
ADAPTIVE FILTERING APPLICATIONS

Edited by **Lino García Morales**

INTECHWEB.ORG

Adaptive Filtering Applications

Edited by Lino García Morales

Published by InTech

Janeza Trdine 9, 51000 Rijeka, Croatia

Copyright © 2011 InTech

All chapters are Open Access articles distributed under the Creative Commons Non Commercial Share Alike Attribution 3.0 license, which permits to copy, distribute, transmit, and adapt the work in any medium, so long as the original work is properly cited. After this work has been published by InTech, authors have the right to republish it, in whole or part, in any publication of which they are the author, and to make other personal use of the work. Any republication, referencing or personal use of the work must explicitly identify the original source.

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published articles. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

Publishing Process Manager Petra Zobic

Technical Editor Teodora Smiljanic

Cover Designer Jan Hyrat

Image Copyright Vicente Barcelo Varona, 2010. Used under license from Shutterstock.com

First published June, 2011

Printed in Croatia

A free online edition of this book is available at www.intechopen.com
Additional hard copies can be obtained from orders@intechweb.org

Adaptive Filtering Applications, Edited by Lino García Morales

p. cm.

ISBN 978-953-307-306-4

INTECH OPEN ACCESS
PUBLISHER

INTECH open

free online editions of InTech
Books and Journals can be found at
www.intechopen.com

Contents

Preface IX

Part 1 Noise and Echo Cancellation 1

- Chapter 1 **Applications of Adaptive Filtering 3**
J. Gerardo Avalos, Juan C. Sanchez and Jose Velazquez
- Chapter 2 **Applications of Adaptive Filtering:
Recent Advancements in Active Noise Control 21**
Akhtar Muhammad Tahir,
Mitsuhashi Wataru and Nishihara Akinori
- Chapter 3 **Active Noise Cancellation:
The Unwanted Signal and the Hybrid Solution 49**
Edgar Omar López-Caudana
- Chapter 4 **Perceptual Echo Control and Delay Estimation 85**
Kirill Sakhnov, Ekaterina Verteletskaya and Boris Simak

Part 2 Medical Applications 121

- Chapter 5 **Adaptive Noise Removal of ECG Signal Based
On Ensemble Empirical Mode Decomposition 123**
Zhao Zhidong, Luo Yi and Lu Qing
- Chapter 6 **Application of Adaptive Noise Cancellation
in Transabdominal Fetal Heart Rate Detection
Using Photoplethysmography 141**
Kok Beng Gan, Edmond Zahedi
and Mohd. Alauddin Mohd. Ali
- Chapter 7 **Adaptive Filtering by Non-Invasive Vital
Signals Monitoring and Diseases Diagnosis 157**
Omar Abdallah and Armin Bolz

- Chapter 8 **Noise Removal from EEG Signals in Polysomnographic Records Applying Adaptive Filters in Cascade 173**
M. Agustina Garcés Correa and Eric Laciár Leber

- Chapter 9 **Fast Extraction of Somatosensory Evoked Potential Based on Robust Adaptive Filtering 197**
Yuxian Zou, Yong Hu and Zhiguo Zhang

Part 3 Communication Systems 211

- Chapter 10 **A LEO Nano-Satellite Mission for the Detection of Lightning VHF Sferics 213**
Ghulam Jaffer, Hans U. Eichelberger,
Konrad Schwingenschuh and Otto Koudelka

- Chapter 11 **Adaptive MIMO Channel Estimation Utilizing Modern Channel Codes 239**
Patric Beinschob and Udo Zölzer

- Chapter 12 **An Introduction to ANFIS Based Channel Equalizers for Mobile Cellular Channels 255**
K. C. Raveendranathan

- Chapter 13 **Adaptive Channel Estimation in Space-Time Coded MIMO Systems 285**
Murilo B. Loiola, Renato R. Lopes and João M. T. Romano

- Chapter 14 **Adaptive Filtering for Indoor Localization using ZIGBEE RSSI and LQI Measurement 305**
Sharly Joana Halder, Joon-Goo Park and Wooju Kim

Part 4 Other Applications 325

- Chapter 15 **Adaptive Filters for Processing Water Level Data 327**
Natasa Reljin, Dragoljub Pokrajac and Michael Reiter

- Chapter 16 **Nonlinear Adaptive Filtering to Forecast Air Pollution 343**
Luca Mesin, Fiammetta Orione and Eros Pasero

- Chapter 17 **A Modified Least Mean Square Method Applied to Frequency Relaying 365**
Daniel Barbosa, Renato Machado Monaro,
Ricardo A. S. Fernandes, Denis V. Coury
and Mário Oleskovicz

- Chapter 18 **Anti-Multipath Filter with Multiple Correlators in GNSS Receivers 381**
Chung-Liang Chang

Preface

Adaptive filtering is useful in any application where the signals or the modeled system vary over time. The configuration of the system and, in particular, the position where the adaptive processor is placed generate different areas or application fields such as: prediction, system identification and modeling, equalization (deconvolution, reverse filtering, inverse modeling), cancellation of interference, etc. which are very important in many disciplines such as control systems, communications, signal processing, acoustics, voice, sound and image, etc. This book consists of a compendium of applications in three areas of great interest in scientific research: noise and echo cancellation, medical applications, communications systems and others hardly joined by their heterogeneity. There is no a structure and/or algorithm better than other; It all depends on the implementation and the performance target. In all these chapters, each application is a case study with rigor that shows the weakness-strength of the method used (in many cases compared with other methods), assesses its suitability and suggests new forms and areas of use. The problems are becoming increasingly complex and applications must be adapted to solve them. The adaptive filters have proven to be useful in these environments of multiple input/output, variant-time behaviors, and long and complex transfer functions effectively but fundamentally to be still evolving. There are many "variables" to take into account and how to combine them, optimize them and achieve the desired outcome. This book is a demonstration of this and a small illustration of everything that is to come.

Dr. Prof. Lino García Morales
Prof. Titular Dpto. Electrónica y Comunicaciones
Coord. Grado en Arte Electrónico y Digital
Escuela Superior Politécnica
Universidad Europea de Madrid
Spain

Part 1

Noise and Echo Cancellation

Applications of Adaptive Filtering

J. Gerardo Avalos, Juan C. Sanchez and Jose Velazquez
*National Polytechnic Institute
 Mexico*

1. Introduction

Owing to the powerful digital signal processors and the development of advanced adaptive algorithms there are a great number of different applications in which adaptive filters are used. The number of different applications in which adaptive techniques are being successfully used has increased enormously during the last two decades. There is a wide variety of configurations that could be applied in different fields such telecommunications, radar, sonar, video and audio signal processing, noise reduction, between others.

The efficiency of the adaptive filters mainly depends on the design technique used and the algorithm of adaptation. The adaptive filters can be analogical designs, digital or mixed which show their advantages and disadvantages, for example, the analogical filters are low power consuming and fast response, but they represent offset problems, which affect the operation of the adaptation algorithm (Shoval et al., 1995). The digital filters are offset free and offer an answer of greater precision. Also the adaptive filters can be a combination of different types of filters, like single-input or multi-input filters, linear or nonlinear, and finite impulse response FIR or infinite impulse response IIR filters.

The adaptation of the filter parameters is based on minimizing the mean squared error between the filter output and a desired signal. The most common adaptation algorithms are, Recursive Least Square (RLS), and the Least Mean Square (LMS), where RLS algorithm offers a higher convergence speed compared to the LMS algorithm, but as for computation complexity, the LMS algorithm maintains its advantage. Due to the computational simplicity, the LMS algorithm is most commonly used in the design and implementation of integrated adaptive filters. The LMS digital algorithm is based on the gradient search according to the equation (1).

$$w(n+1) = w(n) + \mu e(n)x(n) \quad (1)$$

Where $w(n)$ is the weights vector in the instant n , $w(n+1)$ is equal to the weights vector in $n+1$, $x(n)$ is the input signal sample vector which is stored in the filter delayed line, where $e(n)$ corresponds to the filter's error, which is the difference between the desired signal and the output filter's signal, and μ is the filter's convergence factor. The convergence factor μ determines the minimum square average error and the convergence speed. This factor is directly proportional to the convergence speed and indirectly proportional to the minimal error. Then a convergence speed and minimal error relation is established.

The application depends on the adaptive filter configuration used. The classical configurations of adaptive filtering are system identification, prediction, noise cancellation,

and inverse modeling. The differences between the configurations are given by the way the input, the desired and the output signals are used. The main objective of this chapter is to explain the typical configurations and it will focus on recent applications of adaptive filtering that are used in the real world.

2. System identification

The system identification is an approach to model an unknown system. In this configuration the unknown system is in parallel with an adaptive filter, and both are excited with the same signal. When the output MSE is minimized the filter represents the desired model. The structure used for adaptive system identification is illustrated in figure 1, where $P(z)$ is an unknown system to be identified by an adaptive filter $W(z)$. The signal $x(n)$ excites $P(z)$ and $W(z)$, the desired signal $d(n)$ is the unknown system output, minimizing the difference of output signals $y(n)$ and $d(n)$, the characteristics of $P(z)$ can be determined.

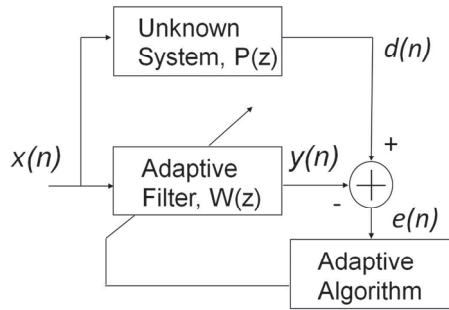


Fig. 1. Adaptive filter for system identification

The estimation error is given as (2)

$$e(n) = d(n) - y(n) = \sum_{l=0}^{L-1} [p(l) - w_1(n)] x(n-l) \quad (2)$$

Where $p(l)$ is the impulse response of the unknown plant, By choosing each $w_1(n)$ close to each $p(l)$, the error will be minimized. For using white noise as the excitation signal, minimizing $e(n)$ will force the $w_1(n)$ to approach $p(l)$, that is,

$$w_1(n) \approx p(l), l = 0, 1, \dots, L-1 \quad (3)$$

When the difference between the physical system response $d(n)$ and the adaptive model response $y(n)$ has been minimized, the adaptive model approximates $P(z)$ from the input/output viewpoint. When the plant is time varying, the adaptive algorithm has the task of keeping the modelling error small by continually tracking time variations of the plant dynamics.

Usually, the input signal is a wideband signal, in order to allow the adaptive filter to converge to a good model of the unknown system. If the input signal is a white noise, the best model for the unknown system is a system whose impulse response coincides with the $N + 1$ first samples of the unknown system impulse response. In the cases where the impulse response of the unknown system is of finite length and the adaptive filter is of sufficient order, the MSE becomes zero if there is no measurement noise (or channel noise).

In practical applications the measurement noise is unavoidable, and if it is uncorrelated with the input signal, the expected value of the adaptive-filter coefficients will coincide with the unknown-system impulse response samples. The output error will of course be the measurement noise (Diniz, 2008). Some real world applications of the system identification scheme include control systems and seismic exploration.

3. Linear predictor

The linear prediction estimates the values of a signal at a future time. This model is wide usually in speech processing applications such as speech coding in cellular telephony, speech enhancement, and speech recognition. In this configuration the desired signal is a forward version of the adaptive filter input signal. When the adaptive algorithm converges the filter represents a model for the input signal, this model can be used as a prediction model. The linear prediction system is shown in figure 2.

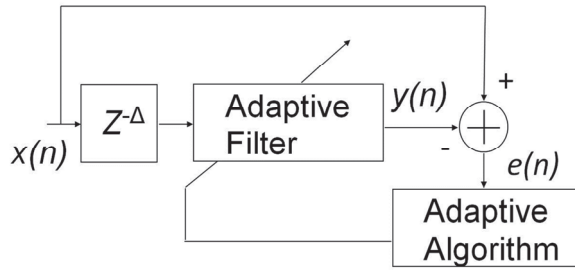


Fig. 2. Adaptive filter for linear prediction

The predictor output $y(n)$ is expressed as

$$y(n) = \sum_{l=0}^{L-1} w_l(n) x(n - \Delta - l) \quad (4)$$

Where Δ is the number of delay samples, so if we are using the LMS algorithm the coefficients are updated as

$$w(n+1) = w(n) + \mu x(n - \Delta) e(n) \quad (5)$$

Where $x(n - \Delta) = [x(n - \Delta) \ x(n - \Delta - 1) \ \dots \ x(n - \Delta - L + 1)]^T$ is then delayed reference signal vector, and $e(n) = x(n) - y(n)$ is the prediction error. Proper selection of the prediction delay Δ allows improved frequency estimation performance for multiple sinusoids in white noise. A typical predictor's application is in linear prediction coding of speech signals, where the predictor's task is to estimate the speech parameters. These parameters are part of the coding information that is transmitted or stored along with other information inherent to the speech characteristics, such as pitch period, among others.

The adaptive signal predictor is also used for adaptive line enhancement (ALE), where the input signal is a narrowband signal (predictable) added to a wideband signal. After convergence, the predictor output will be an enhanced version of the narrowband signal. Yet another application of the signal predictor is the suppression of narrowband interference in a wideband signal. The input signal, in this case, has the same general characteristics of the ALE.

4. Inverse modeling

The inverse modeling is an application that can be used in the area of channel equalization, for example it is applied in modems to reduce channel distortion resulting from the high speed of data transmission over telephone channels. In order to compensate the channel distortion we need to use an equalizer, which is the inverse of the channel's transfer function.

High-speed data transmission through channels with severe distortion can be achieved in several ways, one way is to design the transmit and receive filters so that the combination of filters and channel results in an acceptable error from the combination of intersymbol interference and noise; and the other way is designing an equalizer in the receiver that counteracts the channel distortion. The second method is the most commonly used technology for data transmission applications.

Figure 3 shows an adaptive channel equalizer, the received signal $y(n)$ is different from the original signal $x(n)$ because it was distorted by the overall channel transfer function $C(z)$, which includes the transmit filter, the transmission medium, and the receive filter.

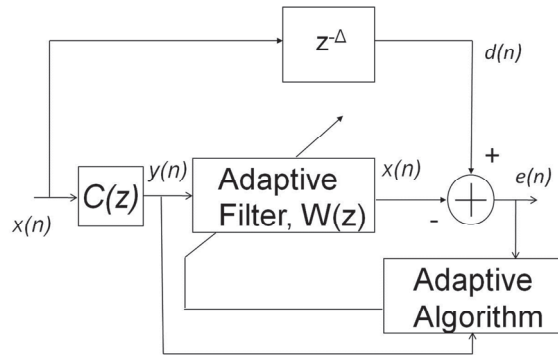


Fig. 3. Adaptive Channel equalizer

To recover the original signal $x(n)$, $y(n)$ must be processed using the equalizer $W(z)$, which is the inverse of the channel's transfer function $C(z)$ in order to compensate for the channel distortion. Therefore the equalizer must be designed by

$$W(z) = \frac{1}{C(z)} \quad (6)$$

In practice, the telephone channel is time varying and is unknown in the design stage due to variations in the transmission medium. Thus it is needed an adaptive equalizer that provides precise compensation over the time-varying channel. The adaptive filter requires the desired signal $d(n)$ for computing the error signal $e(n)$ for the LMS algorithm. An adaptive filter requires the desired signal $d(n)$ for computing the error signal $e(n)$ for the LMS algorithm.

The delayed version of the transmitted signal $x(n - \Delta)$ is the desired response for the adaptive equalizer $W(z)$. Since the adaptive filter is located in the receiver, the desired signal generated by the transmitter is not available at the receiver. The desired signal may be generated locally in the receiver using two methods. During the training stage, the adaptive equalizer coefficients are adjusted by transmitting a short training sequence. This known transmitted sequence is also generated in the receiver and is used as the desired signal $d(n)$ for the LMS algorithm.

After the short training period, the transmitter begins to transmit the data sequence. In the data mode, the output of the equalizer $x(n)$ is used by a decision device to produce binary data. Assuming that the output of the decision device is correct, the binary sequence can be used as the desired signal $d(n)$ to generate the error signal for the LMS algorithm.

5. Jammer suppression

Adaptive filtering can be a powerful tool for the rejection of narrowband interference in a direct sequence spread spectrum receiver. Figure 4 illustrates a jammer suppression system. In this case the output of the filter $y(n)$, is an estimate of the jammer, this signal is subtracted from the received signal $x(n)$, to yield an estimate of the spread spectrum.

To enhance the performance of the system a two-stage jammer suppressor is used. The adaptive line enhancer, which is essentially another adaptive filter, counteracts the effects of finite correlation which leads to partial cancellation of the desired signal. The number of coefficients required for either filter is moderate, but the sampling frequency may be well over 400 KHz.

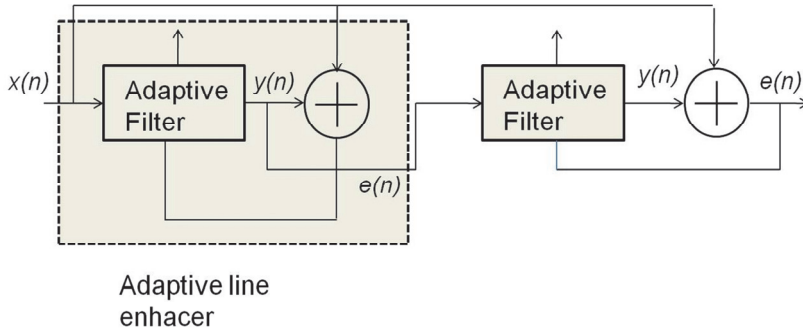


Fig. 4. Jammer suppression in direct sequence spread spectrum receiver

6. Adaptive notch filter

In certain situations, the primary input is a broadband signal corrupted by undesired narrowband (sinusoidal) interference. The conventional method of eliminating such sinusoidal interference is using a notch filter that is tuned to the frequency of the interference (Kuo et al., 2006). To design the filter, we need the precise frequency of the interference. The adaptive notch filter has the capability to track the frequency of the interference, and thus is especially useful when the interfering sinusoid drifts in frequency. A single-frequency adaptive notch filter with two adaptive weights is illustrated in figure 5, where the input signal is a cosine signal as

$$x(n) = x_0(n)A\cos(\omega_0 n) \quad (7)$$

A 90° phase shifter is used to produce the quadrature signal

$$x_1(n) = A\sin(\omega_0 n) \quad (8)$$

For a sinusoidal signal, two filter coefficients are needed. The reference input is used to estimate the composite sinusoidal interfering signal contained in the primary input $d(n)$.

The center frequency of the notch filter is equal to the frequency of the primary sinusoidal noise. Therefore, the noise at that frequency is attenuated. This adaptive notch filter provides a simple method for eliminating sinusoidal interference.

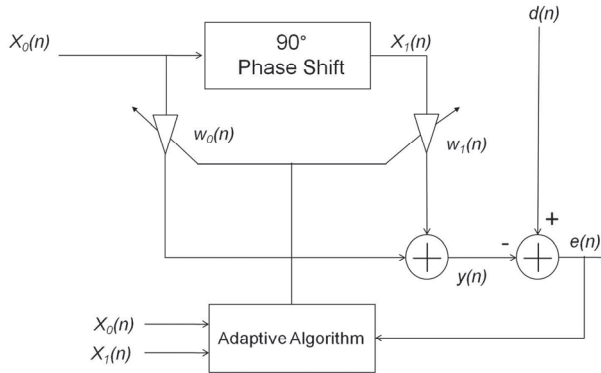


Fig. 5. Adaptive Notch Filter

7. Noise canceller

The noise cancellers are used to eliminate intense background noise. This configuration is applied in mobile phones and radio communications, because in some situations these devices are used in high-noise environments. Figure 6 shows an adaptive noise cancellation system.

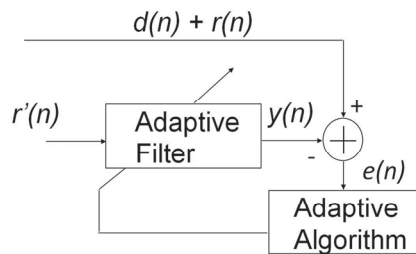


Fig. 6. Adaptive noise canceller system

The canceller employs a directional microphone to measure and estimate the instantaneous amplitude of ambient noise $r'(n)$, and another microphone is used to take the speech signal which is contaminated with noise $d(n) + r(n)$. The ambient noise is processed by the adaptive filter to make it equal to the noise contaminating the speech signal, and then is subtracted to cancel out the noise in the desired signal. In order to be effectively the ambient noise must be highly correlated with the noise components in the speech signal, if there is no access to the instantaneous value of the contaminating signal, the noise cannot be cancelled out, but it can be reduced using the statistics of the signal and the noise process. Figure 7 shows a voice signal with noise; those signals were used in noise canceller system implemented on a digital signal processor. The desired signal is a monaural audio signal with sampling frequency of 8 KHz. The noise signal is an undesired monaural musical piece with a sampling frequency of

11 KHz. As it can be seen in the image the desired signal is highly contaminated, so in this structure it must be used a fast adaptation algorithm in order to reach the convergence and eliminate all the unwanted components from the desired signal.

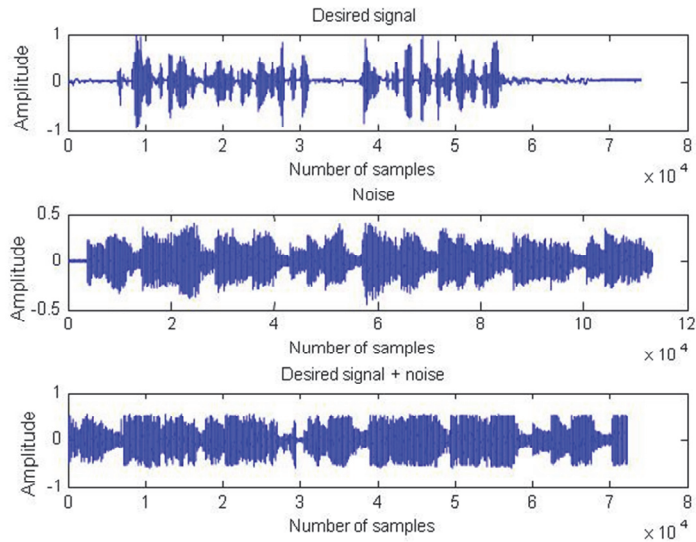


Fig. 7. Signals used in the noise canceller system

The frequency analysis of the signals used in the noise canceller system can be seen on the spectrograms of the figure 8. The figure shows that the output signal has some additional frequency components with respect to the input signal.

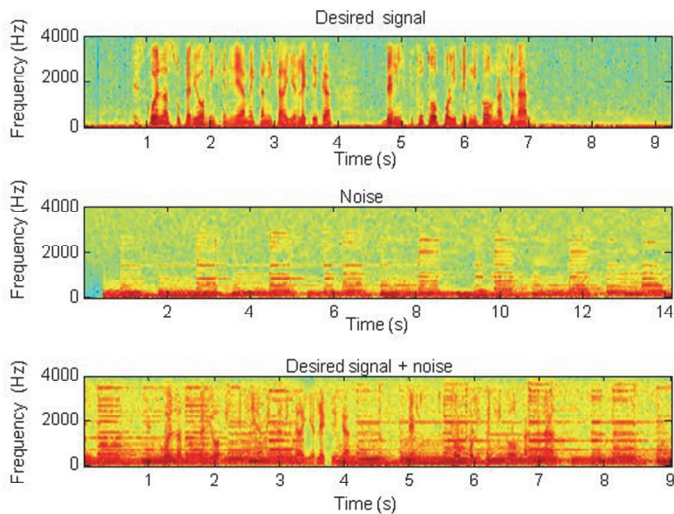


Fig. 8. Spectrograms of the signals used in the noise canceller system

The output of the noise canceller is the error signal, the figure 9 shows the error signal obtained when it is used an LMS algorithm. With the spectrogram of the signal it is shown that all the undesired frequency components were eliminated.

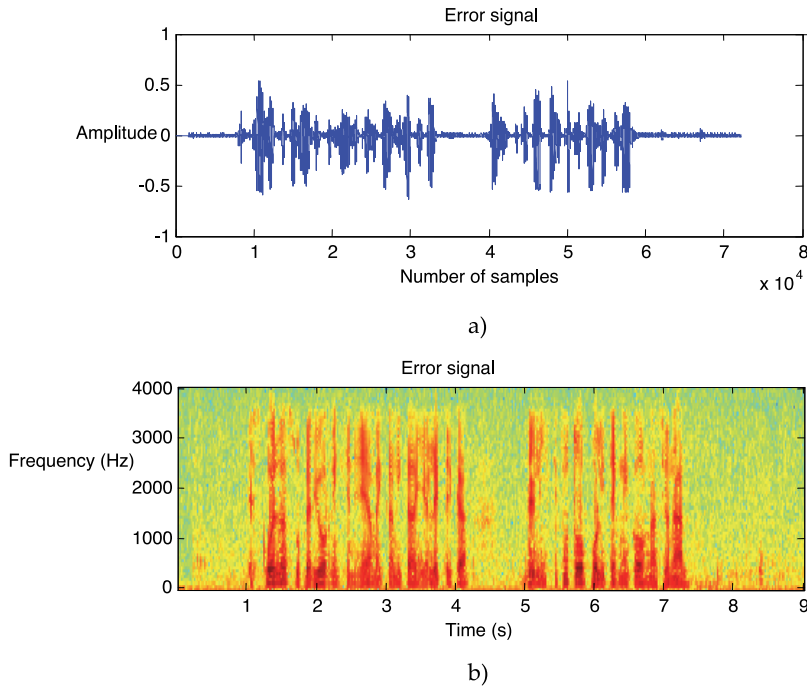


Fig. 9. a) Time waveform of the output signal b) Spectrogram of the output signal

The adaptive noise canceller system is used in many applications of active noise control (ANC), in aircrafts is used to cancel low-frequency noise inside vehicle cabins for passenger comfort. Most major aircraft manufacturers are developing such systems, mainly for noisy propeller-driven airplanes. In the automobile industry there are active noise cancellation systems designed to reduce road noise using microphones and speakers placed under the vehicle's seats.

Another application is active mufflers for engine exhaust pipes, which have been in use for a while on commercial compressors, generators, and such. With the price for ANC solutions dropping, even automotive manufacturers are now considering active mufflers as a replacement of the traditional baffled muffler for future production cars. The resultant reduction in engine back pressure is expected to result in a five to six percent decrease in fuel consumption for in-city driving.

Another application that has achieved widespread commercial success are active headphones to cancel low-frequency noise. The active headphones are equipped with microphones on outside of the ear cups that measure the noise arriving at the headphones. This noise is then being cancelled by sending the corresponding "anti-noise" to the headphones' speakers. For feedforward ANC, the unit also includes a microphone inside each ear cup to monitor the error - the part of the signal that has not been canceled by the

speakers in order to optimize the ANC algorithm. Very popular with pilots, active headphones are considered essential in noisy helicopters and propeller-powered airplanes.

7.1 Echo cancellation

In telecommunications, echo can severely affect the quality and intelligibility of voice conversation in telephone, teleconference or cabin communication systems. The perceived effect of an echo depends on its amplitude and time delay. In general, echoes with appreciable amplitudes and a delay of more than 1 ms can be noticeable. Echo cancellation is an important aspect of the design of modern telecommunications systems such as conventional wire-line telephones, hands-free phones, cellular mobile (wireless) phones, teleconference systems and in-car cabin communication systems.

In transmission networks the echoes are generated when a delayed and attenuated version of the signal sent by the local emitter to the distant receiver reaches the local receiver. These echo signals have their origin in the hybrid transformers which perform the two/four-wire conversion, in the impedance mismatches along the two-wire lines, and in some cases in acoustic couplings between loudspeakers and microphones in the subscriber sets.

The echo cancellation consists in modelling these unwanted couplings between local emitters and receivers and subtracting a synthetic echo from the real echo. According to the nature of the signals involved, the system will work as echo data canceller or voice echo canceller.

7.1.2 Voice echo canceller

Due to the characteristics of the speech signal, the voice echo cancellation system is somewhat different from the data echo canceller. The speech is a high level nonstationary signal, and due to the signal bandwidth and the velocity of the acoustic waves in the open air, the filters must have a very long number of coefficients. Also in order to reach a high level of performance and meet the expectations of the user, the voice echo canceller may have several other functions, like speech detection and denoising.

Figure 10 illustrates the operation of an adaptive line echo canceller. The speech signal on the line from speaker A to speaker B is input to the four/two-wire hybrid B and to the echo canceller. The echo canceller monitors the signal on line from B to A and attempts to model the echo path and synthesise a replica of the echo of speaker A. This replica is used to subtract and cancel out the echo of speaker A on the line from B to A. The echo canceller is basically an adaptive linear filter. The coefficients of the filter are adapted so that the energy of the signal on the line is minimised.

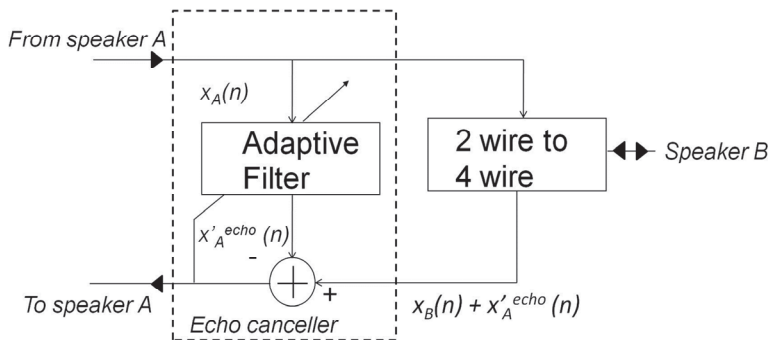


Fig. 10. Adaptive echo cancellation system

Assuming that the signal of the line from speaker B to speaker A, $y_B(n)$, is composed of the speech of speaker B, $x_B(n)$, plus the echo of speaker A, $x_A^{echo}(n)$,

$$y_B(n) = x_B(n) + x_A^{echo}(n) \quad (9)$$

Speech and echo signals are not simultaneously present on a phone line unless both speakers are speaking simultaneously. Assuming that the truncated impulse response of the echo path is modelled by an FIR filter, the output estimate of the synthesised echo signal can be expressed as

$$x_A'^{echo}(n) = \sum_{l=0}^P h_l(n) x_A(n-l) \quad (10)$$

Where $h_l(n)$ are the time varying coefficients of an adaptive FIR filter model of the echo path and $x_A'^{echo}(n)$ is an estimate of the echo of speaker A on the line from speaker B to speaker A. The residual echo signal, or the error signal, after echo subtraction is given by

$$e(n) = y_B(n) - x_A'^{echo}(n) = x_B(n) + x_A^{echo}(n) - \sum_{l=0}^P h_l(n) x_A(n-l) \quad (11)$$

For those time instants when speaker A is talking and speaker B is listening and silent, and only echo is present from line B to A, we have

$$e(m) = x_A'^{echo}(n) = x_A^{echo}(n) - x_A'^{echo}(n) = x_A^{echo}(n) - \sum_{l=0}^P h_l(n) x_A(n-l) \quad (12)$$

Where $x_A'^{echo}(n)$ is the residual echo.

In some cases it may happen the double talk situation, in this case both users talk at the same time, and simultaneous bidirectional transmission takes place. In this way it could be produced misalignment of the coefficients and a drop in echo attenuation, one way to solve this problem is holding the coefficients during double talk, but for this it is needed a double-talk detector. The performance of double-talk detectors is crucial for the comfort of the users.

7.1.3 Data echo canceller

Echo cancellation becomes more complex with the increasing integration of wireline telephone systems and mobile cellular systems, and the use of digital transmission methods such as asynchronous transfer mode (ATM) for integrated transmission of data, image and voice.

Those systems use full-duplex transmission data signals that are transmitted simultaneously in two directions and in the same frequency bands, meanwhile in half-duplex transmission just one direction are used at a time. The figure 11 shows the principle of full-duplex transmission. The signal $x_A(N)$ is sent from terminal A to terminal B through a two wire line. The signal $y(n)$ at the input of the receiver of terminal A consists of two components, a signal from the terminal B ($y_B(n)$), which is the useful data signal, and the returned unwanted echo generated from $x_A(n)$. $H(z)$ is a filter that is going to generate a synthetic echo $y'(n)$ as close as possible to $x_A(n)$, after subtraction, the output error $e(n)$ is kept sufficiently close to $y_B(n)$ to make the transmission of data from terminal B to terminal A satisfactory.

The number of coefficients (N) of the adaptive filter is derived from the duration of the echo impulse response that has to be compensated, taking into account the sampling frequency. In order to calculate the number of coefficients we could use

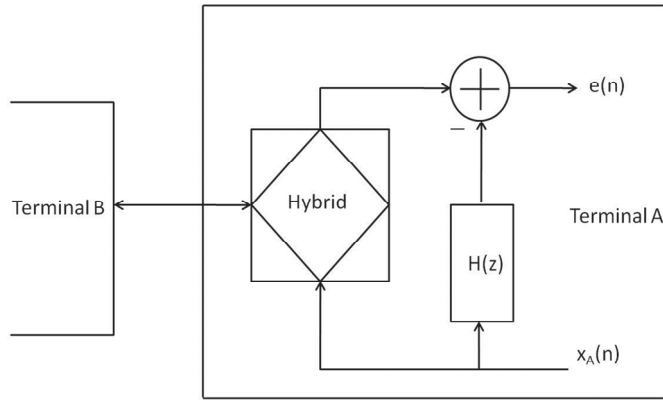


Fig. 11. Echo cancellation for full-duplex transmission

$$N = (2D/v) f_s \quad (1)$$

Where N is the number of coefficients, D is the length of the line, v is the electrical signal velocity over the subscriber line and f_s is the sampling frequency (Bellanger, 2001). Since the characteristics of the transmission line may change with time it is necessary to implement an adaptive filter.

7.1.4 Acoustic echo

Acoustic echo results from a feedback path set up between the speaker and the microphone in a mobile phone, hands-free phone, teleconference or hearing aid system. Acoustic echo is reflected from a multitude of different surfaces, such as walls, ceilings and floors, and travels through different paths. If the time delay is not too long, then the acoustic echo may be perceived as a soft reverberation, and may add to the artistic quality of the sound; concert halls and church halls with desirable reverberation characteristics can enhance the quality of a musical performance.

Acoustic echo can result from a combination of direct acoustic coupling and multipath effect where the sound wave is reflected from various surfaces and then picked up by the microphone. In its worst case, acoustic feedback can result in howling if a significant proportion of the sound energy transmitted by the loudspeaker is received back at the microphone and circulated in the feedback loop.

The most effective method of acoustic feedback removal is the use of an adaptive feedback cancellation system (AFC). Fig. 12 illustrates a model of an acoustic feedback environment, comprising a microphone, a loudspeaker and the reverberating space of a room (Vaseghi, 2006). The z transfer function of a linear model of the acoustic feedback environment may be expressed as

$$H(z) = \frac{G(z)}{1 - G(z)A(z)} \quad (13)$$

Where $G(z)$ is the z transfer function model for the microphone loudspeaker system and $A(z)$ is the z transfer function model of reverberations and multipath reflections of a room environment. Assuming that the microphone loudspeaker combination has a flat frequency response with a gain G , the equation can be simplified to

$$H(z) = \frac{G}{1 - GA(z)} \quad (14)$$

Owing to the reverberation character of the room, the acoustic feedback path $A(z)$ is itself a feedback system. The reverberating characteristics of the acoustic environment may be modelled by an all-pole linear predictive model, or alternatively a relatively long FIR model. The equivalent time-domain input/output relation for the linear filter model of equation (4) is given by the following difference equation

$$y(n) = \sum_{l=0}^P a_l(n)y(n-l) + Gx(n) \quad (15)$$

Where $a_l(n)$ is the coefficient of an all pole linear feedback model of the reverberating room environment, G is the microphone loudspeaker amplitude gain factor, and $x(n)$ and $y(n)$ are the time domain input and output signals of the microphone loudspeaker system.

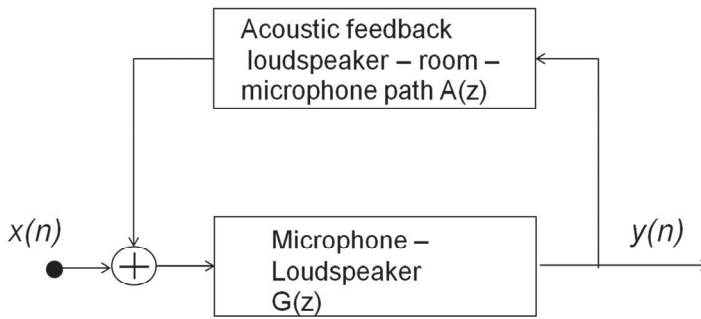


Fig. 12. Acoustic feedback model

The most successful acoustic feedback control systems are based on adaptive estimation and cancellation of the feedback signal. As in a line echo canceller, an adaptive acoustic feedback canceller attempts to synthesise a replica of the acoustic feedback. The problem of acoustic echo cancellation is more complex than line echo cancellation for a number of reasons. First, acoustic echo is usually much longer (up to a second) than terrestrial telephone line echoes. In fact, the delay of an acoustic echo is similar to or more than a line echo routed via a geostationary satellite system. The large delay of an acoustic echo path implies that impractically large filters on the order of a few thousand coefficients may be required. An important application of acoustic feedback cancellation is in hearing aid systems.

7.1.5 Multiple-input multiple-output (MIMO) echo cancellation

Multiple-input multiple-output (MIMO) echo-cancellation systems have applications in car cabin communications systems, stereophonic teleconferencing systems and conference halls. Stereophonic echo cancellation systems have been developed relatively recently and MIMO systems are still the subject of ongoing research and development. In a typical MIMO system there are P speakers and Q microphones in the room. As there is an acoustic feedback path set up between each speaker and each microphone, there are altogether $P \times Q$ such acoustic feedback paths that need to be modelled and estimated. The truncated impulse response of each acoustic path from loudspeaker i to microphone j is modelled by an FIR filter h_{ij} . The truncated impulse response of each acoustic path from a human speaker

i to microphone j is modelled by an FIR filter, g_{ij} . For a large number of speakers and microphones, the modelling and identification of the numerous acoustic channels becomes a major problem due to the correlations of the echo signals, from a common number of sources, propagating through different channels, as discussed below.

7.2 Adaptive feedback cancellation in hearing aids

The hearing-aid processing amplifies the input signal to compensate for the hearing loss of the users. When this amplification is larger than the attenuation of the feedback path, instability occurs and usually results in feedback whistling, which limits the maximum gain that can be achieved.

Acoustic feedback in hearing aids refers to the acoustical coupling between the loudspeaker (also known as the receiver) and the microphone of the hearing aid. Because of this coupling, the hearing aid produces a severe distortion of the desired signal and an annoying howling sound when the gain is increase.

If the Feedback transfer function was known, it can be compensated for in the hardware, but the problem here is the time variability of the dynamics, caused by a change in interference characteristics. Some possible causes of this problem are hugs or objects like a telephone coming close to the ear.

There are several techniques to reduce the negative effects introduced by acoustic feedback. They can be broadly classified into feedforward suppression and feedback cancellation techniques. In feedforward suppression techniques, the regular signal processing path of the hearing aid is modified in such a way that it is stable in conjunction with the feedback path.

The most common technique is the use of a notch filter. In a notch filter, the gain is reduced in a narrow frequency band around the critical frequencies whenever feedback occurs. Nevertheless feedforward suppression techniques all compromise the basic frequency response of the hearing aid, and, hence, may seriously affect the sound quality (Spriet et al., 2006). A more promising solution for acoustic feedback is the use of a feedback cancellation system.

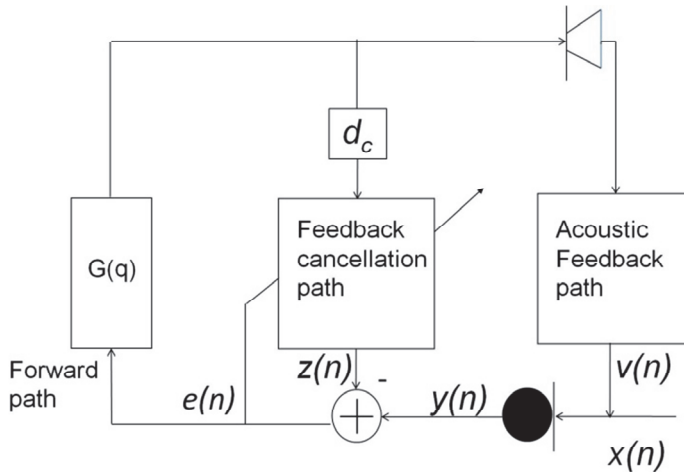


Fig. 13. Adaptive feedback canceller

Figure 13 illustrates an adaptive feedback canceller, which produces an estimate $z(n)$ of the feedback signal $v(n)$ and subtracts this estimate $z(n)$ from the microphone signal, so that, ideally, only the desired signal is preserved at the input of the forward path. Since the acoustic path between the loudspeaker and the microphone can vary significantly depending on the acoustical environment, the feedback canceller must be adaptive.

When the external input signal is correlated with the receiver input signal, the estimate of the feedback path is biased. This so-called “bias problem” results in a large modeling error and a cancellation of the desired signal (Ma, 2010).

7.3 Foetal monitoring, cancelling of maternal ECG during labour

Information derived from the foetal electrocardiogram (ECG), such as the foetal heart rate pattern, is valuable in assessing the condition of the baby before or during the childbirth. The ECG derived from electrodes placed on the mother’s abdomen is susceptible to contamination from much larger background noise (for example muscle activity and foetal motion) and the mother’s own ECG.

Considering the problem as an adaptive noise cancellation, where foetal ECG is a desired signal $d(n)$, corrupted by the maternal signal $r(n)$, a kind of additive noise. The measured foetal signal (MFECG(n)) from foetal lead can be expressed as

$$\text{MFECG}(n) = d(n) + r(n) \quad (16)$$

Another measurement MMECG(n) from maternal lead is given as a reference signal, that is correlated with $r(n)$ and uncorrelated with $d(n)$. MMECG can be used to estimate the noise $r(n)$ by minimizing the mean square error. Figure 14 shows the block diagram for the enhancement of foetal ECG.

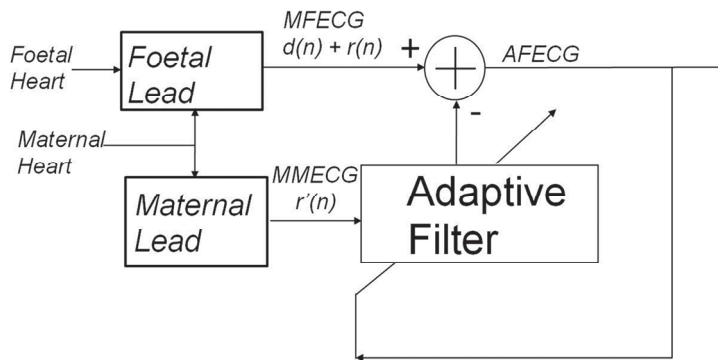


Fig. 14. Adaptive cancelling of maternal ECG in foetal ECG

An adaptive filter is used to estimate maternal components in measured foetal ECG (MFECG) from measured maternal ECG (MMECG). The estimated components then are subtracted from the MFECG to obtain adaptive filtered foetal ECG (AFECG), in which maternal components are suppressed. Other artefacts, such as muscular contraction from maternal body and foetal movement, will induce baseline drift in the MFECG (Chen et al., 2000).

7.4 Removal of ocular artifacts from electro-encephalogram by adaptive filtering

The eye forms an electric dipole, where the cornea is positive and the retina is negative. When the eye moves (saccade, blink or other movements), the electric field around the eye changes, producing an electrical signal known as the electro-oculogram (EOG). As this signal propagates over the scalp, it appears in the recorded electro-encephalogram (EEG) as noise or artifacts that present serious problems in EEG interpretation and analysis. There are at least two kinds of EOG artifact to be removed: those produced by the vertical eye movement (the corresponding EOG is called VEOG) and those produced by the horizontal eye movement (HEOG). Consequently, a noise canceller with two reference inputs is used in this application (He et al., 2004).

Fig. 15 shows the EOG noise canceller. The primary input to the system is the EEG signal $s(n)$, picked up by a particular electrode. This signal is modelled as a mixture of a true EEG $x(n)$ and a noise component $r(n)$. $v(n)$ and $v'(n)$ are the two reference inputs, VEOG and HEOG, respectively. $v(n)$ and $v'(n)$ are correlated, in some unknown way, with the noise component $r(n)$ in the primary input. The desired output from the noise canceller $e(n)$ is the corrected, or clean, EEG.

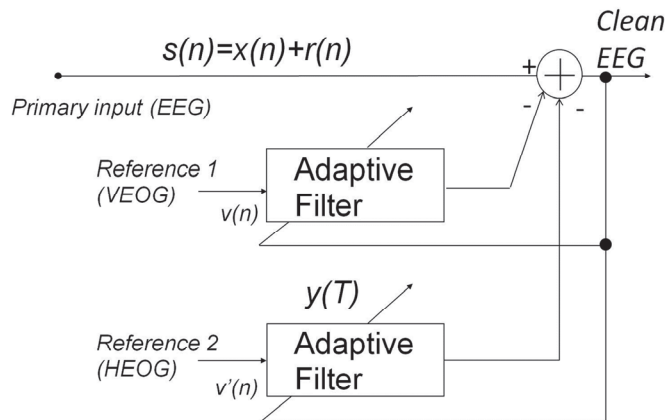


Fig. 15. EOG noise canceller

7.5 Application of adaptive noise cancelling filters in AC electrical measurements

Through adaptive noise cancellation it could be improved the ac electrical measurements. Often ac measurement circuits are influenced by noise caused by line frequency beat. The figure 16 shows a system that cancels the line frequency beat. An ADC is used to sample the suitably divided down line voltage in order to determine the phase relative to the signal channel, which is sampled with a second ADC. The phase data is used as the noise input to an adaptive noise-cancelling filter used to cancel the effect on the transconductance amplifier output data (Wright et al., 2010).

Another common interference in ac measurement circuits is the coupling of the magnetic field generated by a nearby source. In such situations it may be possible to use an adaptive interference cancelling system with a simple coil system to measure the ambient magnetic field that causes the unwanted interference and then remove this interference from data obtained from a measurement circuit.

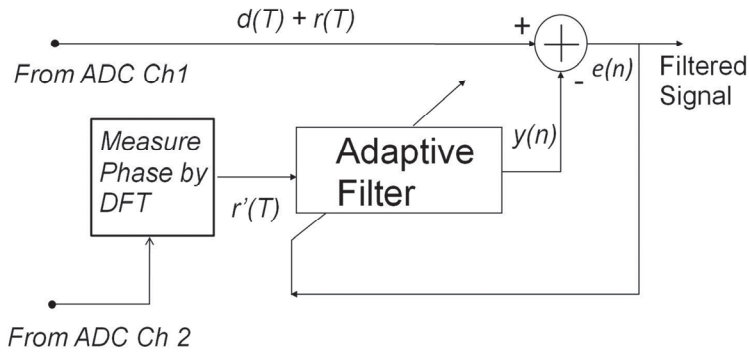


Fig. 16. Line beat Adaptive canceller.

