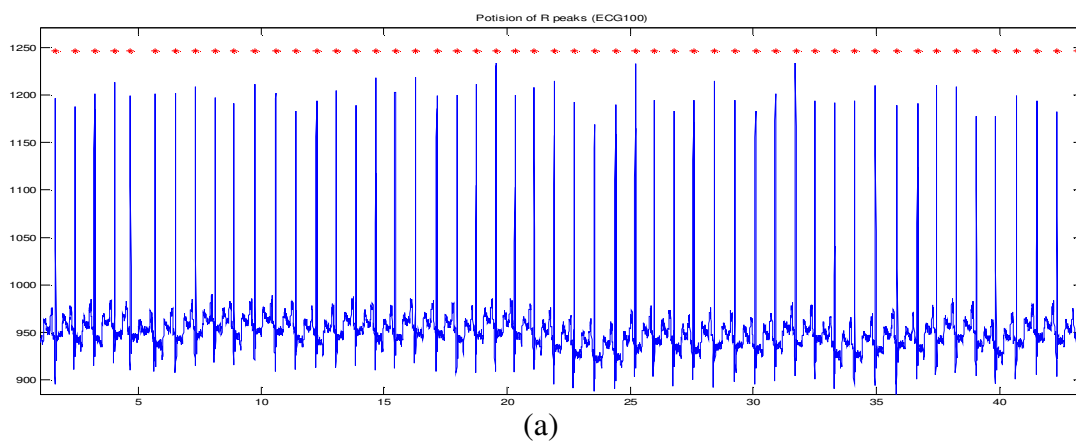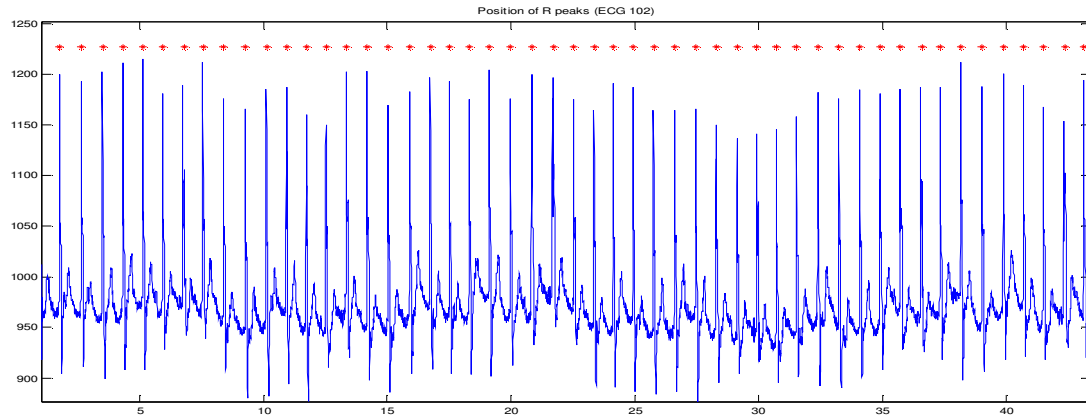
(b)
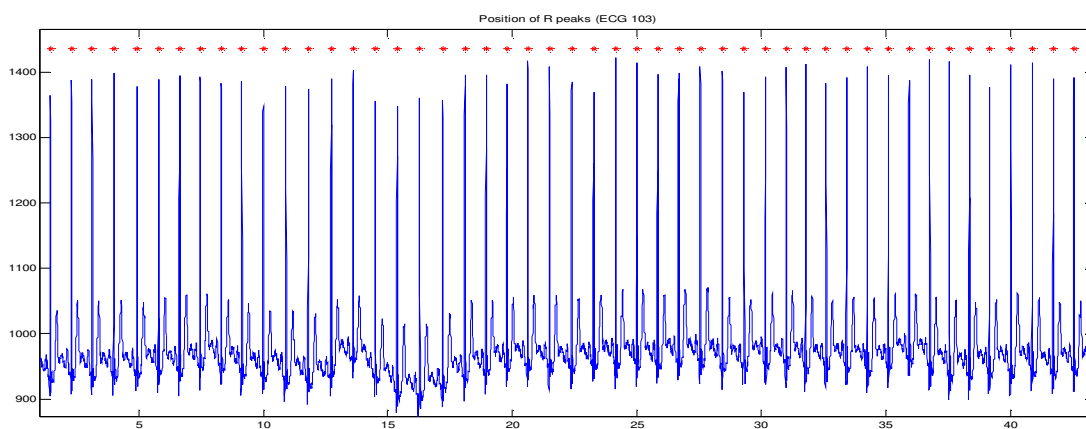
(c)

FIGURE 3: Original ECG signals and positions of R peaks (a) 100 (b) 102 (c) 103.

Those examples show the efficiency of the proposed R-wave detection technique. When we especially compare our proposed technique to the technique of Awadhesh Pachauri et al [1], we see clearly that the proposed technique outperforms the second technique. The proposed technique gives 99.96% as a result of S_e computation and about 99.84% for P^+ whereas the achieved overall accuracy of detection using db6 and sym11 are 96.65% and 84.37% respectively and this for the second technique of Awadhesh Pachauri et al[1]. Moreover, when we use the thechnique of Awadhesh Pachauri et al [1], we can see clearly in figure 4, that there is a great difference between some detected R-peak positions and the real positions of those peaks. This shifting in R-peaks positions is particularly absent when we use our proposed technique.