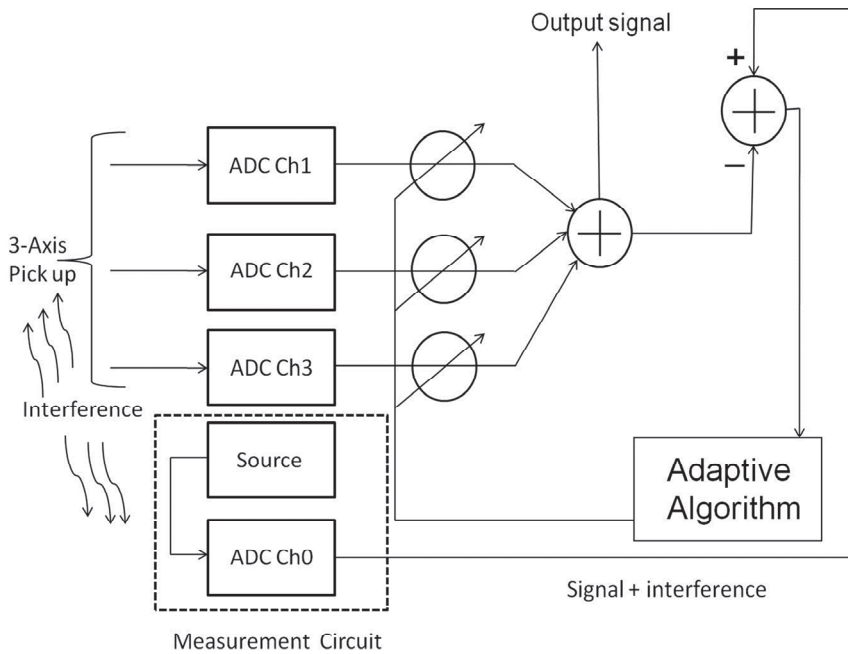


Fig. 17. Three Axis linear combiner for interference cancellation

Figure 17 shows a 3-axis magnetic field sensor which is connected to a separate analogue to digital converter (ADC). A fourth ADC is used to sample the "signal" simultaneously with the 3-axis data. The three "noise" channel ADCs are the inputs to the three channels of a three-way linear combiner (Wright et al., 2010).

8. Conclusion

In recent years, the development and commercial availability of increasingly powerful and affordable digital computers has been accompanied by the development of advanced digital signal processing algorithms for a wide variety of applications; therefore the use of adaptive filters is bigger every day.

Adaptive filters are used for estimation of nonstationary signals and systems, or in applications where a sample-by sample adaptation of a process and/or a low processing delay is required.

In this chapter, we described some of the most used adaptive filtering applications. The material presented here forms the basis to understand the behavior of most adaptive-filtering structures in practical implementations. The main objective was to illustrate how the adaptive-filtering is applied to solve practical problems.

The distinctive feature of each application is the way the adaptive filter input signal and the desired signal are chosen. Once these signals are determined, any known properties of them can be used to understand the expected behavior of the adaptive filter when attempting to minimize the chosen objective function. The efficiency of the adaptive filters mainly depends on the used technique of design and the algorithm of adaptation.

9. References

- Bellanger, M. (2001). *Adaptive Digital Filters*(Second edition). Marcel Dekker, ISBN 0-8247-0563-7, New York
- Chen, W. ; Nemoto, T. ; Kobayashi, T. ; Saito, T. ; Kasuya, E., & Honda, Y. (2000). ECG and Heart RateDetection of Prenatal Cattle Foetus Using Adaptive Digital Filtering, *Engineering in Medicine and Biology Society, 2000. Proceedings of the 22nd Annual International Conference of the IEEE*, Volume 2, (July 2000), pp. 1-23
- Diniz, P. (2008). *Adaptive Filtering, Algorithms and Practical Implementation* (Third edition). Springer, ISBN 978-0-387-31274-3, Rio de Janeiro
- He, P. ; Wilson, G., & Rusell, C. (2004). Removal of ocular artifacts from electro-encephalogram by adaptive filtering, *Medical & Biological Engineering & Computing*, Volume 42, Issue 3, (May 2004), pp. 407-412
- Kuo, S. ; Lee, B., & Tian, W. (2006). *Real Time Digital Signal Processing* (Second edition). John Wiley & Sons Ltd, ISBN 0-470-01495-4, West Sussex
- Ma, G.; Gran, F. ; Jacobsen, F., & Agerkvist, F., (2010). Adaptive feedback cancellation with band-limited LPC vocoder in digital hearing aids , *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. pp, Issue 99, (July 2010), pp. 1, ISSN 1558-7916
- Spriet, A ; Rombouts, G. ; Moonen, M., & Wouters, J. (2006). Adaptive Feedback Cancellation in hearing aids, *Journal of the Franklin Institute* Volume 343, Issue 6, (September 2006), pp. 545-573
- Shoval, A. ; Johns, D., & Snelgrove, W. (1995). Comparison of DC Offset Effects in Four LMS Adaptive Algorithms, *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, Volume 42, No. 3, (March 1995), pp. 176- 185.
- Vaseghi, S. (2006). *Advanced Digital Signal Processing and noise reduction* (Third edition). John Wiley & Sons Ltd, ISBN 0-470-09494-X, West Sussex.

Wright, P.S. ; Clarkson, P., & Hall, M. J. (2010). Application of adaptive noise cancelling filters in ac electrical measurements, *2010 Conference on Precision Electromagnetic Measurements (CPEM)*, ISBN 978-1-4244-6795-2, Daejeon, June 2010

Applications of Adaptive Filtering: Recent Advancements in Active Noise Control

Akhtar Muhammad Tahir¹, Mitsuhashi Wataru² and Nishihara Akinori³

^{1,2}*The University of Electro-Communications, Tokyo*

³*Tokyo Institute of Technology, Tokyo
Japan*

1. Introduction

ACTIVE NOISE CONTROL (ANC) is based on the simple principle of destructive interference of propagating acoustic waves, (Elliot, 2001). The basic idea was proposed in 1936 (Lueg, 1936), however, real applications were quite limited till recently. Thanks to advancement in the algorithms for adaptive signal processing and their implementation using digital signal processors (DSPs); many successful applications of ANC have been reported, the most famous being noise reduction headsets (Gan & Kuo, 2002; Kuo et al., 2006).

The most popular adaptive algorithm used for ANC applications is the filtered-x least mean square (FxLMS) algorithm (Kuo & Morgan, 1996) which is a modified version of the LMS algorithm (Widrow & Stearns, 1985). The FxLMS algorithm is computationally simple, but its convergence speed is slow. Different ANC algorithms, with improved convergence properties, have been proposed, viz., 1) lattice-ANC systems (Park & Sommerfeldt, 1996); 2) infinite impulse response (IIR) filter-based LMS algorithms called filtered-u recursive LMS (FuRLMS) (Eriksson et al., 1987), and filtered-v algorithms (Crawford & Stewart, 1997); 3) recursive least squares (RLS) based algorithms called filtered-x RLS (FxRLS) (Kuo & Morgan, 1996) and filtered-x fast-transversal-filter (FxFTF) (Bouchard & Quednau, 2000); and 4) frequency-domain-ANC systems (see (Kuo & Tahernezehadi, 1997) and references there in). There are the following problems with these approaches: 1) IIR-based structures have inherent stability problems; 2) other approaches mentioned above increase the computational burden substantially; and 3) RLS-based ANC systems have numerical instability problems. These reasons make FxLMS still a good choice for ANC applications, and hence, in this chapter we describe various concepts and methods using FxLMS algorithm.

The main objective of this chapter is to provide a comprehensive review of adaptive filtering algorithms developed and employed for ANC systems. We also provide some recent results for two challenging problems: ANC of impulsive-like noise sources, and mitigating effect of the uncorrelated disturbances for which a correlated reference signal is not available. We see that simple modifications and extensions of the existing algorithms and methods improve robustness of the ANC systems.

The outline of the chapter is as follows. **Section 2** details **FxLMS algorithm** for feedforward and feedback type ANC systems. It also highlights signal processing issues and open problems for further research. **Section 3** describes development of various adaptive algorithm for **ANC for Impulsive Noise Sources**, and **Section 4** addresses issue of **Mitigating**

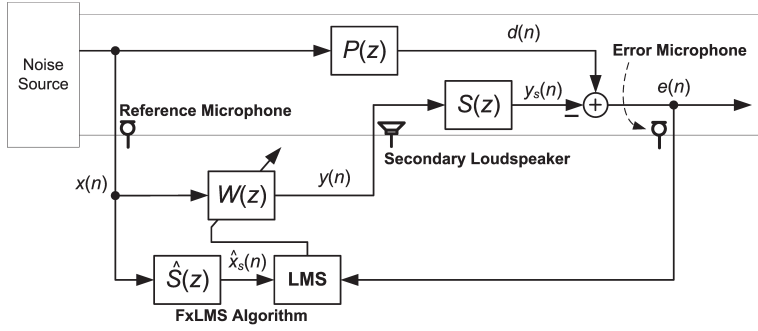


Fig. 1. Block diagram of FxLMS algorithm-based single-channel feedforward ANC system.

Uncorrelated Disturbance appearing at the error microphone of feedforward ANC system. **Section 5** presents results of **Computer Simulations** for two case studies discussed in this chapter, viz., ANC for impulsive sources, and mitigating effect of uncorrelated disturbance. **Section 6** is an **Outlook on Recent ANC Applications** and **Section 7** gives the **Concluding Remarks**.

2. FxLMS algorithm

In this section we give description of FxLMS algorithm for single-channel feedforward and feedback type ANC systems. Furthermore, a brief review on various signal processing issues, solved and unsolved, is also detailed.

2.1 Feedforward ANC

The block diagram for a single-channel feedforward ANC system using the FxLMS algorithm is shown in Fig. 1, where $P(z)$ is primary acoustic path between the reference noise source and the error microphone. The reference noise signal $x(n)$ is filtered through $P(z)$ and appears as a primary noise signal at the error microphone. The objective of the adaptive filter $W(z)$ is to generate an appropriate antinoise signal $y(n)$ propagated by the secondary loudspeaker. This antinoise signal combines with the primary noise signal to create a zone of silence in the vicinity of the error microphone. The error microphone measures the residual noise $e(n)$, which is used by $W(z)$ for its adaptation to minimize the sound pressure at error microphone. Here $\hat{S}(z)$ accounts for the model of the secondary path $S(z)$ between the output $y(n)$ of the controller and the output $e(n)$ of the error microphone. The filtering of the reference signal $x(n)$ through $\hat{S}(z)$ is demanded by the fact that the output $y(n)$ of the adaptive filter is filtered through $S(z)$ (Kuo & Morgan, 1996).

Assuming that $W(z)$ is an FIR filter of tap-weight length L_w , the secondary signal $y(n)$ is expressed as

$$y(n) = \mathbf{w}^T(n) \mathbf{x}(n). \quad (1)$$

where

$$\mathbf{w}(n) = [w_0(n), w_1(n), \dots, w_{L_w-1}(n)]^T \quad (2)$$

is the tap-weight vector, and

$$\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-L_w+1)]^T \quad (3)$$

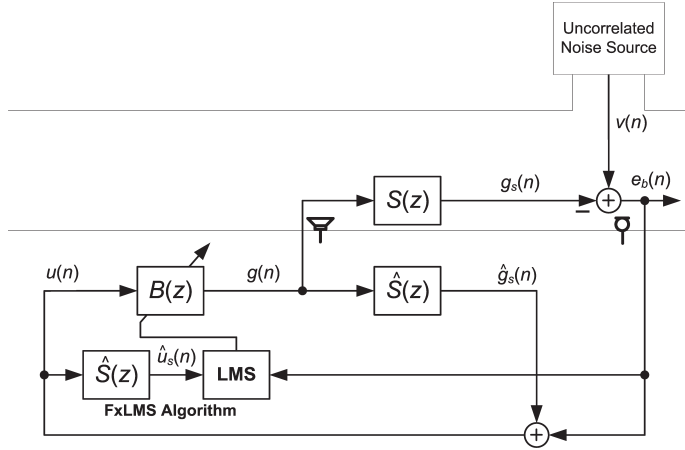


Fig. 2. Block diagram of FxLMS algorithm-based single-channel feedback ANC systems.

is an L_w -sample vector the reference signal $x(n)$. The residual error signal $e(n)$ is given as

$$e(n) = d(n) - y_s(n) \quad (4)$$

where $d(n) = p(n) * x(n)$ is the primary disturbance signal, $y_s(n) = s(n) * y(n)$ is the secondary canceling signal, $*$ denotes linear convolution, and $p(n)$ and $s(n)$ are impulse responses of the primary path $P(z)$ and secondary path $S(z)$, respectively.

Minimizing the mean squared error (MSE) cost function; $J(n) = E\{e^2(n)\} \approx e^2(n)$, where $E\{\cdot\}$ is the expectation of quantity inside; the FxLMS update equation for the coefficients of $W(z)$ is given as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu_w e(n) \hat{\mathbf{x}}_s(n) \quad (5)$$

where μ_w is the step size parameter,

$$\hat{\mathbf{x}}_s(n) = [\hat{x}_s(n), \hat{x}_s(n-1), \dots, \hat{x}_s(n-L_w+1)]^T \quad (6)$$

is filtered-reference signal vector being generated as

$$\hat{\mathbf{x}}_s(n) = \hat{s}(n) * \mathbf{x}(n), \quad (7)$$

where $\hat{s}(n)$ is impulse response of the secondary path modeling filter $\hat{S}(z)$.

2.2 Feedback ANC

The feedforward strategy as described above is widely used in ANC systems, where an independent reference signal $x(n)$ is available and is well correlated with the primary noise $d(n)$. Whenever the reference signal related to the primary noise source is unavailable or several reference signals are in the enclosure, the use of feedforward control becomes impractical. Under such circumstances, feedback control may be envisaged, in which measured residual error signals are used to derive the secondary sources. The block diagram for feedback ANC system is shown in Fig. 2, where $v(n)$ represents a noise source for which a correlated reference signal is not available. As shown, the feedback ANC system comprises

only error microphone and secondary loudspeaker. The output $g(n)$ of the feedback ANC $B(z)$ passes through $S(z)$ to generate the residual error signal $e_b(n)$ as

$$e_b(n) = v(n) - g_s(n), \quad (8)$$

where $g_s(n) = s(n) * g(n)$ is the cancelling signal for $v(n)$. The residual error signal $e_b(n)$ is picked by the error microphone and is used in the adaptation of the FxLMS algorithm for $B(z)$. The reference signal for $B(z)$ is internally generated by filtering $g(n)$ through secondary path model $\hat{S}(z)$ and adding it to the residual error signal $e_b(n)$ as

$$u(n) = e_b(n) + \hat{g}_s(n) = [v(n) - g_s(n)] + \hat{g}_s(n), \quad (9)$$

where $\hat{g}_s(n) = \hat{s}(n) * g(n)$ is the estimate of cancelling signal $g_s(n)$. Assuming that the secondary path is perfectly identified; which can be obtained by using offline (Kuo & Morgan, 1996) and/or online modeling techniques (Akhtar et al., 2005; 2006); $\hat{g}_s(n) \approx g_s(n)$, and hence Eq. (9) simplifies to give estimate of uncorrelated noise source as $u(n) \rightarrow v(n)$. Using this internally generated reference signal¹, the output $g(n)$ of feedback ANC $B(z)$ is computed as

$$g(n) = \mathbf{b}^T(n) \mathbf{u}(n). \quad (10)$$

where

$$\mathbf{b}(n) = [b_0(n), b_1(n), \dots, b_{L_b-1}(n)]^T \quad (11)$$

is the tap-weight vector for $B(z)$,

$$\mathbf{u}(n) = [u(n), u(n-1), \dots, u(n-L_b+1)]^T \quad (12)$$

is the corresponding reference signal vector for $u(n)$, and L_b is the tap-weight length of $B(z)$. Finally the FxLMS algorithm for updating $B(z)$ is given as

$$\mathbf{b}(n+1) = \mathbf{b}(n) + \mu_b e_b(n) \hat{\mathbf{u}}_s(n) \quad (13)$$

where μ_b is the step size parameter for $B(z)$, and filtered-reference signal vector $\hat{\mathbf{u}}_s(n) = [\hat{u}_s(n), \hat{u}_s(n-1), \dots, \hat{u}_s(n-L_b+1)]^T$ is generated as

$$\hat{\mathbf{u}}_s(n) = \hat{s}(n) * \mathbf{u}(n). \quad (14)$$

In feedback ANC, hence, the basic idea is to estimate the primary noise $v(n)$, and use it as a reference signal $u(n)$ for the feedback ANC filter $B(z)$. It is worth mentioning that the feedforward ANC provides wider control bandwidth within moderate controller gain than the feedback ANC, whereas feedback ANC gives significant performance for narrowband or predictable noise sources.

2.3 Review on signal processing challenges

The FxLMS algorithm appears to be very tolerant of errors made in the modeling of $S(z)$ by the filter $\hat{S}(z)$. As shown in (Elliott et. al., 1987; Morgan, 1980), with in the limit of slow

¹ This is why FxLMS algorithm for feedback ANC systems is sometimes referred as internal model control (Kuo & Morgan, 1996)

adaptation, the algorithm will converge with nearly 90° of phase error between $\hat{S}(z)$ and $S(z)$. Therefore, offline modeling can be used to estimate $S(z)$ during an initial training stage for ANC applications (Kuo & Morgan, 1999). For some applications, however, the secondary path may be time varying, and it is desirable to estimate the secondary path online when the ANC is in operation (Saito & Sone, 1996).

There are two different approaches for online secondary path modeling. The first approach, involving the injection of additional random noise into the ANC system, utilizes a system identification method to model the secondary path. The second approach attempts to model it from the output of the ANC controller, thus avoiding the injection of additional random noise into the ANC system. A detailed comparison of these two online modeling approaches can be found in (Bao et al., 1993a), which concludes that the first approach is superior to the second approach on convergence rate, speed of response to changes of primary noise, updating duration, computational complexities, etc.

The basic additive random noise technique for online secondary path modeling in ANC systems is proposed by (Eriksson & Allie, 1989). This ANC system comprises two adaptive filters; FxLMS algorithm based noise control filter $W(z)$, and LMS algorithm based secondary path modeling filter $\hat{S}(z)$. Improvements in the Eriksson's method have been proposed in (Bao et al., 1993b; Kuo & Vijayan, 1997; Zhang et al., 2001). These improved methods introduce another adaptive filter into the ANC system of (Eriksson & Allie, 1989), which results in increased computational complexity. The methods proposed in (Akhtar et al., 2005; 2006) suggest modifications to Eriksson's method such that improved performance is realized without introducing a third adaptive filter. The development of robust and efficient online secondary path modeling algorithm, without requiring additive random noise, is critical and demands further research.

The feedforward ANC system shown in Fig. 1 uses the reference microphone to pick up the reference noise $x(n)$, processes this input with an adaptive filter to generate an antinoise $y(n)$ to cancel primary noise acoustically in the duct, and uses an error microphone to measure the error $e(n)$ and to update the adaptive filter coefficients. Unfortunately, a loudspeaker on a duct wall will generate the antinoise signal propagating both upstream and downstream. Therefore, the antinoise output to the loudspeaker not only cancels noise downstream, but also radiates upstream to the reference microphone, resulting in a corrupted reference signal $x(n)$. This coupling of acoustic waves from secondary loudspeaker to the reference microphone is called *acoustic feedback*. One simple approach to neutralize the effect of acoustic feedback is to use a separate feedback path modeling filter within the controller. This electrical model of the feedback path is driven by the antinoise signal, $y(n)$, and its output is subtracted from the reference sensor signal, $x(n)$. The feedback path modeling filter may be obtained offline prior to the operation of ANC system when the reference noise $x(n)$ does not exist. In many practical cases, however, $x(n)$ always exists, and feedback may be time varying as well. For these cases, online modeling of feedback path is needed to ensure the convergence and stability of the FxLMS algorithm for ANC systems. For a detailed review on existing signal processing methods and various other techniques for feedback neutralization in ANC systems, the reader is referred to (Akhtar et al., 2007) and references therein.

In the case of narrowband noise sources with signal energy being concentrated at a few representative harmonics, the reference microphone in Fig. 1 can be replaced with a non-acoustic sensor, e.g., a tachometer in the case of rotating machines. The output from non-acoustic sensor is used to internally generate the reference signal, which may be an impulse train with a period equal to the inverse of the fundamental frequency of periodic

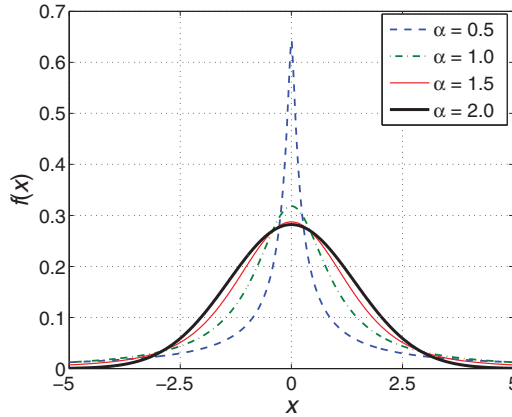


Fig. 3. The PDFs of standard symmetric α -stable (S α S) process for various values of α .

noise, or sinusoids that have the same frequencies as the corresponding harmonic components (Kuo & Morgan, 1996). Essentially, a narrowband ANC system would assume the reference signal $x(n)$ has the same frequency as the primary noise $d(n)$ at the error microphone. In many practical situations, the reference sinusoidal frequencies used by the adaptive filter may be different than the actual frequencies of primary noise. This difference is referred to as frequency mismatch (FM), and will degrade the performance of ANC systems. The effects of FM and solution to the problems have been recently studied in (Jeon et al., 2010; Kuo & Puvvala, 2006; Xiao et al., 2005; 2006).

Another signal processing challenge is ANC for sources with nonlinear behavior. It has been demonstrated that the FxLMS algorithm gives very poor performance in the case of nonlinear processes (Strauch & Mulgrew, 1998). For efficient algorithms for ANC of non linear source, see (Reddy et al., 2008) and references there in.

In many practical situations, it is desirable to shift the quiet zone away from the location of error microphones to a virtual location where error microphone cannot be installed (Bonito et al., 1997). One interesting example is recently investigated snore ANC system, where headboard of bed is mounted with loudspeakers and microphones (Kuo et al., 2008). In this case, the error microphone cannot be placed at the ears of the bed partner, where maximum cancellation is required, and hence an efficient virtual sensing technique is required to improve the noise reduction around ears using error microphones installed on the headboard. There has been a very little research on active control of moving noise sources. It is obvious that acoustic paths will be highly time varying in such cases, and hence the optimal solution for ANC would also vary when the positions of primary noise source change (Guo & Pan, 2000). The behavior of adaptive filters for ANC of moving noise sources is studied in (Omoto et al., 2002), and further researcher is needed to investigate the effects of time varying paths and developing efficient control algorithms that can cope with the Doppler effects.

In the following sections we discuss challenging task of ANC for impulsive noise sources, and mitigating effect of uncorrelated disturbance. We demonstrate that proposed algorithms and methods can greatly improve the convergence and performance of ANC systems for these tasks.

3. ANC for impulsive noise sources

There are many important ANC applications that involve impulsive noise sources (Kuo et al., 2010). In practice, the impulsive noises are often due to the occurrence of noise disturbance with low probability but large amplitude. There has been a very little research on active control of impulsive noise, at least up to the best knowledge of authors. In practice the impulsive noises do exist and it is of great meaning to study its control.

An impulsive noise can be modeled by stable non-Gaussian distribution (Nikias, 1995; Shao & Nikias, 1993). We consider impulse noise with symmetric α -stable (SaS) distribution $f(x)$ having characteristic function of the form (Shao & Nikias, 1993)

$$\varphi(t) = e^{jat - \gamma|t|^a} \quad (15)$$

where $0 < \alpha < 2$ is the characteristics exponent, $\gamma > 0$ is the scale parameter called as dispersion, and a is the location parameter. The characteristics exponent α is a shape parameter, and it measures the “thickness” of the tails of the density function. If a stable random variable has a small value for α , then distribution has a very heavy tail, i.e., it is more likely to observe values of random variable which are far from its central location. For $\alpha = 2$ the relevant stable distribution is Gaussian, and for $\alpha = 1$ it is the Cauchy distribution. An SaS distribution is called *standard* if $\gamma = 1$, $a = 0$. In this paper, we consider ANC of impulsive noise with standard SaS distribution, i.e., $0 < \alpha < 2$, $\gamma = 1$, and $a = 0$. The PDFs of standard SaS process for various values of α are shown in Fig. 3. It is evident that for small value of α , the process has a peaky and heavy tailed distribution.

In order to improve the robustness of adaptive algorithms for processes having PDFs with heavy tails (i.e. signals with outliers), one of the following solution may be adopted:

1. A robust optimization criterion may be used to derive the adaptive algorithm.
2. The large amplitude samples may be ignored.
3. The large amplitude samples may be replaced by an appropriate threshold value.

The existing algorithms for ANC of impulsive noise are based on the first two approaches. In the proposed algorithms, we consider combining these approaches as well as borrow concept of the normalized step size, as explained later in this section. The discussion presented is with respect to feedforward ANC of Fig. 1, where noise source is assumed to be of impulse type. It is important to note that the feedback type ANC works as a predictor and hence cannot be employed for such types of sources.

3.1 Variants of FxLMS algorithm

Consider feedforward ANC system of Fig. 1, where we assume that noise source is impulsive and follows SaS distribution as explained earlier. The reference signal vector; used in the update equation of the FxLMS algorithm and in generating the cancelling signal $y(n)$; is given in Eq. (3) which shows that the samples of the reference signal $x(n)$ at different time are treated “equally”. It may cause the FxLMS algorithm to become unstable in the presence of impulsive noise. To overcome this problem, a simple modification to FxLMS algorithm is proposed in (Sun et al., 2006). In this algorithm, hereafter referred as Sun’s algorithm, the samples of the reference signal $x(n)$ are ignored, if their magnitude is above a certain threshold set by

statistics of the signal (Sun et al., 2006). Thus the reference signal is modified as

$$x'(n) = \begin{cases} x(n), & \text{if } x(n) \in [c_1, c_2] \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

Here, the thresholding parameters c_1 and c_2 can be obtained offline for ANC systems. A few comments on choosing these parameters are given later. Thus Sun's algorithm for ANC of impulsive noise is given as (Sun et al., 2006)

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu_w e(n) \hat{\mathbf{x}}'_s(n), \quad (17)$$

where $\hat{\mathbf{x}}'_s(n) = [\hat{x}'_s(n), \hat{x}'_s(n-1), \dots, \hat{x}'_s(n-L_w+1)]^T$ is generated as

$$\hat{\mathbf{x}}'_s(n) = \hat{s}(n) * \mathbf{x}'(n), \quad (18)$$

where

$$\mathbf{x}'(n) = [x'(n), x'(n-1), \dots, x'(n-L_w+1)]^T \quad (19)$$

is a modified reference signal vector with $x'(n)$ being obtained using Eq. (16). The main advantage is that the computational complexity of this algorithm is same as that of the FxLMS algorithm.

In our experience, however, Sun's algorithm becomes unstable for $\alpha < 1.5$, when the PDF is peaky and the reference noise is highly impulsive. Furthermore, the convergence speed of this algorithm is very slow. The main problem is that ignoring the peaky samples in the update of FxLMS algorithm does not mean that these samples will not appear in the residual error $e(n)$. The residual error may still be peaky, and in the worst case the algorithm may become unstable. In order to improve the stability of the Sun's algorithm, the idea of Eq. (16) is extended to the error signal $e(n)$ as well, and a new error signal is obtained as (Akhtar & Mitsunashi, 2009a)

$$e'(n) = \begin{cases} e(n), & \text{if } e(n) \in [c_1, c_2] \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

Effectively, the idea is to freeze the adaptation of $W(z)$ when a large amplitude is detected in the error signal $e(n)$. Thus modified-Sun's algorithm for ANC of impulse noise is proposed as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu_w e'(n) \hat{\mathbf{x}}'_s(n). \quad (21)$$

In order to further improve the robustness of the Sun's algorithm; instead of ignoring the large amplitude sample; we may clip the sample by a threshold value, and thus the reference signal is modified as

$$x''(n) = \begin{cases} c_1, & x(n) \leq c_1 \\ c_2, & x(n) \geq c_2 \\ x(n), & \text{otherwise} \end{cases} \quad (22)$$

As stated earlier, ignoring (or even clipping) the peaky samples in the update of FxLMS algorithm does not mean that peaky samples will not appear in the residual error $e(n)$. The residual error may still be so peaky, that in the worst case might cause ANC to become unstable. We extend the idea of Eq. (22) to the error signal $e(n)$ as well, and a new error

signal is obtained as

$$e''(n) = \begin{cases} c_1, & e(n) \leq c_1 \\ c_2, & e(n) \geq c_2 \\ e(n), & \text{otherwise} \end{cases} \quad (23)$$

and hence proposed modified FxLMS (MFxLMS) algorithm for ANC of impulsive noise sources is as given below

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu_w e''(n) \hat{\mathbf{x}}_s''(n), \quad (24)$$

where $\hat{\mathbf{x}}_s''(n) = [\hat{x}_s''(n), \hat{x}_s''(n-1), \dots, \hat{x}_s''(n-L_w+1)]^T$ is generated as

$$\hat{\mathbf{x}}_s''(n) = \hat{\mathbf{s}}(n) * \mathbf{x}''(n), \quad (25)$$

where

$$\mathbf{x}''(n) = [x''(n), x''(n-1), \dots, x''(n-L_w+1)]^T \quad (26)$$

is a modified reference signal vector with $x''(n)$ being obtained using Eq. (22).

It is worth mentioning that all algorithms discussed so far; Sun's algorithm (Sun et al., 2006) and its variants; require an appropriate selection of the thresholding parameters $[c_1, c_2]$. As stated earlier, the basic idea of Sun's algorithm is to ignore the samples of the reference signal $x(n)$ beyond certain threshold $[c_1, c_2]$ set by the statistics of the signal (Sun et al., 2006). Here the probability of the sample less than c_1 or larger than c_2 are assumed to be 0, which is consistent with the fact that the tail of PDF for practical noise always tends to 0 when the noise value is approaching $\pm\infty$. Effectively, Sun's algorithm assumes the same PDF for $x'(n)$ (see Eq. (16)) with in $[c_1, c_2]$ as that of $x(n)$, and neglects the tail beyond $[c_1, c_2]$. The stability of Sun's algorithms depends heavily on appropriate choice of $[c_1, c_2]$. We have extended this idea, that instead of ignoring, the peaky samples are replaced by the thresholding values c_1 and c_2 . Effectively, this algorithm adds a saturation nonlinearity in the reference and error signal paths. Thus, the performance of this algorithm also depends on the parameters c_1 and c_2 .

In order to overcome this difficulty of choosing appropriate thresholding parameters, we propose an FxLMS algorithm that does not use modified reference and/or error signals, and hence does not require selection of the thresholding parameters $[c_1, c_2]$. Following the concept of normalized LMS (NLMS) algorithm (Douglas, 1994), the normalized FxLMS (NFxLMS) can be given as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(n) e(n) \hat{\mathbf{x}}_s(n), \quad (27)$$

where normalized time-varying step size parameter $\mu(n)$ is computed as

$$\mu(n) = \frac{\tilde{\mu}}{\|\hat{\mathbf{x}}_s(n)\|_2^2 + \delta} \quad (28)$$

where $\tilde{\mu}$ is fixed step size parameter, $\|\hat{\mathbf{x}}_s(n)\|_2$ is l_2 -norm of the filtered-reference signal vector that can be computed from current available data, and δ is small positive number added to avoid division by zero. When the reference signal has a large peak, its energy would increase, and this would in turn decrease the effective step size of NFxLMS algorithm. As stated earlier, the error signal is also peaky in nature and its effect must also be taken into account. We

propose following modified normalized step size for FxLMS algorithm of Eq. (27)

$$\mu(n) = \frac{\tilde{\mu}}{\|\hat{\mathbf{x}}_s(n)\|_2^2 + E_e(n) + \delta} \quad (29)$$

where $E_e(n)$ is energy of the residual error signal $e(n)$ that can be estimated online using a lowpass estimator as

$$E_e(n) = \lambda E_e(n-1) + (1-\lambda)e^2(n), \quad (30)$$

where λ is the forgetting factor ($0.9 < \lambda < 1$). It is worth mentioning that the proposed modified normalized FxLMS (MNFxLMS) algorithm, comprising Eqs. (27), (29) and (30), does not require estimation of thresholding parameters $[c_1, c_2]$.

3.2 FxLMP Algorithm and proposed modifications

For stable distributions, the moments only exist for the order less than the characteristic exponent (Shao & Nikias, 1993), and hence the MSE criterion which is bases for FxLMS algorithm, is not an adequate optimization criterion. In (Leahy et al., 1995), the filtered-x least mean p -power (FxLMP) algorithm has been proposed, which is based on minimizing a fractional lower order moment (p -power of error) that does exist for stable distributions. For some $0 < p < \alpha$, minimizing the p th moment $J(n) = E\{|e(n)|^p\} \approx |e(n)|^p$, the stochastic gradient method to update $W(z)$ is given as (Leahy et al., 1995)

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu_w p(e(n))^{<p-1>} \hat{\mathbf{x}}_s(n), \quad (31)$$

where the operation $(z)^{<a>}$ is defined as

$$(z)^{<a>} \equiv |z|^a \text{sgn}(z), \quad (32)$$

where $\text{sgn}(z)$ is sign function being defined as

$$\text{sgn}(z) = \begin{cases} 1, & z > 0 \\ 0, & z = 0 \\ -1, & z < 0 \end{cases} \quad (33)$$

It has been shown that FxLMP algorithm with $p < \alpha$ shows good robustness to ANC of impulsive noise. Our objective in this contribution is to improve the convergence performance of the FxLMP algorithm proposed in (Leahy et al., 1995). Based on our extensive simulation studies, we propose two modified versions of the FxLMP algorithm.

The first proposed algorithm attempts to improve the robustness of FxLMP algorithm by using the modified reference and error signals as given in Eqs. (22) and (23), respectively. Thus considering the FxLMP algorithm (Leahy et al., 1995) given in Eq. (31), a modified FxLMP (MFxLMP) algorithm for ANC of impulse noise is given as²

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu_w p(e''(n))^{<p-1>} \mathbf{x}_s''(n). \quad (34)$$

As done with the FxLMS algorithm, the second modification is based on normalizing the step size parameter and hence, it avoids selection of the thresholding parameters $[c_1, c_2]$. In

² Some preliminary results regarding this algorithm were presented at *IEEE ICASSP 2009* (Akhtar & Mituhahsi, 2009b).

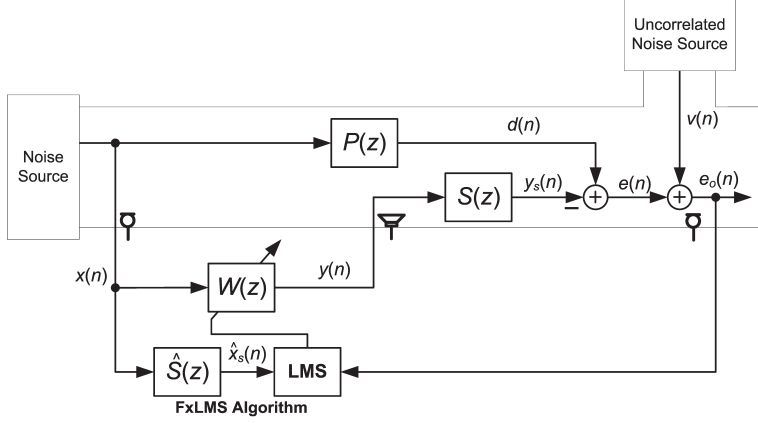


Fig. 4. Block diagram of FxLMS algorithm based single-channel feedforward ANC systems in the presence of uncorrelated disturbance $v(n)$ at the error microphone.

(Aydin et al., 1999), the concept of NLMS algorithm (Douglas, 1994) has been extended to LMP algorithm and a normalized LMP (NLMP) algorithm has been proposed where step size is normalized by the energy of reference signal vector. By extending this idea to FxLMP algorithm (Leahy et al., 1995), the normalized FxLMP (NFxLMP) can be given as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(n)p(e(n))^{<p-1>} \hat{\mathbf{x}}_s(n), \quad (35)$$

where normalized time-varying step size parameter $\mu(n)$ is computed as

$$\mu(n) = \frac{\tilde{\mu}}{\|\hat{\mathbf{x}}_s(n)\|_p^p + \delta'}, \quad (36)$$

where $\|\hat{\mathbf{x}}_s(n)\|_p$ is p th norm computed from current filtered-reference signal vector. Since the error signal $e(n)$ is also peaky in nature and its effect must also be taken into account, we propose following modified normalized step size for FxLMP algorithm of Eq. (35):

$$\mu(n) = \frac{\tilde{\mu}}{\|\hat{\mathbf{x}}_s(n)\|_p^p + E_e(n) + \delta'}, \quad (37)$$

where $E_e(n)$ is energy of the residual error signal $e(n)$. Thus a modified normalized FxLMP (MNFxLMP) algorithm is suggested comprising Eqs. (35), (37) and (30).

In this section we have suggested ad hoc modifications to the existing adaptive algorithms for ANC of impulsive noise. The simulation results presented later in Section 5.1 demonstrate that these modifications greatly improve robustness of ANC system for the impulsive noise sources.

4. Mitigating uncorrelated disturbance

The FxLMS algorithm is widely used in ANC systems, however performance of the FxLMS algorithm in steady state will be degraded due to presence of an uncorrelated disturbance

at the error microphone, shown as $v(n)$ in Fig. 4. This situation arises in many real-world applications. For example, in electronic mufflers for automobiles (Kuo & Gan, 2004), the undesired disturbance such as the noises generated by other passing-by automobiles will affect the stability and performance of the ANC systems. In industrial installations, neighboring machinery near to the location of error microphone may generate uncorrelated disturbance. In the presence of uncorrelated disturbance, $v(n)$, the error signal picked-up by the error microphone is given as

$$e_o(n) = e(n) + v(n) = [d(n) - y_s(n)] + v(n), \quad (38)$$

and hence, the update equation for FxLMS algorithm for $W(z)$ can be written as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu_w e(n) \hat{\mathbf{x}}_s(n) + \mu_w v(n) \hat{\mathbf{x}}_s(n). \quad (39)$$

It is evident that the adaptation is perturbed by the uncorrelated noise component $v(n)$, and as shown in (Sun & Kuo, 2007), the steady-state performance of the FxLMS algorithm will be degraded significantly. Furthermore, $v(n)$ appearing uncontrolled at the error microphone degrades the noise reduction performance of the ANC system.

Up to the best knowledge of Authors, a little research has been done to cope with the uncorrelated disturbance problem. In (Kuo & Ji, 1996), an adaptive algorithm consisting of two interconnected adaptive notch filters is proposed to reduce the disturbance problem. However, this algorithm is effective only for narrowband, single-frequency ANC systems. In (Sun & Kuo, 2007), this algorithm has been generalized to multifrequency narrowband feedforward ANC systems using a single high-order adaptive filter, and a cascaded ANC system is proposed. This method improves the convergence of the FxLMS algorithm, however, cannot mitigate the effect of the uncorrelated disturbance $v(n)$ from the residual noise $e(n)$. One solution to this problem of uncorrelated disturbance is offered by a hybrid ANC comprising feedforward and feedback control strategies (Esmailzadeh et al., 2002). The feedforward ANC attenuates the primary noise that is correlated with the reference signal, whereas the feedback ANC takes care of the narrowband components of noise that are not observed by the reference sensor. We observe that the performance of the hybrid ANC system degrades in certain situations, as explained later in this section.

4.1 Existing solutions for uncorrelated disturbance

The main idea of cascading ANC system (Sun & Kuo, 2007) is to update the adaptive filter $W(z)$ using estimate of the desired error signal $e(n)$ instead of using a disturbed error signal $e_o(n)$. The block diagram of cascading ANC system is shown in Fig. 5, where the adaptive filter $H(z)$ is introduced to estimate the desired error signal $e(n)$.

It is evident that $H(z)$ is excited by the reference signal $x(n)$, and the error signal $e_o(n)$ is used as a desired response for its adaptation. Thus output of $H(z)$, $y_h(n)$, converges to that part in $e_o(n)$ which is correlated with $x(n)$. From Eqs. (1), (3), and (4), it is clear that in $e_o(n)$ given in Eq. (38), $e(n) = [d(n) - y_s(n)]$ is correlated with $x(n)$ and $v(n)$ is the uncorrelated part. Hence, when $H(z)$ converges, its output converges to $y_h(n) \approx e(n) = [d(n) - y_s(n)]$, which is the desired error signal for the adaptation of $W(z)$. Thus FxLMS algorithm for this cascading ANC is given as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu_w y_h(n) \hat{\mathbf{x}}_s(n). \quad (40)$$

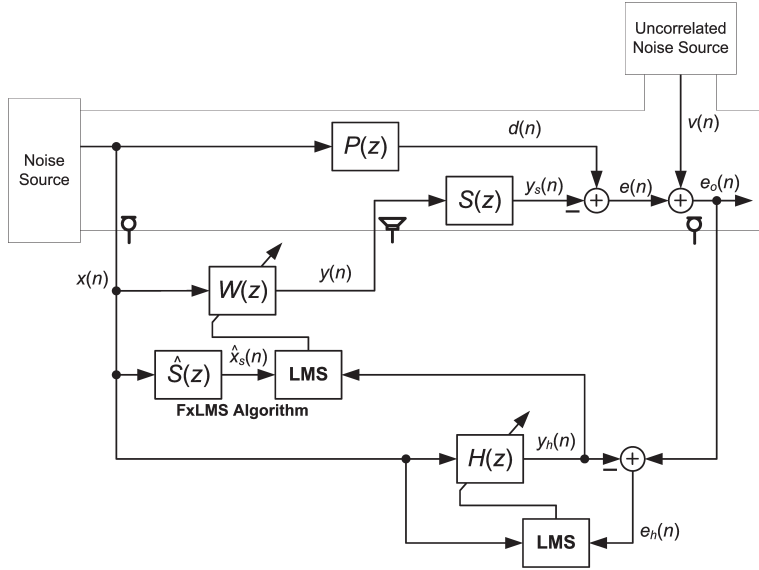


Fig. 5. Block diagram of the cascading ANC system for improving adaptation of FxLMS algorithm in the presence of uncorrelated disturbance $v(n)$ (Sun & Kuo, 2007).

Since a disturbance free error signal is used, cascading ANC improves the convergence of the FxLMS algorithm. However, it cannot mitigate effect of the uncorrelated disturbance $v(n)$ from the residual noise $e_o(n)$.

One solution for ANC of correlated and uncorrelated disturbances would be to consider a hybrid ANC system comprising feedforward ANC $W(z)$ and feedback ANC $B(z)$ as shown in Fig. 6 (Kuo & Morgan, 1996). We assume that the two noise sources are independent from each other, and thus the primary disturbances $d(n)$ and $v(n)$ are uncorrelated with each other. The reference signal $x(n)$ from the reference microphone is correlated with $d(n)$ and is input to feedforward ANC $W(z)$. The total cancelling signal is sum of outputs of $W(z)$ and $B(z)$ and is passed through $S(z)$ to generate the residual error signal $e_o(n)$ as

$$e_o(n) = [d(n) - y_s(n)] + [v(n) - g_s(n)]. \quad (41)$$

This error signal is used in the FxLMS algorithm for both $W(z)$ and $B(z)$. Comparing $e_o(n)$ in Eq. (41) with $e(n)$ in Eq. (4) and with $e_b(n)$ in Eq. (8), we see that $e_o(n)$ comprises two components. The first component is required for the adaptation of feedforward ANC $W(z)$ and acts as a disturbance for feedback ANC $B(z)$. The second component plays exactly the reverse role, i.e., a disturbance for $W(z)$ and desired error signal for adaptation of $B(z)$.

The reference signal for $W(z)$, $x(n)$, is given by the reference microphone, and the reference signal for $B(z)$, $u(n)$, is internally generated as

$$\begin{aligned} u(n) &= e_o(n) + \hat{y}_s(n) + \hat{g}_s(n) \\ &= [d(n) - y_s(n) + \hat{y}_s(n)] + [v(n) - g_s(n) + \hat{g}_s(n)] \\ &\approx \hat{d}(n) + \hat{v}(n). \end{aligned} \quad (42)$$

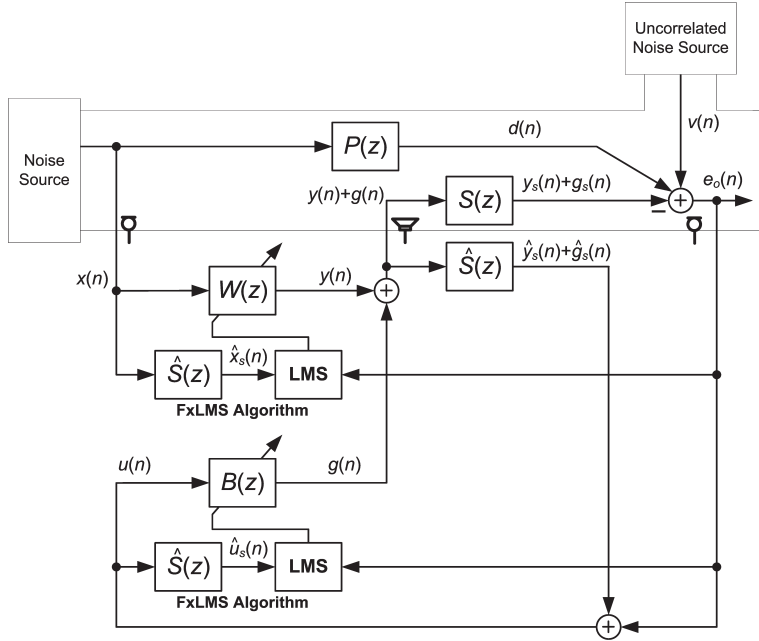


Fig. 6. Block diagram of conventional hybrid ANC system with combination of feedforward ANC $W(z)$ and feedback ANC $B(z)$.

Thus the reference signal $u(n)$ comprises two parts; estimates of disturbances $d(n)$ and $v(n)$. Since objective of the feedback ANC $B(z)$ is to cancel only uncorrelated primary noise $v(n)$, the presence of $d(n)$ (which may be broadband and unpredictable in general) gives a corrupted reference signal for $B(z)$. Thus, both $W(z)$ and $B(z)$ are adapted using inappropriate error signals and may converge slowly. Furthermore, $B(z)$ is excited by a corrupted reference signal and might not converge at all, making whole ANC system unstable. From above discussion, we conclude that

- the cascading ANC (Sun & Kuo, 2007) improves the convergence of the FxLMS algorithm, however, it cannot remove the effect of the uncorrelated disturbance from the residual noise, and that
- the conventional hybrid ANC (Kuo & Morgan, 1996) can provide control over correlated and uncorrelated noise sources, however, its performance might be poor, as ANC filters are using inappropriate error and/or reference signals.

In order to solve these limitations of the existing methods, a modified hybrid ANC is developed as explained in the next section.

4.2 Modified hybrid ANC System

The block diagram of modified hybrid ANC system is shown in Fig. 7 (Akhtar & Mituhahsi, 2011), and as shown, this method comprises three adaptive filters: 1) a feedforward ANC filter $W(z)$ to cancel the primary noise $d(n)$, 2) a feedback ANC filter $B(z)$ to cancel the uncorrelated disturbance $v(n)$, and 3) a supporting filter $H(z)$. The $W(z)$ is excited by the reference signal

computed as

$$\begin{aligned} e_h(n) &= e_o(n) - y_h(n) \\ &= [d(n) - y_s(n)] + [v(n) - g_s(n)] - y_h(n), \end{aligned} \quad (44)$$

and $H(z)$ is adapted using LMS algorithm as

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \mu_h e_h(n) \mathbf{x}(n), \quad (45)$$

where μ_h is the step size for $H(z)$. Since $H(z)$ is excited by $x(n)$, minimizing the error signal $e_h(n)$ means that the output of $H(z)$, $y_h(n)$, would converge to that part in Eq. (44) which is correlated with $x(n)$, thus

$$y_h(n) \rightarrow [d(n) - y_s(n)], \quad (46)$$

and hence $e_h(n)$ would converge to the remaining uncorrelated part

$$e_h(n) \rightarrow [v(n) - g_s(n)]. \quad (47)$$

Comparing Eq. (46) with Eq. (4), we see that $y_h(n)$ can be used as an error signal for adaptation of $W(z)$, and hence FxLMS algorithm for $W(z)$ is given as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu_w y_h(n) \hat{\mathbf{x}}_s(n) \quad (48)$$

and similarly, comparing Eq. (47) with Eq. (8), we observe that $e_h(n)$ can be used as an error signal for feedback ANC filter $B(z)$, and corresponding FxLMS algorithm for $B(z)$ is thus given as

$$\mathbf{b}(n+1) = \mathbf{b}(n) + \mu_b e_h(n) \hat{\mathbf{u}}_s(n) \quad (49)$$

where $\hat{\mathbf{u}}_s(n) = \hat{s}(n) * \mathbf{u}(n)$, and the reference signal $u(n)$ is generated as

$$\begin{aligned} u(n) &= e_h(n) + \hat{g}_s(n) \\ &\approx [v(n) - g_s(n)] + \hat{g}_s(n) \\ &\rightarrow \hat{v}(n). \end{aligned} \quad (50)$$

Comparing Eq. (50) with Eq. (42), we see that the input $u(n)$ for feedback ANC filter $B(z)$ in the modified hybrid ANC of Fig. 7 is equal to estimate of only uncorrelated noise source $v(n)$. A comparison between the modified hybrid ANC and existing approaches is as given below:

- The modified hybrid ANC provides control over both correlated and uncorrelated disturbances, where as cascading ANC of (Sun & Kuo, 2007) can only improve the convergence of $W(z)$, but cannot reduce the uncorrelated disturbance.
- In modified hybrid ANC, role of $H(z)$ is partly same as that in Sun's method. It generates desired error signal for adaptation of $W(z)$ to provide cancellation for correlated disturbance signal $d(n)$. Furthermore, it is used to generate appropriate signals for adaptation of $B(z)$ to cancel uncorrelated disturbance $v(n)$. Therefore, as compared with the conventional hybrid ANC, proposed structure results in generating appropriate signals for two adaptive filter, and hence, an improved convergence and residual noise reduction is expected.
- The modified hybrid ANC comprises three adaptive filters and its computational complexity is higher than the existing approaches. This increased computational

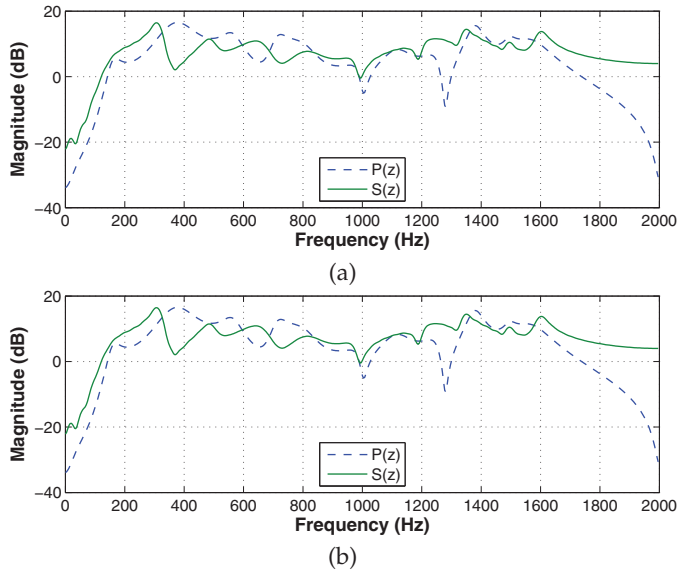


Fig. 8. Frequency response of the primary path $P(z)$ and secondary path $S(z)$. (a) Magnitude response and (b) phase response.

complexity may be considered as the price paid for improved performance. In fact, the simulation results presented in Section 5.2 demonstrate that modified hybrid ANC achieves the performance which is not possible with neither cascading ANC nor conventional hybrid ANC working alone.

5. Computer simulations

In this section results of computer simulation are presented for two case studies discussed in this chapter, viz., ANC for impulsive noise sources, and mitigating uncorrelated disturbance. The acoustic paths are modeled using data provided in the disk attached with (Kuo & Morgan, 1996). Using this data $P(z)$ and $S(z)$ are modeled as FIR filter of length 256 and 128 respectively. The characteristics of the acoustic paths are shown in Fig. 8. It is assumed that the secondary path modeling filter $\hat{S}(z)$ is exactly identified as $S(z)$.

5.1 ANC for impulsive noise sources

The simulation setup is same as shown in Fig. 1, where noise source is assumed to be impulsive and the ANC filter $W(z)$ is selected as an FIR filter of tap-weight length $L_w = 192$. The performance comparison is done on the basis of mean noise reduction (MNR), being defined as

$$\text{MNR}(n) = E \left\{ \frac{A_e(n)}{A_d(n)} \right\}, \quad (51)$$

where $E\{\cdot\}$ denotes expectation or ensemble averaging of quantity inside, and $A_e(n)$ and $A_d(n)$ are estimates of absolute values of residual error signal $e(n)$ and disturbance signal

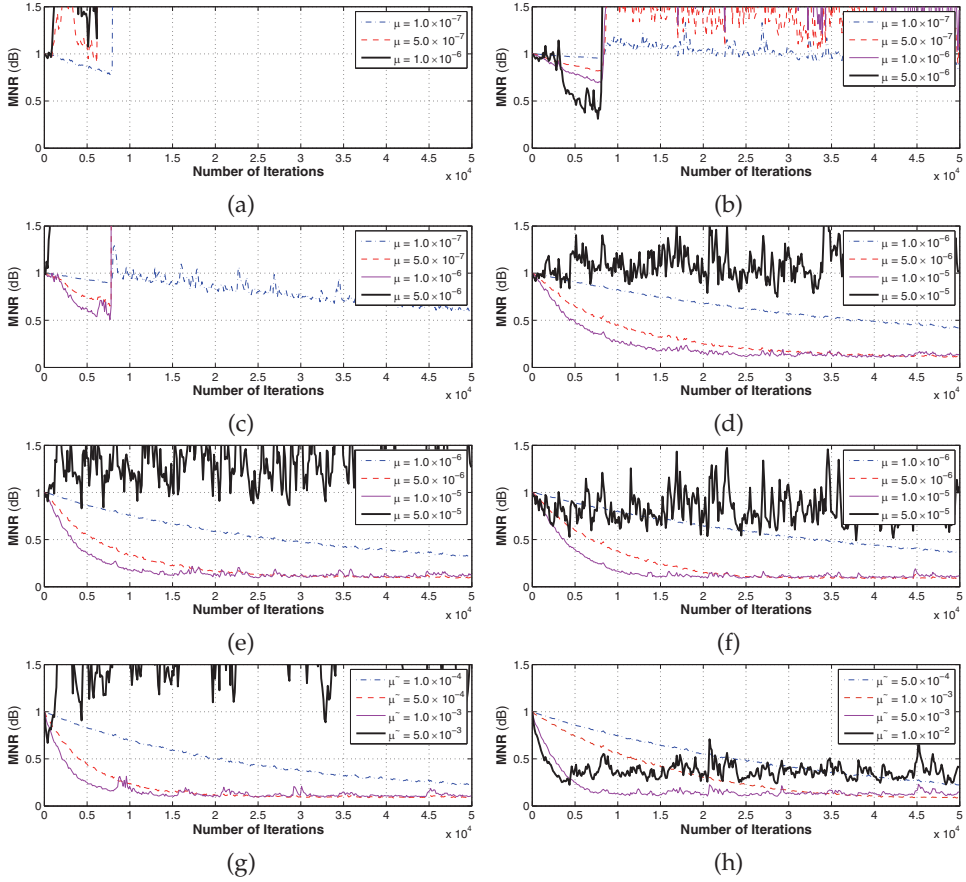


Fig. 9. Mean noise reduction (MNR) curves for various algorithms for ANC of impulsive noise with $\alpha = 1.45$. (a) FxLMS algorithm, (b) FxLMP algorithm, (c) Sun's algorithm, (d) Modified-Sun's algorithm, (e) MFxLMS algorithm, (f) MFxLMP algorithm, (g) MNFxLMS algorithm, and (h) MNFxLMP algorithm.

$d(n)$, respectively, at the location of error microphone. These estimates are obtained using lowpass estimators as

$$A_r(n) = \lambda A_r(n-1) + (1-\lambda)|r(n)| \quad (52)$$

where $|\cdot|$ is the absolute value of quantity, and λ is same as defined in Eq. (30).

The reference noise signal $x(n)$ is modeled by standard S α S process with $\alpha = 1.45$ (which corresponds to a very peaky distribution more toward Cauchy distribution), and $\alpha = 1.65$ (corresponding to distribution towards Gaussian distribution). All simulation results presented below are averaged over 25 realization of the process. Extensive simulations are carried out to find appropriate values for the thresholding parameters $[c_1, c_2]$, and are selected as: $[0.01, 99.99]$ for Sun's algorithm in Eq. (17), $[0.5, 99.5]$ for modified-Sun's algorithm in Eq. (21), and $[1, 99]$ for MFxLMS algorithm in Eq. (24). The detailed simulation results

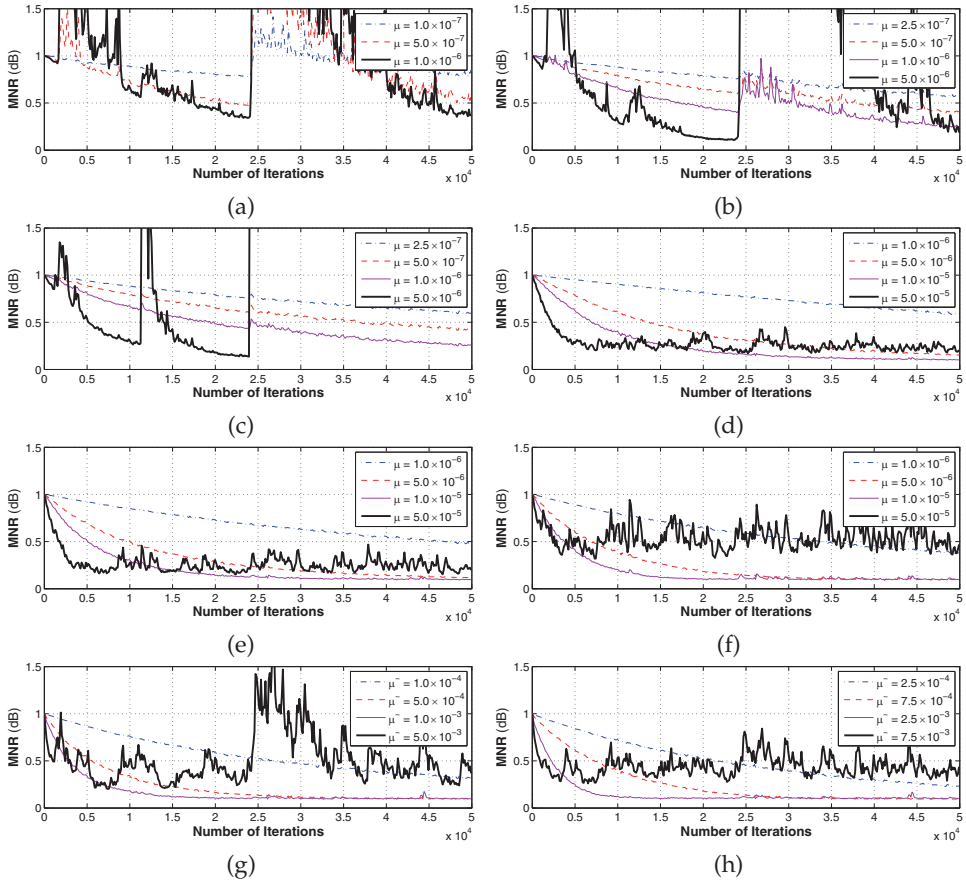


Fig. 10. Mean noise reduction (MNR) curves for various algorithms for ANC of impulsive noise with $\alpha = 1.65$. (a) FxLMS algorithm, (b) FxLMP algorithm, (c) Sun's algorithm, (d) Modified-Sun's algorithm, (e) MFxLMS algorithm, (f) MFxLMP algorithm, (g) MNFxLMS algorithm, and (h) MNFxLMP algorithm.

for two cases are given in Figs. 9 and 10, respectively, where the objective is to study the effect of step size parameter. It is seen that, the FxLMS algorithm is not able to provide ANC for impulsive noise, even for a very small step size. Furthermore, in comparison with the Authors' algorithms, the performance of Sun's algorithm and FxLMP algorithm is very poor. On the basis of best results for the respective algorithms, the performance comparison for two cases is shown in Figs. 11 and 12, respectively. These results show that the proposed algorithms outperform the existing algorithms and, among the algorithms discussed in Section 3, appears as a best choice for ANC of SaS impulsive noise.

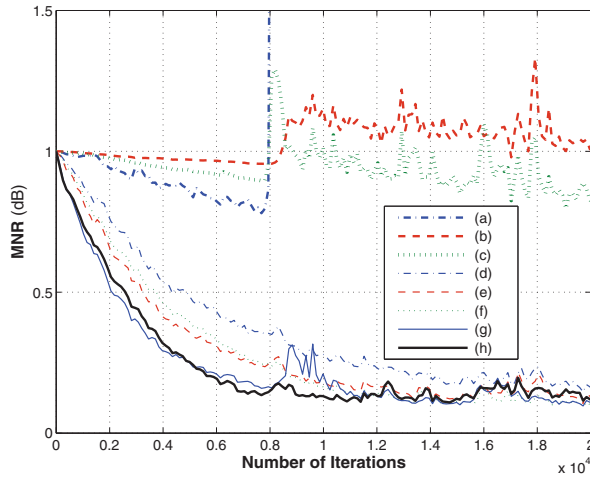


Fig. 11. Performance comparison between various algorithms for ANC of impulsive noise with $\alpha = 1.45$. (a) FxLMS algorithm ($\mu = 1 \times 10^{-7}$), (b) FxLMP algorithm ($\mu = 1 \times 10^{-7}$), (c) Sun's algorithm ($\mu = 1 \times 10^{-7}$), (d) Modified-Sun's algorithm ($\mu = 1 \times 10^{-5}$), (e) MFxLMS algorithm ($\mu = 1 \times 10^{-5}$), (f) MFxLMP algorithm ($\mu = 1 \times 10^{-5}$), (g) MNFfLMS algorithm ($\tilde{\mu} = 1 \times 10^{-3}$), and (h) MNFfLMP algorithm ($\tilde{\mu} = 5 \times 10^{-3}$).

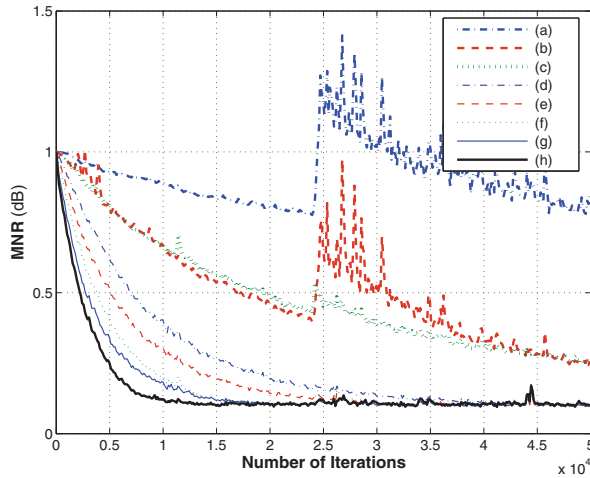


Fig. 12. Performance comparison between various algorithms for ANC of impulsive noise with $\alpha = 1.65$. (a) FxLMS algorithm ($\mu = 1 \times 10^{-7}$), (b) FxLMP algorithm ($\mu = 1 \times 10^{-6}$), (c) Sun's algorithm ($\mu = 1 \times 10^{-6}$), (d) Modified-Sun's algorithm ($\mu = 1 \times 10^{-5}$), (e) MFxLMS algorithm ($\mu = 1 \times 10^{-5}$), (f) MFxLMP algorithm ($\mu = 1 \times 10^{-5}$), (g) MNFfLMS algorithm ($\tilde{\mu} = 1 \times 10^{-3}$), and (h) MNFfLMP algorithm ($\tilde{\mu} = 2.5 \times 10^{-3}$).

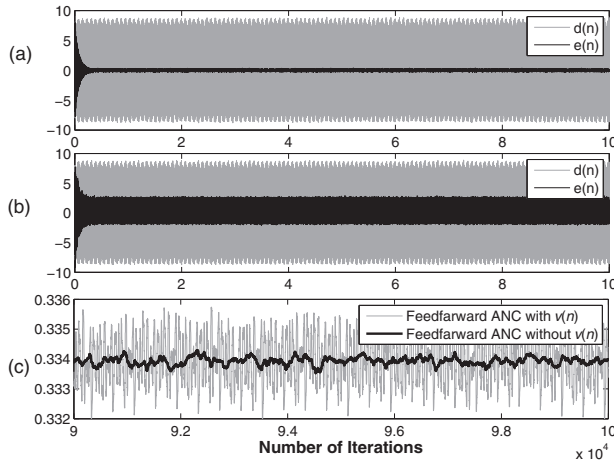


Fig. 13. Simulation results for mitigating uncorrelated disturbance $v(n)$. (a) Feedforward ANC without uncorrelated disturbance $v(n)$. (b) Feedforward ANC in the presence of uncorrelated disturbance $v(n)$. (c) Norm of weight vector, $\|w(n)\|$, at steady state.

5.2 Mitigating uncorrelated disturbance

In this section we present simulation results for various methods studied for mitigating uncorrelated disturbances. Essentially, we consider following methods in our study:

1. feedforward ANC shown in Fig. 4,
2. cascading ANC system of Fig. 5,
3. conventional hybrid ANC of Fig. 6, and
4. modified hybrid ANC of Fig. 7.

The noise source $x(n)$ is assumed as a unit variance narrowband signal composed of three sinusoids with frequencies of 165, 290, and 410 Hz. A white noise with variance 0.001 is added to count for measurement noise at the reference microphone. The uncorrelated disturbance $v(n)$ is another unit variance narrowband signal comprising three sinusoids with frequencies of 250, 350, and 500 Hz, and a white noise with variance 0.001 is added to count for measurement noise at the error microphone. The sampling frequency is 4 kHz, and the results shown are average of 10 realizations. The adaptive filters $W(z)$, $B(z)$, and $H(z)$ are selected as FIR filters of tap-weight lengths 192, 192, and 32, respectively. All adaptive filters are initialized by null vectors of an appropriate order. The step sizes are selected experimentally, such that fast and stable performance is obtained and are adjusted as, feedforward ANC: $\mu_w = 1 \times 10^{-5}$, cascading ANC: $\mu_w = 1 \times 10^{-5}$, $\mu_h = 1 \times 10^{-3}$, conventional hybrid ANC: $\mu_w = 1 \times 10^{-5}$, $\mu_b = 1 \times 10^{-6}$, and modified hybrid ANC: $\mu_w = 1 \times 10^{-5}$, $\mu_b = 1 \times 10^{-5}$, $\mu_h = 5 \times 10^{-3}$.

The effect of the uncorrelated disturbance $v(n)$, on the performance of feedforward ANC of Fig. 4, is studied in Fig. 13. We see that feedforward ANC cannot control the uncorrelated disturbance $v(n)$ appearing at the error microphone, and hence, its noise reduction performance is degraded, as indicated by increased level of residuals noise $e(n)$ (Fig. 13 (a) and (b)). Furthermore, the uncorrelated disturbance disturbs the convergence of ANC

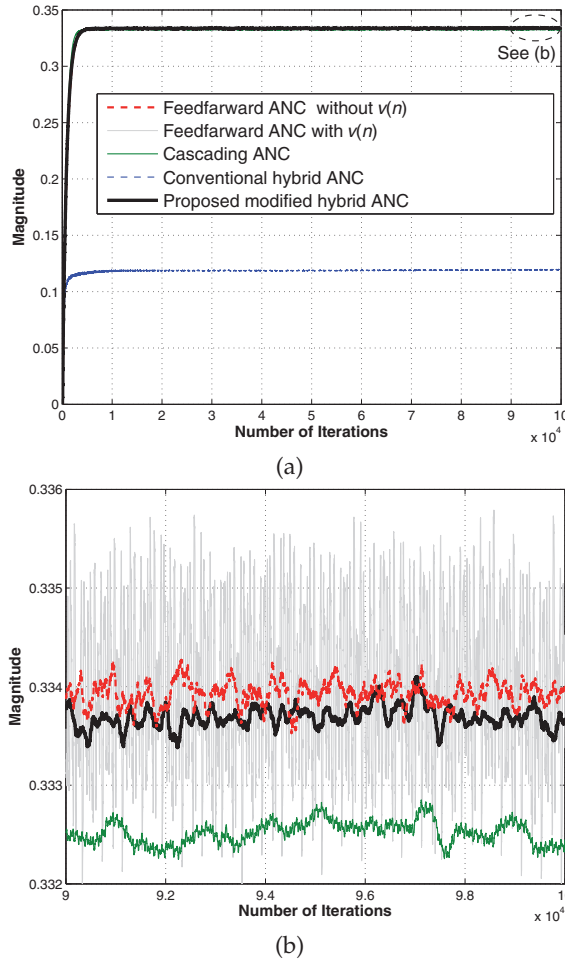


Fig. 14. Simulation results for mitigating uncorrelated disturbance $v(n)$. (a) Convergence of norm of tap-weights for ANC filter $W(z)$, $\|\mathbf{w}(n)\|$. (b) Zoomed curves for region marked by dashed ellipse in (a).

filter, as shown in Fig. 13 (c). Our objective is to improve the noise reduction performance and have stable steady state convergence of tap-weights.

The adaptation of adaptive filter $W(z)$, in terms of $\|\mathbf{w}(n)\|$, which is Euclidean norm of weight vector $\mathbf{w}(n)$, is shown in Fig. 14(a). An enlarged view of curves in steady state is shown in Fig. 14(b). We see that the modified hybrid ANC can provide control over the uncorrelated noise source $v(n)$, and hence can remove its effect from the convergence of $W(z)$. In fact, the steady state solution achieved by the modified hybrid ANC is close to that obtained by feedforward ANC without uncorrelated disturbance. We see that, this is a far better performance as compared with the conventional hybrid ANC.

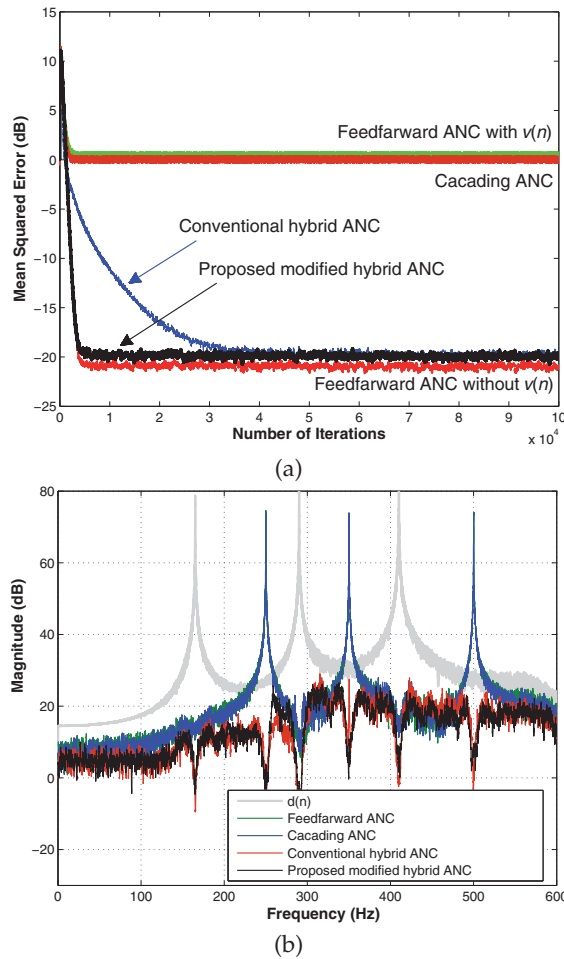


Fig. 15. Simulation results for mitigating uncorrelated disturbance $v(n)$. (a) Curves for mean squared error (MSE) for various methods. (b) Magnitude spectrum of residual error signal $e(n)$ at steady-state for various methods in comparison with the magnitude spectrum of primary disturbance $d(n)$.

The noise reduction performance, in terms of MSE, for various methods is shown in Fig. 15(a). We see that the uncorrelated disturbance $v(n)$ appearing at the error microphone degrades the noise reduction performance of the feedforward ANC system. The conventional and modified hybrid ANC systems, incorporating a separate ANC filter $B(z)$ for uncorrelated disturbance, give significantly improved noise reduction performance, with the proposed method converging at a faster rate. Fig. 15(b) shows the power spectral density of the residual error signal $e(n)$ in the steady state. The spectrum of the primary disturbance $d(n)$ is also shown as a reference. We see that, as compared with the feedforward ANC, the conventional and modified hybrid ANCs are more effective in removing the uncorrelated disturbance.

6. An outlook on recent ANC applications

In this section, we give a brief overview of new applications of ANC systems in consumer electronics and medical instruments. The first example is active control of speech signals for privacy-phone handsets (Kondo & Nakagawa, 2007). The idea is to generate out-of-phase speech to cancel the original speech in space, thus allowing private and quiet voice communication in public areas. Developing efficient algorithms and methods for efficient speech emission control in 3D environment requires further research.

In hospitals, there are a lot of life-saving equipment such as breathing and IV pumps that generate impulse-like noises. For example, infant incubators are used in neonatal intensive care units (NICU) to increase the survival of premature and ill infants. The application of ANC for reducing incubator noise in NICU was reported in (Liu et al., 2008), where a nonlinear filtered-X least mean M-estimate algorithm is developed for reducing impulse-like noise in incubators. In Section 3 we have presented some robust algorithms for ANC of impulsive noise sources, and theoretical performance analysis, real-time experiments, and development of more effective ANC algorithms is open for further research.

Recently very interesting results have been reported concerning head mounted ANC for the noise generated during magnetic resonance imaging (MRI) (Kida et al., 2009). The noise generated during MRI is found to be of a narrowband nature, and work presented in (Kida et al., 2009) considers feedback type ANC. It would be interesting to investigate, whether we can get better performance by employing proposed hybrid ANC system for MRI noise.

In the recent years, traffic noise coming from streets, highways, railways, and airports has been of increasing concern. In such situations the positions of noise sources are time varying, and it is necessary to study and develop dynamic ANC systems for moving noise sources relative to the ANC installation. One challenging, yet a very interesting, application would be to study an efficient ANC system for a quiet car interior even when the window or sunroof is open.

In some applications, it is desirable to retain a low-level residual noise with a desired spectral shape or changed noise signature. Active sound quality control (ASQC), which changes amplitudes of noise components with predetermined values, is a useful and important extension of ANC, see (Kuo & Ji, 1995) for narrowband ASQC and (Kuo & Yang, 1996) for broadband ASQC. The broadband ASQC algorithm uses a shaping filter to control the residual noise spectrum, and further research is needed to design an appropriate shaping filter. Recently noise reduction for motorcycle helmets is evaluated and some interesting results using hybrid ANC have been reported in (Castañé-Selga & Sánchez Peña, 2010). In such applications, ASQC systems must be employed so that noise and horn-sounds from the neighboring vehicles are not completely removed.

7. Concluding remarks

In this chapter we have provided a comprehensive review of adaptive filtering algorithms developed for ANC systems. We mentioned several signal processing challenges with open problems for further research and development. Furthermore, we presented efficient algorithms and methods for two challenging problems in ANC systems: 1) active control of impulsive-like noise sources, and 2) controlling uncorrelated disturbance for which a correlated reference signal is not available. Finally an outlook of some challenging applications of ANC is provided.

8. Acknowledgments

This research is carried out with funding from Japan Science and Technology (JST) Agency.

9. References

- Aydin, G. & Arikan, O. & Centin, E. (1999). Robust adaptive filtering algorithms for α -stable random process, *IEEE Transactions Circuits Systems-II: Analog Digital Signal Processing*, Vol. 46, No. 2, pp. 198–202, 1999.
- Akhtar, M. T. & Abe, M. & Kawamata, M. (2005). A new structure for feedforward active noise control systems with improved online secondary path modeling. *IEEE Transactions Speech Audio Processing*, Vol. 13, No. 5, 2005, pp. 1082–1088.
- Akhtar, M. T. & Abe, M. & Kawamata, M. (2006). A new variable step size LMS algorithm-based method for improved online secondary path modelling in active noise control systems. *IEEE Transactions Audio Speech Language Processing*, Vol. 14, No. 2, 2006, pp. 720–726.
- Akhtar, M. T. & Abe, M. & Kawamata, M. (2007). On active noise control systems with online acoustic feedback path modeling. *IEEE Transactions Audio Speech Language Processing*, Vol. 15, No. 3, 2007, pp. 593–600.
- Akhtar, M. T. & Mitsunashi, W. (2009a). Improving performance of FxLMS algorithm for active noise control of impulsive noise. *Journal of Sound and Vibration*, Vol. 327, No. 3–5, 2009, pp. 647–656.
- Akhtar, M. T. & Mitsunashi, W. (2009b). Robust adaptive algorithms for active noise control of impulse noise. *Proceedings of IEEE International Conference Acoustic Speech Signal Processing (ICASSP)*, pp. 261–264, April 19–24, 2009.
- Akhtar, M. T. & Mitsunashi, W. (2011). Improving performance of hybrid active noise control systems for uncorrelated narrowband disturbances. *IEEE Transactions Audio Speech Language Processing*, 2011, (in press).
- Bao, C. & Sas, P. & Brussel, H. V. (1993a). Comparison of two online identification algorithms for active noise control. *Proceedings of Recent Advances in Active Control of Sound Vibration*, 1993, pp. 38–51.
- Bao, C. & Sas, P. & Brussel, H. V. (1993b). Adaptive active control of noise in 3-D reverberant enclosure. *Journal of Sound and Vibration*, Vol. 161, No. 3, Mar. 1993, pp. 501–514.
- Bonito, G. & Elliott, S. J. & Boucher, C. C. (1997). Generation of zones of quiet using a virtual microphone arrangement. *Journal of Acoustical Society of America*, Vol. 101, No. 6, June 1997, pp. 3498–3516.
- Bouchard, M. & Quednau, S. (2000). Multichannel recursive-least-squares algorithms and fast-transversal-filter algorithms for active noise control and sound reproduction systems. *IEEE Transactions Speech Audio Processing*, Vol. 8, 2000, pp. 606–618.
- Crawford, D. H. & Stewart, R. W. (1997). Adaptive IIR filtered-v algorithms for active noise control. *Journal of Acoustical Society of America*, Vol. 101, No. 4, 1997, pp. 2097–2103.
- Castañé-Selga, R. & Sánchez Peña, R. S. (2010). Active noise hybrid time-varying control for motorcycle helmets. *IEEE Transactions Control Systems Technology*, Vol. 18, No. 3, May 2010, pp. 602–612.
- Douglas, S. C. (1994). A family of normalized LMS algorithms. *IEEE Signal Processing Letters*, Vol. 1, No. 3, 1994, pp. 49–51.

- Elliott, S. J. & Stothers, I. M. & Nelson, P. A. (1987) A multiple error LMS algorithm and its application to the active control of sound and vibration. *IEEE Transactions Acoustic Speech Signal Processing*, Vol. ASSP-35, Oct. 1987, pp. 1423–1434.
- Elliot, S. J. (2001). *Signal Processing for Active Control*, London, U.K.: Academic Press.
- Eriksson, L. J.; Allie, M. C. & Greiner, R. A. (1987). The selection and application of an IIR adaptive filter for use in active sound attenuation. *IEEE Transactions Acoustic Speech Signal Processing*, Vol. 35, No. 1, 1987, pp. 433–437.
- Eriksson, L. J. & Allie, M. C. (1989). Use of random noise for on-line transducer modeling in an adaptive active attenuation system. *Journal of Acoustical Society of America*, Vol. 85, No. 2, Feb. 1989, pp. 797–802.
- Esmailzadeh, E. & Alasty, A. & Ohadi, A. R. (2002). Hybrid active noise control of a one-dimensional acoustic duct. *Transactions of the ASME*, Vol. 124, 2002, pp. 10–18.
- Gan, W. S. & Kuo, S. M. (2002). An integrated audio and active noise control headsets. *IEEE Transactions Consumer Electronics*, Vol. 48, No. 2, May 2002, pp. 242–247.
- Guo, J. N. & Pan, J. (2000). Active control of moving noise source: effects of off-axis source position. *Journal of Sound and Vibration*, Vol. 215, 2000, pp. 457–475.
- Jeon, H. J. & Chang, T. G. & Kuo, S. M. (2010). Analysis of frequency mismatch in narrowband active noise control. *IEEE Transactions Audio Speech Language Processing*, Vol. 18, No. 6, Aug. 2010, pp. 1632–1642.
- Kida, M. & Hirayama, R. & Kajikawa, Y. & Tani, T. & Kurumi, Y. (2009). Head-mounted active noise control system for MR noise. *Proceedings IEEE International Conference Acoustic Speech Signal Processing*, 2009, pp. 245–248.
- Kondo, K. & Nakagawa, K. (2007). Speech emission control using active cancellation. *Speech Communication*, Vol. 49, No. 9, September 2007, pp. 687–696.
- Kuo, S. M. & Ji, M. J. (1995). Development and analysis of an adaptive noise equalizer. *IEEE Transactions Speech Audio Processing*, Vol. 3, May 1995, pp. 217–222.
- Kuo, S. M. & Yang, Y. (1996). Broadband adaptive noise equalizer. *IEEE Signal Processing Letters*, Vol. 3, No. 8, August 1996, pp. 234–235.
- Kuo, S. M. & Ji, M. (1996). Passband disturbance reduction in periodic active noise control systems. *IEEE Transactions Speech Audio Processing*, Vol. 4, No. 2, 1996, pp. 96–103.
- Kuo, S. M. & Morgan, D. R. (1996). *Active Noise Control Systems-Algorithms and DSP Implementations*, New York: Wiley.
- Kuo, S. M. & Tahernezahdi, M. (1997). Frequency-domain periodic active noise control and equalization. *IEEE Transactions Speech Audio Processing*, Vol. 5, 1997, pp. 348–358.
- Kuo, S. M. & Vijayan, D. (1997). A secondary path modeling technique for active noise control systems. *IEEE Transactions Speech Audio Processing*, Vol. 5, No. 4, July 1997, pp. 374–377.
- Kuo, S. M. & Morgan, D. R. (1999). Active noise control: A tutorial review. *Proceedings of IEEE*, Vol. 87, June 1999, pp. 943–973.
- Kuo, S. M. & Gan, W. S. (2004). Active noise control systems with optimized secondary path. *Proceedings IEEE International Conference Control Applications*, pp. pp. 765–770, September 2004.
- Kuo, S. M. & Mitra, S. & Gan, W. S. (2006). Active noise control system for headphone applications. *IEEE Transactions on Control Systems Technology*, Vol. 14, No. 2, March 2006, pp. 331–335.

- Kuo, S. M. & Puvvala, A. B. (2006). Effects of frequency separation in periodic active noise control systems. *IEEE Transactions Audio Speech Language Processing*, Vol. 14, No. 5, Sept. 2006, pp. 1857–1866.
- Kuo, S. M. & Gireddy, R. (2007). Real-time experiment of snore active noise control. *Proceedings of IEEE International Conference on Control Applications*, 1-3 Oct. 2007, pp. 1342–1346.
- Kuo, S. M. & Chakravarthy, S. R. & Gireddy, R. & Gupta, A. (2008). Experiment of active snore noise control systems. *Noise Control Engineering Journal*, Vol. 56, No. 1, Jan-Feb., 2008, pp. 16–24.
- Kuo, S. M. & Kuo, K. & Gan, W. S. (2010). Active noise control: Open problems and challenges. *Proceedings of International Conference on Green Circuits and Systems (ICGCS)*, pp. 164–169, Shanghai, 21-23 June 2010.
- Leahy, R. & Zhou, Z. & Hsu, Y. C. (1995). Adaptive filtering of stable processes for active attenuation of impulsive Noise. *Proceedings of IEEE International Conference Acoustic Speech Signal Processing (ICASSP)*, Vol. 5, pp. 2983–2986, May 1995.
- Liu, L. & Gujjula, S. & Thanigai, P. & Kuo, S. M. (2008). Still in womb: intrauterine acoustic embedded active noise control for infant incubators. *Advances in Acoustics and Vibration*, Vol. 2008, Article ID 495317, 9 pages, doi:10.1155/2008/495317.
- Lueg, P. (1936). Process of silencing sound oscillations. *US Patent*, 2043416, June 9, 1936.
- Morgan, D. R. (1980). An analysis of multiple correlation cancellation loops with a filter in auxiliary path. *IEEE Transactions Acoustic Speech Signal Processing*, Vol. ASSP-28, Aug. 1980, pp. 454–467.
- Nikias, C. L. (1995). *Signal Processing with Alpha-stable Distribution and Applications*, Wiley, New York.
- Omoto, A. & Morie, D. & Fujiwara, K. (2002). Behavior of adaptive algorithms in active noise control systems with moving noise sources. *Acoustical Science and Technology*, Vol. 23, No. 2, 2002, pp. 84–89.
- Park, Y. C. & Sommerfeldt, S. D. (1996). A fast adaptive noise control algorithm based on lattice structure. *Appl. Acoust.*, Vol. 47, No. 1, 1996, pp. 1–25.
- Reddy, E. P. & Das, D. P. & Prabhu, K. M. M. (2008). Fast adaptive algorithms for active control of nonlinear noise processes. *IEEE Transactions Signal Processing*, Vol. 56, No. 9, Sep. 2008, pp. 4530–4536.
- Saito, N. & Sone, T. (1996). Influence of modeling error on noise reduction performance of active noise control systems using filtered-x LMS algorithm. *Journal of Acoustical Society of Japan (E)*, Vol. 17, No. 4, Apr. 1996, pp. 195–202.
- Shao, M. & Nikias, C. L. (1993). Signal processing with fractional lower order moments: Stable processes and their applications. *Proceedings of IEEE*. Vol. 81, No. 7, 1993, pp. 986–1010.
- Strauch, P. & Mulgrew, B. (1998). Active control of nonlinear noise processes in a linear duct. *IEEE Transactions Signal Processing*, Vol. 46, No. 9, Sep. 1998, pp. 2404–2412.
- Sun, X. & Kuo, S. M. & Meng, G. (2006). Adaptive algorithm for active control of impulsive noise. *Journal of Sound and Vibration*, Vol. 291, No. 1–2, 2006, pp. 516–522.
- Sun, X. & Kuo, S. M. (2007). Active narrowband noise control systems using cascading adaptive filters. *IEEE Transactions Audio Speech Language Processing*, Vol. 15, No. 2, 2007, pp. 586–592.
- Widrow, B. & Stearns, S. D. (1985). *Adaptive Signal Processing*, Prentice Hall, New Jersey.

- Xiao, Y. & Ward, R. K. & Ma, L. & Ikuta, A. (2005). A new LMS-based Fourier analyzer in the presence of frequency mismatch and applications. *IEEE Transactions Circuits and Systems-I: Regular Papers*, Vol. 52, No. 1, Jan. 2005, pp. 230–245.
- Xiao, Y. & Ma, L. & Khorasani, K. & Ikuta, A. (2006). A new robust narrowband active noise control system in the presence of frequency mismatch. *IEEE Transactions Audio Speech Language Processing*, Vol. 14, No. 6, Nov. 2006, pp. 2189–2200.
- Zhang, M. & Lan, H. & Ser, W. (2001). Cross-updated active noise control system with online secondary path modeling. *IEEE Transactions Speech Audio Processing*, Vol. 9, No. 5, July 2001, pp. 598–602.

Active Noise Cancellation: The Unwanted Signal and the Hybrid Solution

Edgar Omar López-Caudana
Tecnológico de Monterrey, Campus Ciudad de México
Mexico

1. Introduction

The need to eliminate unwanted sounds is greater, as this is an expression of the limited tolerance that we have as individuals to the perception of sounds generated by industrial equipment, appliances and some general properties that are unpleasant for most people. Vibrations mechanical engines in operation, digging machinery and electricity generators, are just a few examples of processes or equipments that can produce signals nuisance to human ear.

While mitigating these unwanted sounds already exists today, the methods used are based on passive elements that offer a poor response to low frequency sounds. This drawback happens (Kuo & Morgan, 1996), when the wavelength of the signal is long compared to the size of the muffler liabilities. The relevance in the treatment of low-frequency sounds is that they produce fatigue and loss of concentration, thus affecting the performance of people, machinery and equipment present. That is, low-frequency sounds produced very intense vibrations that can fracture structures in the team very long periods of exposure.

ANC Systems must respond to changes in frequency of the primary noise with which they are working. In other words, the primary non-stationary noise vary, hence to get an acceptable cancellation we must use some kind of adaptive system, that carried out many operations at a high speed. To satisfy the requirements those ANC systems, it should be used digital signal processors (PDSs) compute based on adaptive algorithms and adaptive filters. The ability of an adaptive filter to operate and respond satisfactorily to an unknown environment, and variations that may be involved in signal reference, to make a powerful adaptive filter for signal processing and control applications. There are various types of adaptive filters but generally all share the characteristic of working with an input signal (input vector), and a desired response (output vector), these two signals are used to compute an estimate of error (error signal), which allows to control the coefficients of the adjustable filter.

In this chapter, it's proposed a new Hybrid active noise control (ANC) system. The main feature of the proposed system is the consideration of both, the online secondary path modeling and the acoustic feedback. The objective is to realize improved performance at a reasonable computational cost in a Hybrid ANC system that considers two of the more important troubles of the ANC. In the proposed system, two least mean square (LMS) adaptive filters are used in the noise control process, one for the Feedforward stage and the other for the Feedback stage, both of them use the same error signal as used in the

adaptation of the modeling filter (Romero et al, 2008). Moreover, the method of online secondary path modeling is an adaptation of the method proposed by (Akhtar et al, 2006) in which a variable step size (VSS-LMS) algorithm is used to adapt the modeling filter. With the VSS-LMS algorithm a small step size is required initially for the modeling filter and later its value is increased accordingly. The combination of the Feedback and Feedforward stages, gives to the system a solid robustness in frequency domain. Finally, the computer simulations are presented to show the success of the proposed system.

2. Adaptive filters

2.1 What is an adaptive filter?

An adaptive filter responds to changes in its parameters, like for example: its resonance frequency, input signal or transfer function that varies with time. This behavior is possible due to the adaptive filter coefficients vary over time and are updated automatically by an adaptive algorithm.

Therefore, these filters can be used in applications where the input signal is unknown or not necessarily stationary. An adaptive filter is made up of two parts: a digital filter and an adaptive algorithm. The block diagram of an adaptive filter is shown in Figure 1, we can see that the adaptive algorithm needs, two input signals, $x(n)$ and $e(n)$ as its references to set the parameters of the digital filter and update its coefficients.

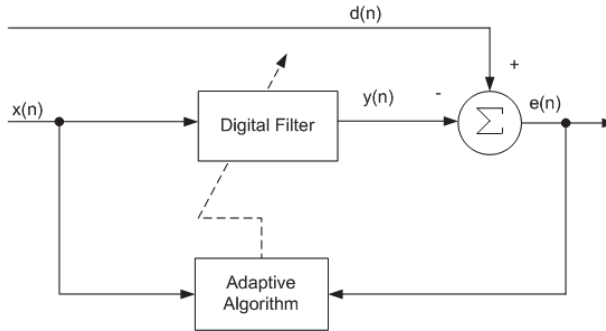


Fig. 1. Adaptive Filter

2.2 Least Mean Square algorithm

The Least Mean Square (LMS) algorithm is part of the stochastic gradient algorithm family that does not require the knowledge of autocorrelation function or inverse function arrays to update the coefficients values of an adaptive filter.

From Figure 1 we define that a Finite Impulse Response (FIR) adaptive filter with LMS algorithm can describes is update equation as

$$\bar{\mathbf{w}}(n+1) = \bar{\mathbf{w}}(n) + \mu \bar{\mathbf{x}}(n) e(n) \quad (1)$$

where $\bar{\mathbf{x}}(n) \equiv [x(n), x(n-1), \dots, x(n-L+1)]^T$, is the L sampled reference signal vector, and $x(n)$ is the reference signal obtained by the reference microphone, μ is the step size of the adaptive algorithm (Kuo & Morgan, 1996), $e(n)$ is the residual noise signal given as

$$e(n) = d(n) - y(n) \quad (2)$$

and finally $\bar{w}(n) \equiv [w_0(n), w_1(n), \dots, w_{L-1}(n)]^T$, is the tap-weight vector. LMS algorithm also has the advantage of being simple and requires no differences or averages from the input signal (Kuo & Morgan, 1996).