Mourad Talbi, Akram Aouinet, Lotfi Salhi & Adnane Cherif
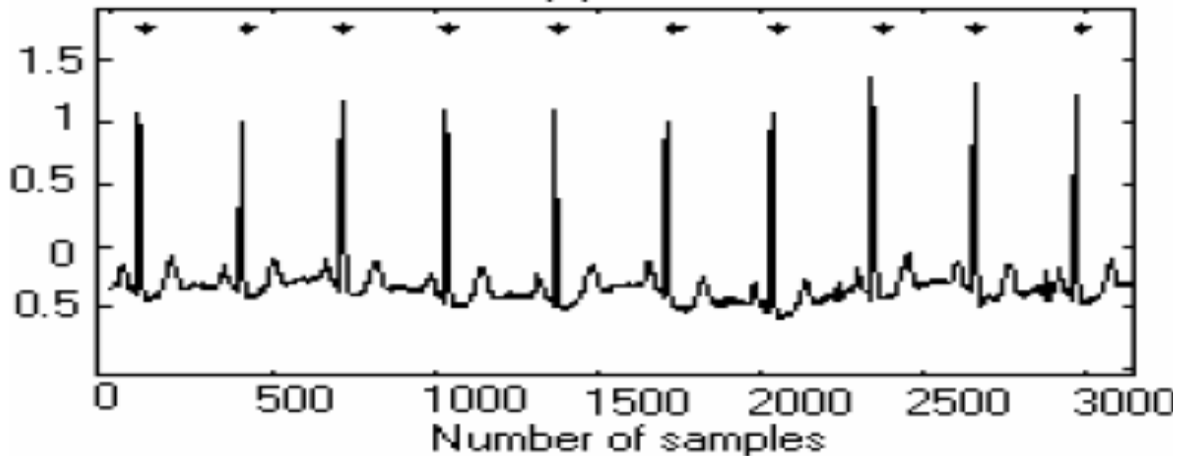
FIGURE 4: Shifting in R-peak positions marked by the technique of Awadhesh Pachauri et al[1].

The performance of the proposed technique can be seen as a result of the use of the discretised continuous wavelet transform which is modified (MMycwt) according to the characteristics of the ECG signal. The latter has less dynamics than a speech signal for example. Therefore it is more suitable to use a discrete transform than a continuous transform. Moreover the length of each coefficient obtained from the MMycwt application to an ECG signal, is the same length of that signal so this fact permits to facilitate the detection of the R-wave positions.

5. CONCLUSION

In this paper we have presented a new method for R wave detection using discretised continuous wavelet transform used by the bionic wavelet transform (BWT). This transform was modified according the ECG signal characteristics in order to obtain the MMycwt. The mother wavelet associated to this transform is the Morlet wavelet. We have decomposed the ECG signal into 8 scales and we have chosen the fourth coefficient in order to detect the R-peaks. This detection is performed by applying a hard thresholding to the fourth coefficient obtained from the application of the MMycwt to the ECG signal. The algorithm has been validated using MIT-BIH standard database and is compared to some others techniques. The obtained results from S_e and P^+ computation, show that the proposed technique outperforms the others techniques used in our evaluation.

6. REFERENCES

[1] A. Pachauri, and M. Bhuyan, "Robust Detection of R-Wave Using Wavelet Technique", World Academy of Science, Engineering and Technology 56 2009.

[2] S.Z.Mahmoodabadi, A.Ahmadian, and M.D.Abolhasani, "ECG Feature Extraction Using Daubechies Wavelets", Proceeding of the Fifth IASTED International Conference.

[3] http://www.physionet.org/physiobank/database/SVdb/.MIT-BIH

[4] Supraventricular Arrhythmia Database. Available from Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA02139, USA.

[5] H. Khorrami, and M. Moavenian, "Acomparative study of DWT, CWT, and DCT transformations in ECG arrhythmias classification", Expert systems with Applications 37 (2010) 5751-5757.

[6]   Yao, J. and Zhang, Y.T. (2001) "Bionic wavelet transform: a new time-frequency method based on an auditory model", IEEE Trans. On Biomedical Engineering. Vol. 48, No. 8, pp.856-863.

[7]   P.S.Addison "Wavelet Transform and the ECG": a review. Physiological Measurement 2005; 26:155-199.

[8]   F. E.Olevera, Jr., and Student Member, IEEE, "Electrocardiogram Wave Feature Extraction Using the Matched Filter", ECE 510: STATISTICAL SIGNAL PROCESSING II.

[9]   Omid Sayadi, Mohammad BagherShamsollahi, "ECG Denoising with Adaptive Bionic Wavelet Transform",

[10]   X. Yuan, "Auditory Model-based Bionic Wavelet Transform For Speech Enhancement", M.Sc. Thesis, Marquette University, Speech and Signal Processing Lab Milwaukee, Wisconsin, May 2003

[11]   Johnson, M.T., Yuan, X. and Ren, Y. (2007), "Speech signal enhancement through adaptive wavelet thresholding", Science Direct, Speech Communication, Vol. 49, pp.123-133.

[12]   N. M. Arzeno, Z.-D. Deng, and C.-S. Poon, "Analysis of first derivative based QRS detection algorithms," IEEE Transactions on Biomedical Engineering, vol. 55, no. 2, pp. 478–484, 2008.

[13]   Z. Huabin and W. Jiankang, "Real-time QRS detection method," in Proceedings of the 10th International Conference on E-Health Networking, Applications and Services, pp. 169–170, Singapore, July 2008.

[14]   R. Bessrour, Z. Lachiri and N. Ellouze, "Using Multiscale Product for ECG Characterization, Hindawi Publishing Corporation, Research Letters in Signal Processing, Volume 2009, Article ID 209395, 5 pages

[15]   A. Josko, "Discrete wavelet transform in automatic ECG signal analysis," in IEEE Instrumentation and Measurement Technology Conference,Warsaw, Poland, 2007.

[16]   J.P. Martinez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, "A wavelet-based ECG delineator: evaluation on standard databases," IEEE Transactions on Biomedical Engineering, vol. 51, no. 4, pp. 570-581, 2004.