2.3 ANC Systems

2.3.1 Types of systems

2.3.1.1 Feedforward systems

Based on the application of identification of adaptive filters, the Feedforward system described in Figure 2, exemplifies this ANC system in an acoustic media, where unwanted signal emitted by the source noise is monitored by both the reference and error microphones and processed by the adaptive filter in the block of ANC.

Figure 3 shows in a simplified form a Feedforward ANC system where the digital filter $W(z)$ is used to estimate the unknown plant $P(z)$. It is assumed that the plant and the filter have the same input signal $x(n)$.

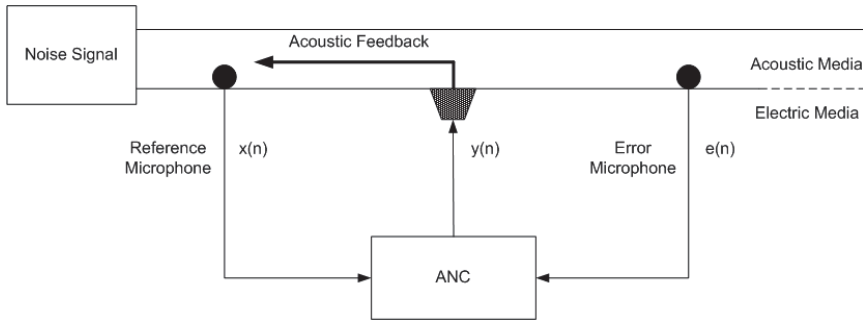


Fig. 2. Feedforward ANC Process

Moreover, we introduce the algorithm Filtered-X Least Mean Square (FXLMS), a variant of the LMS algorithm discussed in (Kuo & Morgan, 1996); FXLMS solve the problem of secondary path described as the set of transformations experienced by the signal response of the filter and adaptive signal mistake of passing a domain electric acoustic one. In this electroacoustic process, the signal can be delayed or altered, so that's why is necessary to minimize these effects. The technique of the algorithm FXLMS consists of placing a filter with the same characteristics as the secondary path in the reference signal that leads towards the adaptive filter, like is shown in Fig 3.

The filter $\hat{S}(z)$ in Figure 3, is the estimation of the secondary path defined by the filter $S(z)$. By taking this consideration, the actualization of the filter $W(z)$ becomes as:

$$\bar{w}(n+1) = \bar{w}(n) + \mu \bar{x}(n)e(n) \quad (3)$$

where:

$$\bar{x}(n) = \hat{s}(n) * x(n) \quad (4)$$

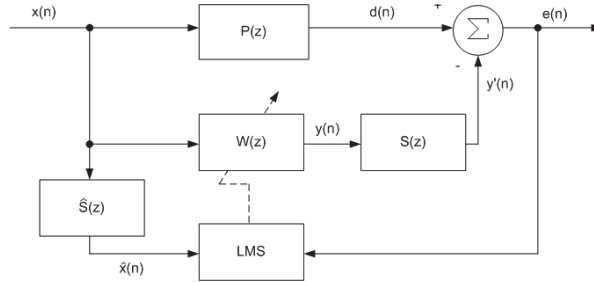


Fig. 3. Feedforward ANC System with FXLMS Algorithm

2.3.1.2 Feedback systems

There are some applications where it is not possible to take into account the reference signal from the primary source of noise in a Feedforward ANC system, perhaps because it is difficult to access to the source, or there are several sources that make it difficult to identify a specific one by the reference microphone. One solution to this problem is the one that introduced a system that predicts the input signal behavior, this system is known as the Feedback ANC system which is characterized by using only one error sensor and a secondary source (speaker) to achieve the noise control process.

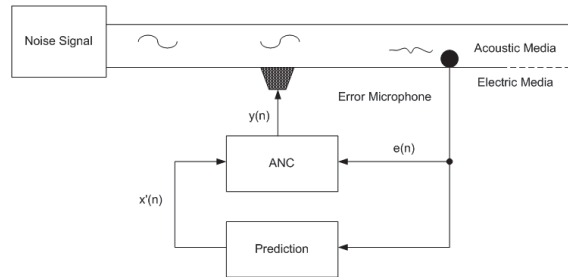


Fig. 4. Feedback ANC Process

Figure 5 describes a Feedback ANC system with FXLMS algorithm, in which $d(n)$ is the noise signal, $e(n)$ is the error signal defined as the difference between $d(n)$ and the $y'(n)$,

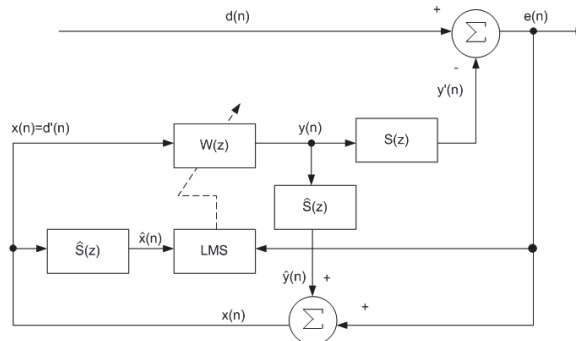


Fig. 5. Feedback ANC System with FXLMS Algorithm

the output signal of the adaptive filter once it already has crossed the secondary path. Finally, the input signal of the adaptive filter is generated by the addition of the error signal and the signal resulting from the convolution between the secondary path $\hat{S}(z)$ and the estimated output of adaptive filter $y(n)$.

2.3.1.3 Hybrid system

A Hybrid system consists of one identification stage (Feedforward) and one prediction (Feedback) stage. This combination of both Feedback and Feedforward systems needs two reference sensors: one related to the primary source of noise and another with the residual error signal.

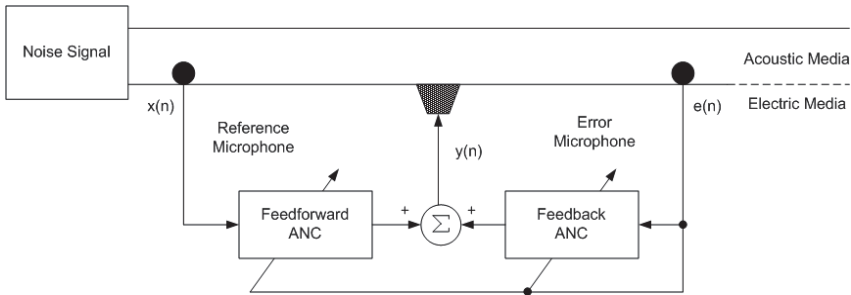


Fig. 6. Hybrid ANC Process

Figure 7 shows the detailed block diagram of an ANC Hybrid System in which it is possible to observe the basic systems (Feedforward, Feedback) involved in this design. The attenuation signal resulting from the addition of the two outputs $W(z)$ and $M(z)$ of adaptive filters is denoted by $y(n)$. The filter $M(z)$ represents the adaptive filter Feedback process, while the filter $W(z)$ represents the Feedforward process. The secondary path consideration in the basic ANC systems is also studied in the design of the Hybrid system and is represented by the transfer function $S(z)$.

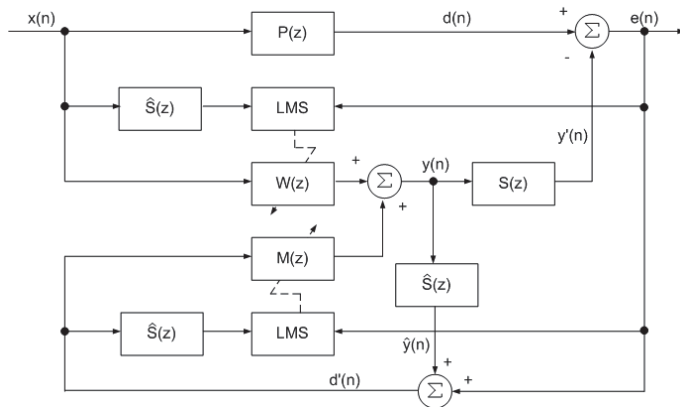


Fig. 7. Hybrid ANC System with FXLMS Algorithm

As we can see, the block diagram of the Hybrid ANC system from Figure 8 also employs the FXLMS algorithm to compensate the possible delays or troubles that the secondary path provokes.

2.3.2 ANC problematic

2.3.2.1 Acoustic feedback

This characteristic is present in an ANC Feedforward system; Figure 2 shows that the contribution of the attenuation signal $y(n)$, causes a degradation of the system response because this signal is present in the microphone reference. Two possible solutions to this problem are: the neutralization of acoustic feedback and the proposal for a Hybrid system that by itself has a better performance in the frequency range of work and the level of attenuation. To solve this issue we analyze a Hybrid system like shown in the Figure 8, where $F(z)$ represents the transfer function of the Feedback process.

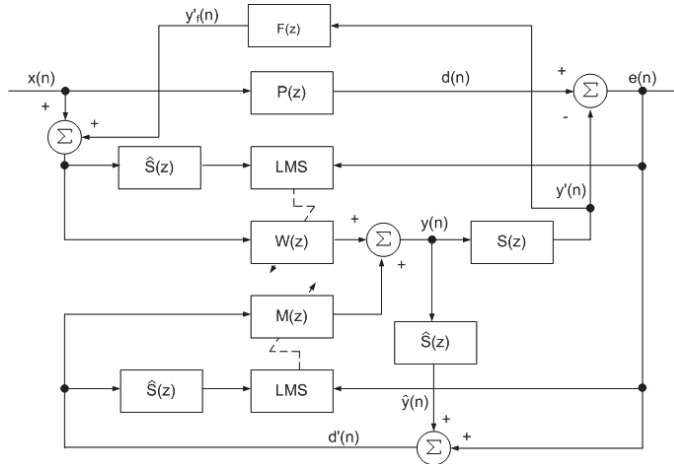


Fig. 8. Hybrid ANC System with Acoustic Feedback

2.3.2.2 Secondary path modeling

As previously mentioned, the process that makes the signal resulting from the adaptive filter $y(n)$ into $e(n)$, is defined as a secondary path. This feature takes in consideration, digital to analog converter, reconstruction filter, the loudspeaker, amplifier, the trajectory of acoustic loudspeaker to the sensor error, the error microphone, and analog to digital converter. There are two techniques for estimating the secondary path, both techniques have their tracks that offer more comprehensive and sophisticated methods in certain aspects, these techniques are: offline secondary path modeling and the online secondary path modeling. The first one is done by a Feedforward system where the plant now is $S(z)$ and the coefficients of the adaptive filter are the estimation of the secondary path, like shown in Figure 9:

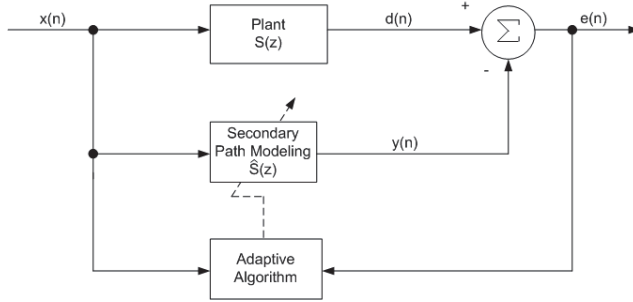


Fig. 9. Offline Secondary Path Modeling

For online secondary path modeling we study two methods: Eriksson's method (Eriksson et al, 1988) and Akhtar's method (Akhtar et al, 2006). Figure 10 shows the Eriksson's Method where first a zero mean white noise $v(n)$, which is not correlated with the primary noise is injected at the entrance to the secondary loudspeaker. Secondly, $x(n)$ represents the discrete output form reference microphone, also known as reference signal; $\bar{p}(n) = [p(n), p(n-1), \dots, p(n-L-N+1)]^T$ is the vector containing the impulse response of the primary path from the digital output microphone reference to the exit of the microphone error. The vector composed of the impulse response of the secondary path of the digital output of the loudspeaker secondary to the exit discrete microphone error is defined as $\bar{s}(n) = [s(n), s(n-1), \dots, s(n-L-N+1)]^T$. Moreover, the adaptive filter $\bar{w}(n)$ is in charge of the noise control process, and it is defined as $\bar{w}(n) \equiv [w(0), w(1), \dots, w(L-1)]^T$ where L represents the length of the filter. The signal $d(n)$ is output $p(n)$ due to $x(n)$; the signal that cancels, $y(n)$, is output of the noise control process due $x(n)$. It is important to consider the update of the coefficients of the secondary path filter defined as:

$$\bar{s}(n+1) = \bar{s}(n) + \mu_s \bar{v}(n) [v'(n) - \hat{v}(n)] + \mu_s \bar{v}(n) \eta(n) \quad (5)$$

where $v'(n) = v(n) * s(n)$ and $\hat{v}(n) = v(n) * \hat{s}(n)$; $*$ denotes convolution.

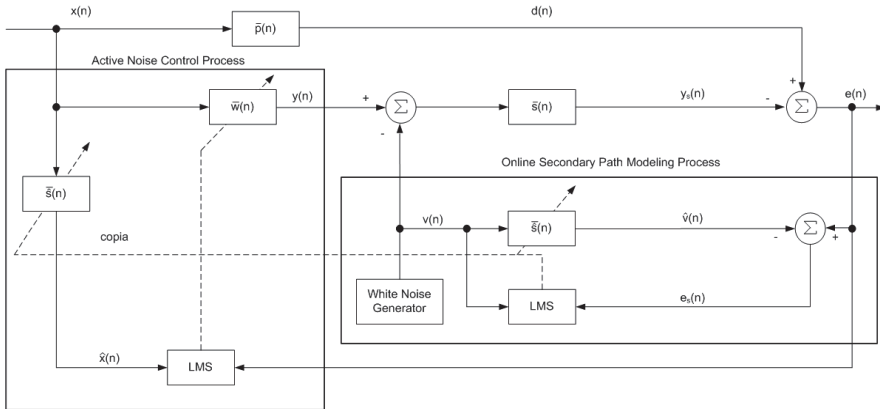


Fig. 10. ANC System with Online Secondary Path Modeling (Eriksson's Method)

For the Akhtar's method the noise control adaptive filter is updated using the same error signal that the adaptive filter that estimated the secondary path. At the same time, an algorithm LMS variable sized step (VSS-LMS) is used to adjust the filter estimation of the secondary path. The main reason for using an algorithm VSS-LMS responds to the fact that the distorted signal present at the desired filter response of the secondary path decreases in nature, ideally converge to zero. Ec. 6 describes the coefficients vector of the noise control filter as:

$$\begin{aligned}\bar{w}(n+1) = \bar{w}(n) + \mu_w [d(n)\bar{\hat{x}}(n) - y'(n)\bar{\hat{x}}(n)] \\ + \mu_w [v'(n) - \hat{v}(n)]\end{aligned}\quad (6)$$

Is important to realize that the contribution of the white noise, $v'(n)$ and $\hat{v}(n)$ is uncorrelated with the input signal $x(n)$, so the Akhtar's method reduces this perturbation in the coefficients vector of the filter $W(z)$ when the process of secondary path modeling is such that $\hat{S}(z) \approx S(z)$, in this moment, $v'(n) - \hat{v}(n) \approx 0$ and the noise control process is completely correlated.

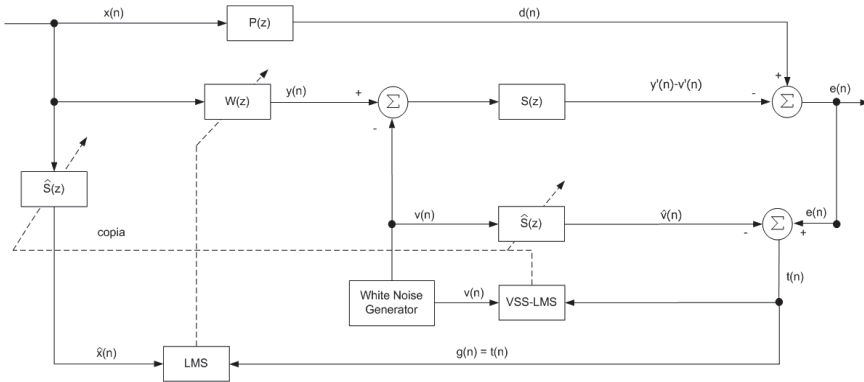


Fig. 11. ANC System with Online Secondary Path Modeling (Akhtar's Method)

2.3.3 Proposed Hybrid system

As a result of both considerations, the acoustic feedback and the online secondary path modeling, here we suggest a Hybrid ANC system with online secondary path modeling and acoustic feedback. The idea is to conceive a new robust system like the block diagram of the Figure 12 shows.

Its possible to observe from Figure 12 that the same signal, $a(n)$, is used as the error signal of the adaptive filter $W(z)$ which intervenes in the identification stage of the Feedforward system present in the proposed configuration. Also it's important to realize that in our design we have three FIR adaptive filters $W(z)$, $M(z)$ and $\hat{S}(z)$. The first one intervenes in the Feedforward process, $M(z)$ is part of the Feedback process; $\hat{S}(z)$ represents the online secondary path modeling adaptive filter. Finally the block $F(z)$ is the consideration of the acoustic feedback.

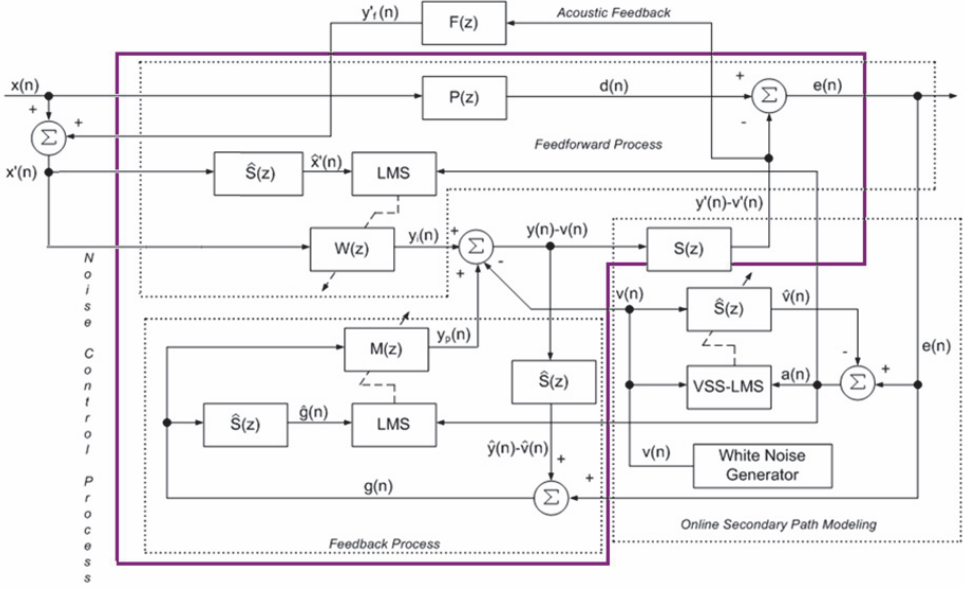


Fig. 12. A Hybrid Active Noise Control System with Online Secondary Path Modeling and Acoustic Feedback (Proposed System)

On the basis of the Figure 12, we can see that the error signal of all the ANC system is defined as:

$$e(n) = d(n) + [v(n) - y(n)] * s(n) \quad (7)$$

where $d(n)$ is the desired response, $v(n)$ is the white noise signal, $s(n)$ is the finite impulse response of the secondary path filter $S(z)$ and $y(n)$ is the resultant signal of the acoustic noise control process that achieves attenuate the primary noise signal and is defined as:

$$y(n) = y_i(n) + y_p(n) \quad (8)$$

where $y_i(n) = \bar{\mathbf{w}}^T(n) \bar{\mathbf{x}}'(n)$ represents the signal resultant of the Feedforward process, once again $\bar{\mathbf{w}}(n) \equiv [w_0(n), w_1(n), \dots, w_{L-1}(n)]^T$, is the tap-weight vector, $\bar{\mathbf{x}}'(n) \equiv [x'(n), x'(n-1), \dots, x'(n-L+1)]^T$ is the L sample reference signal vector of the Feedforward stage and $x'(n) = x(n) + y_f(n) - v_f(n)$ is the reference signal that already considers the effects of the acoustic feedback. By the way, as a result of the acoustic feedback consideration we expressed:

$$y_f(n) = y_i(n) * f(n) \quad (9)$$

$$v_f(n) = v(n) * f(n) \quad (10)$$

Both Ec. 9 and Ec. 10 contain $f(n)$, the finite impulse response of the acoustic feedback filter; moreover $y'(n)$ and $v'(n)$ are the signals that already have cross $S(z)$, the secondary path filter.

In the other and, for the Feedback stage we have that $y_p(n) = \bar{m}^T(n)\bar{g}(n)$ is the noise control signal for this process, where $\bar{m}(n) \equiv [m_0(n), m_1(n), \dots, m_{M-1}(n)]^T$ is the tap-weight vector of length M of the filter $M(z)$; $\bar{g}(n) \equiv [g(n), g(n-1), \dots, g(n-M+1)]^T$ is the sample reference signal for this adaptive filter and $g(n) = e(n) + \hat{y}(n) - \hat{v}(n)$ is the reference signal, where:

$$\hat{v}(n) = v(n) * \hat{s}(n) \quad (11)$$

$$\hat{y}(n) = y(n) * \hat{s}(n) \quad (12)$$

Once again as a result of the FXLMS algorithm, the Ec. 11 and Ec. 12 consider the signals $y(n)$ and $v(n)$ once both already have cross the estimation of the secondary path defined by $\hat{S}(z)$.

The advantages of using the Akhtar's method (Akhtar et al, 2006 and Akhtar et al, 2004), for the secondary path modeling in our proposed system are reflected in the VSS-LMS algorithm that allows the modeling process to select initially a small step size, $\mu_s(n)$, and increases it to a maximum value in accordance with the decrease in $[d(n) - y'(n)]$. If the filter $W(z)$ is slow in reducing $[d(n) - y'(n)]$, then step size may stay to small value for more time. Furthermore, the signal $a(n) = e(n) - \hat{v}(n)$ is the same error signal for all the adaptive filters involved in our system, $W(z)$, $M(z)$ and $\hat{S}(z)$, the reason to use this signal is that for $W(z)$, $[v'(n) - v(n)] < v'(n)$ compared with the Eriksson's method, so when $\hat{S}(z)$ converges as $\hat{S}(z) \approx S(z)$, ideally $v'(n) \approx v(n) \Rightarrow v'(n) - v(n) \rightarrow 0$. The bottom equations describe the update vector equations for the three adaptive filters:

$$\begin{aligned} \bar{w}(n+1) = & \bar{w}(n) + \mu_w \bar{\hat{x}}(n)[d(n) - y'(n)] \\ & + \mu_w \bar{\hat{x}}(n)[v'(n) - \hat{v}(n)] \end{aligned} \quad (13)$$

$$\begin{aligned} \bar{m}(n+1) = & \bar{m}(n) + \mu_m \bar{\hat{g}}(n)[d(n) - y'(n)] \\ & + \mu_m \bar{\hat{g}}(n)[v'(n) - \hat{v}(n)] \end{aligned} \quad (14)$$

$$\begin{aligned} \bar{s}(n+1) = & \bar{s}(n) + \mu_s \bar{v}(n)[v'(n) - \hat{v}(n)] \\ & + \mu_s \bar{v}(n)[d(n) - y'(n)] \end{aligned} \quad (15)$$

Although the Ec. 13 shows that when $\hat{S}(z)$ converges the whole control noise process of the system is not perturbed by the estimation process of $\hat{S}(z)$, it is significant to identify that the online secondary path modeling is degraded by the perturbation of $\eta(n) = \mu_s \bar{v}(n)[d(n) - y'(n)]$.

3. Performance indicators

3.1 Classical analysis

This section presents the simulation experiments performed to verify the proposed method. The modeling error was defined by Akhtar (Akhtar et al, 2006), as:

$$\Delta S(dB) = 10 \log_{10} \left[\frac{\sum_{i=0}^{M-1} [s_i(n) - \hat{s}_i(n)]^2}{\sum_{i=0}^{M-1} [s_i(n)]^2} \right] \quad (16)$$

First, an offline modeling was used to obtain FIR representations of tap weight length 20 for $P(z)$ and of tap weight length 20 for $S(z)$. The control filter $W(z)$ and the modeling filter $\hat{S}(z)$ are FIR filters of tap weight length of $L = 20$ both of them. A null vector initializes the control filter $W(z)$. To initialize $\hat{S}(z)$, offline secondary path modeling is performed which is stopped when the modeling error has been reduced to -5dB. The step size parameters are adjusted by trial and error for fast and stable convergence.

Case	Step Size: μ_w, μ_m	Step Size: μ_s
Case 1	0.01	(0.01 - 0.10)
Case 2	0.01	(0.01 - 0.15)
Case 3	0.01	(0.01 - 0.20)

Table 1. Filters Step Size Used in Classical Analysis

3.2 Proposed analysis

It is important to mention that the system is considered within the limitations of a duct, or one-dimensional waveguide, whose limitations are relatively easy to satisfy, as the distance between the control system and the primary sources is not very important. A duct is the simplest system, since it only involves one anti-noise source and one error sensor. (Kuo & Morgan, 1999). The amount of noise reduction will depend on the physical arrays of the control sources and the error sensors. Moving their positions affects the maximum possible level of noise reduction and the system's stability (the rate at which the controller adapts to system changes).

In order to decide which control system is the best, the properties of the noise to be cancelled must be known. According to (Kuo & Morgan, 1999), it is easier to control periodic noise; practical control of random or transitory noise is restricted to applications where sound is confined, which is the case of a duct.

The noise signals used for the purposes of this work are sorted into one of three types, explained next. This classification is used by several authors, amongst whom are (Kuo & Morgan, 1999) and (Romero et al, 2005), as well as companies such as (Brüel & Kjaer Sound & Vibration Measurement, 2008).

1. *Continuous or constant*: Noise whose sound pressure level remains constant or has very small fluctuations along time.
2. *Intermittent or fluctuant*: Noise whose level of sound pressure fluctuates along time. These fluctuations may be periodic or random.
3. *Impulsive*: Noise whose level of sound pressure is presented by impulses. It is characterized by a sudden rise of noise and a brief duration of the impulse, relatively compared to the time that passes between impulses.

Various articles on the subject of ANC were taken into consideration before establishing three main analysis parameters to determine the hybrid system's performance:

- a. Nature of the test signals; as far as the test signals are concerned, the system was tested with several real sound signals taken from an Internet database (Free sounds effects &

music, 2008). The sound files were selected taking into consideration that the system is to be implemented in a duct-like environment.

- b. Filter order; it is important to evaluate the system under filters of different orders. In this case, 20 and 32 coefficients were selected, which are low numbers given the fact that the distance between the noise source and the control system is not supposed to be very large. For 20th order filters, two cases were considered.
- c. Nature of the filter coefficients; on a first stage, the coefficients were normalized; this means that they were set randomly with values from -1 to 1. Next, the coefficients were changed to real values taken from a previous study made on a specific air duct (Kuo & Morgan, 1996).

Thus, the tests were carried out on three different stages:

1. Analysis with real signals and filters with 20 random coefficients;
2. Analysis with real signals and filters with 32 random coefficients; and
3. Analysis with real signals and filters with 20 real coefficients.

The simulation results are presented according to the following parameters:

1. Mean Square Error (MSE); and
2. Modeling error from online secondary path modeling.

Equation 17 shows the MSE calculation, given by the ratio between the power of the error signal, and the power of the reference signal.

$$MSE(dB) = 10 \log_{10} \left[\frac{\sum_{i=0}^{M-1} [e_i(n)]^2}{\sum_{i=0}^{M-1} [x_i(n)]^2} \right] \quad (17)$$

Equation 18 is the calculation for the Modeling error, given by the ratio of the difference between the secondary path and its estimation, and the secondary path as defined by Akthar (Akthar et al, 2006):

$$\Delta S(dB) = 10 \log_{10} \left[\frac{\sum_{i=0}^{M-1} [s_i(n) - \hat{s}_i(n)]^2}{\sum_{i=0}^{M-1} [s_i(n)]^2} \right] \quad (18)$$

4. Analysis of results

4.1 Classical references

In this cases, according bibliography, three sceneries are explained.

4.1.1 Case 1

Here the reference signal is a senoidal signal of 200Hz. A zero mean uniform white noise is added with SNR of 20dB, and a zero mean uniform white noise of variance 0.005 is used in the modeling process. Figure 13a shows the curves for relative modeling error ΔS , the corresponding curves for the cancellation process is shows in Figure 13b. In iteration 1000 it is performed a change on the secondary path.

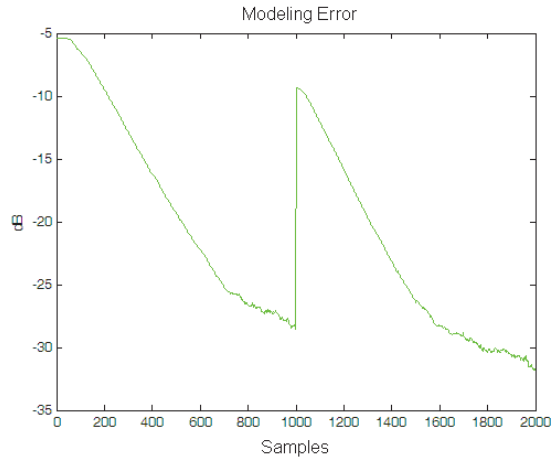


Fig. 13.a Relative Modeling Error

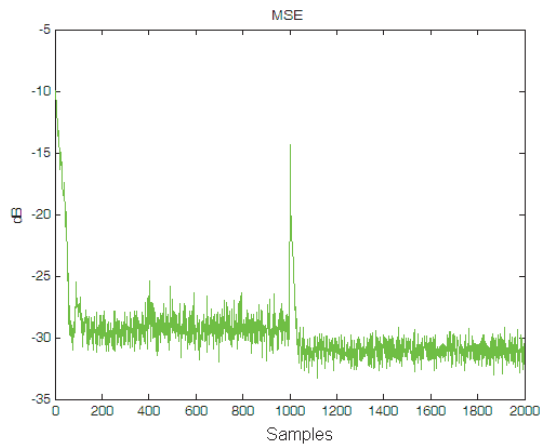


Fig. 13.b Attenuation Level

4.1.2 Case 2

In this case the reference signal is a narrow band sinusoidal signal with frequencies of 100, 200, 400, 600 Hz. A zero mean uniform white noise is added with SNR of 20dB, and a zero mean uniform white noise of variance 0.005 is used in the modeling process. The simulations results are shown in Figure 14a. In iteration 1000 it is performed a change con the secondary path.

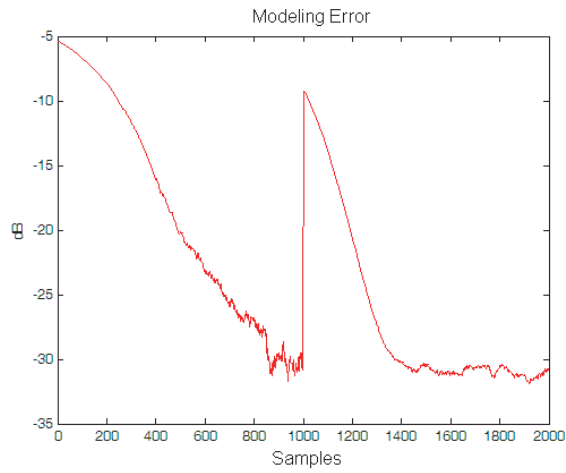


Fig. 14.a Relative Modeling Error

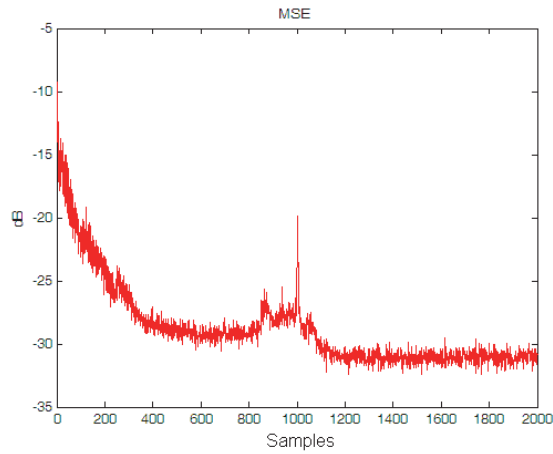


Fig. 14.b Attenuation Level

4.1.3 Case 3

Here we consider a motor signal for the reference signal. A zero mean uniform white noise of variance 0.005 is used in the modeling process. The simulation results are shown in Figure 15a. In iteration 1000 it is performed a change on the secondary path.

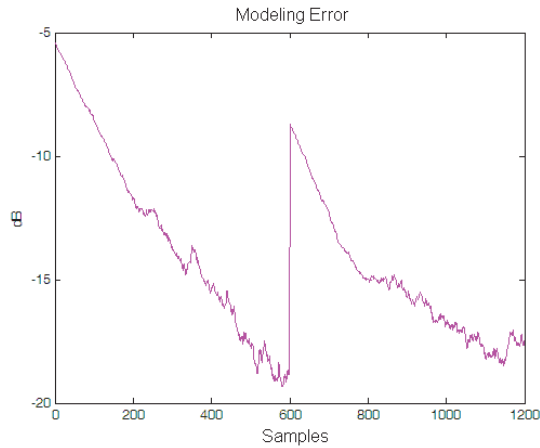


Fig. 15.a Relative Modeling Error

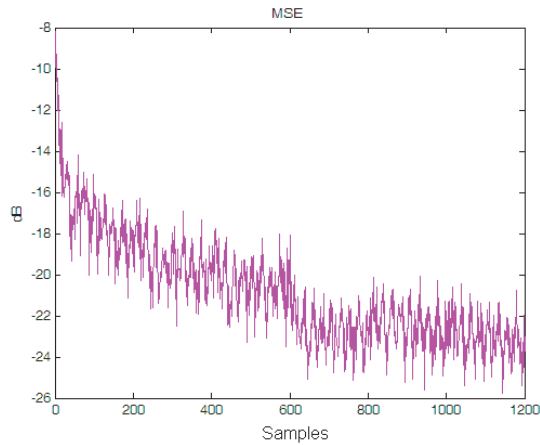


Fig. 15.b Attenuation Level

4.2 Proposed evaluation set

4.2.1 Test signal characterization

In order to characterize the hybrid system, several simulation tests were made with different real signals of each type described before. One signal of each type was selected to show the simulation results in this in this work. These three signals are the most representative case for each noise type.

First, each signal characterization will be shown, obtained through a program written in the simulation environment Matlab®. The graphs shown for each signal are: 1) Amplitude vs. Number of samples; 2) Amplitude vs. Frequency; and 3) Power vs. Frequency. Figure 16 shows the continuous signal, which corresponds to the audio of a vacuum cleaner in use. This signal has mainly low frequency components, and the power distribution is also found within low frequencies.

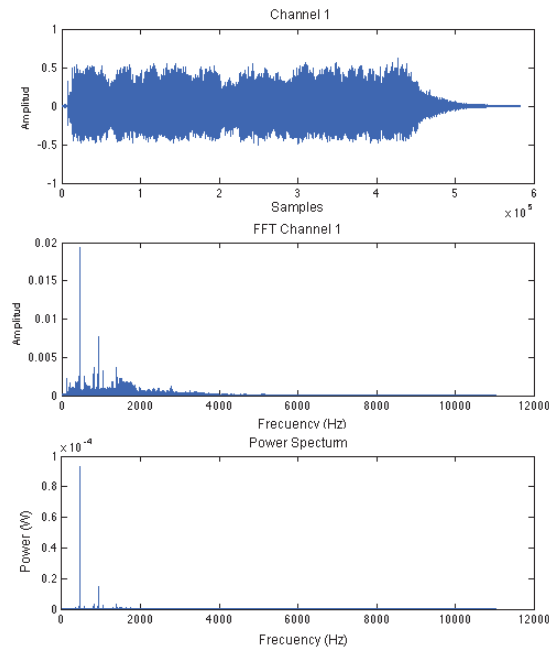


Fig. 16. Continuous Test Signal

Figure 17 shows the intermittent signal, which is the audio of a hand blender in use. This signal has relatively periodic fluctuations of different lengths. It could be considered a broadband signal because of the distribution of its frequency components, and its power is concentrated in low frequencies.

Finally, figure 18 presents the impulsive signal, given by the recording of some metallic objects falling down (a “crash” sound). There is an especially abrupt impulse by the end of the signal, which has mainly low frequency components and whose power is concentrated on low frequencies as well.

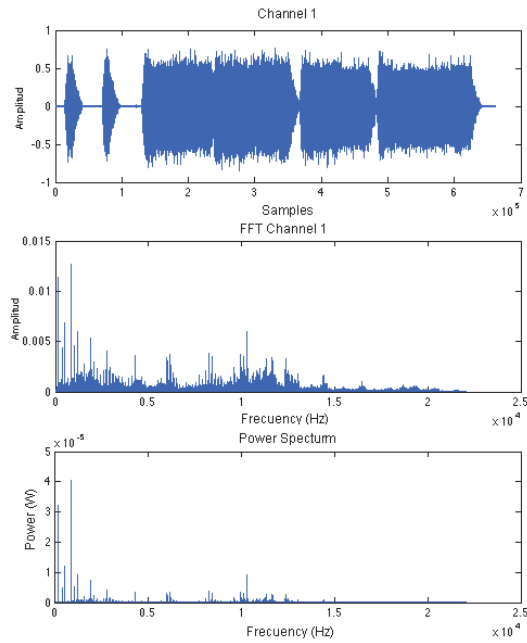


Fig. 17. Intermittent Test Signal

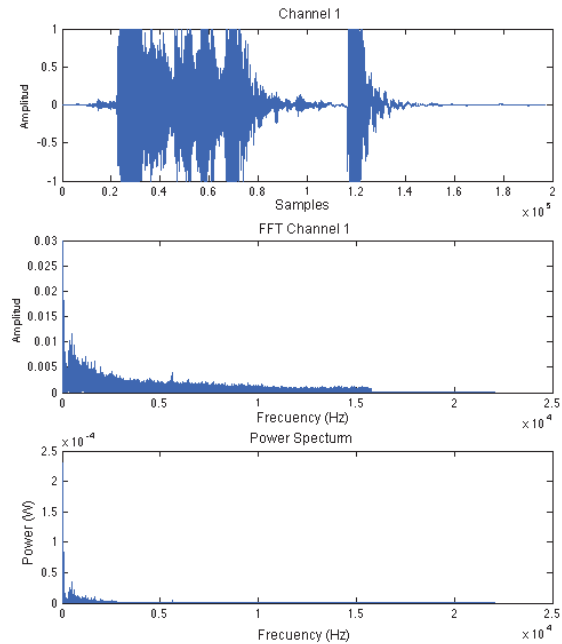


Fig. 18. Impulsive Test Signal

4.2.2 Filters with 20 random coefficients

The first tests were for 20th order filters with random coefficients. The optimum values of the step sizes μ_w and μ_m , belonging to the feedforward and feedback sections respectively, were established by trial and error. Table 2 shows the values used for each section's step size, as well as the range of values used for the step size in the secondary path filter (Lopez-Caudana et al, 2008).

Signal	Step size μ_w, μ_m	Step size μ_s
Continuous	0.001	0.01 – 0.15
Intermittent	0.001	0.01 – 0.15
Impulsive	0.0001	0.01 – 0.15

Table 2. Filters Step Size Used in Proposed Analysis

On each case, a white noise with mean zero and variance equal to 0.005 was used. Also, an abrupt change in secondary path was implemented on iteration 1000 out of 2000, to test the response of the system to such changes.

The signal that gave the best response was the continuous signal. Figure 19 shows the Modeling error, while Figure 20 shows the MSE for this case. These Figures show that the system successfully achieved stability and cancelled part of the input noise signal.

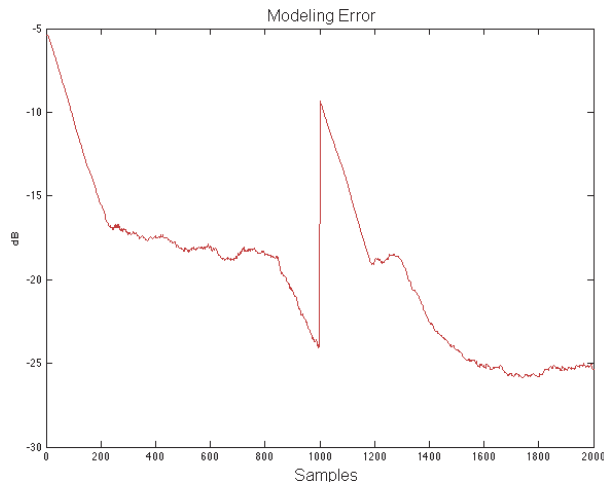


Fig. 19. Relative Modeling Error for Continuous Signal - Filters with 20 Random Coefficients

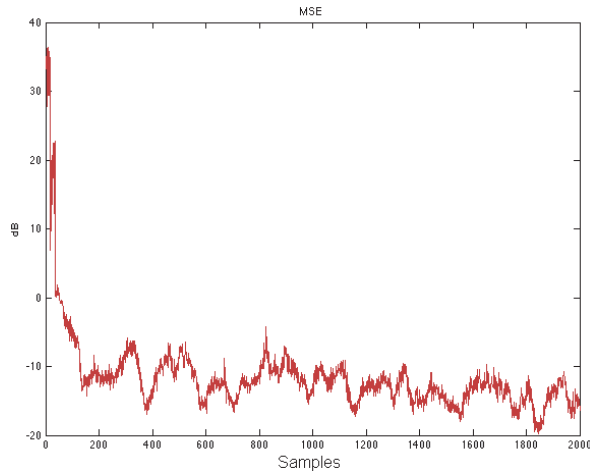


Fig. 20. MSE for Continuous Signal - Filters with 20 Random Coefficients

The response for the intermittent signal also achieved stability, despite the peaks that the signal presented at some samples, and managed to cancel part of the input noise signal as well. Figure 21 shows the Modeling error for the intermittent signal, while Figure 22 shows the MSE.

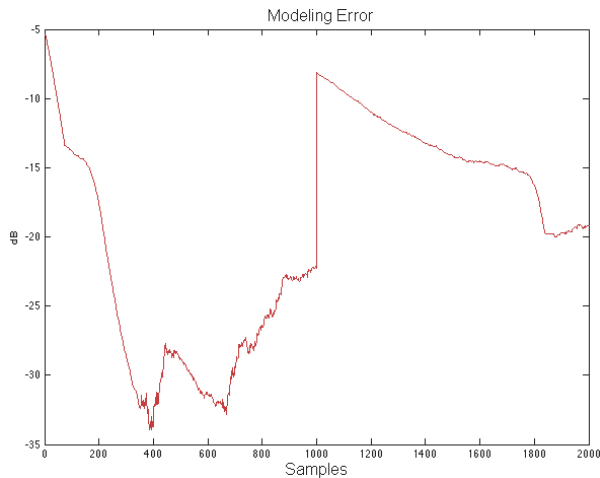


Fig. 21. Relative Modeling Error for Intermittent Signal - Filters with 20 Random Coefficients.

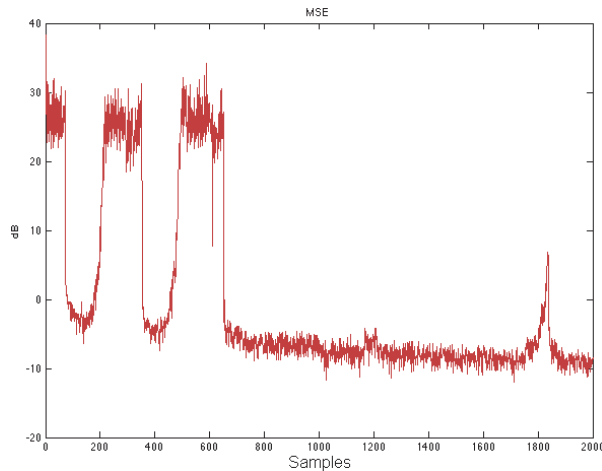


Fig. 22. MSE for Intermittent Signal - Filters with 20 Random Coefficients

However, the system presented more trouble stabilizing after the most abrupt impulse on the impulsive signal, and although it started converging, it could not cancel noise past that significant change.

4.2.3 Filters with 32 random coefficients

The next step in our set of tests was to increase the order of the filters, which means the system is taking into consideration a larger number of the duct's properties. Once again, the values of the coefficients are random from -1 to 1.

Table 3 shows the values used for the feedforward and feedback step sizes, as well as the range of step sizes for the secondary path filter. This values were established parting from the previous test's values and, if necessary, were adjusted by trial and error.

Signal	Step size μ_w, μ_m	Step size μ_s
Continuous	0.0001	0.001 – 0.05
Intermittent	0.001	0.01 – 0.15
Impulsive	0.00005	0.0001 – 0.05

Table 3. Filters Step Size Used in Proposed Analysis

Once again, a white noise with zero mean and variance equal to 0.005 was used for the three cases. An abrupt change in secondary path was done on iteration 1000 out of 2000.

Figure 23 shows the Modeling error response for the continuous signal, whereas Figure 24 shows the MSE for the same case.

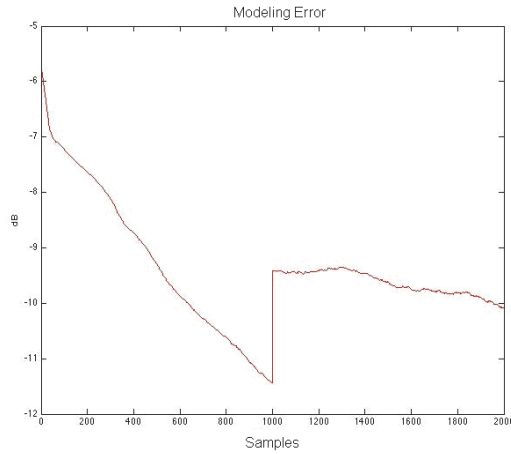


Fig. 23. Relative Modeling Error for Continuous Signal - Filters with 32 Random Coefficients

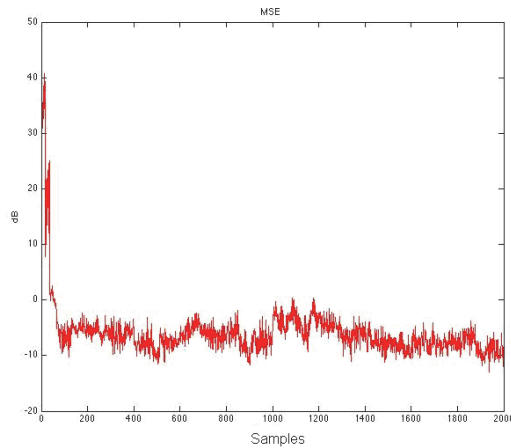


Fig. 24. MSE for Continuous Signal - Filters with 32 Random Coefficients

It can be observed that the system was, again, able to achieve stability as well as cancel noise. The step size was reduced by an order of ten in this case, probably due to the fact that it is a more accurate analysis than the previous test, because of the larger filter order. In the case of the intermittent input signal, the step size values did not need to be altered, and the

hybrid system achieved both stability and noise cancellation. Figures 25 and 26 show the response for the Modeling error and the MSE of the intermittent signal, respectively.