A Meter Classification System for Spoken Persian Poetries

Saeid Hamidi   *saeidhamidi@iau-maragheh.ac.ir*
Department of Electrical Engineering,
Maragheh Branch
Islamic Azad University
Maragheh,Iran

Farbod Razzazi   *razzazi@srbiau.ac.ir*
Department of Electrical and Computer Engineering,
Science and Research Branch
Islamic Azad University
Tehran,Iran

Abstract

In this article, a meter classification system has been proposed for Persian poems based on features that are extracted from uttered poem. In the first stage, the utterance has been segmented into syllables using three features, pitch frequency and modified energy of each frame of the utterance and its temporal variations. In the second stage, each syllable is classified into long syllable and short syllable classes which is a historically convenient categorization in Persian literature. In this stage, the classifier is an SVM classifier with radial basis function kernel. The employed features are the syllable temporal duration, zero crossing rate and PARCOR coefficients of each syllable. The sequence of extracted syllables classes is then softly compared with classic Persian meter styles using dynamic time warping, to make the system robust against syllables insertion, deletion or classification and poems authorities. The system has been evaluated on 136 poetries utterances from 12 Persian meter styles gathered from 8 speakers, using k-fold evaluation strategy. The results show 91% accuracy in three top meter style choices of the system.

Keywords: Syllable Classification, Utterance Syllabification, Automatic Meter Detection, Support Vector Machines, Dynamic Time Warping, Poetries Categorization

1. INTRODUCTION

Poem is the vital part of literature of all cultures and reflects the specifications and maturity of a cultural society. Rhyme and meter are considered as inseparable elements of the poetry and meter extraction is a historically exquisite subject for literary scholars which have been extracted intuitively by now.

There is a rich literature on automatic speech recognition systems for general applications in last 30 years; however, automatic extraction of rhyme and meter styles from uttered poems is the focus of some studies in recent years for various languages including Chinese [1-3], Thai [4] and European languages [5,6]. Although there is a very rich treasury of Persian poetries which are created during more than are thousand years, however, most of the studies on these poems are literary studies and they are not well prepared for machinery manipulations. In particular, there is little research, concentrating on machinery Persian meter detection in uttered speech.

In this article, automatic speech recognition utilities are employed to extract an algorithm for automatic meter detection from uttered poetries. The input of this system is a single uttered verse of a poem and the output is the meter style. The organization of the paper is as follows. The theory of Meter extraction in Persian poetries is introduced in section 2. In section 3, the architecture of the proposed system is discussed. The syllabification algorithm is presented in section 3. Section 4 and 5 is devoted to syllable classification and sequence classification of

syllables respectively. In section 6, the implementation of the system is analyzed and evaluated. The paper is concluded in section 7.

2. THEORY OF METER DETECTION IN PERSIAN POETRIES

There are a few studies on extracting and detecting the poetries meter and rhyme in different languages [1-5]. However, the poetry is a specific property of each language and the meter extraction problem should be handled separately in each language.

Thank to the theory of Persian meters, named as Arooz, with more than 700 years old, Persian meter detection is based on syllabification of speech into long and short syllables [7,8]. There is a set of distinguished classes of Persian meter styles in the literature which 12 classes of them covers most of the existent poetries.

The categorization of Persian syllables is tabulated in TABLE 1. As demonstrated, the variety of Persian syllables is limited. There is a vowel in the kernel of each syllable and there is one starting consonant before the vowel. After the vowel, it is possible to have no consonant, one consonant or two consonants. This simple structure motivates us to find out the syllabification sequence by extracting the location of kernel vowel and locating the boundaries of the syllables by moving front and back from the kernel.

#	Syllable Category
1	CV
2	CVC
3	CVCC

TABLE 1: Variety of Persian Syllables

In Arooz Theory, there are two kinds of syllables, short syllables and long syllables. Long and short syllables are distinguished by the kernel vowel used in the syllables. The vowel in short syllables are a member of the set (/ae/,/eh/,/oy/) . In contrast, long syllables consist a vowel in the set (/aa/,/iy/,/ux/) The Nomination short and long syllables is due to the utterance duration of each syllable when it is intended to read the poem in meter style. It is empirically shown that the times spent to utter long syllables are almost similar. Short syllables are uttered in similar duration too. However, the duration of short and long syllables are not the same.

It is revealed that there are common standard meters that are frequently used by poets and seems to be well accepted by Persian speakers. Over 95% of the poetries are covered by 12 standard meters. Therefore, this study is concentrated on these standard meters which are tabulated in TABLE 2.

All of the verses of each poem should employ similar standard meter. However, there is no strict rule in the artistic world. Although, most of the verses employ one standard meter, in fact, sometimes, poets have used the standard meter by slight modifications in some verses. This phenomenon, which is called poetry authorities, will make the approach of the machinery system to be a soft likelihood measurement rather than pattern matching.

#	Standard Meter	Example written in Phonetics	Example written in Persian
1	SSLS LSLL SSLS LSLL	q/ae/l/iy/eh/y/hh/oy/m/aa/y/eh/r/ae/hh/m/ae/t/ oy/ch/eh/aa/y/ae/t/iy/kh/oy/d/aa/r/aa	علی ای همای رحمت تـو چه آیـتی خدا را
2	LSSL SLSL LSSL SLSL	s/ae/r/v/eh/ch/ae/m/aa/n/ae/m/ae/n/ch/eh/r/aa/ m/eh/y/l/eh/ch/ae/m/ae/n/n/ae/m/iy/k/oy/na/d	سرو چمان من چرا مـیل چمن نمیکند
3	SSLL SLSL SSL	d/ae/r/d/eh/q/eh/sh/gh/iy/k/ae/sh/iy/d/eh/q/ae/m/ k/eh/m/ae/p/oy/r/s	درد عشقی کشیده ام کـه مپرس
4	SLSL SSLL SLSL SSL	t/ae/n/ae/t/b/eh/n/aa/z/eh/t/ae/b/iy/b/aa/n/ n/iy/y/aa/z/m/ae/n/d/m/ae/b/aa/d	تـنت بـه نـاز طبیبـان نـیازمـند مبـاد
5	SLL SLL SLL SL	m/ae/g/ae/r/d/aa/n/s/ae/r/ae/z/ d/iy/n/oy/q/ae/z/r/aa/s/t/iy	مـگردان سر از دیـن و از راستی
6	SLL SLL SLL SLL	n/eh/k/ou/h/eh/sh/m/ae/k/oy/n/ ch/ae/r/kh/eh/n/iy/l/oy/f/ae/r/iy/r/aa	نکـوهش مکن چرخ نـیـلـوفـری را
7	LLSL LLSL LLSL LLSL	b/aa/m/ae/n/b/eh/g/oy/t/aa/k/iy/s/t/iy/ m/eh/h/r/iy/b/eh/g/ou/m/aa/h/iy/b/eh/g/ou	بـا مـن بـگو تـا کیستی مـهری بـگو مـا هی بـگو
8	SLLL SLLL SLLL SLLL	b/iy/aa/t/aa/g/oy/l/b/ae/r/ae/f/sh/aa/n/ iy/m/oy/m/eh/iy/d/ae/r/s/aa/gh/ae/r/ ae/n/d/aa/z/iy/m	بـیا تـا گل بـرافشانـیم و مـی در سـاغر انـدازیم
9	LLSL SLL LLSL SLL	y/aa/r/ae/b/t/oy/q/aa/sh/ae/n/aa/r/aa/ m/oy/h/l/ae/t/d/eh/h/oy/s/ae/l/aa/m/ae/t	یـارب تـو آشنا را مـهلت ده و سلامت
10	LLS LLL LLS SLL	q/ae/iy/p/aa/d/ae/sh/ae/h/eh/kh/ou/b/aa/ n/d/aa/d/ae/z/gh/ae/m/eh/t/ae/n/h/aa/q/iy	ای پـادشه خوبـان داد از غم تـنهائی
11	LSLL LSLL LSL	m/ae/r/d/eh/r/aa/d/ae/r/d/iy/ae/g/ae/r/ b/aa/sh/ae/d/kh/oy/sh/ae/s/t	مـرد را دردی اگر بـاشد خوشست
12	SSLL SSLL SSLL SSLL	d/ae/r/d/eh/l/ae/m/b/ou/d/k/eh/b/iy/d/ou/s/t/ n/ae/b/aa/sh/ae/m/h/ae/g/ae/z	در دلم بـود کـه بـی دوست نبـاشم هرگز

TABLE 2: Categories of frequently used Persian standard meters

3. SYSTEM ARCHITECTURE

FIGURE.1 presents the overall block diagram of meter extraction for Persian poetries. As it is shown, after preprocessing and syllabification, some features are extracted from each syllable. The features are zero-crossing rate, PARCOR coefficients and temporal duration for each syllable.

In the next stage, syllables are classified into long and short syllables. Finally the sequence of syllable classes is compared to standard Persian poetry meters using dynamic programming. The best match of the sequence with the standard meters shows the category of the meter.
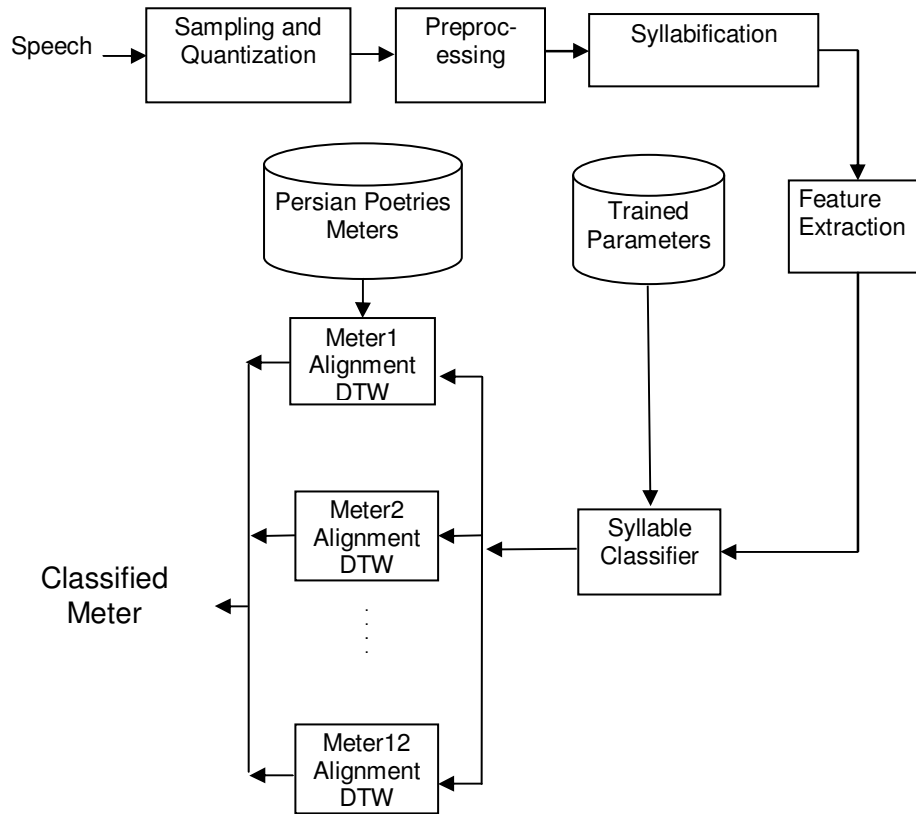
FIGURE 1: Overall Block Diagram

4. SYLLABIFICATION SUBSYSTEM

The meters in Persian poetries are categorized based on the sequence of syllables classes. Therefore, it is essential to have a syllabification stage [9-11]. Syllables in Persian consist of one consonant, one vowel and probably one or two consonants respectively. As a result, syllable segmentation is based on detecting the location of vowels in the utterance. The implemented syllable segmentation is based on three features, the pitch frequency, the energy and estimation of energy derivative during time.