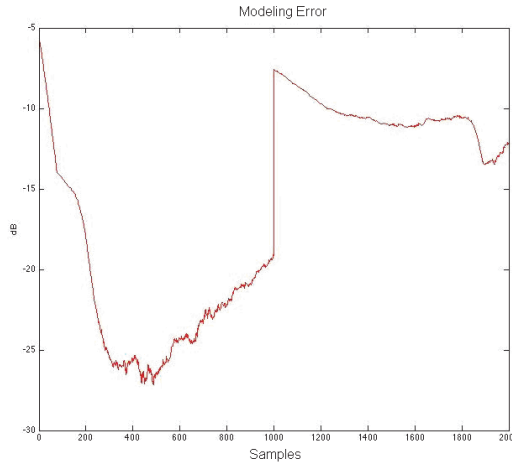


Fig. 25. Relative Modeling Error for Intermittent Signal - Filters with 32 Random Coefficients

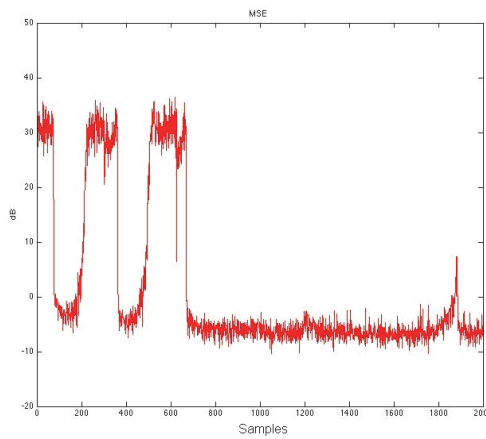


Fig. 26. MSE for Intermittent Signal - Filters with 32 Random Coefficients

Finally, for the impulsive input signal, a similar behavior to the previous test may be observed. The system took a very long time to start converging after the most abrupt impulse and was not able to cancel noise.

4.2.4 Filters with 20 real coefficients

The last set of tests that were made involved the use of 20th order filters with real coefficients. These coefficients were taken from the work done in (Lopez-Caudana et al, 2009) to determine the values of the primary and secondary path filters for an air duct. Table 4 shows the values used for the feedforward and feedback step sizes, as well as the range of step sizes used for the secondary path filter. The values were set by trial and error, starting with the values that were determined with the previous test.

Signal	Step size μ_w, μ_m	Step size μ_s
Continuous	0.000001	0.0001 - 0.001
Intermittent	0.000001	0.0001 - 0.001
Impulsive	0.000001	0.0001 - 0.001

Table 4. Filters Step Size Used in Proposed Analysis

For each of the three cases, a white noise with zero mean and variance equal to 0.05 was used in the system. Since there were not enough resources to implement an abrupt secondary path change (which means there was only one set of values available for the secondary path filter from (Lopez-Caudana et al, 2008)), a gradual change was made, given by the sum of a sinusoidal function to the secondary path coefficients, from iteration 1000 to 1100. Since the best response was shown by the continuous signal, Figure 27 shows the Modeling error for this case, while Figure 28 shows the MSE.

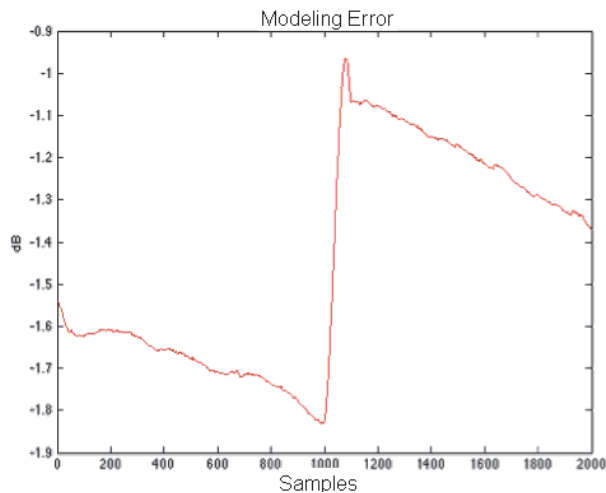


Fig. 27. Relative Modeling Error for Continuous Signal - Filters with 20 Real Coefficients

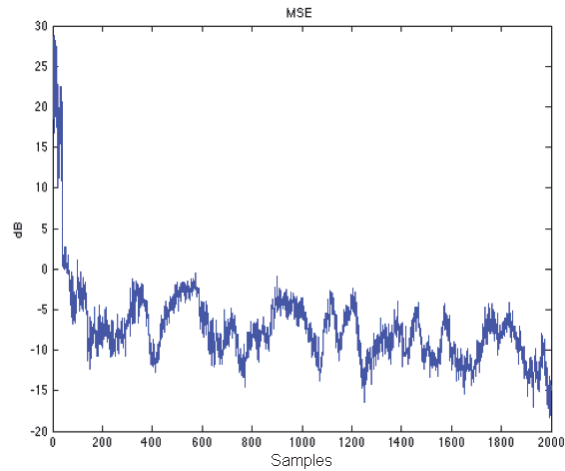


Fig. 28. MSE for Continuous Signal - Filters with 20 Real Coefficients

From Table 4, it is noticeable that the step sizes had to be considerably reduced, about an order of 1000 in comparison to the values established for the tests with 20 random coefficients. This is due to the fact that the coefficient values are not necessarily within a range of -1 to 1, so the secondary path modeling needs a smaller step size to be able to achieve a point of convergence.

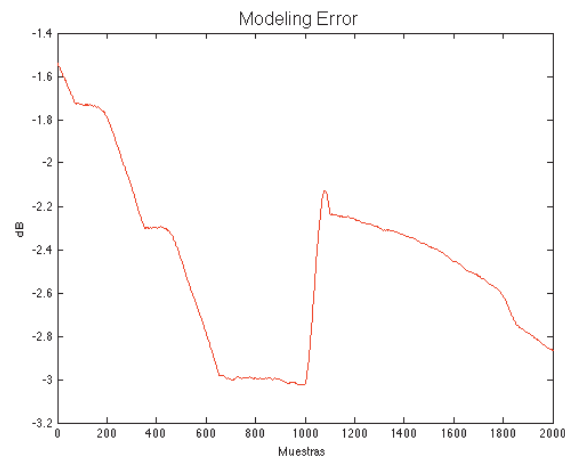


Fig. 29. Relative Modeling Error for Intermittent Signal - Filters with 20 Real Coefficients

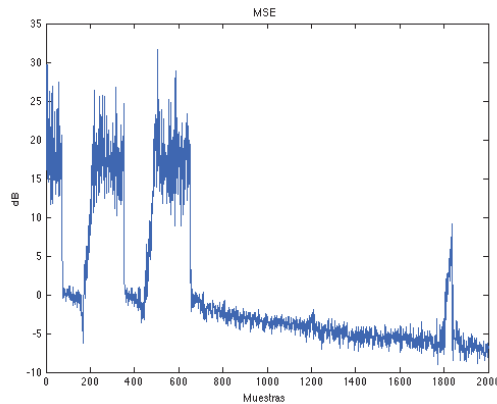


Fig. 30. MSE for Intermittent Signal - Filters with 20 Real Coefficients

According to figure 27 and 28, the system needed more time to converge to a certain value, compared with the results from the previous tests. This is because the values of the step sizes are very small and cause the convergence process to go slower and, also, the level of noise cancellation to be reduced. For the intermittent signal, the effects of the small step sizes were similar: the system took more time to converge and the level of noise cancellation was reduced. Nonetheless, the response did achieve stability at some point during the simulation. Figure 29 and Figure 30 correspond to the Modeling error and MSE for the intermittent signal, respectively.

However, in the case of the impulsive input signal the results were not as good as they had been expected. Due to the fact that there are very abrupt changes in the signal amplitude, and the step size is very small, there comes a point where the values of the coefficients tend to infinity and the simulation stops at about iteration 200.

4.3 An special case: the analysis of hybrid system versus neutralization system

4.3.1 Acoustic feedback path modeling

It is important to bring attention to the most common way to eliminate acoustic feedback, which is to make an online path modeling, like indicated on (Kuo & Morgan, 1999) and, more recently, in relevant work like (Akthar et al, 2007) . However, one of the main characteristics of the hybrid system presented in (Lopez-Caudana et al, 2008), is that it does not take the secondary path modeling into consideration, but instead takes advantage of the inherent robustness of hybrid systems when it comes to acoustic feedback.

The system in Figure 31, proposed by Kuo in (Kuo, 2002), was used to compare the robustness of the HANC system against the neutralization system.

The details of the system in Figure 7 may be consulted in (Kuo, 2002), however, an important fact of this system is that it uses additive noise for modeling. Also, as mentioned in (Akthar et al, 2007), it has some limitations in reference to predictable noise sources.

System	Primary Path μ_P	Secondary Path μ_S	Feedback Path μ_F
Neutralization System	0.000001	0.00005	0.00005
Hybrid System	0.001	0.001	

Table 5. Filters Step Size Used in Proposed Analysis

4.3.3 Experimental results

The performance of the systems is shown next, graphing the mean square error as a typical measurement for these kinds of systems when measuring the power of the error output, in dB (Kuo & Morgan, 1999). The analysis cases are as follows.

4.3.3.1 Analysis with different input signals

This is the longest case due to the fact that it shows the systems' performance for each mentioned signal. All paths have the same order (25 coefficients), taking critical conditions in a real duct for the analyzed phenomenon. Figure 32 to Figure 37 shows the result of the systems analysis with the previously mentioned set of signals. All results are shown in dBs, measuring the error power at the output (Mean Square Error).

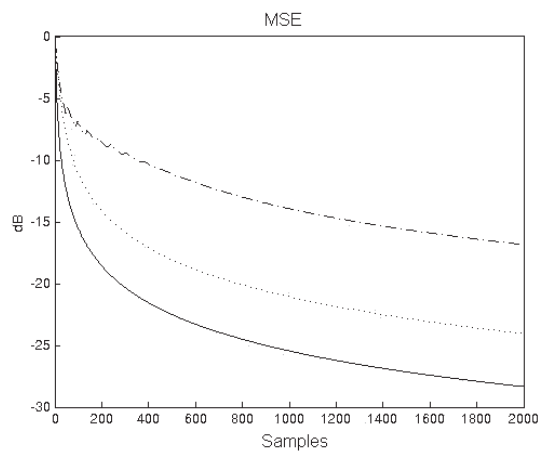


Fig. 32. MSE with “sinusoidal” reference signal: Hybrid System; Neutralization; Feedforward

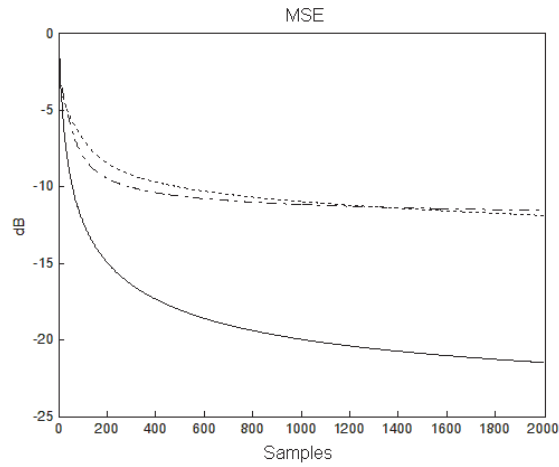


Fig. 33. MSE with “4 tones” reference signal: Hybrid System; ~~Neutralization~~; ~~Feedforward~~

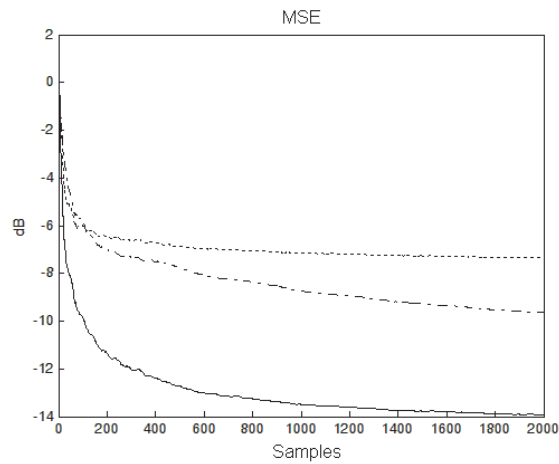


Fig. 34. MSE with “Motor” reference signal: Hybrid System; ~~Neutralization~~; ~~Feedforward~~

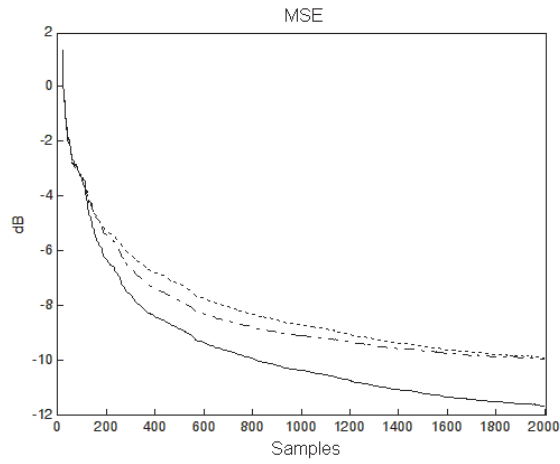


Fig. 35. MSE with “Airplane” reference signal: Hybrid System; ~~Neutralization~~; ~~Feedforward~~

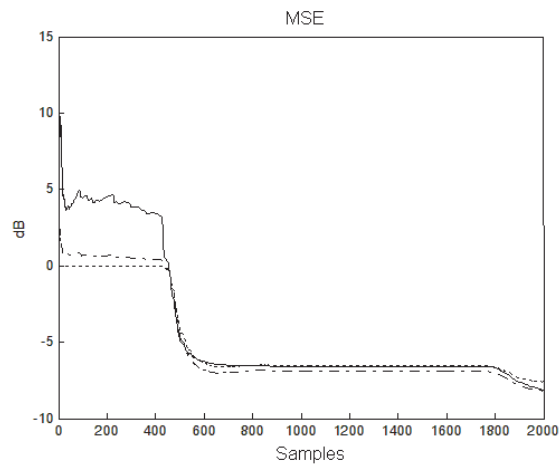


Fig. 36. MSE with “Snoring” reference signal: Hybrid System; ~~Neutralization~~; ~~Feedforward~~

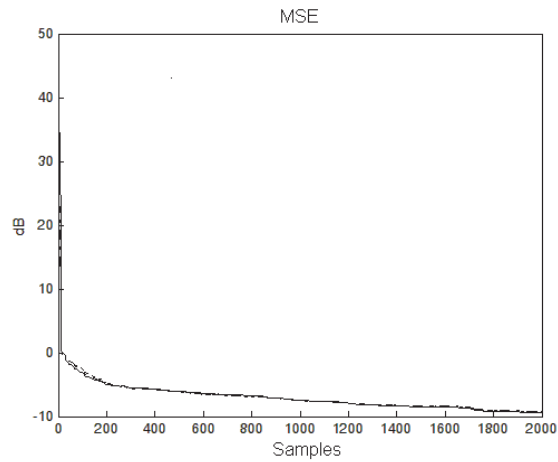


Fig. 37. MSE with “Street” reference signal: Hybrid System; Neutralization; Feedforward

4.3.3.2 Analysis of systems with random path’s coefficient values

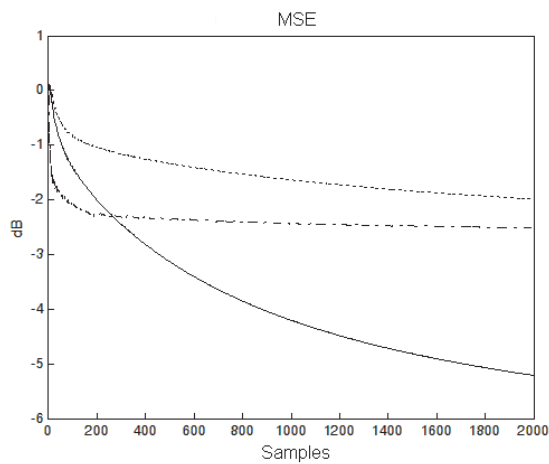


Fig. 38. MSE with “4 tones” reference signal: Hybrid System; Neutralization; Feedforward

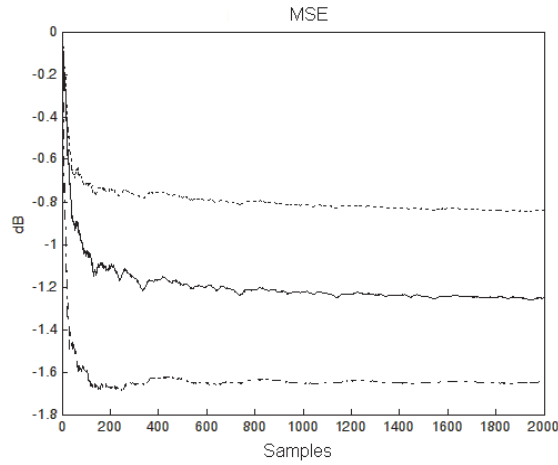


Fig. 39. MSE with “motor” reference signal: Hybrid System; Neutralization; Feedforward

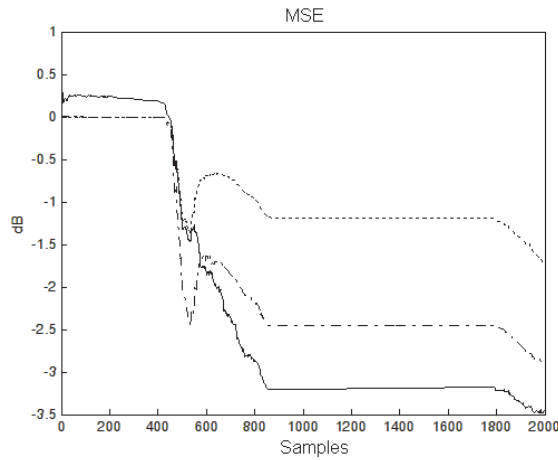


Fig. 40. MSE with “snoring” reference signal: Hybrid System; Neutralization; Feedforward

For this case, the values for the paths proposed by Kuo in (Kuo & Morgan, 1996), were changed for random values between 0.5 and -0.5, as done in (Kuo & Morgan, 1996). These random values generate broader spectra since they do not have such a homogeneous behavior as the experimental values from Kuo (Kuo & Morgan, 1996). Thus, the same conditions from Case 1 were kept, except for the paths' values. We only show the results of

three signals, considered the most representative from the set of tests done for this paper: “4 tones”, “motor”, and “snoring” (classified as continuous and intermittent signals, as mentioned previously).

4.3.3.3 Length of the paths under analysis

In this section the order of the paths is modified in order to analyze the systems’ performance, changing them to 32, 12, and 22 coefficients for the primary, secondary, and feedback path. The values of the paths are the same proposed in (Lopez-Caudana et al, 2009). Only the HANC and neutralization systems are analyzed.

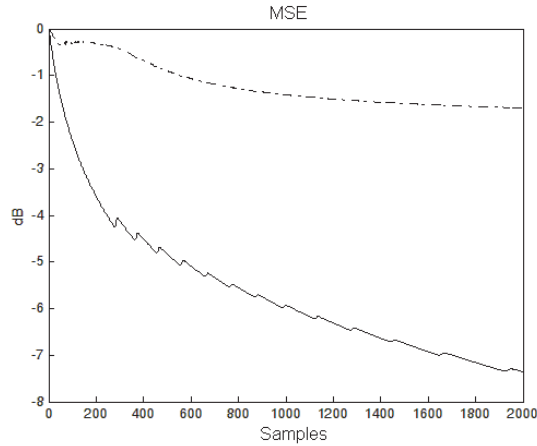


Fig. 41. MSE with “4 tones” reference signal: Hybrid System; Neutralization

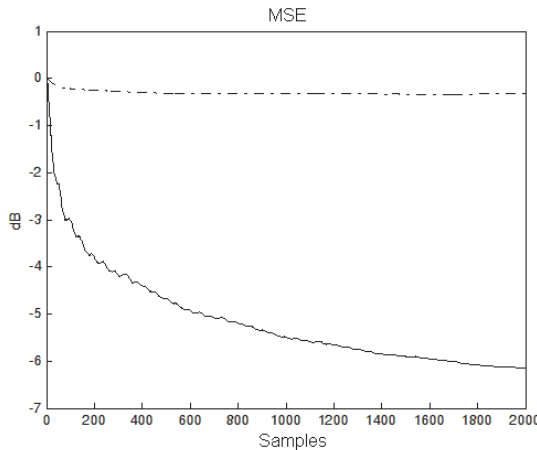


Fig. 42. MSE with “motor” reference signal: Hybrid System; Neutralization

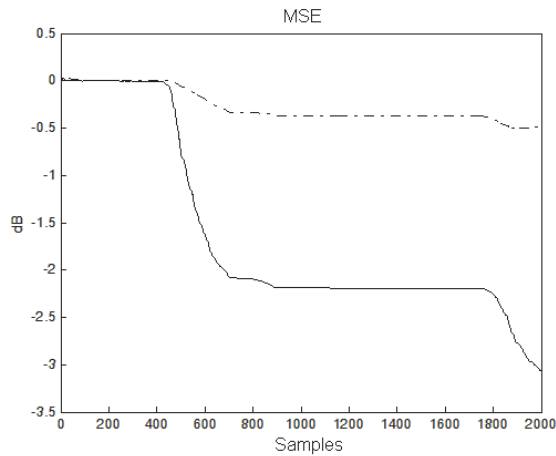


Fig. 43. MSE with “snoring” reference signal: Hybrid System; Neutralization

4.3.3.4 Change in secondary path

An important characteristic of ANC systems is that they must be capable of secondary path online modeling, which is observed in the graphs 44. There is an abrupt secondary path change in the thousandth iteration – taken from (Lopez-Caudana et al, 2008)– which does not causes the behavior of either system to destabilize when the values for a new secondary path appear.

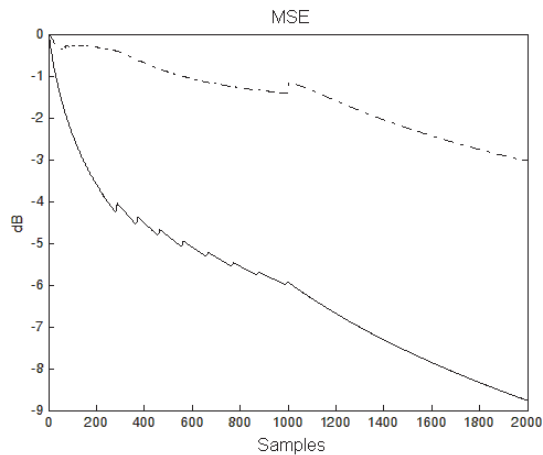


Fig. 44. MSE with “4 tones” reference signal: Hybrid System; Neutralization

5. Conclusions

In this chapter, we have showed a new system with a robustness structure that combines both the Feedforward and Feedback ANC basics systems to improve the frequency range of work. Additionally the consideration of the acoustic feedback and the online secondary path modeling allows the system to be adjustable for any kind of secondary path change (gradual ideally). The computer simulations demonstrate proposed system gives improved performance, at somewhat increased computational cost because the Akhtar's online secondary path method, but this method compensates the noise control process for the Feedforward and Feedback stages.

In the noise control process, two adaptive filters are incorporated with the FXLMS algorithm, one for identification (Feedforward) and other for prediction (Feedback). Both of them achieve a signal that is the result of the noise control process. A VSS-LMS algorithm is used in the secondary path modeling to achieve the fastest optimal solution of the modeling filter. Finally computer simulation results shown support these conclusions and demonstrate the effectiveness of the proposed system.

By other hand, this work successfully showed a set of analysis parameters to characterize the performance of a hybrid ANC system with acoustic feedback and the online secondary path estimation, outsider of an ideal simulation environment.

Different parameters were considered, such as the order of the paths, and the abrupt secondary path change, to achieve a fuller analysis. Also, the hybrid systems was subject not only to those reference signals considered 'classical', but to a set of real noise recordings, which allowed to analyze more complete characteristics from a system, while making it real. In the process of noise control, two adaptive filters were incorporated with the FXLMS algorithm, one for identification (Feedforward), and one for prediction (Feedback). Even though an adaptive filter for the acoustic feedback path was not considered, as done for systems designed to specifically attack this problem, the performance of the hybrid system is more efficient, compared to a system that is set out to solve that particular problem. Also, the, the computer simulation results show the support for these conclusions and confirm the effectiveness of the proposed system.

Based on the presented classification of noise signals we can conclude that:

1. The system has an efficient response to continuous signals, showing a stable response as well as considerable noise cancellation;
2. The response to intermittent signals is acceptable in general, though it depends to a certain extent on the waveform of the input noise signal. However, the system converges to a certain value and is able to cancel noise, though less than in the case of the continuous signals; and
3. In reference to impulsive signals, the system achieves a convergent point slowly, but cannot cancel noise significantly because the change in the signal amplitude is too abrupt for the system to adapt the filter coefficients efficiently enough.

When tested with real-value filter coefficients, the system presented a decrease on its convergence speed, which affected the performance of the system mainly in the case of an impulsive input noise signal. Finally, it must be considered that the results presented for a real-value filter coefficients refer to only one specific kind of duct. This means that the response could probably improve in a different environment or in a duct with different properties. This presents a problem for the designer of a hybrid ANC system, as it means that, for each environment where the system is to be implemented; there would be the need

to accurately individualize the parameters to achieve the desired response. However difficult, this may not be impossible to do, so there is still a lot of work to be done with hybrid ANC systems.

6. Acknowledgement

The contributions of several students from Communications and Electronic Engineering from Tecnológico de Monterrey, Mexico City Campus, are gratefully acknowledged and the guidance from Dr. Hector Perez-Meana from IPN SEPI ESIME CULHUACAN. This work has been supported by Mechatronic's Department of the Engineering and Architecture School from Tecnológico de Monterrey, Mexico City Campus.

7. References

- Akhtar Muhammad Tahir , Masahide Abe, and Masayuki Kawamata (2004). "Modified-filtered-x LMS algorithm based active noise control system with improved online secondary-path modeling" in *Proc. IEEE 2004 Int. Mid. Symp. Circuits Systems (MWSCAS2004), Hiroshima, Japan, Jul. 25-28, 2004*, pp. I-13-I-16, 2004.
- Akhtar, M.T.; Abe, M.; Kawamata, M., (2006). "A new variable step size LMS algorithm-based method for improved online secondary path modeling in active noise control systems". *IEEE Transactions on Audio, Speech, and Language Processing*, Volume 14, Issue 2, March 2006 Page(s):720 – 726.
- Akhtar et al. (2007) Muhammad Tahir Akhtar, M. Tufail, Masahide Abe, y Masayuki Kawamata. "Acoustic feedback neutralization in active noise control systems" *IEICE Electronics Express*. Vol. 4, No. 7, pp. 221 - 226.
- Akhtar et al. (2007) Muhammad Tahir Akhtar, Masahide Abe, y Masayuki Kawamata. "On active Noise Control Systems with Online Acoustic Feedback Path Modeling" in *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 2, February 2007 pp. 593-599.
- Eriksson L. J., Allie M. C., y Bremigan C. D. (1998), Active noise control using adaptive digital signal processing, *Proc. ICASSP*, , pp. 2594-2597.
- Kuo Sen M, Dennis R. Morgan (1999). "Active Noise Control Systems: A tutorial review" *Proc. IEEE*, vol. 87, no. 6, pp. 943-973, Junio 1999.
- Kuo Sen M, Dennis R. Morgan (1996). "Active Noise Control Systems: Algorithms and DSP Implementations" New York: Wiley Series in Telecommunications and Signal Processing Editors, 1996.
- Kuo Sen M, (2002) "Active Noise Control System and Method for On-Line Feedback Path Modeling" *US Patent 6,418,227*, Julio 9, 2002.
- Lopez-Caudana Edgar, Pablo Betancourt, Enrique Cruz, Mariko Nakano Miyatake, Hector Perez-Meana (2008). "A Hybrid Active Noise Canceling Structure", *International Journal of Circuits, Systems and Signal Processing*. Issue 2 Vol 2, 2008. pp 340-346.
- Lopez-Caudana Edgar, Pablo Betancourt, Enrique Cruz, Mariko Nakano Miyatake, Hector Perez-Meana (2008) "A Hybrid Noise Cancelling Algorithm with Secondary Path Estimation" *WSEAS TRANSACTIONS on SIGNAL PROCESSING* Issue 12, Volume 4, December 2008.
- Lopez-Caudana, E.; Betancourt, P.; Cruz, E.; Nakano-Miyatake, M.; Perez-Meana, H., (2008). "A hybrid active noise cancelling with secondary path modeling", *Circuits and*

- Systems, 2008. MWSCAS 2008. 51st Midwest Symposium on* . 10-13 Aug. 2008
Page(s):277 – 280.
- Lopez-Caudana Edgar, Paula Colunga, Alejandro Celis, Maria J. Lopez, and Hector Perez-Meana (2009). "Evaluation of a Hybrid ANC System with Acoustic Feedback and Online Secondary Path Modeling". *19th International Conference on Electronics, Communications and Computers 2009*, Cholula, Puebla. 26-28 Febrero de 2009.
- Lopez-Caudana, Edgar, Paula Colunga, Rogelio Bustamante and Hector Perez-Meana, (2010). "Evaluation for a Hybrid Active Noise Control System with Acoustic Feedback". *53rd IEEE Int'l Midwest Symposium on Circuits & Systems* , tSeattle, Washington from August 1-4, 2010.
- Nakano M., H. Perez (1995), A Time Varying Step Size Normalized LMS Algorithm for Adaptive Echo Canceler Structure, *IEICE Trans. on Fundamentals of Electronics Computer Sciences*, Vol. E-78-A, 1995, pp. 254-258.
- Romero, A; Perez-Meana, H.; Lopez-Caudana, E. (2008); "A Hybrid Active Noise Canceling Structure", *International Journal of Circuits, Systems and Signal Processing*. Issue 2 Vol 2, 2008. pp 340-346.
- Romero, Nakano-Miyatake, Perez-Meana (2008), A Hybrid Noise Canceling Structure with Secondary Path Estimation, *WSEAS Recent Advances in Systems, Communications and Computers*, 2008, pp.194-199.
- <http://www.freesfx.co.uk/soundeffectcats.html>
Free Sound Effects, Samples & Music. *Free Sound Effects Categories*. Visited on July 3, 2008.
- <http://www.nonoise.org/library/envnoise/index.htm>
Brüel & Kjær Sound & Vibration Measurement A/S. *Environmental Noise Booklet*. Visited on March 25, 2008.

Perceptual Echo Control and Delay Estimation

Kirill Sakhnov, Ekaterina Verteletskaya and Boris Simak
*Czech Technical University in Prague
Czech Republic*

1. Introduction

Echo phenomenon has been always existed in telecommunications networks. Generally it has been noticed on long international telephone calls. As technology advances and the data transmission methods tend more to packet-switching concepts, the traditional echo problem remained up-to-date. An important issue in echo analysis is a round-trip delay of the network. This is a time interval required for a signal from speaker's mouth, across the communication network through the transmit path to the potential source of the echo, and then back across the network again on the receive path to the speaker's ear. The main problem associated with IP-based networks is that the round-trip delay can be never reduced below its fundamental limit. There is always a delay of at least two to three packet sizes (50 to 80 ms) (Choi et al., 2004) that can make the existing network echo more audible (Gordy & Goubran, 2006). Therefore, all Voice over IP (VoIP) network terminals should employ echo cancellers to reduce the amplitude of returning echoes. A main parameter of each echo canceller is a length of its coverage. The coverage means the length of time that the echo canceller stores its approximation in memory. The adaptive filter should be long enough to model an unknown system properly, especially in case of VoIP applications (Nisar et al., 2009; Youhong et al., 2005). On the other hand, it is known that an active part of the network echo path is usually much smaller compared to the whole echo path that has to be covered by the adaptive filtering algorithm. That is why the knowledge of the echo delay is important for using echo cancellers in packet-switching networks. Today, there is a wide family of adaptive filtering algorithms that can exploit sparseness of the echo path to reduce high computational complexity associated with long echo paths (Dyba, 2008; Hongyang & Dyba, 2008; Khong & Naylor, 2006; Hongyang & Dyba, 2009). In this chapter, we discuss numerous methods used for estimation of echo delay. Algorithms based on cross-correlation function and adaptive filters are used in the art. We will consider both types of them, discuss their advantages and drawbacks. Afterwards, we will pay our attention to the adaptive filtering techniques. We provide a study on different partial, proportionate, sparseness-controlled time- and frequency-domain adaptive filters. The readers will get closer to an issue of echo cancellation, which is relevant in nowadays telecommunications networks. Ones will able to recognize important features and particular areas of implementation of various adaptive algorithms. Further, we are giving a short introduction to the issue of echo control for telecommunications networks. This description emphasises on two most important aspects of perceptual echo control, which are echo loudness and echo delay.

1.1 Echo control issue

In the very beginning of the telephone age, all calls were made through an analog pair of copper wires. The technology has progressively moved to digital circuit switched networks over the past several decades. Today most of the phone traffic is handled by the Public Switched Telephone Network (PSTN), which provides end-to-end dedicated circuits. During the last years a move to packet-switched networks has been initiated to support voice traffic over Internet Protocol (IP). The main reason for the move from circuit-switched voice networks to packet-switched networks is to enable convergence between data services and voice services. It is of economical interest to be able to use the same equipment for voice and data traffic. Reduced cost of placing a phone call is another reason, since the voice-packet is treated and routed much in the same way as any other data packet (note that Quality of Service plays a vital role in this process). Thus, conventional long distance tariffs have a tendency to be completely eliminated in Voice over IP (VoIP) networks as well.

Echo issue has long been recognized as a problem on telecommunications networks, though generally it has been noticed mostly on international telephone calls or when using speaker phones. As technology advances and the information transmission methods tend more to packet-switching concepts, the traditional echo problem should be reviewed and updated. Previously unconsidered factors now play an important part in the echo characteristics. This section describes the echo delay problem, which is often encountered in packet-switched networks. This problem is highlighted in relation to VoIP networks. More specific details on the process of locating and eliminating echoes are included in conclusion to the chapter.

Consider a simple voice telephone call, where an echo occurs when you hear your own voice repeated. An echo is the audible leak-through of your own voice into your own receive path. Every voice conversation has always at least two participants. From the perspective of each participant, there are two voice paths in every call:

- Transmit path – The transmit path is usually depicted as Tx path. In a conversation, the transmit path is created when any person begins speaking. The sound is transmitted from the mouth of the speaker to the ear of the listener.
- Receive path – The receive path is also called the return and depicted as Rx path. In a conversation, the receive path is created when a person hears the conversation coming from the mouth of another speaker.

Fig. 1 illustrates a simple diagram of a voice call between two persons A (Kirill) and B (Kate). From the user A's perspective, the Tx path carries his voice to the user B's ear, and the Rx path carries the user B's voice to the user A's ear.

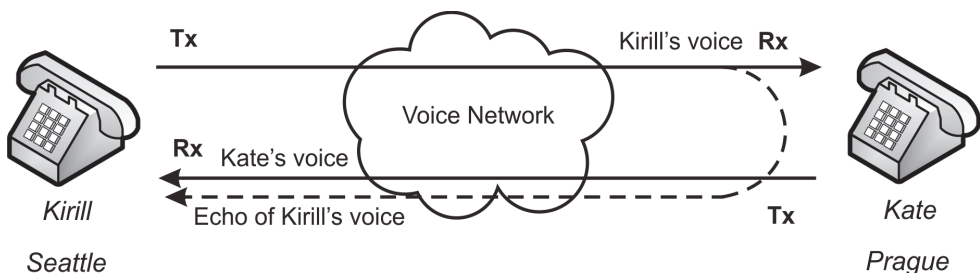


Fig. 1. A simple telephone call scenario

There is one significant factor in the echo analysis, and especially for the packet-switching networks. It is a round-trip delay of the voice network. The round-trip delay is the length of time required for an utterance from the user A's mouth, across the network on the Tx path to the source of the leak, and then back across the network again on the Rx path to the user A's ear. Let's define two important statements about echo nature, which are the following:

- The louder the echo (echo amplitude), the more annoying it is,
- The longer the round-trip delay (the "later" the echo), the more annoying it is.

Table 1 shows how time delay can affect the quality of a voice conversation.

One-Way Delay Range (ms)	Effect on Voice Quality
0-25	This is the expected range for national calls. There are no difficulties during conversation.
25-150	This is the expected range for international calls using a terrestrial transport link and IP telephony, which includes only one instance of IP voice. This range is acceptable for most users, assuming the use of echo control devices.
150-400	This is the expected range for a satellite link. Delays in this range can interrupt the normal flow of a conversation. A high-performance echo canceller must be used and careful network planning is necessary.
Greater than 400	This is excessive delay and must be avoided by network planning.

Table 1. Effect of Delay on Voice Quality

Fig. 2 shows how the echo disturbance influenced by the two parameters: delay and echo level. The metric called Talker Echo Loudness Rating (TELR) denotes the level difference between the voice and echo signals. The "acceptable" curve represents the limit for acceptable talker echo performance for all digital networks. The fact that the speaker A, in Fig. 1, hears an echo illustrates one of the basic characteristics of echo: perceived echo most likely indicates a problem at the other end of the call. The problem that is producing the echo that A hears, the leakage source, is somewhere on B's side of the network. If the person B was experiencing echo, the problem would be on the user A's side.

The perceived echo usually originates in the terminating side of the network for the following two reasons:

- Leakage happens only in analog circuits. Voice traffic in the digital portions of the network does not pass from one path to another.
- Echo arriving after very short time, about 25 milliseconds, is generally imperceptible, because it is masked by the physical and electrical side-tone signal.

A hybrid transformer is often main source of the electrical signal leakage. The typical analog telephone terminal is 2-wire device: a single pair of conductors is used to carry both the Tx and Rx signals. For analog trunk connections, known as 4-wire transmission, two pairs of conductors carry separate Tx and Rx signals. Digital trunks (T1/E1) can be virtual 4-wire links because they also carry separate Tx and Rx signals. A hybrid is a transformer that is used to interface 4-wire links to 2-wire links. Fig. 4 shows a hybrid transformer in an analog tail circuit. Because a hybrid transformer is a non-ideal physical device, a certain fraction of the 4-wire incoming (Rx) signal will be reflected into 4-wire outgoing (Tx) signal. A typical

fraction for a properly terminated hybrid in a PBX is about -25 decibels (dB), meaning that the reflected signal (the echo) will be a version of the Rx signal attenuated by about 25 dB. For a PSTN POTS (Plain Old Telephone Service) termination, the expected value is between 12 and 15 dB. Echo strength is expressed in dB as a measurement called Echo Return Loss (ERL). Therefore, an ERL of 0 dB indicates that the echo is the same amplitude as the original source. A large ERL indicates a negligible echo. Remember that an echo must have both sufficient amplitude and sufficient delay to be perceived. For local calls with one-way delay from 0 to 25 ms, an echo of strength of -25 dB relative to the speech level of the talker is generally quiet enough to not be annoying. For a one-way delay in the range of 25 to 150 ms, the ERL should exceed 55 dB to eliminate the perception of echo from the end-user perspective, as recommended in ITU-T recommendation G.168 on echo cancellation (ITU-T G.168, 2002). In this case echo cancellation is required.

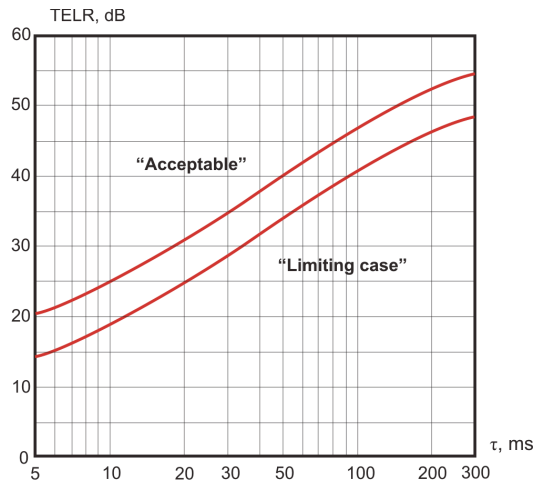


Fig. 2. Talker echo tolerance curves (ITU-T G.131, 2003)

2. Echo delay estimation using cross-correlation

The following section presents a study of cross-correlation-based Time Delay Estimation (TDE) algorithms. The main purpose is to analyze a number of methods, in order to find the most suitable one for real-time speech processing. As TDE is an important topic during transmission of voice signals over packet-switching telecommunication systems, it is vital to estimate the true time delay between Tx and Rx speech signals. We consider algorithms processing both in time- and frequency domains. An echo delay problem associated with IP-based transport networks is also included into the discussion. An experimental comparison of the performance of numerous methods based on cross-correlation, normalized cross-correlation and a generalized cross-correlation function is presented.

2.1 General scenario of delay estimation using cross-correlation functions

The known problem associated with IP-based networks is that the round-trip delay can be never reduced below its fundamental limit. There is always a delay of at least two to three packet sizes (50 to 80 ms) that can make the existing network echo more audible. Therefore, all

Voice over IP (VoIP) network terminals should employ echo cancellers to reduce the amplitude of returning echoes. A main parameter of each echo canceller is a length of coverage. Echo canceller coverage specifies the length of time that the echo canceller stores its approximation in memory. The adaptive filter should be long enough to model an unknown system properly, especially in case of VoIP applications. On the other hand, it is known that the active part of the network echo path is usually much smaller compared to the whole echo path that has to be covered by the adaptive filtering algorithm inside the echo canceller. That is why the knowledge of the echo delay is important for using echo cancellers in packet-switching networks successfully. In general, every communications system includes a communications network and communications terminals on the both sides of the network. The communications terminals could be telephones, soft phones, and wireless voice communication devices. Fig. 3 illustrates how an echo assessment device can be arranged into the defined system. The echo delay estimator has to monitor two parallel channels. An outgoing voice channel transmits an original voice waveform from the first terminal through the communications network to the second terminal. An incoming voice channel receives an echo waveform of the original signal returning from the second terminal through the communications network back. This is a delayed and attenuated version of the original voice signal.

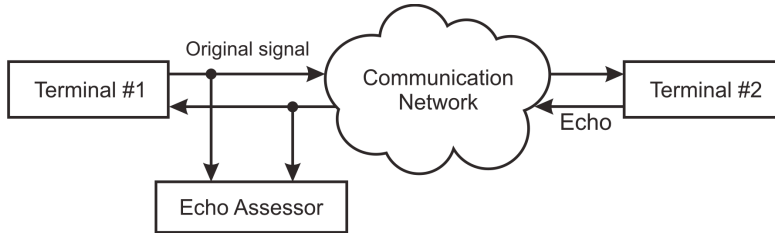


Fig. 3. Arrangement of echo assessment module in the network

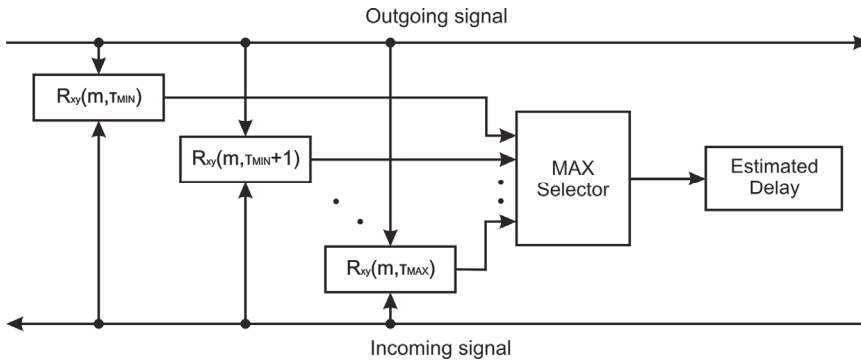


Fig. 4. General block diagram of delay estimator

Fig. 4 illustrates a general block diagram of the echo delay estimator. The echo delay estimator computes correlation between two voice channels for different set of delays in parallel manner (Carter, 1976). The delay-shift with the largest cross-correlation coefficient is selected as the delay estimate. Fig. 5 illustrates, in a flowchart form, steps performed when implementing a method of echo delay estimation utilizing cross-correlation algorithms. Once started from block 1, block 2 calculates the cross-correlation function for a buffer of

input samples of the Rx and Tx signals. Block 3 utilizes cross-correlation coefficients to compute the similarities between the transmitted signal and the received signal over a range of delays. For each particular delay, the similarity is obtained. Once the similarities have been determined for each delay within the range of delays, block 4 chooses a delay that produces the greatest similarity metric for the given input frames. Consequently, block 5 indicates that the estimation process is completed.

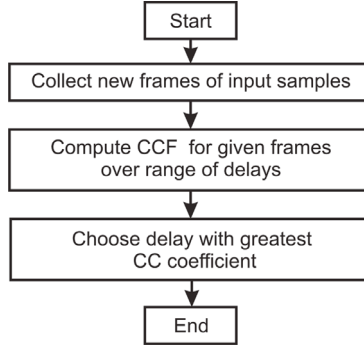


Fig. 5. Flowchart for estimating echo delay value

2.2 Algorithms proceeding in time-domain

Time domain implementation of Cross-Correlation Function (CCF) and Normalized CCF (NCCF) is presented. The cross-correlation function for a successive pair of speech frames can be estimated by (Mueller, 1975)

$$\begin{aligned}
 R_{xy\tau\text{MIN}}(m) &= \sum_{n=D}^{D+L-1} x(n) \cdot y(n-m) & n = 0, \dots, L-1, m = 0, \dots, L-1 \\
 &\dots & \\
 R_{xy\tau\text{MAX}}(m) &= \sum_{n=D}^{D+L-1} x(n) \cdot y(n-m) & n = 0, \dots, L-1, m = 0, \dots, L-1
 \end{aligned} \tag{2}$$

Here, $x(n)$ simply denotes a frame of the outgoing signal, $y(n)$ is related to a frame of the incoming signal. According to Fig. 4, the estimation of the CCF is done for a supposed range of delays. The time-shift, τ , which is always in range of $[\tau_{\min}; \tau_{\max}]$ and causes the maximal peak value of the CCF is declared as an estimate of the true echo delay T_D . Similarly to the CCF, an estimate of the NCCF is done (Buchner et al., 2006)

$$\begin{aligned}
 R_{xy\tau\text{MIN}}^n(m) &= \frac{\sum_{n=D}^{D+L-1} x(n) \cdot y(n-m)}{\sqrt{E_x \cdot E_y}} & n = 0, \dots, L-1, m = 0, \dots, L-1 \\
 &\dots & \\
 R_{xy\tau\text{MAX}}^n(m) &= \frac{\sum_{n=D}^{D+L-1} x(n) \cdot y(n-m)}{\sqrt{E_x \cdot E_y}} & n = 0, \dots, L-1, m = 0, \dots, L-1
 \end{aligned} \tag{3}$$

Here, E_x and E_y denotes a short-term energy of the outgoing and the incoming frames. These values are calculated using the following equations

$$E_x = \sum_{n=D}^{D+L-1} x^2(n) \quad (4)$$

$$E_y = \sum_{n=D}^{D+L-1} y^2(n-m) \quad (5)$$

Let us further consider generalized cross-correlation algorithms, which operate in the frequency domain (Youn et al., 1983; Zetterberg et al., 2005).

2.3 Algorithms proceeding in frequency-domain

More sophisticated way how to provide TDE is to compute the cross-correlation function in the frequency domain. This process in literature is called Generalized Cross-Correlation (GCC) (Hertz, 1986). The idea behind this method is to perform pre-filtering of the input signals before calculating CCF. It makes possible to improve the accuracy of delay estimation. Note that the filtering procedure is performed in the frequency domain. Let us describe this process in more details. It is well known, that the simple cross-correlation function, R_{xy} , between signals $x(n)$ and $y(n)$ is related to the cross-power density function (cross-power spectrum), G_{xy} , by the general inverse Fourier transform relationship, as

$$R_{xy}(m) = \int_{-\infty}^{\infty} G_{xy}(f) \cdot e^{j2\pi fm} df \quad (6)$$

When $x(n)$ and $y(n)$ have been filtered with filters having transfer functions $H_1(f)$ and $H_2(f)$, the cross-power spectrum between the filter out-puts is given by

$$G_{xy}^g(f) = H_1(f) \cdot H_2^*(f) \cdot G_{xy}(f) \quad (7)$$

Consequently, the Generalized Cross-Correlation Function (GCCF) between $x(n)$ and $y(n)$ is given by (Knapp & Carter, 1976)

$$R_{xy}^g(m) = \int_{-\infty}^{\infty} \Psi_g(f) \cdot G_{xy}(f) \cdot e^{j2\pi fm} df \quad (8)$$

$$\Psi_g(f) = H_1(f) \cdot H_2^*(f) \quad (9)$$

Here, Ψ_g is a generalized weighting function. Table 2 represents weighting functions that were used for experiments with speech signals (Wilson & Darrell, 2006).

The parameter γ_{xy} denotes a complex coherence function. It can be calculated as (Tianshuang & Hongyu, 1996)

$$\gamma_{xy}(f) = \frac{G_{xy}(f)}{\sqrt{G_{xx}(f) \cdot G_{yy}(f)}} \quad (10)$$

$$G_{xx}(f) = \int_{-\infty}^{\infty} R_{xx}(m) \cdot e^{j2\pi fm} dm \quad (11)$$

$$G_{yy}(f) = \int_{-\infty}^{\infty} R_{yy}(m) \cdot e^{j2\pi fm} dm \quad (12)$$

Here, $G_{xx}(f)$ and $G_{yy}(f)$ are auto-power spectra of the outgoing and the incoming signal; $R_{xx}(m)$ and $R_{yy}(m)$ are auto-correlation functions of the same signals. Fig. 6 illustrates a block diagram of the implemented generalized cross-correlation algorithm, where the Fast Fourier Transform (FFT) is used for auto-spectra and cross-spectrum calculation. After the cross-power spectrum is estimated, it is multiplied by the corresponding GCC weighting function. The inverse FFT is used for obtaining the time domain generalized-cross correlation function. This operation is repeated for the specified range of possible delays. After the whole process has completed, the time shift with maximum corresponding peak value is declared as an estimation of the true delay.

Processor Name	Weighting Function
Cross-correlation	1
ROTH	$1/G_{yy}(f)$
SCOT	$1/\sqrt{G_{xx}(f) \cdot G_{yy}(f)}$
PHAT	$1/ G_{xy}(f) $
Cps-m	$1/\sqrt[m]{G_{xx}(f) \cdot G_{yy}(f)}$
HT (ML)	$\frac{ \gamma_{xy}(f) ^2}{ G_{xy}(f) \cdot (1 - \gamma_{xy}(f) ^2)}$
Eckart	$\frac{ G_{xy}(f) }{(G_{xx}(f) - G_{xy}(f)) \cdot (G_{yy}(f) - G_{xy}(f))}$
HB	$ G_{xy}(f) /G_{xx}(f) \cdot G_{yy}(f)$
Wiener	$ \gamma_{xy}(f) ^2$

Table 2. Various GCC weighting functions

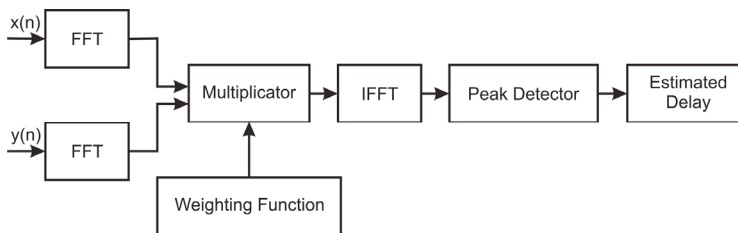
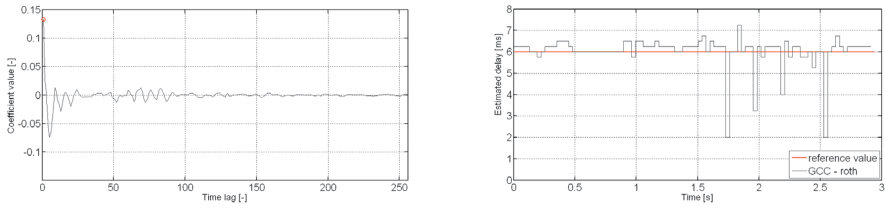


Fig. 6. Diagram of the implemented generalized cross-correlation algorithm

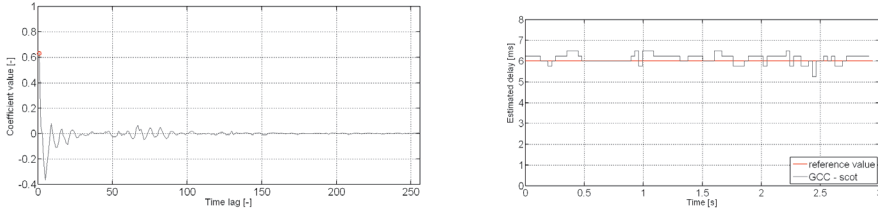
2.4 Discussion over experimental results

We used MATLAB software as a simulation environment. The time difference between time when the outgoing signal leaves the voice terminal and moment when the incoming signal containing the echo of the original signal arrives back from the network is referred to as a true echo delay. This value for the first three figures that are presented below equals 6ms (48 samples). For the purpose of TDE it is also necessary to specify time interval through which

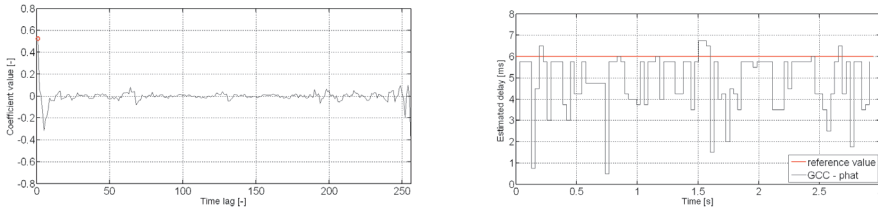
the value of the true delay is searched. To cover the 6ms delay we choose the interval between 0 and 10ms what corresponds to the maximum delay value of 60 samples. Afterwards we present the estimation results for a group of different delays. It helps to understand better performance of the algorithms. Unfortunately, because of the non-stationary nature of human speech, the CCF is not reliable for all situations. Its performance highly depends on numerous factors, i.e. signal strength, signal-to-noise ratio (SNR), etc (Chen et al., 2006). The NCCF is not so sensitive to the sudden changes in the signal's amplitude. It outperforms the CCF when we work with low level signals. The advantages of the algorithms proceeding in the frequency domain compared to the algorithms operating in the time domain are accuracy and reduced computational complexity. Fig. 7 illustrates the outputs of the GCC algorithms, which were presented in Table 2.



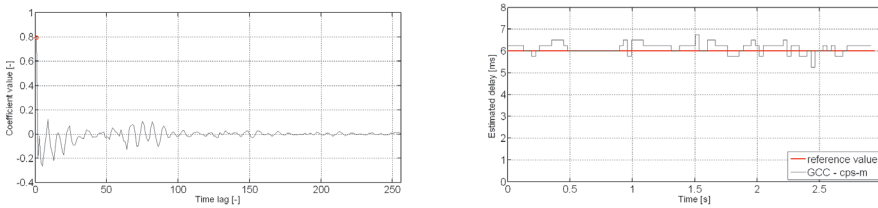
(a) ROTH - weighting function



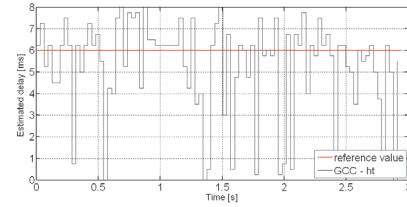
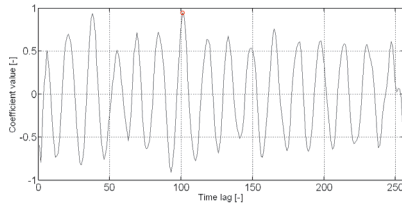
(b) SCOT - weighting function



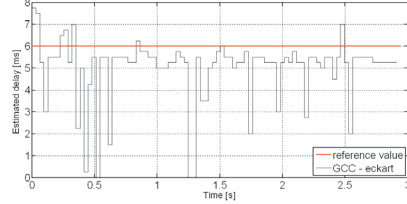
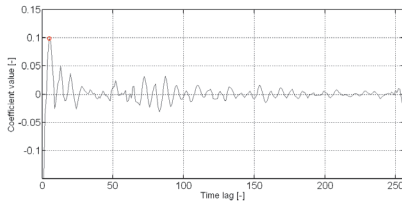
(c) PHAT - weighting function



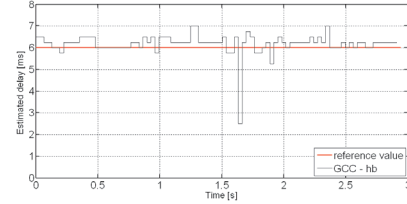
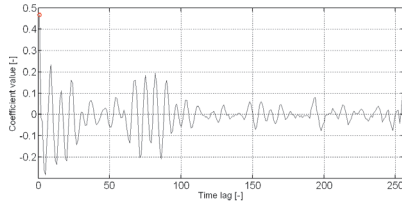
(d) CPS-M (M=2) - weighting function



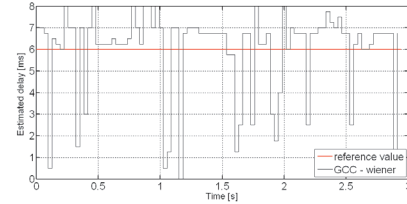
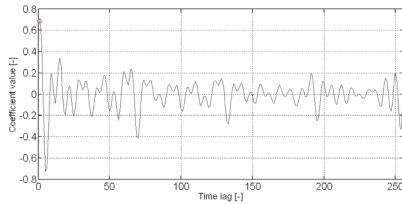
(e) HT - weighting function



(f) ECKART - weighting function



(g) HB - weighting function



(h) WIENER - weighting function

Fig. 7. Time delay estimation using GCCF

Table 3 and 4 provides us along with the following results. The joint comparison was done in terms of the estimation accuracy of the algorithms. The group of delays was chosen for this experiment. Delay values are consistent with the ones referenced in the corresponding ITU-T recommendation G.131 (ITU-T G.131, 2003). Once the respective cross-correlation function was calculated, its maximum peak value is detected using the searching procedure described in Fig. 4. SCC is related to the Standard CC function.

[ms]	SCC	ROTH	SCOT	PHAT	CPS-2	HT	ECKART	HB	WIENER
5	4,9	5,1	5,2	3,7	5,2	4,4	4,2	5,2	5,2
10	9,7	10,3	10,3	7,5	10,3	8,8	8,3	10,4	10,3
20	19,5	20,6	20,7	15,0	20,7	17,7	16,7	20,8	20,7
30	29,2	30,9	31,0	22,4	31,0	26,5	25,0	31,2	31,0
50	48,7	51,4	51,6	37,4	51,6	44,2	41,7	52,1	51,7
100	97,3	102,9	103,3	74,8	103,3	88,5	83,3	104,2	103,4
200	194,6	205,7	206,6	149,6	206,6	177,0	166,6	208,3	206,8
300	292,0	308,6	309,9	224,4	309,9	265,4	249,9	312,5	310,3

Table 3. Mean values of estimated delays

[ms]	SCC	ROTH	SCOT	PHAT	CPS-2	HT	ECKART	HB	WIENER
5	4,9	5,1	5,2	3,7	5,2	4,4	4,2	5,2	5,2
10	9,7	10,3	10,3	7,5	10,3	8,8	8,3	10,4	10,3
20	19,5	20,6	20,7	15,0	20,7	17,7	16,7	20,8	20,7
30	29,2	30,9	31,0	22,4	31,0	26,5	25,0	31,2	31,0
50	48,7	51,4	51,6	37,4	51,6	44,2	41,7	52,1	51,7
100	97,3	102,9	103,3	74,8	103,3	88,5	83,3	104,2	103,4
200	194,6	205,7	206,6	149,6	206,6	177,0	166,6	208,3	206,8
300	292,0	308,6	309,9	224,4	309,9	265,4	249,9	312,5	310,3

Table 4. Root mean square deviation of estimated delays

The abscissa of the largest peak value is the estimated delay. Note that 50 trial speech records for each processor were evaluated to obtain the mean value and the Root Mean Square Deviation (RMSD) parameter (Anderson & Woessner, 1992). Not only different speech signals, but various hybrid impulse response models have been used. The results for delays from 5 to 300 ms are presented in the corresponding tables. Table 3 contents the mean values, whether Table 4 illustrates the estimated RMSD values.

3. Echo delay estimation using adaptive filters

In this section, we introduce methods for extracting an echo delay between speech signals using adaptive filtering algorithms. We know that time delay estimation is an initial step for many speech processing applications. Conventional techniques that estimate a time difference of arrival between two signals are based on the peak determination of the generalized cross-correlation between these signals. To achieve a good precision and stability in estimation, the input sequences have to be multiplied by an appropriate weighting function. Regularly, the weighting functions are dependent on the signals power spectra. The spectra are generally unknown and have to be estimated in advance. An implementation of the time delay estimation via the adaptive least mean squares is analogous to estimating the Roth generalized cross-correlation weighting function. The estimated parameters using the adaptive filter have a smaller variance, because it avoids the need for the spectrum estimation. In the following, we discuss proportionate and partial-update adaptive techniques and consider their performance in term of delay estimation.

Time delay estimation (TDE) has always been and remains a popular research topic. It is known, that the main problem associated with IP-based networks is the round-trip delay, which can never be reduced below its fundamental limit. A number of efforts were made in order to improve the TDE precision. Various methods based on the Generalized Cross-Correlation (GCC) were proposed. The GCC algorithms arrange a pre-filter to obtain the modified signal spectrum for optimal time delay estimation. To specify the filter's characteristic, it requires a priori knowledge of the statistics of the received signals. However, the efficiency of the algorithms decreases considerably when little or no prior knowledge about the signal statistics is known. From the time when Widrow proposed an adaptive filtering technique based on Least Mean Squares (LMS) (Widrow, 2005; Haykin, 2001), an adaptive theory also found an application to delay estimation. An adaptive implementation of the time delay estimation via Widrow's LMS algorithm is usually referred to as TDLMS. Comparing to the GCC algorithms, the adaptive filtering techniques do not require a priori information of the signal statistics, because the estimation of the signal spectrum is no longer needed. The adaptive filtering algorithms determine the time delay in an iterative manner. There are comparative studies, which provide comparison of the LMS versus the generalized cross-correlation (Zetterberg et al., 2005). Generally, the time domain implementation of any adaptive filter is associated with high computational complexity. It directly depends on the length of the adaptive filter. In order to reduce the computational load of the TDLMS (Emadzadeh et al., 2008), we offer using adaptive filtering algorithms with reduced computational complexity. They provide savings comparing to the conventional adaptive algorithms. In the following, we discuss each of the algorithms in greater details. First, a general scenario for the adaptive time delay estimation using a simple Normalized Least Mean Squares (NLMS) adaptive filtering algorithm is presented. Afterwards, we introduce the proportionate and partial-updated algorithms proceeding in the time domain. A new partial-updated proportionate NLMS algorithm is outlined. A comparison between the TDE algorithms is made in context of the network echo delay estimation.

3.1 General scenario of delay estimation using adaptive filters

Traditionally, the NLMS algorithm is used for the echo canceller implementation. It applies a Finite Impulse Response (FIR) adaptive filter with adjustable weights for modelling the unknown echo path's impulse response. The NLMS algorithm minimizes the least mean squares difference between two signals: the reference incoming signal and the filtered original (outgoing) signal (see Fig. 7).

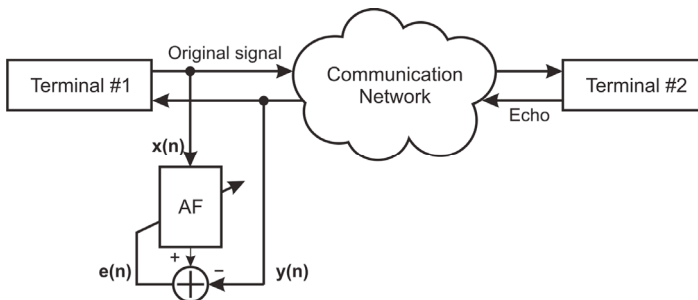


Fig. 7. Time delay estimation using adaptive filter

Basically, the NLMS algorithm is a simple extension of the Widrow's LMS algorithm. According to Fig. 7, the adaptive filter weights are tuned iteratively using a special rule. The important parameter controlling the whole adaptation process is referred to as a step-size parameter. It should be varied in time in order the algorithm to be able to track non-stationary changes in the echo path's impulse response. Unlike the LMS algorithm, the NLMS algorithm's step-size is adjusted according to the instantaneous short-time energy of the input signal. The adaptive filtering process using the NLMS algorithm can be described by the following set of equations

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{1}{\mathbf{x}^T(n) \cdot \mathbf{x}(n)} \mu_0 \cdot e(n) \cdot \mathbf{x}(n) \quad (13)$$

$$\begin{aligned} e(n) &= y(n) - \mathbf{w}^T(n) \cdot \mathbf{x}(n) = y(n) - \sum_{i=0}^{L-1} w(n) \cdot x(n-i) \\ \mathbf{w}(n) &= [w_1(n), w_2(n), \dots, w_L(n)]^T \\ \mathbf{x}(n) &= [x(n), x(n-1), \dots, x(n-L+1)]^T \\ \boldsymbol{\mu}(n) &= [\mu_1(n), \mu_2(n), \dots, \mu_L(n)]^T \end{aligned} \quad (14)$$

where $\mathbf{w}$ is a vector containing L coefficients of the adaptive filter; $\mathbf{x}$ is a vector consisting of L samples of the input signal $x(n)$; $e(n)$ is a difference between the reference signal, $y(n)$, and the adaptive filter output during the n^{th} iteration; μ_0 is the fixed NLMS step-size parameter from the interval (0;1). Fig. 8 illustrates the basic principle of the NLMS adaptive algorithm. While the signal $x(n)$ corresponding to the outgoing voice signal (that notation is used in the previous section) is referred to as the far-end signal, the signal $y(n)$ corresponding to the incoming voice signal is referred to as the near-end signal. Basically, the near-end signal $y(n)$ is the delayed and attenuated version of the far-end signal $x(n)$.

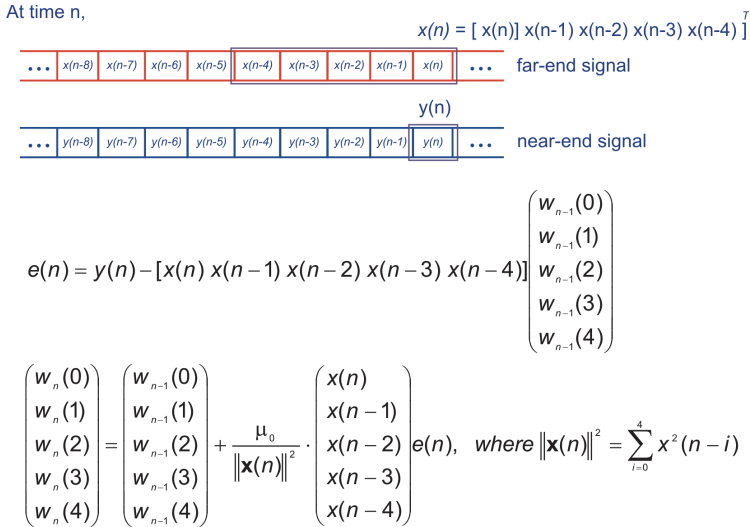


Fig. 8. Principle of the normalized least mean squares algorithm

3.2 Time-domain adaptive algorithms

Knowing the adaptive theory, it is trivial that the delay estimation can be achieved by selecting the largest value from the adaptive filter weights vector, w . There is only one issue that has to be taken into account. The adaptive filter needs some time in order to converge to the optimal performance. The existing adaptive algorithms differ from each other with different convergence properties and computational memory requirements. The robust fast converging algorithms are primarily used in the acoustical echo cancellation applications. They take a lot of computational resources. In our case, it is not necessary to apply the complex algorithms, because the adaptive filter is not directly used for the purpose of echo cancellation, but for the delay estimation. Therefore, the reduced complexity adaptive filtering algorithms became the subject of our interest.

3.2.1 Proportionate adaptive filtering algorithms

In the following, we provide a reader along with a principal of the proportionate adaptive technique. The proportionate normalized least mean squares (PNLMS) algorithm proposed in (Duttweiler, 2000) has been developed for use especially in the telephone network environment. For hybrid echo cancellers, it is reasonable to assume that the echo path has a sparse character (i.e., many IR's (Impulse Response) coefficients are close to zero). Although there are studies and research on the multiple reflection echo paths, a typical echo path impulse response in the practical communication networks has only one reflection, which means all the active coefficients are occupied in a continuous area of the whole echo span. Therefore, we do not need to update all the filter weights to achieve a nominal performance. One can adjust and operate only with non-zero active coefficients. Basically, this is a main idea behind the PNLMS algorithm and other subsequently derived proportionate adaptive filters (Hongyang & Doroslovacki, 2006; Paleologu et al., 2006). Proportionate approaches achieve their higher convergence rate by using the fact that the active part of network echo path is usually much smaller (4-8ms) compared to 64-128 ms of the whole echo path that has to be covered by the adaptive filter. In case of voice transmission over the packet-switching network, these numbers may be more considerable. In the PNLMS algorithm, the adaptive step-size parameters are assigned to all the filter coefficients. They are calculated from the last estimate of the filter weights in such a way that a larger coefficient receives a larger increment. As a result, the convergence rate can be increased the fact that the active taps are adjusted faster than non-active coefficients. Therefore for the sparse IR, the PNLMS algorithm converges much faster comparing to the NLMS. This feature is an advantage especially when it is necessary to estimate the long echo delays. The PNLMS algorithm can be described using the following equations:

$$w(n+1) = w(n) + \frac{1}{x^T(n) \cdot G(n-1) \cdot x(n)} \mu_0 \cdot G(n-1) \cdot e(n) \cdot x(n) \quad (15)$$

$$G(n-1) = \text{diag}\{g_0(n-1), \dots, g_{L-1}(n-1)\} \quad (16)$$

where $G(n-1)$ is a diagonal matrix adjusting the step-size parameters, μ_0 is an overall step-size parameter. The diagonal elements of $G(n)$ are estimated as follows:

$$\gamma_i(n) = \max\left\{\rho \max\left[\delta_p, |w_0(n)|, \dots, |w_{L-1}(n)|\right], |w_i(n)|\right\} \quad (17)$$

$$g_l(n) = \frac{\gamma_l(n)}{\sum_{i=0}^{L-1} \gamma_i(n)}, \quad 0 \leq l \leq L-1 \quad (18)$$

Parameters δ_p and ρ are positive numbers with typical values for $\delta_p = 0.01$ and for $\rho = 5/L$. The first term in (17), ρ , prevents $w_l(n)$ from stalling when it is much smaller than the largest coefficient and δ_p regularizes the updating when all coefficients have zero values at initialization. In spite of the sparse system identification, which is a vital requirement for the fast converging adaptive filters, there is another requirement. It is directly addressed to the adaptive filter implementation. The algorithm should have reasonable power concerns. Unfortunately, the PNLMS algorithm has drawbacks. One of them is an increase in the computational complexity by 50% compared to the NLMS algorithm. Furthermore, the PNLMS algorithm shows the slow convergence rate after the fast initial start. It is because of the slow convergence rate dedicated to the small coefficients (Gay, 1998). The increased computational complexity can be reduced by the way of selective partial-updating. In turn, the slow convergence of the PNLMS in the stable state can be improved by switching from the PNLMS to NLMS equations after the fast initial convergence has been achieved (Benesty & Gay, 2002). Recently researchers proposed the partial-update techniques for using along with the adaptive algorithms. The partial-update algorithms reduce computational complexity by updating only a subset of the filter weights per iteration. They can be relevant for applications requiring fast real-time processing. Unfortunately, there is another side to the coin. The fewer coefficients you are going to update per iteration, there is more misalignment presented in the algorithm. Therefore, a certain trade-off should be made when selecting the number of coefficients to update. The following subsection presents several partial-update algorithms along with the proposed modification of the partial-update PNLMS algorithm.

3.2.2 Partial-update adaptive filtering algorithms

The partial-update algorithms can be seen to exploit the sparseness of the echo path in two different ways. It is known that when the unknown system's impulse response is sparse, many of the adaptive filter's weights can be approximated to zero. Alternatively, the sparseness may be present in the weight update vector as a consequence of the distribution of the input samples in the $(L \times 1)$ input vector, $\mathbf{x}_n = [x(n), x(n-1), \dots, x(n-L+1)]^T$. In both these cases, exploiting the sparseness properties can reduce complexity and improve performance of the adaptive algorithm (Fevrier & Gelfand, 1999). Some of the first work on the partial-update algorithms was done by Douglas (Douglas, 1997). It presents the periodic and the sequential updating schemes for the Max-NLMS algorithm. However, these partial-update algorithms show slow convergence 2properties compared to the full-update algorithms. The reason is inconsistent updating schemes. More recently, the partial-updating concept was developed by Aboulnasr (Aboulnasr, 1999). It leads to the M-Max NLMS algorithm and supporting convergence analysis (Aboulnasr & Mayyas, 1998). Another block-updating scheme for the NLMS algorithm was studied by Schertler (Schertler, 1998). The latter work was published by Dogancay and Tanrikulu. They consider approaches for more robust Affine Projection Algorithm (APA) (Dogancay & Tarinkulu, 2001, 2002). Further, we give a summary of the general partial-update algorithms along with the proposed one.

M-Max NLMS

The algorithm selects a specified number of the coefficients providing the largest reduction in the mean squared error per iteration (Naylor & Sherliker, 2003). Only M out of the total L filter coefficients are updated. Those M coefficients are the ones associated with the M largest values within the following vector $|x(n - i + 1)|; i = 1; \dots; L$. The update equations for this algorithm are

$$w_i(n+1) = w_i(n) + \frac{1}{x^T(n) \cdot x(n)} \mu_0 \cdot e(n) \cdot x_i(n) \quad (19)$$

$$w_i(n) = \begin{cases} 1, & i \in \{M \text{ maxima of } |x(n-i+1)|\} \\ 0, & \text{otherwise} \end{cases}, \quad 0 \leq i \leq L \quad (20)$$

One of the features of the M-MAX-NLMS algorithm is that it reduces the complexity of the adaptive filter by selectively updating the coefficients while maintaining the closest performance to the full-update NLMS algorithm. We present misalignment curves for the algorithm in the up-following section. The misalignment value was calculated using the equation formula

$$\text{Misalignment}(n) = 10 \cdot \log_{10} \left(\frac{\sum_{j=1}^L (h_j(n) - w_j(n))^2}{\sum_{j=1}^L (h_j(n))^2} \right) \quad (21)$$

Here, w_j is the j -th coefficient of the adaptive filter, while h_j is the j -th coefficient of the simulated echo path's IR.

Selective-partial-update NLMS

This algorithm opposed to the M-Max NLMS has a block structure. An objective behind the latter is the same: it reduces computational costs by updating a subset of the filter coefficients. But first, the vector $x(n)$ and the coefficient vector $w(n)$ are arranged into K blocks of length $M = L/K$, where L is an integer as in

$$x(n) = [x_1^T(n) \ x_2^T(n) \ \dots \ x_K^T(n)]^T \quad (22)$$

$$w(n) = [w_1^T(n) \ w_2^T(n) \ \dots \ w_K^T(n)]^T \quad (23)$$

The coefficient vector's blocks $w_1(n), w_2(n), \dots, w_K(n)$ represent candidate subsets that can be updated during the current iteration. For a single-block updating scheme, the constrained minimization problem, which is solved by the NLMS algorithm, can be written as

$$w_i(n+1) = w_i(n) + \frac{1}{\|x_i(n)\|^2} \mu_0 \cdot e(n) \cdot x_i(n) \quad (24)$$

The selection of the block that has to be updated is made by determining the block with the smallest squared-Euclidian-norm update. According to Eq. (24), that justification can be described by the following terms

$$i = \arg \min_{1 \leq j \leq M} 1/\|x_j(n)\|^2 = \arg \max_{1 \leq j \leq M} \|x_j(n)\|^2 \quad (25)$$

Generalization from the single-block to the multiple-block updating scheme is done through the following. Suppose that only the B ($B < K$) blocks with the largest magnitudes are selected to be updated. Let a vector $I_B = \{i_1, i_2, \dots, i_B\}$ denote the subset of B blocks out of $\{1, 2, \dots, K\}$ to be updated. Thus, the equation for the B -block updating scheme is

$$w_{I_B}(n+1) = w_{I_B}(n) + \frac{1}{\|x_{I_B}(n)\|^2} \mu_0 \cdot e(n) \cdot x_{I_B}(n) \quad (26)$$

where x_{I_B} and w_{I_B} are defined as follows

$$x_{I_B}(n) = [x_{i_1}^T(n) \ x_{i_2}^T(n) \ \dots \ x_{i_B}^T(n)]^T \quad (27)$$

$$w_{I_B}(n) = [w_{i_1}^T(n) \ w_{i_2}^T(n) \ \dots \ w_{i_B}^T(n)]^T \quad (28)$$

The computational and memory requirements of the selective-partial-update NLMS algorithm are almost identical to those of the selective-block-update algorithm proposed in (Aboulnasr & Mayyas, 1998). Nevertheless, simulation results illustrated in the next section shows that this approach does not lead to the reasonable trade-off between performance and simplicity. The algorithm's efficiency is weaker than the one of the M-Max NLMS algorithm. As an alternative approach, a sparse-partial-update NLMS algorithm applies more relevant selection criterion.

Sparse-partial-update NLMS

This algorithm utilize a so-called sparse-partial (SP) weight selection criterion (Jinhong & Doroslovacki, 2008). The adaptive filter weights are updated based on the largest product of the multiplication of $x(n)$ and $w(n)$. The SP-NLMS single-block update equations are given by

$$w_i(n+1) = w_i(n) + \frac{1}{x^T(n) \cdot x(n)} \mu_0 \cdot e(n) \cdot x_i(n) \quad (29)$$

$$w_i(n) = \begin{cases} 1, & i \in \{M \text{ maxima of } |x(n-i+1)| \cdot w_i(n)\} \\ 0, & \text{otherwise} \end{cases}, \quad 0 \leq i \leq L \quad (30)$$

Hongyang and Dyba recently suggested a generalization for updating B blocks out of K , i.e.

$$w_{I_B}(n+1) = w_{I_B}(n) + \frac{1}{x_{I_B}^T(n) \cdot x_{I_B}(n)} \mu_0 \cdot e(n) \cdot x_{I_B}(n) \quad (31)$$

$$I_B = \{i : \text{is one of the } B \text{ largest among } w_{i_1}^T(n) \cdot x_{i_1}(n), \dots, w_{i_B}^T(n) \cdot x_{i_B}(n)\} \quad (32)$$

Simple-partial-update PNLMS

This proposed algorithm exploits the sparseness of the communication channel to speed up the initial convergence and employs the partially updating scheme to reduce the

computational complexity. A selection procedure is performed in accordance with the estimated magnitude of the channel's impulse response. The S-PNLMS algorithm for single-block update is defined as follows. Arrange $x(n)$ and $w(n)$ into K blocks of length $M = L/K$ in the same way as it is done in Eq. (22, 23). Then let $G_i(n)$ denote the corresponding $M \times M$ block of the diagonal weighing matrix, $G(n)$. The recursion for updating adaptive filter weights is given by

$$w_i(n+1) = w_i(n) + \frac{1}{x_i^T(n) \cdot G_i(n-1) \cdot x_i(n)} \mu_0 \cdot G_i(n-1) \cdot e(n) \cdot x_i(n) \quad (33)$$

where the block selection is done according to the following

$$i = \arg \min_{1 \leq j \leq M} G_j(n) \quad (34)$$

It is different to

$$i = \arg \min_{1 \leq j \leq M} x_j \cdot G_j(n) \cdot x_j \quad (35)$$

which is used with the SPU-PNLMS algorithm. It is apparently from the simulations that the S-PNLMS has similar performance to the SP-NLMS and outperforms the SPU-PNLMS algorithm. The S-PNLMS for updating B blocks out of M has these update equations

$$w_{I_B}(n+1) = w_{I_B}(n) + \frac{1}{x_{I_B}^T(n) \cdot G_{I_B}(n-1) \cdot x_{I_B}(n)} \mu_0 \cdot G_{I_B}(n-1) \cdot e(n) \cdot x_{I_B}(n) \quad (36)$$

$$I_B = \{ i : G_i(n) \text{ is one of the } B \text{ largest among } G_1(n), \dots, G_M(n) \} \quad (37)$$

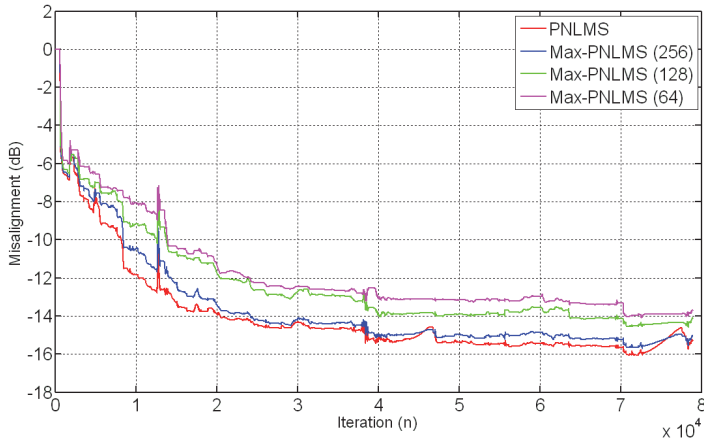
Further, we provide the comparison results for the presented algorithms and demonstrate their performance while estimating the predefined echo delay. Table 5 illustrates the computational complexity of the full-update algorithms and shows saving achieved by the partially updating schemes.

ALG.	MULT.	ADD.	DIV.
NLMS	$3L+1$	$3L-1$	1
M-Max-NLMS	$3M+1$	$3M-1$	1
SPU-NLMS	$3M^*B+1$	$3M^*B-1$	1
SP-NLMS	$3M^*B+1$	$3M^*B-1$	1
PNLMS	$6L+1$	$4L-2$	$L+1$
M-Max-PNLMS	$6M+1$	$4M-2$	$M+1$
SPU-PNLMS	$6M^*B+1$	$4M^*B-2$	M^*B+1
SP-PNLMS	$6M^*B+1$	$4M^*B-2$	M^*B+1
S-PNLMS	$6M^*B+1$	$4M^*B-2$	M^*B+1

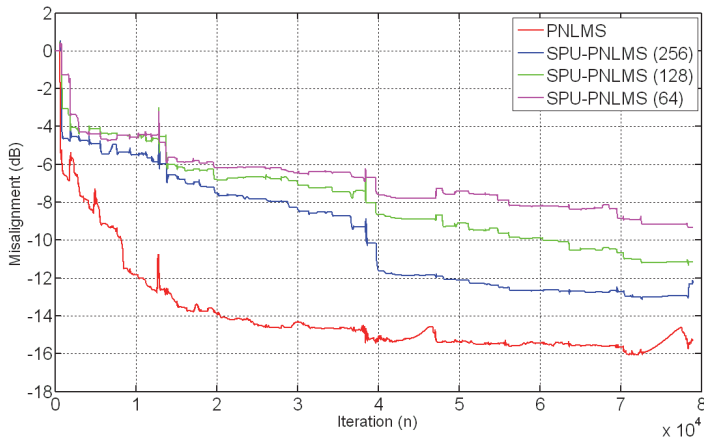
Table 5. Comparison of computational complexity

3.3 Discussion over experimental results

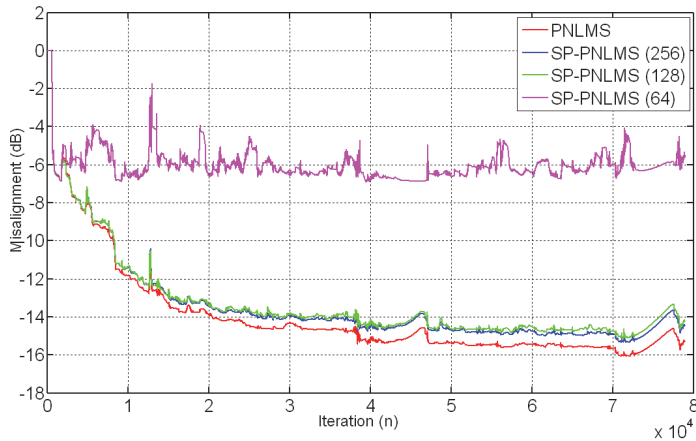
To evaluate the performance of the adaptive filtering algorithms, we implemented them in MATLAB environment. The working scenario is as follows. The filters have to estimate the predefined echo path's impulse responses specified in the ITU-T Recommendation G.168 (ITU-T G.168, 2002). They consist of 1024 taps. The overall step-size parameter, μ_o , is chosen to be 0.1. The control parameters ρ and δ_p are chosen to be 0.001 and 0.01 respectively. In the first part of the experiment, we look at the misalignment curves of the M-Max-, SPU-, SP- and S-PNLMS algorithms. They are illustrated in Fig. 9 below. The SPU-updating scheme produces the worst results. The proposed S-criterion considerably outperforms it especially in terms of the initial convergence speed. The rest of the algorithms have nearly the same convergence and tracking performance.



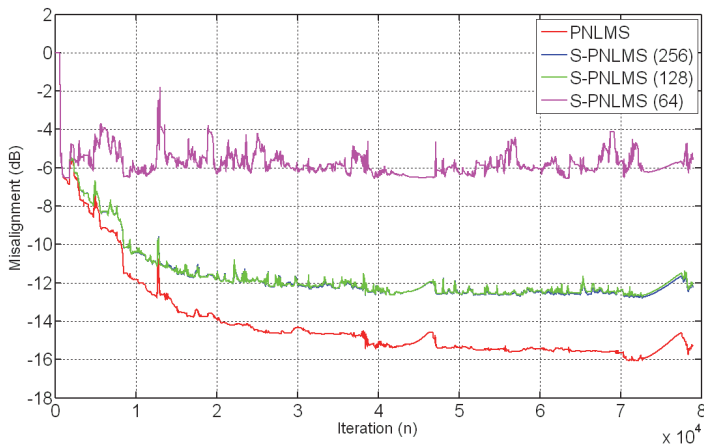
(a) Max-PNLMS algorithm



(b) SPU-PNLMS algorithm



(c) SP-PNLMS algorithm

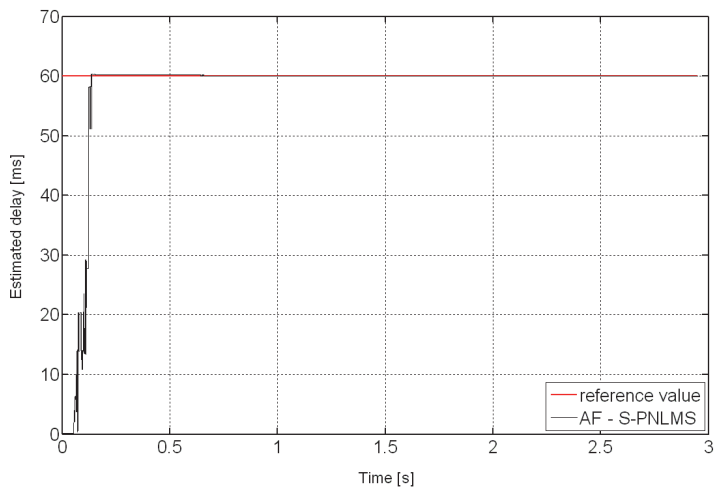


(d) S-PNLMS algorithm

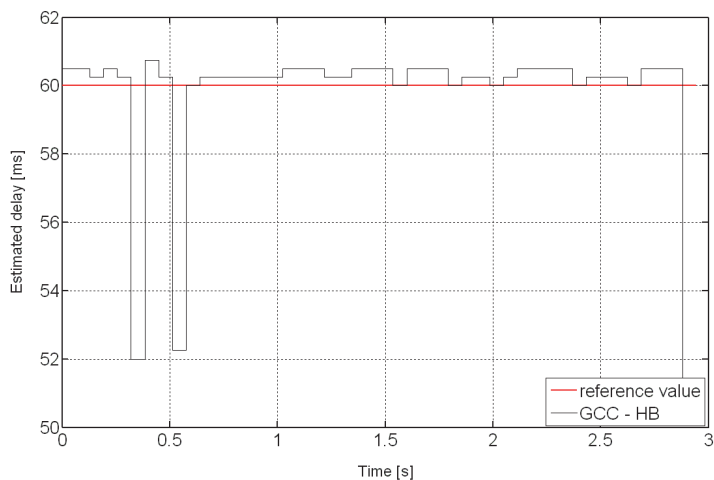
Fig. 9. Misalignment curves of the partial-update algorithms

All the algorithms, except the M-Max-PNLMS, show poor results when the M value equals 64. It can be explained by the fact that the active part of the IR is approximately 16ms long. This value corresponds to 128 samples for sampling frequency of 8kHz, therefore, 64 samples are not enough to cover the active region completely. Regarding to the dissimilar selection criterion, the M-Max-PNLMS algorithm can deal relatively well with that problem. The Max-updating formula does not count with the sparse character of the IR. It performs selection according to the distribution of the values of the input vector. Otherwise, its drawback is lower initial convergence speed comparing to the SP-PNLMS algorithm. The second part of our experiment concerns the performance of the adaptive algorithms versus the ones based on the generalized cross-correlation function. They are compared in the context of the time delay estimation. The time difference between time when the outgoing

signal leaves the voice terminal and the moment when the incoming signal containing the echo of the original signal arrives back from the network is referred to as a true echo delay. Table 6 concludes the results for the AF and GCC algorithms for different delay values. Fig. 10 shows the estimated echo delay diagram for the echo path that has a 60 ms pure delay. It can be observed that the outputs from the adaptive filter in the steady state have much smaller variance than the results obtained from the algorithm based on the generalized cross-correlation function.



(a) TDE using S-PNLMS algorithm



(b) TDE using S-PNLMS algorithm

Fig. 10. Estimated echo delay curves

[ms]	ROTH	SCOT	CPS-2	HB	M-Max	SPU	SP	S	samples
5	5,1	5,2	5,2	5,2	5,1	5,1	5,0	5,0	40
10	10,3	10,3	10,3	10,4	10,3	10,3	10,4	10,1	80
20	19,6	19,7	20,0	20,0	20,5	20,5	20,1	20,3	160
30	29,3	29,6	30,1	30,0	30,8	30,8	30,1	30,2	240
50	49,0	49,7	49,4	49,8	51,3	51,3	50,3	50,1	400
100	98,1	99,4	98,8	99,6	102,5	102,5	100,5	100,6	800
200	196,2	198,8	197,6	199,1	205,1	205,1	201,3	201,2	1600
300	294,2	298,1	296,4	298,7	307,6	307,6	301,8	301,9	2400

Table 6. Mean values of estimated echo delays

4. Partial, proportionate and sparse control for multi-delay filters

Taking into the account the fact that the generalized cross-correlation algorithms operate in the frequency domain and use advantages of the fast Fourier transform, reasonable computational savings for adaptive filtering algorithms can be achieved as well. The basic operation underlying frequency domain adaptive filters is the transformation of the input signal into a more desirable wave form before its adaptive processing. This is accomplished by the Discrete Fourier Transform (DFT) whereby the input signal is transformed to the frequency domain as shown in Fig. 11.

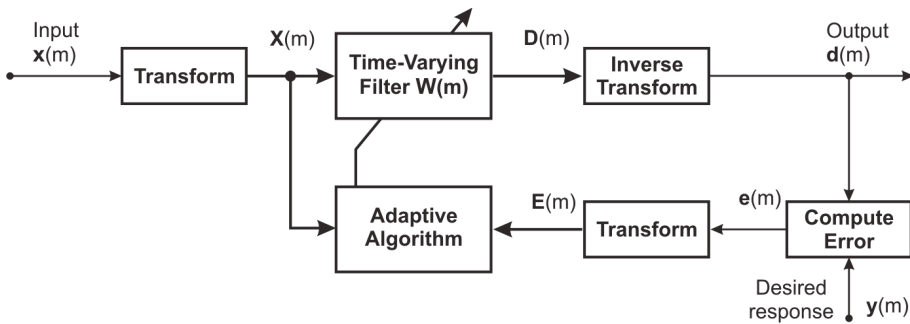


Fig. 11. Frequency domain adaptive filter configuration

Frequency Domain (FD) adaptive filters have primarily two advantages compared to time domain implementations (Dohnal, 1995). The first advantage is the potentially large savings in the computational complexity. The Fast Fourier Transform (FFT) is an efficient implementation of the DFT, which provides these savings. A second advantage is that the DFT structures generate signals that are approximately uncorrelated (orthogonal) (Shynk, 1992). As a result, a time-varying step-size parameter is used for each adaptive weight, thereby allowing a more uniform convergence rate across the adaptive filter. Thus, the overall convergence rate of the algorithm may be improved, sometimes approaching that achievable with Recursive Least Squares (RLS) algorithms without a similar increase in the computational complexity (Widrow & Stearns, 1985). Many other attractive features and variation of the FD adaptive filters can be found in the literature (Khong et al., 2007).