FIGURE 2 demonstrates the detail of mentioned syllable segmentation procedure. After a 32ms, 50% overlapped framing, the utterance is filtered to have low pass frequency components and detect the pitch frequency more accurately. The energy and pitch frequency of the frame are then extracted. These two features are used to extract the boundary frames of syllables.
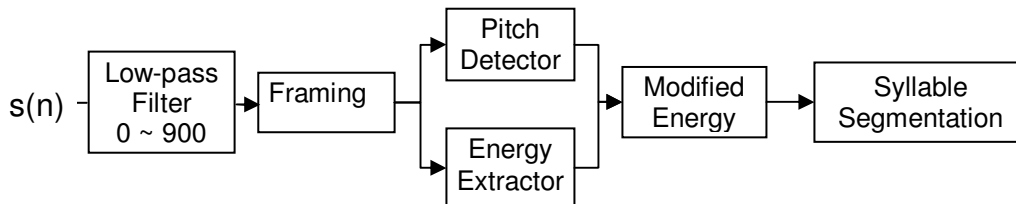

**FIGURE 2:** Syllable Segmentation Block Diagram

The segmentation algorithm requires locating the voiced frames. Hence, the focus of the pitch detection algorithm is the extraction of the frames where pitch frequency can be recovered in

them. As demonstrated in FIGURE 3, the voiced and unvoiced frames of the utterance are segmented based on Dubnowski - Rabiner algorithm [12].

To extract the modified energy, the energy of the frame is computed as the sum of squared amplitude of frame samples.

$$E[n] = \sum_{i=1}^{L} x^2(i)$$

where E[n] is the nth frame energy and L is the length of the nth frame. x represents the frame samples amplitude. The frame energies are then normalized to maximum frame energy.

$$E_{norm}[n] = \frac{E[n]}{Max(E[n])}$$
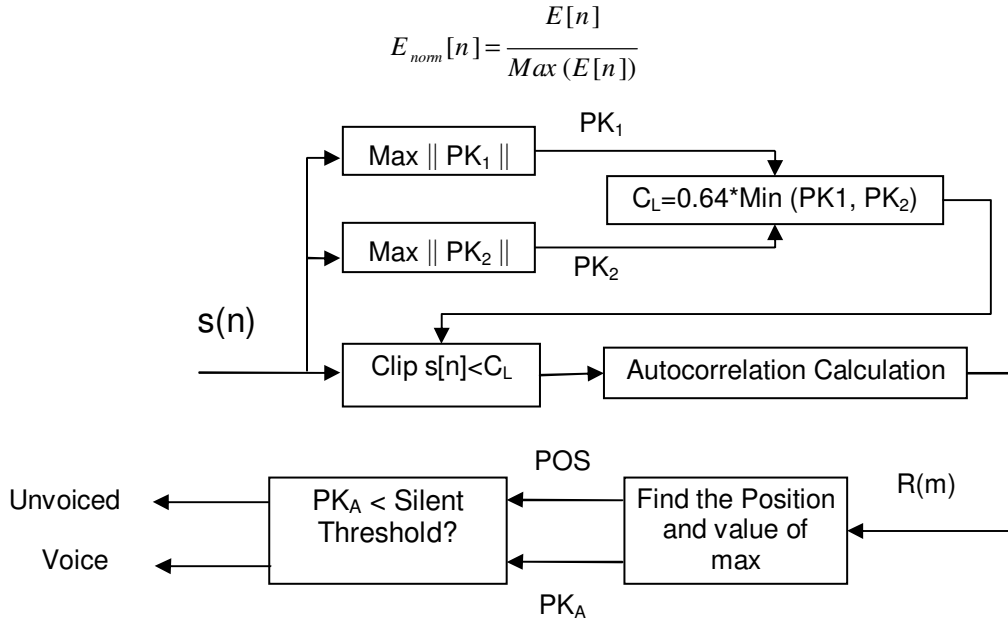
FIGURE 3: Voiced/ Unvoiced Classification

The energy contour of the utterance is then filtered by a nonlinear median filter to enhance the locating procedure of peaks and valleys in energy contour and therefore enhancing the performance of syllable segmentation stage. Modified energy is then extracted as the windowed version of smoothed energy by voiced framed locations.

FIGURE 4 demonstrates a sample of segmented signals into voiced and unvoiced segments. The modified energy is the clipped version of smoothed energy. Pitch and modified energy contours are nonzero in voiced frames of the utterance, while the smoothed energy is nonzero for all frames.
Modified Energy is the input of syllabification stage. As it is shown in FIGURE 5, in this stage, the first valley in nonzero modified energies is the primary estimation of syllable boundary. To extract the subtle boundary, it is better to trace back the signal to find out the first zero in modified energy contour. To avoid undesired over-segmentation, the value of modified energy in the valley should be less than two thresholds derived by two adjacent peaks.