4.1 Sparse partial-update proportionate multi-delay filter

The multi-delay frequency domain adaptation algorithm generates its individual step-size control factors that are mutually orthogonal. This advantage put the Multi-Delay block Frequency domain (MDF) algorithm closer to the analytical RLS algorithm. It differs from the algorithms processing in the time domain with better convergence properties and lower computational complexity. The MDF algorithm decreases the processing delay associated with frequency domain adaptive filters. The delay value is directly proportional to the number of the filter coefficients. The multi-delay approach solves this problem by partitioning the time-domain filter of length L into K sub-filters each of length N with $L=KN$. Consequently, the delay value is reduced by a factor of L/N compared to the full length approach. The diagram of the MDF algorithm is shown in Fig. 12. This algorithm uses the FFT of size N_{FFT} . It equals to the smallest power of two integers larger than or equal to $2L/K=2KN/K=2N$. The MDF algorithm works as follows. Let us first define m , as the frame index. The speech signals are operated on the frame-by-frame basis.

The first step of the MDF algorithm is to convert the most recent overlapped input samples to the frequency domain as follows

$$\mathbf{X}_f(K, m) = \text{diag}\left\{ \text{FFT}\left[x_0(m-1), x_1(m-1), \dots, x_{N-1}(m-1), x_0(m), x_1(m), \dots, x_{N-1}(m) \right]^T \right\} \quad (38)$$

The input vectors incoming to each of the sub-filter blocks can be obtained via the frame index shifting without invoking any computation as follows:

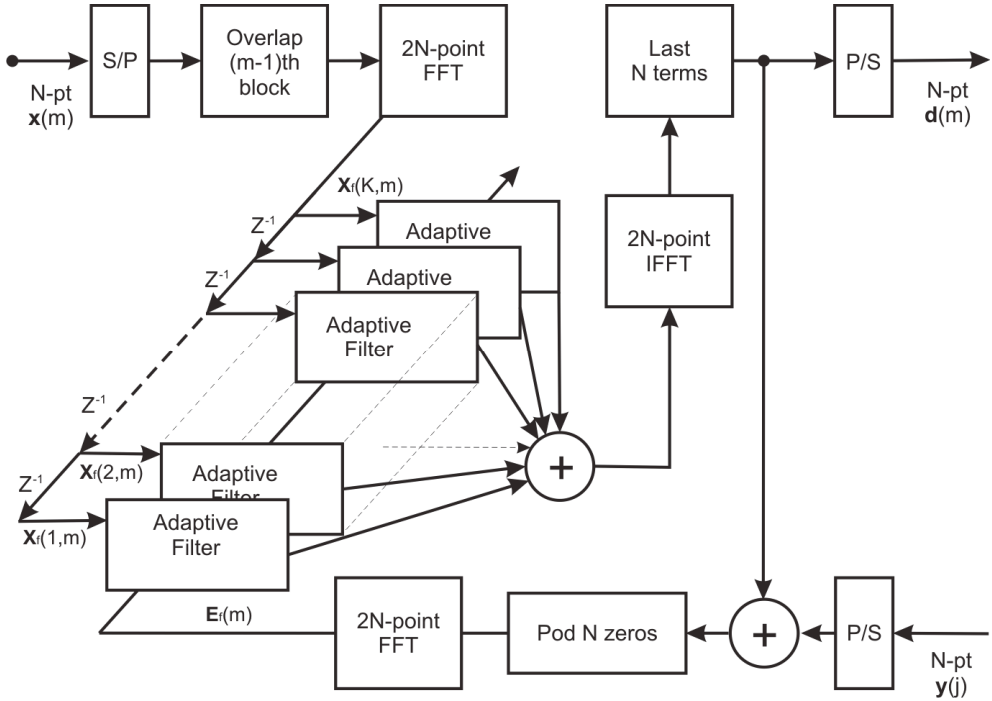


Fig. 12. Block diagram of the MDF algorithm

$$\mathbf{X}_f(k, m) = \mathbf{X}_f(k+1, m-1), \quad k = 1, \dots, K-1 \quad (39)$$

The described approach suggests that only one FFT is needed per frame iteration in order to transform the input vector into the frequency domain. It implies a significant computation saving. The output and the error frequency-domain vectors can be expressed and calculated as

$$\mathbf{d}(m) = \text{last } N \text{ terms of } \left\{ FFT^{-1} \left[\sum_{k=1}^K \mathbf{X}_f(k, m) \cdot \mathbf{W}_f(k, m) \right] \right\} \quad (40)$$

$$\mathbf{E}_f(m) = FFT \left\{ \underbrace{0, 0, \dots, 0}_{N \text{ zeros}}, \underbrace{\mathbf{y}(m) - \mathbf{d}(m)}_{N \text{ terms}} \right\} \quad (41)$$

where $\mathbf{W}_f(k, m)$ is the k^{th} coefficient vector and $\mathbf{d}(m)$ is the desired vector. The coefficient update equations to minimize the Mean Square Error (MSE) are the following

$$\boldsymbol{\varphi}_i(k, m) = \text{first half of } \left\{ FFT^{-1} \left[\mathbf{X}_f(k, m)^* \cdot [S_{MDF}(m) + \delta]^{-1} \cdot \mathbf{E}_f(m) \right] \right\} \quad (42)$$

$$S_{MDF}(m) = \lambda \cdot S_{MDF}(m-1) + (1-\lambda) \cdot \mathbf{X}_f(k, m)^* \cdot \mathbf{X}_f(k, m) \quad (43)$$

$$\boldsymbol{\Phi}_f(k, m) = FFT \left[\boldsymbol{\varphi}_i(k, m), \underbrace{0, 0, \dots, 0}_{N \text{ zeros}} \right]^T \quad (44)$$

$$\mathbf{W}_f(k, m+1) = \mathbf{W}_f(k, m) + K \cdot \mu_k \cdot \boldsymbol{\Phi}_f(k, m) \quad (45)$$

where $k=1, 2, \dots, K$ and μ_k is the block step-size control parameter, $0 < \lambda < 1$ is the forgetting factor. Further, we examine how the partial update technique may be incorporated into the MDF algorithm. For frequency-domain selection, the known approach is to select frequency bins corresponding to the largest magnitude Fourier transform of the tap-input over all the sub-filter blocks $k=1, 2, \dots, K$ (Khong et al, 2008). However, the selection is better to perform on the coefficient-block basis (Deng & Dyba, 2008). It is faster and takes less computation power. Therefore, first we need to calculate metrics for each of the sub-filter in order to perform selection procedure. The number of metrics equals the number of sub-filter blocks K . One particular frequency-domain metric, $\mathbf{Q}_k$, may, for example, represent the sum of the frequency bins in the corresponding sub-filter block

$$\mathbf{Q}_f(k, m) = \sum_{j=1}^{2N} |\mathbf{X}_{fj}(k, m)| \quad (46)$$

There is a little bit different selection approach, which shows better convergence performance compared to the last one. Define the $2L \times 1$ vector $\chi(m)$, which consists of the concatenated Fourier transform of the input signal across all sub-filters, as

$$\chi(m) = [\mathbf{X}_f^T(0, m), \dots, \mathbf{X}_f^T(K-1, m)] = [\chi_1(m), \dots, \chi_{2L}(m)]^T \quad (47)$$

Each element in the $2L \times 2L$ diagonal selection matrix $\mathbf{P}(m)$

$$p_l(m) = \begin{cases} 1, & \text{if } l \text{ corresponds to } M_f \text{ maxima of } |\chi_l(m)|, l = 0, \dots, 2L-1 \\ 0, & \text{otherwise} \end{cases} \quad (48)$$

The tap-selection routine is done by multiplying this matrix with the corresponding input in every k^{th} sub-filter, as it is shown below

$$\tilde{X}_f(k, m) = P_k(m) \cdot X_f(k, m) \quad (49)$$

Otherwise, if it is supposed that the algorithm models a sparse network impulse response, the sparse partial updating scheme should be involved. The control metrics are calculated as follows ($k=1, 2, \dots, K$)

$$Q_f(k, m) = X_f(k, m)^* \cdot W_f(k, m) \quad (50)$$

Here, we suggest using a new type of metric, in order to choose sub-filter blocks for updating. Note, that it should be calculated in the time-domain as follows

$$Q_i(k, m) = \frac{\sum_{i=0}^{N-1} |w_i(k, m)|}{\sqrt{L} \cdot \sqrt{\sum_{j=0}^{L-1} (w_j(m))^2 + \varepsilon}} \quad (51)$$

We have to remember that the minimum number, M , of the sub-filter blocks to be updated must equal or be greater than two ($2 \leq M < K$). This is because the pure delay of the echo path is never known. Thus, if only one sub-filter is chosen, it can not cover all the possibilities of the variation of echo path. Another important question for frequency domain adaptive filters is the FFT length? – From our point of view, it is reasonable using the FFT of 256 samples long. This is an optimal length, which can provide satisfactory results for the high range of MDF algorithms. We have tried to apply different values, but they did not bring sufficient improvements. For example, if the echo path consists of 1024 taps, the adaptive filter should be divided in $1024/(256/2)$ sub-filters. The number K of sub-filters will equal 8.

Before we present a proposed frequency domain algorithm, let us resume the following. The partial update frequency-domain algorithms may use the values of the calculated block metrics as selection criteria. Each type of metrics has its own features. In case of (13, 18, 19), only K ($K < 2L$) metrics are sorted in order to select M ($2 \leq M < K$) active subfilters. Since K is a small number compared to the entire number $2L$ of subfilters coefficients in (14, 15, 16, 17), the sorting overhead may be reduced by factor $2L/N$. We have spoken about the features of sparse echo paths and the possibility of improving their convergence rate by assigning the individual step-size parameters proportionally to the coefficient magnitudes. In the art, there is a big family of various MDF algorithms. Researchers are trying to improve algorithms performance from one hand and to reach computational savings from the other hand. Further we present our approach how to deal with this contradictory problem. First, we describe the basic idea of sparseness measure, and show how to incorporate it into the MDF framework. The concept of sparseness refers to a representational scheme where only a few units (out of a large massive) are effectively used to represent typical data vectors. In effect, this implies most units taking values close

to zero while only few take significantly non-zero values. Numerous sparseness measures have been proposed and used in the literature to date (Benesty et al, 2006). Such measures quantify how much energy of a vector is packed into only a few components. On a normalized scale, the sparsest possible vector (only a single component is non-zero) should have a sparseness of one, whereas a vector with all elements equal should have a sparseness of zero. In our approach, we use a sparseness measure based on the relationship between the L_1 norm and the L_2 norm:

$$\xi(m) = 1 - \frac{\|w_j(m)\|_1}{\sqrt{L} \cdot \|w_j(m)\|_2 + \varepsilon} = 1 - \frac{\sum_{j=0}^{L-1} |w_j(m)|}{\sqrt{L} \cdot \sqrt{\sum_{j=0}^{L-1} (w_j(m))^2} + \varepsilon} \quad (52)$$

Here, L is the length of the adaptive filter in the time domain. This function evaluates to unity if and only if w contains only a single non-zero component, and takes a value of zero if and only if all components are equal (up to signs), interpolating smoothly between the two extremes. We use this sparseness measure ξ of the time domain estimated filter coefficients to adaptively control the number of taps that need to be updated

$$M_t = L \cdot (1 - \xi(m)) \quad (53)$$

Note that the total number of corresponding frequency domain coefficients is $2L$ for an L -long time domain adaptive filter, therefore

$$M_f = 2 \cdot M_t = 2 \cdot L \cdot (1 - \xi(m)) \quad (54)$$

If selection is performed on the block-based basis, then the corresponding number of sub-filters in frequency domain equals

$$M = \text{round} \left\{ \frac{2 \cdot L \cdot (1 - \xi(m))}{2 \cdot N} \right\} = \text{round} \{ K \cdot (1 - \xi(m)) \} \quad (55)$$

Fig. 13 shows sparseness measure for sparse and dispersive impulse responses. It is also shown the number of adaptive filter coefficients that should be updated, if partial updating scheme is considered.

The convergence process, by itself, can be divided into two stages, i.e. before and after the convergence of the impulse response (or some time, before and after the convergence of the considerable/large coefficients). It has been showed in the time domain adaptive filtering that it is better using PNLMS/IPNLMS at the initial stage for fast convergence, and afterwards switching to NLMS for assuring better misalignment. Here, for recognizing between two stages in adaptation process, we propose to use the following parameter

$$\eta(m) = 10 \cdot \log_{10} \left(\frac{\sum_{j=0}^{L-1} (w_j(m))^2}{L \cdot 0,001} \right) \quad (56)$$

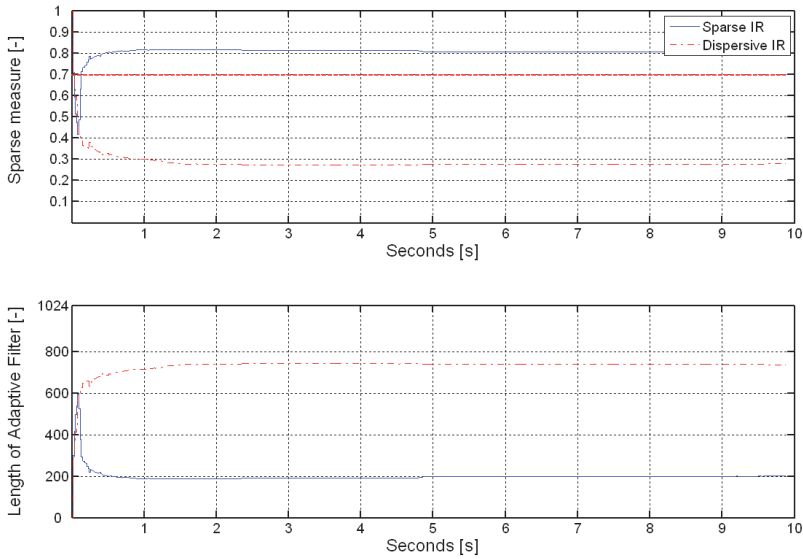


Fig. 13. Sparseness measure and its impact on length of adaptive filter

The quantity in Eq. (56) represents an energy measure in [dBm] within an estimated impulse response. Fig. 14 demonstrates the curve for this parameter for two frequency domain algorithms. It can be observed that after 0.5 seconds estimation of IR energy measure for IPMDF stops fluctuating. Consequently, this fact can be used for switching between different adaptation schemes.

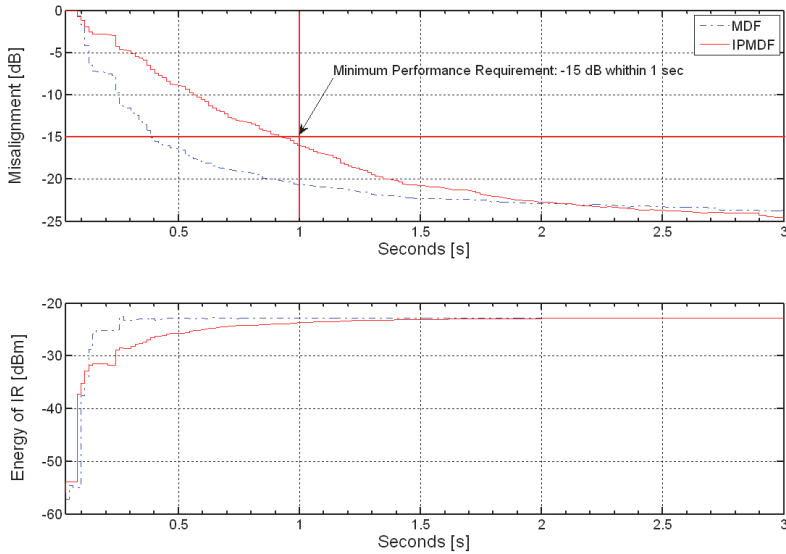


Fig. 14. Estimated misalignment and energy measure for MDF and IPMDF algorithms

Before representing the proposed MDF scheme, declare the following statements:

I. Utilization of sparseness measure, $0 < \xi(m) < 1$ (for real IR)

- a. $\xi(m) < 0.7$
 - i. IR is considered dispersive
 - ii. $M_t = L$, if fully updated scheme is chosen (mostly during initial period)
 - iii. $L/K \leq M_t < L$, if partial updated scheme is chosen
- b. $\xi(m) \geq 0.7$
 - i. IR is considered sparse
 - ii. $M_t = L \cdot (1 - \xi(m))$

II. Utilized updating schemas

- a. non-partial
- b. χ - based selection (coefficient-based), $2L$ -vector
- c. μ - based selection (block-based), K -vector

III. Utilization of IR energy measure, η

- a. $|\eta(m) - \eta(m-1)| \leq \Delta_\eta$, switching to the MDF algorithm
- b. $|\eta(m) - \eta(m-1)| > \Delta_\eta$, switching to the IPMDF (SC-IPMDF) algorithm

Finally, our proposed algorithm can be described as:

$$\begin{aligned}
 & \text{1st stage: } |\eta(m) - \eta(m-1)| \leq \Delta_\eta \\
 & \quad \text{if } \xi(m) < 0.7, \\
 & \quad \quad M_f = 2 \cdot L \quad \text{or} \quad M = K \\
 & \quad \quad \mathbf{w}(k, m+1) = \mathbf{w}(k, m) + L \cdot \mu_k \cdot \mathbf{G}_t(k, m) \cdot \boldsymbol{\varphi}_t(k, m) \\
 & \quad \quad \mathbf{G}_t(k, m) = \text{diag}\{g_{kN}(m), g_{kN+1}(m), \dots, g_{kN+N-1}(m)\} \\
 & \quad \quad g_{kN+l} = \frac{1-\alpha}{2 \cdot L} + \frac{(1+\alpha) \cdot |w_{kN+l}(m)|}{2 \cdot \sum_{j=0}^{L-1} |w_j(m)| + \varepsilon} \\
 & \quad \quad \mu_k = \text{const} \\
 & \quad \quad \text{if } m \cdot N \leq L, \\
 & \quad \quad \quad \alpha = \text{const} \\
 & \quad \quad \text{else } m \cdot N > L, \\
 & \quad \quad \quad \alpha = 1 - 2 \cdot \frac{L}{L - \sqrt{L}} \cdot \left(1 - \frac{\|w_j(m)\|_1}{\sqrt{L} \cdot \|w_j(m)\|_2 + \varepsilon} \right) \\
 & \quad \quad \text{else } \xi(m) \geq 0.7, \\
 & \quad \quad M_f = 2 \cdot L \cdot (1 - \xi(m)) \quad \text{or} \quad M = \text{round}\{K \cdot (1 - \xi(m))\} \\
 & \quad \quad \text{if } \text{mod}(m, T) = 0, \\
 & \quad \quad p_l(m) = \begin{cases} 1, & \text{if } l \text{ corresponds to } M_f \text{ maxima of } |\chi_l(m)|, l = 0, \dots, 2L-1 \\ 0, & \text{otherwise} \end{cases} \\
 & \quad \quad \tilde{\mathbf{X}}_f(k, m) = P_k(m) \cdot \mathbf{X}_f(k, m) \\
 & \quad \quad \boldsymbol{\varphi}_t(k, m) = \text{first half of } \left\{ FFT^{-1} \left[\tilde{\mathbf{X}}_f(k, m)^* \cdot [S_{MDF}(m) + \delta]^{-1} \cdot \mathbf{E}_f(m) \right] \right\} \\
 & \quad \quad \mathbf{w}(k, m+1) = \mathbf{w}(k, m) + L \cdot \mu_k \cdot \mathbf{G}_t(k, m) \cdot \boldsymbol{\varphi}_t(k, m) \\
 & \quad \quad \mu_k = \text{const}
 \end{aligned}$$

To compare the performance of the proposed algorithm with reference MDF, IPMDF, SC-IPMDF adaptive filtering algorithms, we implement all four adaptive filters in MATLAB environment. These MDF-based filters estimate the impulse response of the predefined echo paths, which are specified by the ITU-T Recommendation (ITU-T G.168, 2002). The tested impulse responses are 1024 taps long. All experiments are performed using pre-recorded real speech signals. The number K of sub-filters equals 8. The following values for control parameters are used: $\mu_{initial} = 0.1$, $a_{initial} = -0.75$, $\varepsilon = 10^{-3}$, $\Delta_{\eta} = 0.05$ dBm, $T_{\eta} = 8$. The

performance of each algorithm is studied using the normalized misalignment parameter, which can be estimated as follows

$$MIS = 10 \cdot \log_{10} \left(\frac{\|h - w(m)\|_2}{\|h\|_2} \right) \text{ in } [dB] \quad (57)$$

where h is a true impulse response of length L . Another criterion is Echo Return Loss Enhancement (ERLE), which is used in real-life environment to evaluate performance

$$ERLE = 10 \cdot \log_{10} \left(\frac{\|y(m) - d(m)\|_2}{\|y(m)\|_2} \right), [dB] \quad (58)$$

where $y(m)$ is a desired signal (echo) and $d(m)$ is adaptive filter's output. Note that any reliable adaptive filter with disabled residual echo suppressor has to achieve ERLE of -15dB within 1 second after starting convergence process (ITU-T G.131, 2003). Fig. 15, 16, 17 illustrate an application of μ -based selection metric. This kind of metric is used for sub-filter selection, when estimated sparse measure parameter, ξ_s equals or larger than 0.7. This value was defined experimentally during multiple trials for numerous types of echo path. We suggest using μ -based selection metric as an individual block step-size parameter. It helps accelerating a speed of convergence by allocating larger step-size values for currently updated sub-filters. If you look at the diagram illustrated in Fig. 17 carefully, you will notice that the energy, which is available for adaptation, is concentrated around the sparse region of the echo path. Thus, this fact can be used for selecting sub-filters to be updated along with setting the step-size parameter for these sub-filters. When estimation of sparse measure, ξ_s , is smaller than 0.7, we suggest switching to the χ - based selection metric.

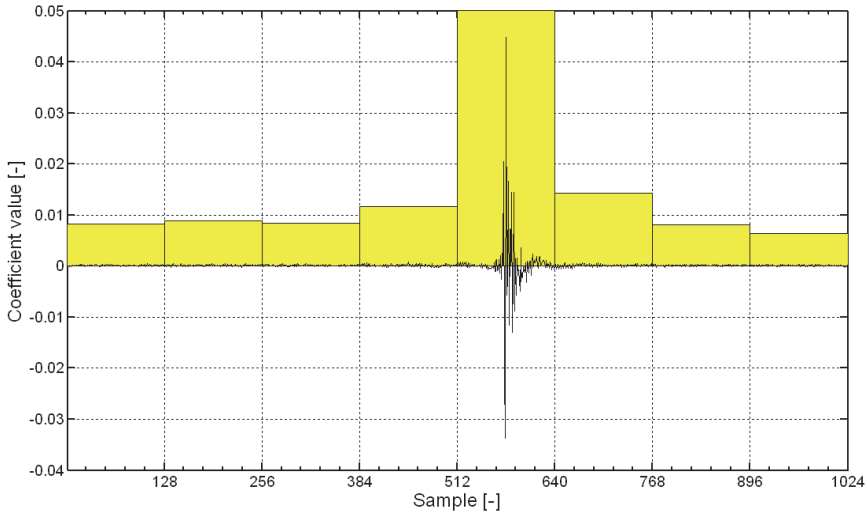


Fig. 15. Sparse impulse response and estimated μ -based selection metrics

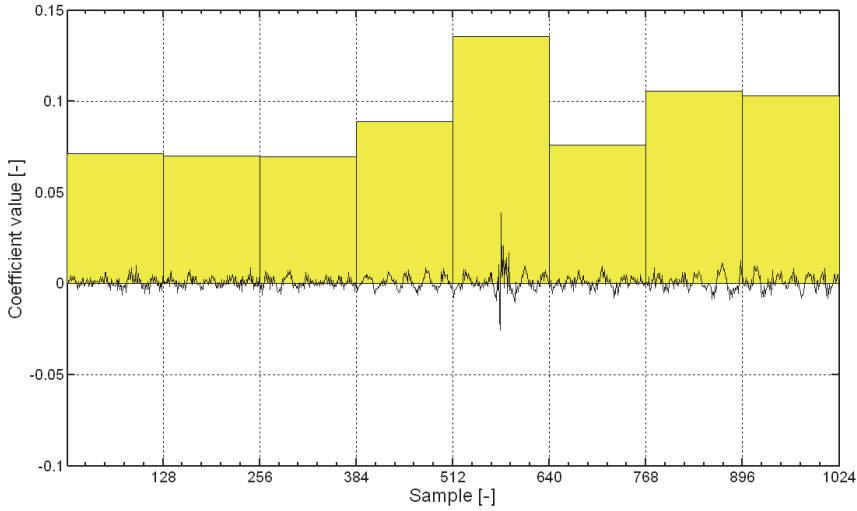


Fig. 16. Dispersive impulse response and estimated μ -based selection metrics

Fig. 18 demonstrates the normalized misalignment and ERLE parameters obtained for real speech signals. The proposed partially updated scheme for MDF shows the similar performance comparing to the other three fully updated frequency domain algorithms. During our future work we are going to enhance the above described algorithm and propose a new class of partial sparse-controlled robust algorithms, which will work reliably, even in double-talk situation. We will apply all the knowledge, which were presented within this particular chapter. Further to conclude the chapter, let us provide summary of material and make several contributions according to the proposed algorithms.

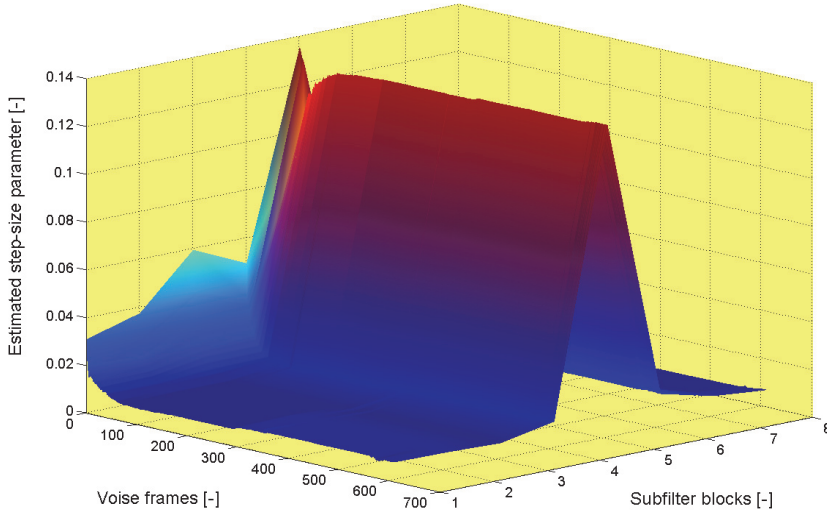


Fig. 17. Step-size parameter estimated using μ -based selection metric

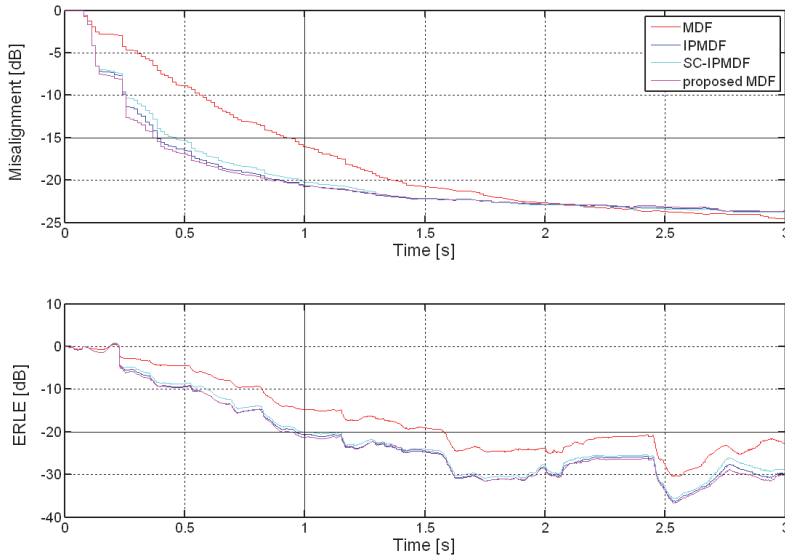


Fig. 18. Misalignment and ERLE curves

5. Conclusion

The first section outlines a basic principle of echo control in packet-based networks. It explains why it is so important to provide monitoring during telephone conversations. When delivering the VoIP service in the packet-switching network, it is important to have the value of the echo delay under control. The increasing transmission delay associated with packet data transmission can make a negligible echo more annoying. Therefore, it is suggested using the echo assessment algorithm. Its purpose is to add an additional attenuation to a particular voice channel (which in terms means to activate an echo canceller), so as to remove the unwanted echo in time. In the second section, we consider an opportunity of using cross-correlation for estimating echo delays. That section provides readers along with up-to-date correlation-based TDE algorithms, which we use to estimate the echo path delays. The problem of long delays taken place in the packet-switching network is considered as a topic of interest. The experiments show that the algorithms precision decreases with increasing transmission delays. The generalized cross-correlation algorithms operating in the frequency domain provide more reliable result comparing to the standard cross-correlation and normalized cross-correlation algorithms. As an alternative to correlation-based methods, techniques, which use adaptive filtering algorithms, can be also applied. Therefore, the third section presents numerous partial-update algorithms and their application to delay estimation. The echo assessment is based on the reduced complexity partial-update adaptive filters. The experiments show a reliable performance of these algorithms. However, their precision suffers during the initial stage of convergence. According to the ITU-T Recommendation G.168, this period should not last more than one second. The Multi-Delay block Frequency domain (MDF) adaptive algorithm can easily

outperform all existing time domain algorithms. Moreover, taking into the account the fact that the generalized cross-correlation algorithms operate in the frequency domain and use advantages of the fast Fourier transform, further computational savings for the adaptive filters are achieved in the frequency domain. Therefore, the fourth section deals with partial, proportionate, and sparse-controlled adaptive filtering algorithms working in the frequency domain. What we claimed, within this section, is: a new metric for performing partial updating; a new approach for designating transitions between MDF and IPMDF-based updating schemas; a method for estimating step-size control parameter; a new partially updated sparseness-controlled improved proportionate multi-delay filter; all the approaches are suitable for implementation whether in time or frequency domains. The proposed algorithm has both a performance compared to the IPMDF and SC-MDF algorithms and reduced computational complexity along with the adjustable step-size parameter. Although the preferred embodiments of the proposed algorithm have been described, it will be understood by those skilled in the art that various changes may be made thereto without departing from the main scope of the invention or the appended claims.

6. Acknowledgment

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS 10/275/OHK3/3T/13 and by Grant The Ministry of Education, Youth and Sports No. MSM6840770014.

7. References

- Choi, B.-K.; Moon, S.; Zhi-Li, Z. (2004) Analysis of Point-To-Point Packet Delay In an Operational Network, *Proceedings of INFOCOM 2004, Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 3, pp.1797-1807, ISBN 0-7803-8355-9, Hong Kong, March 7-11, 2004
- Gordy, J.D.; Goubran, R.A. (2006) On the Perceptual Performance Limitations of Echo Cancellers in Wideband Telephony, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, issue 1, pp.33-42, ISSN: 1558-7916
- Nisar, K.; Hasbullah, H.; Said, A.M. (2009) Internet Call Delay on Peer to Peer and Phone to Phone VoIP Network, *Proceedings of ICCET '09, International Conference on Computer Engineering and Technology*, vol. 2, pp.517-520, ISBN 978-1-4244-3334-6, Singapore, Januar 22-24, 2009
- Dyba, R.A. (2008) Parallel Structures for Fast Estimation of Echo Path Pure Delay and Their Applications to Sparse Echo Cancellers, *Proceedings of CISS 2008, 42nd Annual Conference on Information Sciences and Systems*, pp.241-245, ISBN 978-1-4244-2246-3, Princeton, NJ, March 19-21, 2008
- Hongyang, D.; Dyba, R.A. (2008) Efficient Partial Update Algorithm Based on Coefficient Block for Sparse Impulse Response Identification, *Proceedings of CISS 2008, 42nd Annual Conference on Information Sciences and Systems*, pp.233-236, ISBN 978-1-4244-2246-3, Princeton, NJ, March 19-21, 2008
- Khong, A.W.H.; Naylor, P.A. (2006) Efficient Use Of Sparse Adaptive Filters, *Proceedings of ACSSC '06, Fortieth Asilomar Conference on Signals, Systems and Computers*, pp.1375-1379, ISBN 1-4244-0784-2, Pacific Grove, CA, October 29 -November 1, 2006

- Hongyang, D.; Dyba, R.A. (2009) Partial Update PNLMS Algorithm for Network Echo Cancellation, *Proceedings of ICASSP 2009, IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.1329-1332, ISBN 978-1-4244-2353-8, Taipei, April 19-24, 2009
- ITU-T Recommendation G.131 (2003) *Talker Echo and its Control*
- ITU-T Recommendation G.168 (2002) *Digital network echo cancellers*
- Carter, G.C. (1976) *Time Delay Estimation*, Ph.D. dissertation, University of Connecticut, Storrs, CT, pp.67-70
- Mueller, M. (1975) Signal Delay, *IEEE Transactions on Communications*, vol. 23, issue 11, pp.1375-1378
- Buchner, H.; Benesty, J.; Gansler, T.; Kellermann, W. (2006) Robust Extended Multidelay Filter and Double-talk Detector for Acoustic Echo Cancellation, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, issue 5, pp.1633-1644, ISSN 1558-7916
- Youn, D.H.; Ahmed, N.; Carter, G.C. (1983) On the Roth and SCOTH Algorithms: Time-Domain Implementations, *In Proceedings of the IEEE*, vol. 71, issue 4, pp.536-538, ISSN 0018-9219
- Zetterberg, V.; Pettersson, M.I.; Claesson, I. (2005) Comparison Between Whitened Generalized Cross-correlation and Adaptive Filter for Time Delay Estimation, *Proceedings of MTS/IEEE, OCEANS*, vol. 3, ISBN 0-933957-34-3, Washington D.C., September 17-23, 2005
- Hertz, D. (1986) Time Delay Estimation by Combining Efficient Algorithms and Generalized Cross-correlation Methods, *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, issue 1, pp.1-7, ISSN 0096-3518
- Knapp, C.; Carter, G.C. (1976) The Generalized Correlation Method for Estimation of Time Delay, *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, issue 4, pp.320-327, ISSN 0096-3518
- Wilson, K.W.; Darrell, T. (2006) Learning a Precedence Effect-Like Weighting Function for the Generalized Cross-Correlation Framework, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, issue 6, pp.2156-2164, ISSN 1558-7916
- Tianshuang, Q.; Hongyu, W. (1996) An Eckart-weighted adaptive time delay estimation method, *IEEE Transactions on Signal Processing*, vol. 44, issue 9, pp.2332-2335, ISSN 1053-587X
- Chen, J.; Benesty, J.; Huang, Y.A. (2006) The SCOT Weighted Adaptive Time Delay Estimation Algorithm Based on Minimum Dispersion Criterion, *Journal of EURASIP on Applied Signal Processing*, vol. 2006, ISBN 978-1-4244-7047-1
- Anderson, M.P.; Woessner, W.W. (1992) *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*, Academic Press (2nd Edition ed.), ISBN 978-0120594856, USA
- Widrow, B. (2005) Thinking about thinking: the discovery of the LMS algorithm, *IEEE Magazine on Signal Processing*, vol. 22, issue 1, pp.100-106, ISSN 1053-5888
- Haykin, S. (2001) *Adaptive Filter Theory, Fourth Edition*, Prentice-Hall, ISBN 0130901261, USA
- Zetterberg, V.; Pettersson, M.I.; Claesson, I. (2005) Comparison Between Whitened Generalized Cross-correlation and Adaptive Filter for Time Delay Estimation, *Proceedings of MTS/IEEE, OCEANS*, vol. 3, ISBN 0-933957-34-3, Washington D.C., September 17-23, 2005

- Emadzadeh, A.A.; Lopes, C.G.; Speyer, J.L. (2008) Online time delay estimation of pulsar signals for relative navigation using adaptive filters, *Proceedings of 2008 IEEE/ION, Position, Location and Navigation Symposium*, pp.714-719, ISBN 978-1-4244-1536-6, Monterey, CA, May 5-8, 2008
- Duttweiler, D.L. (2000) Proportionate normalized least-mean-squares adaptation in echo cancellers, *IEEE Transactions on Speech and Audio Processing*, vol. 8, issue 5, pp.508-518, ISSN 1063-6676
- Hongyang, D.; Doroslovacki, M. (2006) Proportionate adaptive algorithms for network echo cancellation, *IEEE Transactions on Signal Processing*, vol. 54, issue 5, pp.1794-1803, ISSN: 1053-587X
- Paleologu, C.; Benesty, J.; Ciochina, S. (2010) An improved proportionate NLMS algorithm based on the l0 norm, *Proceedings of ICASSP '10, 2010 IEEE International Conference on Acoustics Speech and Signal Processing*, pp.309-312, ISSN 1520-6149, Dallas, TX, March 14-19, 2010
- Gay, S.L. (1998) An efficient, fast converging adaptive filter for network echo cancellation, *Proceedings of the Thirty-Second Asilomar Conference on Signals, Systems & Computers*, vol. 1, pp.394-398, ISBN 0-7803-5148-7, Pacific Grove, CA, November 1-4, 1998
- Benesty, J.; Gay, S.L. (2002) An improved PNLMS algorithm, *Proceedings of ICASSP '02, 2002 IEEE International Conference on Acoustics Speech and Signal Processing*, vol. 2, pp.1881-1884, ISSN 1520-6149, Minneapolis, MN, USA, April 27-30, 2002
- Fevrier, I.J.; Gelfand, S.B.; Fitz, M.P. (1999) Reduced complexity decision feedback equalization for multipath channels with large delay spreads, *IEEE Transactions on Communications*, vol. 47, issue 6, pp.927-937, ISSN 0090-6778
- Douglas, S.C. (1997) Adaptive filters employing partial updates, *IEEE Transactions on Circuits and Systems II*, vol. 44, issue 3, pp.209-216, ISSN 1057-7130
- Aboulnasr, T.; Mayyas, K. (1999) Complexity reduction of the NLMS algorithm via selective coefficient update, *IEEE Transactions on Signal Processing*, vol. 47, issue 5, pp.1421-1424, ISSN 1053-587X
- Aboulnasr, T.; Mayyas, K. (1998) MSE analysis of the M-Max NLMS adaptive algorithm, *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing*, vol. 3, article ID 10.1109/ICASSP.1998.681776, pp.1669-1672, ISBN 0-7803-4428-6, Seattle, WA, May 12-15, 1998
- Schertler, T (1998) Selective block update of NLMS type algorithms, *In Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing*, vol. 3, pp.1717-1720, ISBN 0-7803-4428-6, Seattle, WA, May 12-15, 1998
- Dogancay, K.; Tanrikulu, O. (2001) Adaptive filtering algorithms with selective partial updates, *IEEE Transactions on Circuits and Systems II*, vol. 48, issue 8, pp.762-769, ISSN 1057-7130
- Naylor, P.A.; Sherliker, W. (2003) A short-sort M-Max NLMS partial-update adaptive filter with applications to echo cancellation, *Proceedings of ICASSP '03, 2003 IEEE International Conference on Acoustics Speech and Signal Processing*, vol. 5, pp.373-376, ISBN 0-7803-7663-3, Hong Kong, April 7-10, 2003
- Jinhong, W.; Doroslovacki, M. (2008) Partial update NLMS algorithm for sparse system identification with switching between coefficient-based and input-based selection, *Proceedings of CISS 2008, 42nd Annual Conference on Information Sciences and Systems*, vol.3, pp.237-240, ISBN 978-1-4244-2246-3, Princeton, NJ, March 19-21, 2008

- Shynk, J. J. (1992) Frequency-Domain and Multirate Adaptive Filtering, *IEEE Transactions on Signal Processing Magazine*, vol. 9, pp. 474–475, ISSN: 1053-5888
- Widrow, B.; Stearns, S.D. (1985) *Adaptive Signal Processing*, Prentice-Hall, ISBN 0130040290, USA
- Khong, A. W. H.; Naylor, P. A.; Benesty, J. (2007) A low delay and fast converging improved proportionate algorithm for sparse system identification, *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2007, Article ID 84376, 8 pages
- Khong, A. W. H.; Xiang, L.; Doroslovacki, M.; Naylor, P. A. (2008) Frequency domain selective tap adaptive algorithms for sparse system identification, *Proceedings of ICASSP 2008, International Conference on Acoustics, Speech and Signal processing*, pp. 229–233, ISBN 978-1-4244-1483-3, Las Vegas, NV, March 31-April 4, 2008
- Deng, H.; Dyba, R. (2008) Efficient Partial Update Algorithm Based on Coefficient Block for Sparse Impulse Response Identification, *Proceedings of CISS 2008, Conference on Information Sciences and Systems*, pp. 233-236, ISBN 978-1-4244-2246-3, Princeton, NJ, March 19-21, 2008
- Benesty, J.; Huang, Y. A.; Chen, J.; Naylor, P. A. (2006) Adaptive algorithms for the identification of sparse impulse responses, In: *Selected Methods for the Cancellation of Acoustical Echoes, the Reduction of Background Noise, and Speech Processing*, Hänslér, E.; Schmidt, G., Springer, pp 125-153, ISBN 978-3-540-33212-1, Berlin

Part 2

Medical Applications

Adaptive Noise Removal of ECG Signal Based On Ensemble Empirical Mode Decomposition

Zhao Zhidong, Luo Yi and Lu Qing
Hangzhou Dianzi University
China

1. Introduction

The electrocardiogram (ECG) records the electrical activity of the heart, which is a noninvasively recording produced by an electrocardiographic device and collected by skin electrodes placed at designated locations on the body. The ECG signal is characterized by six peaks and valleys, which are traditionally labeled P, Q, R, S, T, and U, shown in figure 1.

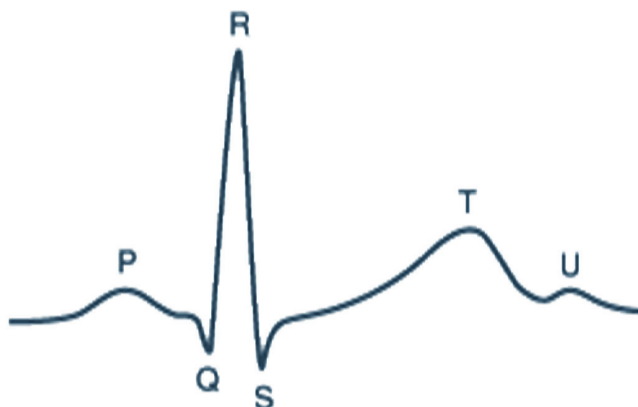


Fig. 1. ECG signal

It has been used extensively for detection of heart disease. ECG is non-stationary bioelectrical signal including valuable clinical information, but frequently the valuable clinical information is corrupted by various kinds of noise. The main sources of noise are: power-line interference from 50–60 Hz pickup and harmonics from the power mains; baseline wanders caused by variable contact between the electrode and the skin and respiration; muscle contraction form electromyogram (EMG) mixed with the ECG signals; electromagnetic interference from other electronic devices and noise coupled from other electronic devices, usually at high frequencies. The noise degrades the accuracy and precision of an analysis. Obtaining true ECG signal from noisy observations can be formulated as the problem of signal estimation or signal denoising. So denoising is the method of estimating the unknown signal from available noisy data. Generally, excellent

ECG denoising algorithms should have the following properties: Ameliorate signal-to-noise ratio (SNR) for obtaining clean and readily observable signals; Preserve the original characteristic waveform and especially the sharp Q, R, and S peaks, without distorting the P and T waves.

A lot of methods have been proposed for ECG denoising. In general both linear and nonlinear filters are presented, such as elliptic filter, median filter, Wiener filter and wavelet transform etc. These methods have some drawbacks. They remove not only noise but also the high frequency components of non-stationary signals. In the worse they can remove the characteristic points of signals that are crucial for successful detection of waveform. In recent years wavelet transform (WT) has become favourable technique in the field of signal processing. Donoho et al proposed the denoising method called "wavelet shrinkage"; it has three steps: forward wavelet transform, wavelet coefficients shrinkage at different levels and the inverse wavelet transform, which work in denoising the signals such as Universal threshold, SureShrink, Minimax. Wavelet shrinkage methods have been successful in denoising ECG signals (Agante, P.M&Marques J.P, 1999; Brij N. Singh & Arvind.K, 2006). A New wavelet shrinkage method for denoising of biological signals is proposed based on a new thresholding filter (Prasad V.V.K.D.V; Siddaiah P; Rao BP,2008).De-noising using traditional DWT has a translation variance problem which results in Pseudo-Gibbs phenomenon in the Q and S waves , so the following algorithms tried to solve this problem: used cyclic shift tree de-noising technique for reducing white Gaussian noise or random noise, EMG noise and power line interference (Kumari, R.S.S. et al ,2008).The selected optimal wavelets basis has been investigated with suitable shrinkage method to de-noise ECG signals, not only it obtains higher SNR, but preserves the peaks of R wave in ECG(Suyi Li. et al ,2009).Scale-dependent threshold methods are successively proposed. A new thresholding procedure is proposed based on wavelet denoising using subband dependent threshold for ECG signals: The S-median-DM and S-median thresholds (Poornachandra.S, 2008).

In this work, in order to enhance ECG, the new adaptive shrunken denoising method based on Ensemble Empirical Mode Decomposition (EEMD) is presented that has a good influence in enhancing the SNR, and also in terms of preserving the original characteristic waveform. The paper is organized as follows: section 2 introduces Empirical Mode Decomposition (EMD) and EEMD is studied in section 3. EMD is a relatively new, data-driven adaptive technique used to decompose ECG signal into a series of Intrinsic Mode Functions (IMFs). The EEMD overcomes largely the mode mixing problem of the original EMD by adding white noise into the targeted signal repeatedly and provides physically unique decompositions. Wavelet shrinkage is studied in section 4; the wavelet shrinkage denoising method is simply signal extraction from noisy signal via wavelet transform. It has been shown to have asymptotic near-optimality properties over a wide class of functions. The crucial points are the selections of threshold value and thresholding function. The generalized threshold function is build. Computationally exact formulas of bias , variance and risk of generalized threshold function are derived. Section 5 concentrates on adaptive threshold values based on EEMD.Noisy signal is decomposed into a series of IMFs, and then the threshold values are derived by the noise energies of each IMFs. To evaluate the performance of the algorithm, Test signal and Clinic noisy ECG signals are processed in section 6. The results show that the novel adaptive threshold denoising method can achieve the optimal denoising of the ECG signal. Conclusions are presented in section7.

2. Empirical mode decomposition

EMD has recently been proposed by N.E.Huang in 1998 which is developed as a data-driven tool for nonlinear and non-stationary signal processing. EMD can decompose signal into a series of IMFs subjected to the following two conditions:

1. In the whole dataset, the number of extrema and the number of zero-crossing must either be equal or differ at most by one.
2. At any time, the mean value of the envelope of the local maxima and the envelope of the local minima must be zero.