The thresholds are set to suitable empirical values to avoid unwanted syllabifications, a minimum length for syllables was considered. The algorithm continues until the end of the utterance. FIGURE 6 is an example of the procedure output. In this example, the utterance is /t ae v aa n aa b oy v ae d/ and the segmentation process succeeded to segment the utterance into (/t/ae/, v/aa, /n/aa, /b/oy, /v/ae/d)(توانا بود).
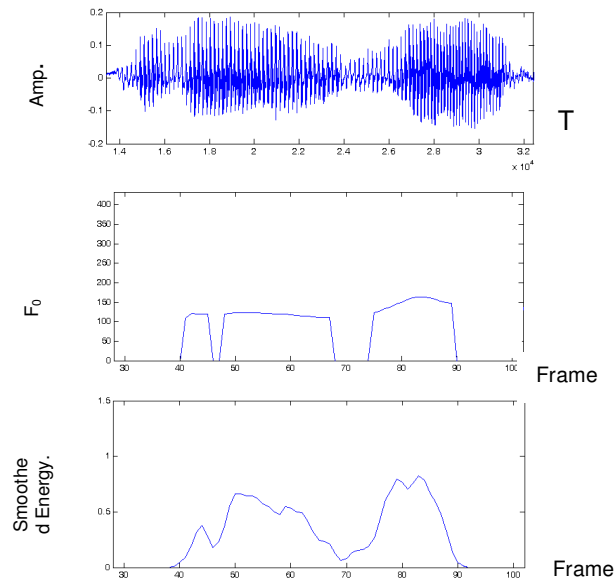
FIGURE 4: Sample of Intermediate Signals of Block Diagram in FIGURE 3 Top: input Signal Middle: Pitch Contour Bottom: Smoothed Energy.

5. SYLLABLE CLASSIFICATION

To classify the syllables into long and short syllables, 12 partial reflection coefficients are extracted from linear prediction analysis of each syllable in addition to zero crossing rate and syllable duration. These features are added up with two previous features extracted in the syllabification stage (pitch frequency, modified energy), make the whole feature vector. Intuitively, it seems that the most effective feature in syllable classification would be the syllable duration. Therefore the tests were designed to check the performance of the system on both single duration feature and the whole vector as the feature vector. Each feature is normalized with respect to mean and variance of the feature in the whole syllables space. Hence, all features become zero mean, unity variance after normalization.

The features are classified using a kernel based support vector machine with RBF kernel [13,14]. Kernel meta-parameters are optimized empirically using grid search [15] evaluated on K-fold evaluation strategy.

The overall block diagram is depicted in FIGURE 5. the modified energy is the input of syllable classification. The output of the block is both the start point and ending point of the syllable. The local maxima of the modified energy are the candidates of starting and ending point of the syllable. The candidates are refined using a minimum energy candidate. A sample of modified energy schematic behavior and real world behavior is demonstrated in FIGURES 6 and 7. The notations b and p demonstrate the central and boundary points of the syllable.

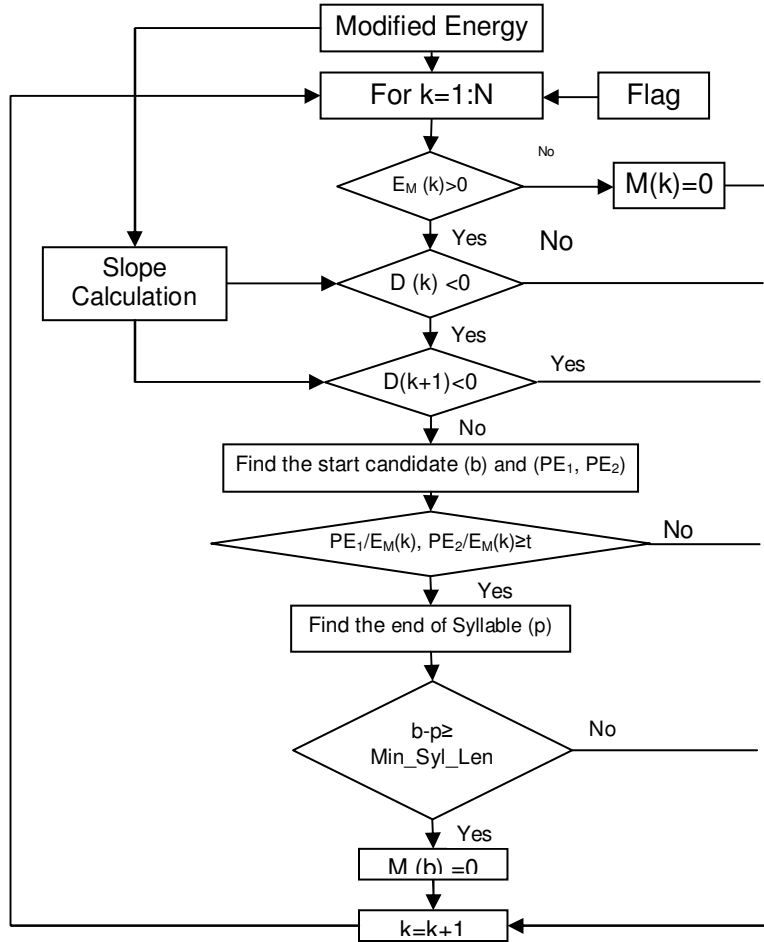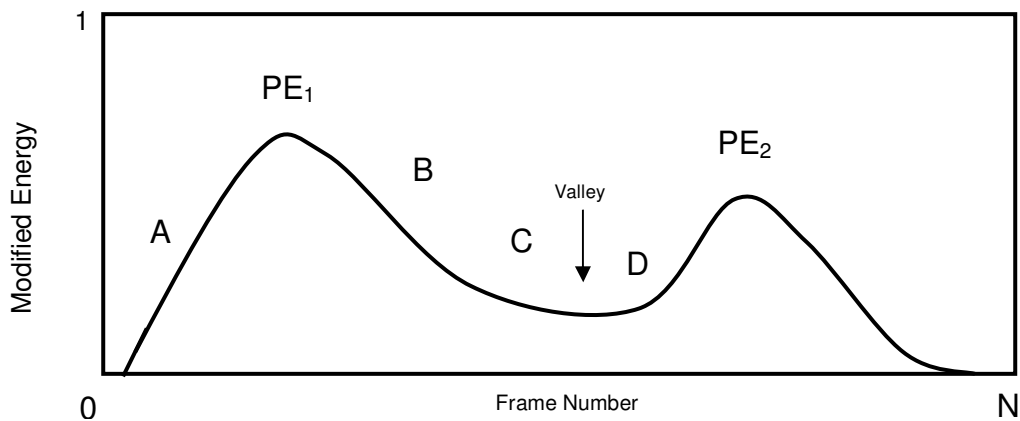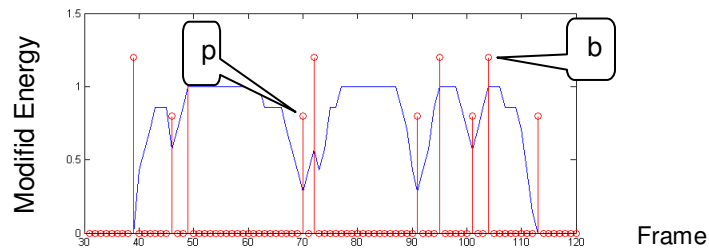**FIGURE 5:** Syllabification Block Diagram

FIGURE 6: Modified Energy Variations in a Syllable

FIGURE 7: A Sample of Syllabification Based on Modified Energy

6. METER MATCHING USING DYNAMIC TIME WARPING

The meters are commonly described by a sequence of syllable classes. Fortunately, the employed meters in Persian poems are selected from a limited number of standard meters. This fact lets the meter classifier system to compensate syllable segmentation and classification errors. In addition, poets are not rigidly loyal to standard meters, however they use long syllables instead of short syllables and vise versa depending on linguistic and semantic constraints. This phenomenon, known as "poetry options" in the literature, should be considered to find out the best match (not exact match) between the extracted syllable classes sequence and the standard meters, including substitution, deletion and insertion errors. This matching was carried out by well known dynamic programming approach in speech utterances matching, named as dynamic time warping [16]. The output of the system is the number of matches, substitutions, insertions and deletions of one verse utterance with respect to each of standard meters. The error is defined as the sum of substituted, inserted and deleted syllables. The standard meter with minimum error is referred as the recognized meter.

7. EXPERIMENTS AND RESULTS

The proposed system was evaluated on 17 versus with 12 distinct Persian standard meters. The meters were selected to cover more than 95% of the Persian poetries. The verses were uttered by 8 native speakers whose are requested to pronounce each verse in correct meter. For evaluation purposes, all verses were manually segmented into syllables and all syllables were labeled by long and short syllable labels.

The first evaluation was made on the syllabification stage. The automatically extracted syllables were compared to manual segmentations and the number of insertions and deletions were evaluated. TABLE 3 explains the results in detail. As it can be observed, the average syllabification error is 10% which is comparable to the literature for other languages. The accuracy is variable with respect to the speaker, due to the pronunciation and accents variations. The standard deviation of this error in the set of speakers is about 19% of the mean value.
The system parameters are optimized to achieve the highest accuracy. In this point, the number of deletions is about twice the number of insertions.

In syllable classification stage, the system was evaluated by the classification rate. To optimize the classifier, two parameters (i.e. the misclassification weight in the training procedure (denoted as C) and the RBF kernel parameter denoted as □) were optimized in the logarithmic grid search basis.

The evaluation was performed based on K-Fold strategy with K=10. The average classification rates in 10 folds are tabulated in TABLE 4 for different C and γ values. It can be concluded that the system is optimized with $(C,\gamma)=(3.16, 0.0316)$. The best syllable classification rate is about 75%. This accuracy rate may cause the sequence classification unusable unless the result is post-processed by a dynamic comparison with the reference meters.

Speaker	Number of Segments	Deletion	Insertion	Error Percentage
Spk1	433	0.1	0.01	11.7
Spk2	455	0.07	0.037	11.2
Spk3	434	0.108	0.02	12.9
Spk4	471	0.036	0.042	7.8
Spk5	448	0.078	0.024	10.2
Spk6	454	0.074	0.026	10.1
Spk7	459	0.074	0.045	11.9
Spk8	467	0.038	0.027	6.6
Ave.	452.6	32.8	13.7	10.2

TABLE 3: Syllable Segmentation Rates For Different Speakers

C	$\gamma=0.01$	$\gamma=0.03$	$\gamma=0.1$	$\gamma=0.31$
0.0001	62.66	62.66	62.66	62.66
0.00033	62.66	62.66	62.66	62.66
0.001	62.66	62.66	62.66	62.66
0.003	62.66	62.66	62.66	62.66
0.01	62.66	62.66	62.66	62.66
0.03	62.66	62.66	62.66	62.66
0.1	61.90	70.07	70.37	63.90
0.31	70.96	73.36	73.93	68.69
1	73.21	74.27	74.80	72.03
3.16	73.85	75.41	74.20	71.65
10	74.19	74.49	72.80	70.04
31.32	74.28	73.69	70.93	69.26
100	74.63	72.68	68.82	69.36
316.22	73.04	72.12	67.59	69.36

TABLE 4: SVM Classifier Meta-Parameter Optimization

In The last stage, the mentioned post-processing is carried out. A dynamic time warping algorithm was employed to compare the meters and select the best among common reference Persian poems meters. TABLE 5 and FIGURE 7 demonstrate the result for this dynamic matching and scoring.
The results in TABLE 5 show that the system achieves 91% in 3 best meter classification rate.

n-Best	Spk 1	Spk 2	Spk 3	Spk 4	Spk 5	Spk 6	Spk 7	Spk 8	Ave
1	64	64	58	64	76	76	76	70	69
2	88	94	76	94	88	100	82	94	89
3	88	100	76	94	94	100	82	94	91

TABLE 5: N-Best Classification Rates For Different Speakers
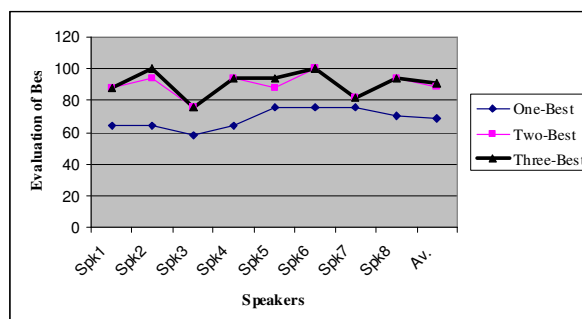

FIGURE 8: Comparison of n-best Rates for Different Speakers

The comparison of speakers in FIGURE 8 shows that there is a significant difference in the meter detection accuracy for different speakers. Probing the result, it was observed that bad results belong to the speakers who did not utter the poems in correct meter. Therefore, in this situation, incorrect detection of meter type is expectable. The system may be generalized to reject the miss-uttered poems by speakers.

8. CONSLUSION & FUTURE WORK
In this paper, an automatic meter detection algorithm was implemented and evaluated. This is the first attempt to analyze the utterance of Persian poetries automatically which may lead the researchers toward a new approach in investigating the literature theoretically and practically. In addition, there is a rich source of human culture in poetries, which can be digitized by tracing this trajectory.

9. REFERENCES
[1] Z. S. He, W.T. Liang, L. Y. Li, Y.F. Tian, *"SVM-Based Classification Method For Poetry Style"*, Proceedings of Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, Aug 2007, PP. 2936-2940.

[2] X.Wang, Y. Liu, L.Cai, *"Entering Tone Recognition in a Support Vector Machine Approach"*, Proceedings of Fourth International Conference on Natural Computation, Aug. 2007, PP. 61-65.

[3] Y.YI, Z. S. He, L. Y. Li, T. Yi, E. Yi, *"Advanced Studies On Traditional Chinese Poetry Style Identification"*, Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhu, Aug. 2005, PP. 3830-3833.

Saeid Hamidi & Farbod Razzazi

[4]   N. Sataravaha. *"Tone classification of syllable segmentation Thai speech based on multilayer   perceptron"* Acoustical Society of America Journal, Volume 111, Issue 5, 2002,pp. 2366-2366.

[5]   H.R. Tizhoosh, R.A. Dara, *"On Poem Recognition"*, Pattern Analysis and Application, Vol. 9, No.4, Nov. 2006, PP. 325-338.

[6]   H. Meinedo *"The use of syllable segmentation information in continuous speech recognition hybrid systems "* Applied to the Portuguese language , IEEE Transaction on Acoustic , speech and signal processing. 0-1623-8, 2000

[7]   Vahidian-Kamkar T, *"Persian Poetries Meter and Rhyme"*, Nashr-e-Daneshgahi, 1990. (Published in Persian)

[8]   Shamisa S, *"An Introduction to Meter and Rhyme"*, Ferdos Publications, 1992, (Published in Persian)

[9]   Liany-Yan Li *, " Poetry stylistic analysis technique based on term connection " .* IEEE Speech and signal processing. 0-7803-8403, 2004

[10]   Hartmut R.Pfitzinger   *" syllable detection in read and spontaneous speech "* 13-4822-13324, 1996

[11]   I. Kopecek: *"Speech recognition and syllable segmentation"*; in Proceedings of the Workshop on Text, Speech and Dialogue - TSD'99, Lectures Notes in Artificial Intelligence 1692, Springer-Verlag, 1999, pp. 203-208.

[12]   Dubnowski, J.J. & Rabiner *"Real time digital hardware Pitch detector"*. IEEE Transactions on Acoustics speech and signal processing. 24(1), 2-8. 1976

[13]   V. Vapnik. The Nature of Statistical Learning Theory, Springer Verlag, New York, 1995.

[14]   C. J. C. Burges. *"A Tutorial on Support Vector Machines for Pattern Recognition"* Knowledge Discovery and Data Mining, 2(2), 1998.

15]   P.H.Chen, C.J. Lin, B.S.ch¨olkopf*, "A Tutorial on v-Support Vector Machines"* Applied Stochastic Models in Business and Industry 21(2), pages 111-136, 2005.

[16]   L. Rabiner and B.H. Juang. Fundamentals of Speech Recognition. Prentice Hall, 1993.

INSTRUCTIONS TO CONTRIBUTORS

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.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Starting with volume 5, 2011, SPIJ appears in more focused issues. Besides normal publications, SPIJ intend to organized special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

We are open to contributions, proposals for any topic as well as for editors and reviewers. We understand that it is through the effort of volunteers that CSC Journals continues to grow and flourish.

SPIJ LIST OF TOPICS
The realm of Signal Processing: An International Journal (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
- Signal Filtering
- Signal Processing Technology
- Software Developments
- Spectral Analysis
- Stochastic Processes

- Acoustic and Vibration Signal Processing
- Data Processing
- Digital Signal Processing
- Geophysical and Astrophysical Signal Processing
- Multi-dimensional Signal Processing
- Pattern Recognition
- Remote Sensing
- Signal Processing Systems
- Signal Theory
- Sonar Signal Processing
- Speech Processing

CONTACT INFORMATION