Figure.2 shows a classical IMF. The IMFs represent the oscillatory modes embedded in signal. Each IMF actually is a zero mean monocomponent AM-FM signal with the following form:

$$x(t) = a(t)\cos\phi(t) \quad (1)$$

with time varying amplitude envelop $a(t)$ and phase $\phi(t)$. The amplitude and phase have both physically and mathematically meaning.

Most signals include more than one oscillatory mode, so they are not IMFs. EMD is a numerical sifting process to disintegrate empirically a signal into a finite number of hidden fundamental intrinsic oscillatory modes, that is, IMFs. The sifting process can be separated into following steps:

1. Finding all the local extrema, including maxima and minima; then connecting all the maxima and minima of signal $x(t)$ using smooth cubic splines to get its upper envelope $x_{up}(t)$ and lower envelope $x_{low}(t)$.
2. Subtracting mean of these two envelopes $m_1(t) = (x_{up}(t) + x_{low}(t)) / 2$ from the signal to get their difference: $h_1(t) = x(t) - m_1(t)$.
3. Regarding the $h_1(t)$ as the new data and repeating steps 1 and 2 until the resulting signal meets the two criteria of an IMF, defined as $c_1(t)$. The first IMF $c_1(t)$ contains the highest frequency component of the signal. The residual signal $r_1(t)$ is given by $r_1(t) = x(t) - c_1(t)$.
4. Regarding $r_1(t)$ as new data and repeating steps (1) (2) (3) until extracting all the IMFs. The sifting procedure is terminated until the Mth residue $r_M(t)$ becomes less than a predetermined small number or becomes monotonic.

The original signal $x(t)$ can thus be expressed as following:

$$x(t) = \sum_{j=1}^M c_j(t) + r_M(t) \quad (2)$$

$c_j(t)$ is an IMF where j represents the number of corresponding IMF and $r_M(t)$ is residue.

The EMD decomposes non-stationary signals into narrow-band components with decreasing frequency. The decomposition is complete, almost orthogonal, local and adaptive. All IMFs form a completely and “nearly” orthogonal basis for the original signal. The basis directly comes from the signal which guarantees the inherent characteristic of signal and avoids the diffusion and leakage of signal energy. The sifting process eliminates

riding waves, so each IMF is more symmetrical and is actually a zero mean AM-FM component.

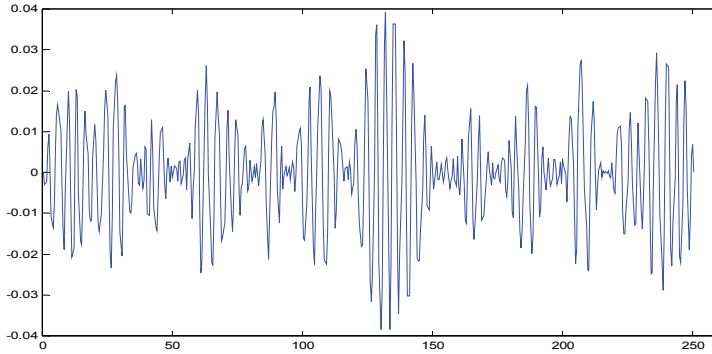


Fig. 2. A classical IMF

The major disadvantage of EMD is the so-called mode mixing effect. For example, the simulated signal is defined as follows:

$$\begin{aligned}
 s(t) &= \sin(2 \times \pi \times t) + 10 \times w(t) * \delta(t - n) \quad (n = \dots, -2, -1, 0, 1, 2, \dots) \\
 w(t) &= \begin{cases} t - 0.2 - 0.015 \times m, & 0.2 + 0.03 \times m < t < 0.215 + 0.03 \times m \\ 0.215 + 0.015 \times m - t, & 0.215 + 0.03 \times m < t < 0.23 + 0.03 \times m \end{cases} \\
 m &= 0, 1, 2, 3
 \end{aligned} \tag{3}$$

The signal is composed of sine wave and impulse functions, shown as figure3. It is decomposed into a series of IMFs by EMD, illustrated as figure 4. The decomposition is polluted by mode mixing, which indicates that oscillations of different time scales coexist in a given IMF, or that oscillations with the same time scale have been assigned to different IMFs.

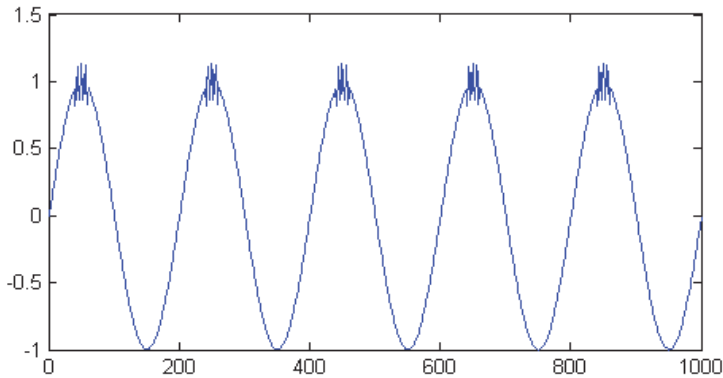


Fig. 3. Simulated signal

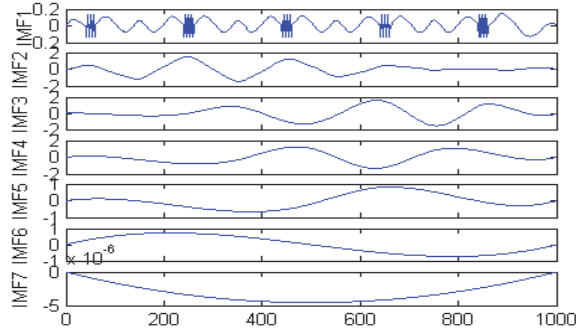


Fig. 4. IMFs obtained by EMD

3. Ensemble empirical mode decomposition

Ensemble EMD (EEMD) was introduced to remove the mode-mixing effect. The EEMD overcomes largely the mode mixing problem of the original EMD by adding white noise into the targeted signal repeatedly and provides physically unique decompositions when it is applied to data with mixed and intermittent scales.

The EEMD decomposing process can be separated into following steps:

1. Add a white noise series $w(t)$ to the targeted data $x(t)$, the noise must be zero mean and variance constant, so $X(t) = x(t) + w(t)$.
2. Decompose the data with added white noise into Intrinsic Mode Functions (IMFs) and residue r_n

$$X(t) = \sum_{j=1}^n c_j + r_n \quad (4)$$

3. Repeat step 1 and step 2 N times, but with different white noise series $w_i(t)$ each time, so

$$X_i(t) = \sum_{j=1}^n c_{ij} + r_{in} \quad (5)$$

4. Obtain the ensemble means of corresponding IMFs of the decompositions as the final result. Each IMF is obtained by decomposed the targeted signal.

$$c_j = \frac{1}{N} \sum_{i=1}^N c_{ij} \quad (6)$$

This new approach utilizes the full advantage of the statistical characteristics uniform distribution of frequency of white noise to improve the EMD method. The above signal is decomposed into a series of IMFs by EEMD, which is shown in figure 5. Through adding white noise into the targeted signal makes all scaled continues to avoid mode mixing phenomenon. Comparing the IMF component of the same level, EEMD has more concentrated and band limited components.

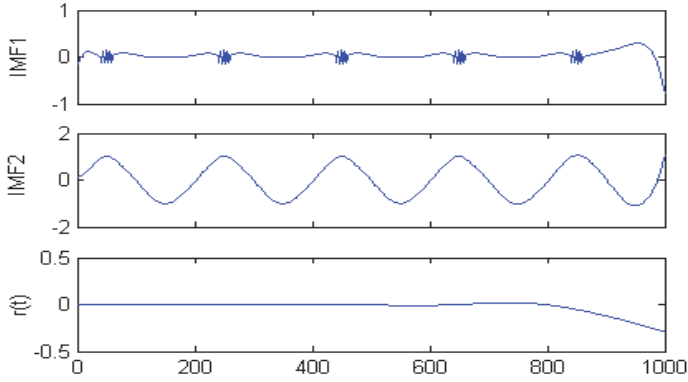


Fig. 5. IMFs obtained by EEMD

4. Wavelet shrinkage method

We consider the following model of a discrete noisy signal:

$$x = \theta + \sigma z \quad (7)$$

The vector x represents noisy signal and θ is an unknown original clean signal. z is independent identity distribution Gaussian white noise with mean zero and unit variance. For simplicity, we assume intensity of noise is one. The step of wavelet shrinkage is defined as follows:

1. Apply discrete wavelet transform to observed noisy signal.
2. Estimate noise and threshold value, thresholding the wavelet coefficients of observed signal.
3. Apply the inverse discrete wavelet transform to reconstruct the signal.

The wavelet shrinkage method relies on the basic idea that the energy of signal will often be concentrated in a few coefficients in wavelet domain while the energy of noise is spread among all coefficients in wavelet domain. Therefore, the nonlinear shrinkage function in wavelet domain will tend to keep a few larger coefficients over threshold value that represent signal, while noise coefficients down threshold value will tend to reduce to zero.

In the wavelet shrinkage, how to select the threshold function and how to select the threshold value are most crucial. Donoho introduced two kinds of thresholding functions: hard threshold function and soft threshold function.

$$\delta_{\lambda}^H(x) = \begin{cases} 0 & |x| \leq \lambda \\ x & |x| > \lambda \end{cases} \quad (8)$$

$$\delta_{\lambda}^S(x) = \begin{cases} 0 & |x| \leq \lambda \\ x - \lambda & x > \lambda \\ x + \lambda & x < -\lambda \end{cases} \quad (9)$$

Hard threshold function (8) results in larger variance and can be unstable because of discontinuous function. Soft threshold function (9) results in unnecessary bias due to shrinkage the large coefficients to zero. We build the generalized threshold function:

$$\delta_{\lambda}^m(x) = x - \frac{\lambda^m}{x^{m-1}}, \quad m=1, 2, \dots, \infty \quad (10)$$

λ is threshold value.

When m is even number:

$$\delta_{\lambda}^m(x) = x - xI(|x| \leq \lambda) - \frac{\lambda^m}{x^{m-1}}I(|x| > \lambda) \quad (11)$$

When m is odd number:

$$\delta_{\lambda}^m(x) = x - xI(|x| \leq \lambda) - \frac{\lambda^m}{x^{m-1}}I(|x| > \lambda)\text{sign}(x) \quad (12)$$

When $m=1$, it is soft threshold function; when $m=\infty$, it is hard threshold function. When $m=2$ it is Non-Negative Garrote threshold function. We show slope signal as an example, Figure.6 graphically shows generalized threshold functions for different m . It can be clearly seen that when the coefficient is small, the smaller m is, the closer the generalized function is to the soft threshold function; when the coefficient is big, the bigger m is, the closer the generalized function is to the hard threshold function. As $1 < m < \infty$, generalized threshold function achieves a compromise between hard and soft threshold function. With careful selection of m , we can achieve better denoising performance.

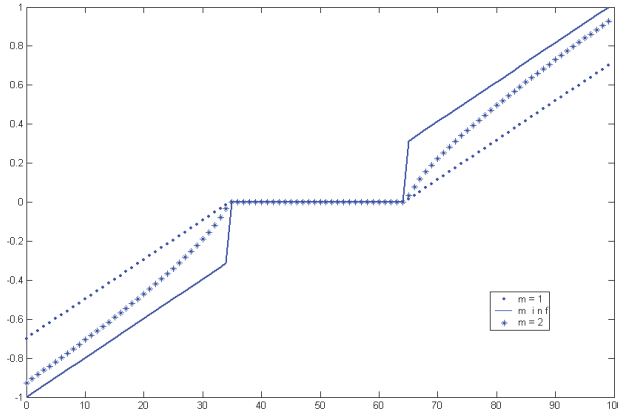


Fig. 6. Generalized threshold function

We derived the exact formula of mean, bias, variance and l_2 risk for generalized threshold function.

Let $x \sim N(\theta, 1)$

$$A_m(\theta) = \int_{-\infty}^{\infty} \frac{\phi(x - \theta) - \phi(x + \theta)}{x^m} dx \quad B_m(\theta) = \int_{-\infty}^{\infty} \frac{\phi(x - \theta) + \phi(x + \theta)}{x^m} dx$$

ϕ and Φ are density and probability function of standard Gaussian random variable respectively. Then:

Mean:

$$M^m(\lambda, \theta) = M^H(\lambda, \theta) - \lambda^m A_{m-1}(\theta) \quad (13)$$

Bias:

$$SB^m(\lambda, \theta) = (M^m(\lambda, \theta) - \theta)^2 \quad (14)$$

Variance:

$$V^m(\lambda, \theta) = V^H(\lambda, \theta) - 2\lambda^m B_{m-2}(\theta) - \lambda^{2m} A_{m-1}^2(\theta) + \lambda^{2m} B_{2m-2}(\theta) + 2\lambda^m M^H(\lambda, \theta) A_{m-1}(\theta) \quad (15)$$

l_2 Risk:

$$\rho_{\lambda}^m(\theta) = E(\delta_{\lambda}^m(x) - \theta)^2 = \rho_{\lambda}^H(\theta) - 2\lambda^m B_{m-2}(\theta) + \lambda^{2m} B_{2m-2}(\theta) + 2\theta \lambda^m A_{m-1}(\theta) \quad (16)$$

Where

$$\begin{aligned} M^H(\lambda, \theta) &= \theta + \theta[1 - \Phi(\lambda - \theta) - \Phi(\lambda + \theta)] + \phi(\lambda - \theta) - \phi(\lambda + \theta) \\ V^H(\lambda, \theta) &= (\theta^2 + 1)(2 - \Phi(\lambda - \theta) - \Phi(\lambda + \theta)) + (\lambda + \theta)\phi(\lambda - \theta) + (\lambda - \theta)\phi(\lambda + \theta) - M^H(\lambda, \theta)^2 \\ \rho_{\lambda}^H(\theta) &= 1 + (\theta^2 - 1)(\Phi(\lambda - \theta) - \Phi(-\lambda - \theta)) + (\lambda + \theta)\phi(\lambda + \theta) + (\lambda - \theta)\phi(\lambda - \theta) \end{aligned}$$

$M^m(\lambda, \theta)$, $SB^m(\lambda, \theta)$, $V^m(\lambda, \theta)$, $\rho_{\lambda}^m(\theta)$ are the mean, bias, variance and risk of generalized threshold function. When m is 1, 2, ∞ , they are the mean, bias, variance and risk of the risk of soft, Non-Negative Garrote, hard threshold functions, respectively.

Soft threshold function provides smoother results in comparison with the hard threshold function; however, the hard threshold function provides better edge preservation in comparison with the soft threshold function. The hard threshold function is discontinuous and this leads to the oscillation of denoised signal. Soft threshold function tends to have bigger bias because of shrinkage, whereas hard threshold function tends to have bigger variance because of discontinuity. Non-Negative Garrote threshold function is the trade-off between the hard and soft threshold function. Firstly it is continuous; secondly the shrinkage amplitude is smaller than the soft threshold function.

5. Adaptive threshold values based on EEMD

Threshold value is a parameter that controls the bias and variance tradeoff of the risk. If it is too small, the estimators tend to overfit the data, then result is close to the input and the estimate bias is reduced but the variance is increased. If the threshold value is too large, a lot

of wavelet coefficients are set zero and the estimators tend to underfit the data; the estimate variance is reduced but the bias is increased. The optimal threshold value is the best compromise between variance and bias and it should minimize the risk of the results as compared with noise-free data.

Several methods have been proposed for the determinations of threshold values. The universal threshold, proposed by Donoho and Johnstone, uses the fixed form threshold equal to the square root of two times the logarithm of the length of the signal. LDT, the level dependent threshold, proposed by I.M. Johnstone, and B.W. Silverman, uses a different threshold for each of the levels based on a single formula. Stein Unbiased Risk Estimate (SURE) is an adaptive threshold selection rule. It is data driven and the threshold value minimizes an estimate of the risk. Other threshold values include minimaxi threshold etc. In this paper, an adaptive threshold method is proposed based on EEMD. The threshold values directly relate to the energy of noise on each IMFs. Next, the derivation of adaptive threshold values is initiated by the characteristic of Fractional Gaussian noise (fGn).

fGn is a generalization of white noise. The statistical properties of fGn are controlled by a single parameter H , and the autocorrelation sequence

$$r_H(k) = E(X_{H,i}X_{H,i+k}) \quad (17)$$

This can also be defined as:

$$r_H[k] = \frac{\sigma^2}{2} (|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}) \quad (18)$$

σ^2 is the variance of fGn. The value of H is in the range of 0 to 1. The Fourier transform of (18) gives the power spectral density of fGn:

$$S_H(f) = C\sigma^2 \left| e^{i2\pi f} - 1 \right|^2 \sum_{k=-\infty}^{\infty} \frac{1}{|f+k|^{2H+1}} \quad (19)$$

In the decomposing of a given fGn, EMD is worked as a dyadic filter. Restricting to the band-pass IMFs, self-similarity would mean that

$$S_{k',H}(f) = \rho_H^{\alpha(k'-k)} S_{k,H}(\rho_H^{k'-k} f) \quad k' > k \geq 2 \quad (20)$$

Given the self-similar relation (6) for PSDs for band-pass IMFs we can deduce how the variance should evolve as a function of k :

$$V_H[k'] = \rho_H^{(\alpha-1)(k'-k)} V_H[k] \quad k' > k \geq 2 \quad (21)$$

$V_H[k]$ is the variance of the IMF index.

According to lots of simulation:

$$\log_2(\log_2(T_H[k] / W_H[k])) = a_H k + b_H \quad (22)$$

$W_H[k]$ denotes the H -dependent variation of the IMF energy. In practice, $W_H[1]$ can be estimated from (23), which also gives the model energy of the noisy signal

$$\hat{W}_H[1] = \sum_{n=1}^N c_1^2(n) \quad (23)$$

c_1 represents the first IMF coefficients.

According to (21)

$$V_H[k] = C_H \rho_H^{2(H-1)k} \quad k \geq 2 \quad (24)$$

$C_H = \hat{W}_H[1] / \beta_H$, the parameter H and β_H are given in table 1. Through (25) we can obtain the model energy of noise only signal.

H	0.2	0.5	0.8
β_H	0.487	0.719	1.025

Table. 1. H and β_H

According to the relationship between energy and variance $W_H[k]$ are given by

$$\hat{W}_H[k] = C_H \rho_H^{-2(1-H)k} \quad k \geq 2 \quad (25)$$

For white noise,

$$H = \frac{1}{2}, \quad \beta_H = 0.719 \quad (26)$$

$$\rho_H = [2.01 + 0.2(H - \frac{1}{2}) + 0.12(H - \frac{1}{2})^2] = 2.01 \quad (27)$$

The energies of each IMFs can be defined as:

$$V_k = \frac{\sigma_n^2}{0.719} 2.01^{-k}, k = 2, 3, 4, \dots \quad (28)$$

σ_n^2 is the noise energy that can be achieved by the first IMF variance, which can be achieved by (23).

The adaptive threshold value of each IMF can be identified as:

$$T_k = \sqrt{\frac{V_k}{N}} 2 \ln N, k = 2, 3, 4, \dots \quad (29)$$

N is the length of signal.

Given these results, a possible strategy for de-noising a signal (with a known H) is generalized as follows:

1. Decompose the noisy signal into IMFs with EEMD.
2. Assuming that the first IMF captures most of the noise, estimate the noise level in the noisy signal by computing V_k from (28).
3. Discarding the first IMF, for other IMFs, calculate the adaptive threshold value T_k from (29); shrink the coefficients using the Non-Negative Garrote threshold function.
4. Reconstruct the signal by the shrunk IMFs, obtain the denoised signal.

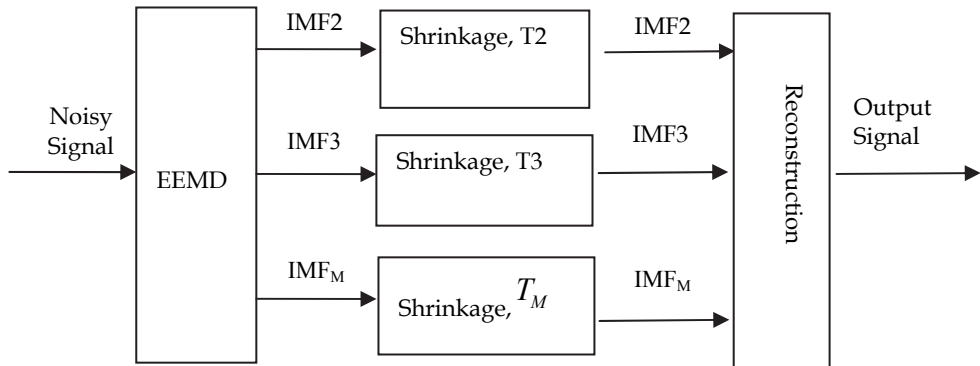


Fig. 7. The block diagram of the denoising algorithm

6. Results and discussions

To evaluate the performance of the algorithm, Test signal and Clinic noisy ECG signals are processed.

6.1 Test signal

We choose time shifted sine signal which shapes similarly to ECG to test above method; Gaussian White Noise is added as noise, which is zero mean and standard deviation change with the SNR. $SNR = 10\log(\text{var}(\text{signal}) / \text{var}(\text{noise}))$, var means standard deviation. The SNR of noisy test signals are 5. Figure8 shows the original clean signal; figure9 shows the noisy signal; figure10 shows the denoised signal by the above algorithm, the SNR of which achieve 14. Furthermore, the original characteristic waveform is preserved.

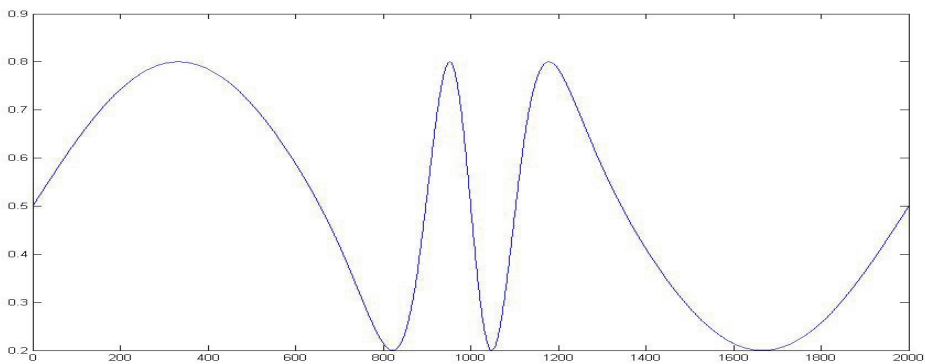


Fig. 8. The clean time shifted sine signal

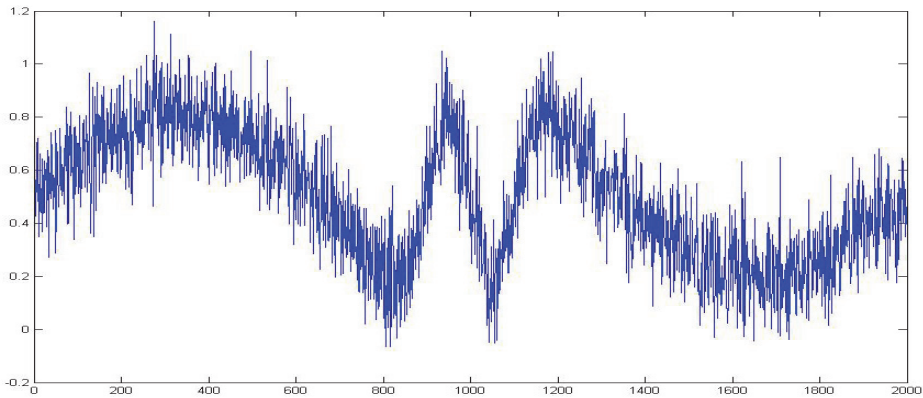


Fig. 9. The noisy time shifted sine signal

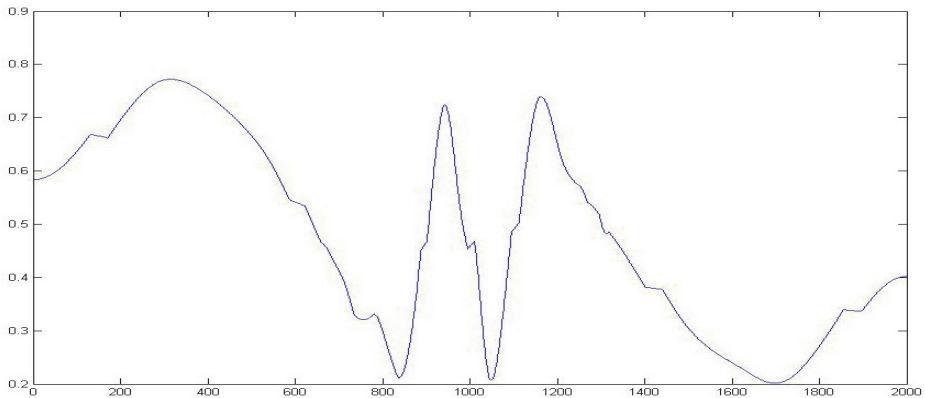


Fig. 10. The denoised time shifted sine signal

6.2 Clinical noisy ECG signal

The ECG signal as Figure.11 illustrates comes from clinical patient. Signal is sampled at 360 Hz; signal length is 1500; the ECG signal is corrupted by noise. Figure12 shows its phase space diagram, which is a plot of the time derivative of the ECG signal against the ECG signal itself. The derivative can accentuate the noisy and high frequency content in signal, so it can better show dramatic improvement after denoising. The noisy ECG signal is processed using the method mentioned above. For the generalized threshold function, m is selected as 2, which is Non-Negative Garrote threshold function. The noisy ECG signal is decomposed into a series of IMFs by EEMD. The first seven IMFs are shown in figure13; the latter seven IMFs are shown in figure 14. The First IMF is discarded owing to predominant noise. Obtain the adaptive threshold value of each IMFs by formula (29). The values are 0.0422, 0.0297, 0.0210, 0.0148, 0.0104, 0.0074, 0.0052, 0.0037, 0.0026, 0.0018, 0.0013, 0.0009, 0.0006. Then shrink the coefficients of each IMFs by the adaptive threshold values and Non-Negative

Garrote threshold function. The first shrunk six IMFs are shown in figure15; the latter shrunk seven IMFs are shown in figure16. Reconstruct the signal by the shrunk 13 IMFs and obtain the denoised signal. The filtered ECG signal is illustrated as figure17. The phase space diagram of filtered ECG signal is shown as figure 18. From visual inspection, the ECG signal is much cleaner after being denoised; the original characteristic waveform, especially the sharp Q, R, and S peaks is preserved, without distorting the P and T waves. The results indicate that the method we have proposed significantly reduces noise and well preserves the characteristics of ECG signal.

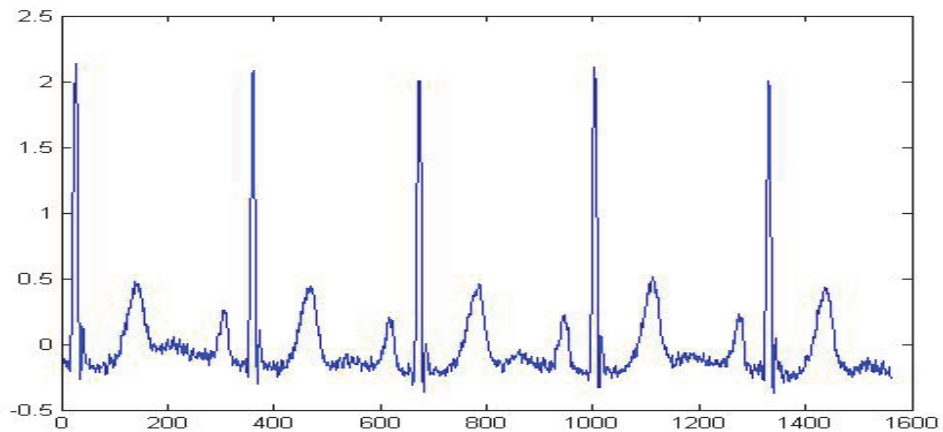


Fig. 11. Noisy ECG signal

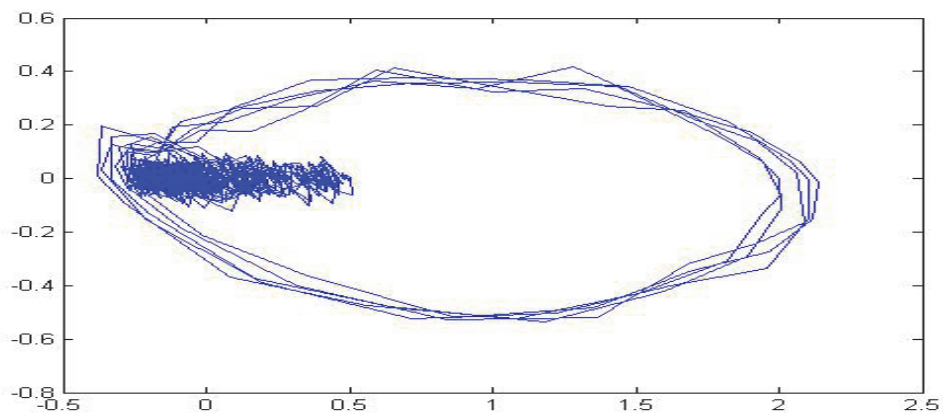


Fig. 12. Phase space diagram of noisy ECG signal

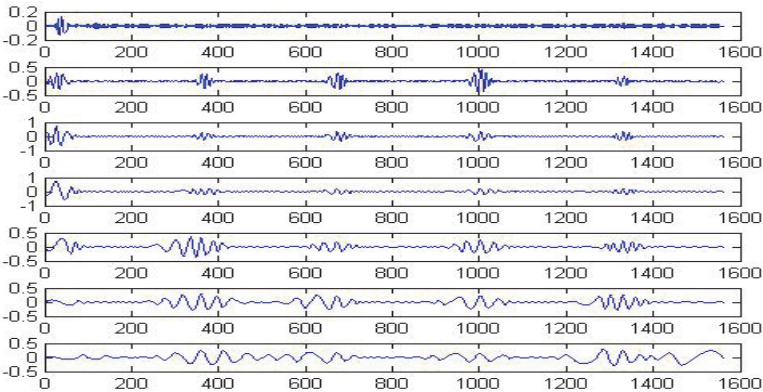


Fig. 13. The first seven IMFs

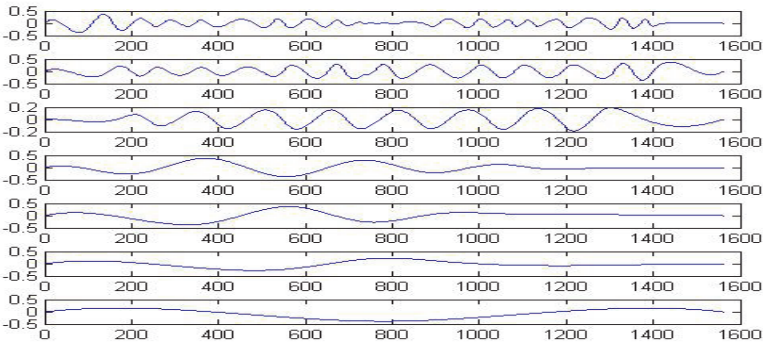


Fig. 14. The latter seven IMFs

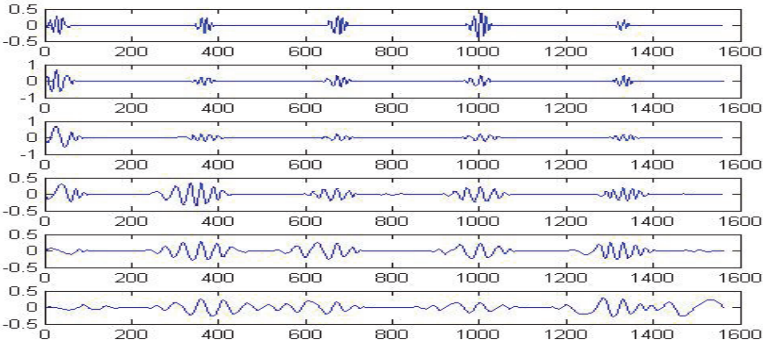


Fig. 15. The first shrunken six IMFs

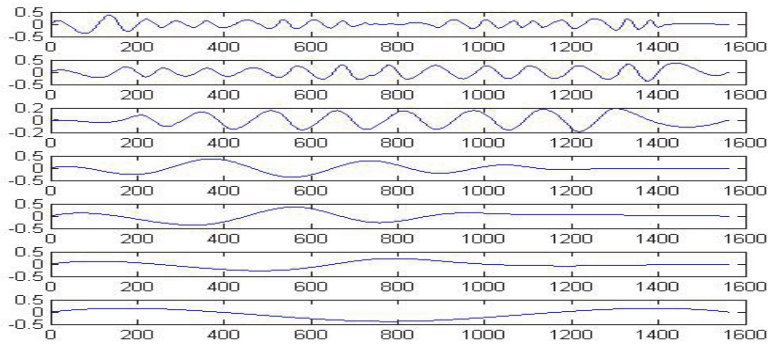


Fig. 16. The latter shrunk seven IMFs

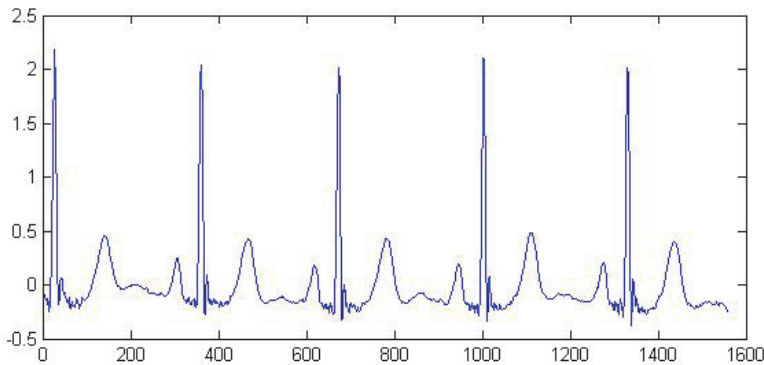


Fig. 17. The denoised ECG signal

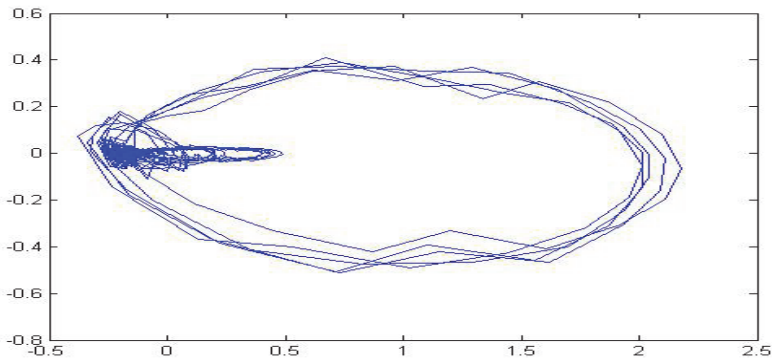


Fig. 18. The Phase space diagram of denoised ECG signal

Another ECG signal as Fig.19 illustrates comes from clinical patient. Signal is sampled at 360 Hz. Length is 1500; the ECG signal is corrupted by noise. Figure20 shows its phase space diagram. The filtered ECG signal is illustrated as figure 21 using the above algorithm. The

phase space diagram of filtered ECG signal is shown as figure 22. It is obvious that the noise is reduced.

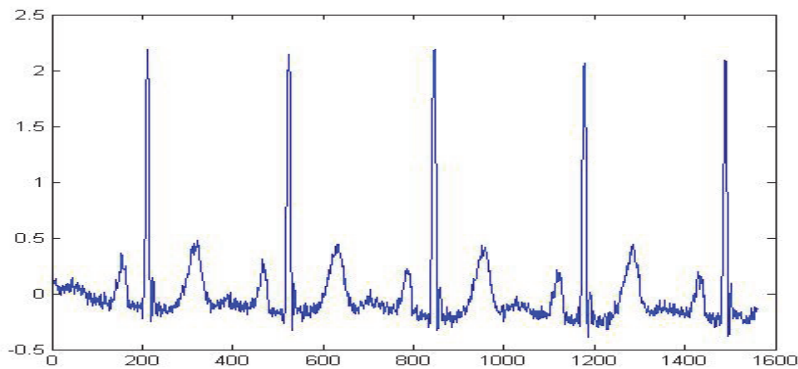


Fig. 19. The noisy ECG signal

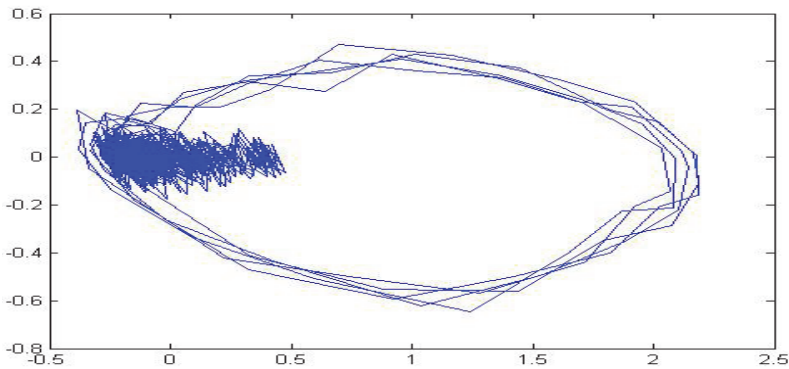


Fig. 20. Phase space diagram of noisy ECG signal

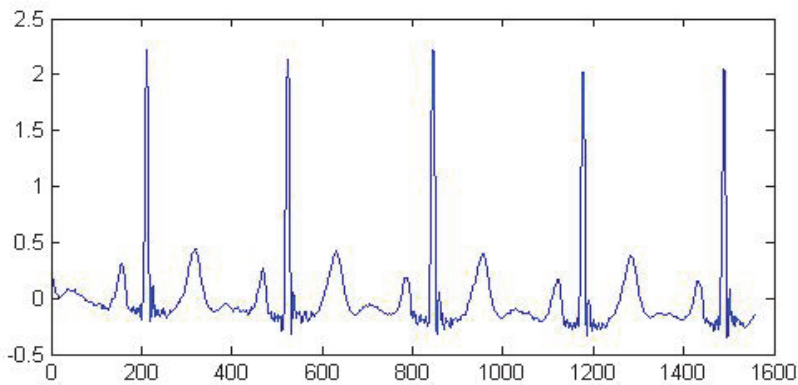


Fig. 21. The denoised ECG signal

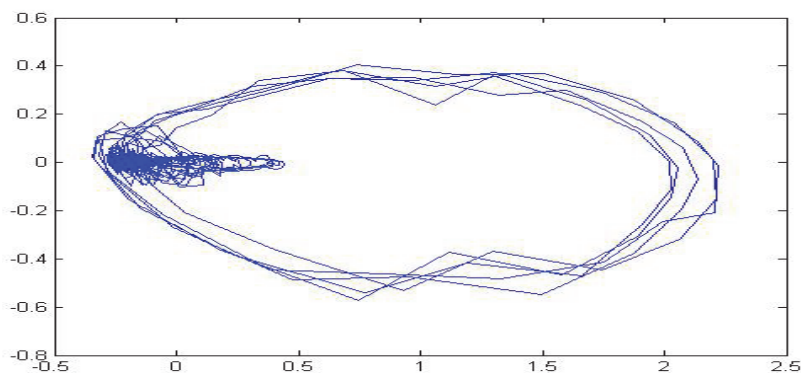


Fig. 22. Phase space diagram of the denoised ECG signal

7. Conclusion

In this paper, the adaptive noise removal scheme based EEMD is studied for ECG signal. EEMD reduces the mode mixing existing in EMD. Better filtering performance for EEMD is achieved. The adaptive threshold values guarantee the better estimation of noise. We have demonstrated that the algorithm is useful for removing noise from clinic ECG signal. It not only decreases the signal noise, but also the ECG waveform is better conserved. Application of EEMD with adaptive threshold value also has potential for other biomedical signals or other fields.

8. Acknowledgment

This work was supported in part by the public welfare Program of Zhejiang Province Science and technology department under the Grant: 2010C31022.

9. References

- Brij N. Singh; Arvind.K(2006). Optimal selection of wavelet basis function applied to ECG signal denoising, *Digital Signal Processing*, Vol.16, No.3, pp.275-287.
- Agante , P.M; Marques J.P(1999). ECG noise filtering using wavelets with soft-thresholding methods, *Computers in Cardiology*, 1999, pp.535-538.
- Enrique.CB; Hartmann.E(2003). ECG Front-End Design is Simplified with MicroConverter, *Analog Dialogue* ,Vol. 37, No. 4.
- Donoho DL; Johnstone IM(1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, Vol.81, No.3, pp. 425-455.
- Donoho DL(1995). De-noising by soft-thresholding, *IEEE Trans Inform Theory*, Vol.14, No.3, pp. 612-627.
- Gao HY (1998). Wavelet shrinkage denoising using the non-negative garrote, *J. Comput. Graph. Statist*, Vol 7, No.4, pp.469-488.
- Prasad V.V.K.D.V; Siddaiah P; Rao BP(2008). Denoising of Biological Signals using a New Wavelet Shrinkage Method, *IEEE region 10 and the Third international Conference on Industrial And Information Systems*, pp.1-5.

- Kumari, R.S.S.;Thilaganimala, A ; Sadasivam, V (2008). ECG signal interferences removal using wavelet based CSTD technique, *Proceedings - International Conference on Computational Intelligence and Multimedia Applications*, pp. 530-534.
- Suyi Li ;Yanju Ji ;Guangda Liu(2009). Optimal wavelet basis selection of wavelet shrinkage for ECG de-noising, *International Conference on Management and Service Science*, pp.1-4.
- Poornachandra.S(2008). Wavelet-based denoising using subband dependent threshold for ECG signals ,*Digital Signal Processing*, Vol.18, No.1,pp.49-55.
- SadAbadi H; Ghasemi M;Ghaffari A (2007). A mathematical algorithm for ECG signal denoising using window analysis, *Biomedical papers- Palacky Univesrisy Inolomouc*, Vol.151, No.1, pp.73-78.
- Kang-Ming Chang (2010). Arrhythmia ECG Noise Reduction by Ensemble Empirical Mode Decomposition, *sensors*, pp.6063-6080.
- Huang.N.E, et al(1998). The empirical mode composition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceeding of R.Soc.Lond.A*, Vol 454,pp. 903-995.
- Liang HL et al(2000). Artifact reduction in electrogastrogram based on the empirical model decomposition method, *Medical and Biological Engineering and Computing*,Vol.38, No.1, pp.35-41.
- Flandrin. P;Rilling.G;Goncalves.P(2003). Empirical mode decomposition as a filter bank, *IEEE Signal Processing Letter*, Vol.11, No.2, pp. 112-114.
- Flandrin. P;Rilling.G;Goncalves.P(2004). Empirical mode decomposition as a filter bank, *IEEE Signal Processing Letters*, Vol11, No. 2, pp. 112-114.
- Flandrin. P;Rilling.G;Goncalves.P(2004). Detrending and denoising with empirical mode decomposition, *European Signal Processing Conference*, pp. 1581-1584.
- Flandrin. P;Rilling.G;Goncalves.P(2005). Empirical mode decomposition, fractional Gaussian noise and Hurst exponent estimation, *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol 4,pp. 489-492..
- Wu .Z.H; Huang.N.E(2009). Ensemble Empirical Mode Decomposition: A Noise Assisted Data Analysis Method, *Advances in Adaptive Data Analysis*, pp. 1-41.
- Kang-Ming Chang (2010). Ensemble empirical mode decomposition based ECG noise filtering method, *International Conference on Machine Learning and Cybernetics*, pp.210-213.
- Binwei Weng; Blanco-Velasco, M.; Earner, K.E (2006). ECG Denoising Based on the Empirical Mode Decomposition, *28th IEEE Annual International Conference of Engineering in Medicine and Biology Society*, pp.1 - 4.
- Lihuang She; Zhongqiang Xu; Shi Zhang; Yuning Song (2010). De-noisng of ECG based on EMD improved-thresholding and mathematical morphology operation, *3rd International Conference on Biomedical Engineering and Informatics*, pp. 838 - 842.

Application of Adaptive Noise Cancellation in Transabdominal Fetal Heart Rate Detection Using Photoplethysmography

Kok Beng Gan¹, Edmond Zahedi² and Mohd. Alauddin Mohd. Ali¹

¹*Institute of Space Science, National University of Malaysia*

²*School of Electrical Engineering, Sharif University of Technology*

¹*Malaysia*

²*Iran*

1. Introduction

Monitoring the fetal heart rate (FHR) throughout pregnancy empowers the clinician to diagnose fetal well being, characterize fetal development and detect abnormality (Freeman et al. 2003). A non-invasive and low cost system would enable monitoring of normal pregnancies and promote large population studies of fetal physiological development (Freeman et al. 2003). FHR monitoring is an ongoing observation of human fetal physiology. The expected outcome of this early detection is a reduced risk of fetal morbidity and mortality (Philip et al. 2002).

Currently, Doppler ultrasound has been extensively used for FHR detection and obstetric purposes (Hershkovitz et al. 2002), where the standard pre-delivery test of fetal health is the fetal non-stress test (NST). These tests are routinely performed at the hospital where continuous-wave instruments are more popular than pulsed ones. The use of Doppler ultrasound in the first trimester is generally not recommended as a routine (Hershkovitz et al. 2002) as it may increase the occurrence of intrauterine growth restriction. Besides that, the FHR measurements using Doppler ultrasound are not always reliable (Karlsson et al. 2000) due to the complexity of the Doppler signal and the effects of fetal and maternal breathing. An alternative to ultrasound is using the fetal electrocardiogram (FECG). In direct (invasive) fetal electrocardiogram (FECG), the FHR could be obtained by attaching scalp electrode to the fetal scalp after the rupture of the membrane (Khandpur 2004). During invasive FECG recording the uterus may be perforated leading to its infection, besides possible scalp injuries to the fetus (Khandpur 2004). The other approach is non-invasive FECG but FECG signals have a low (signal to noise ratio) SNR due to the interference from noise, maternal electrocardiogram (MECG) and electromyogram (EMG). The application of non-invasive FECG requires multiple leads and advanced digital signal processing techniques (Najafabadi et al. 2006). It is worth mentioning that commercial devices operating on non-invasive FECG are not available at this moment.

Optical techniques has received a considerable attention in biomedical diagnostic and monitoring of biological tissues such as brain imaging, breast imaging and for fetal heart rate detection and oxygen saturation measurement due to its theoretical advantages in

comparison with other modalities. Continuous wave near infrared (NIR) spectroscopy has been applied to trans-abdominal fetal pulse oximetry (Ramanujam et al. 2000; Chance 2005; Zourabian et al. 2000; Nioka et al. 2005; Vintzileos et al. 2005). The system consists of NIR sources (halogen lamps) and a photomultiplier as detection unit (Ramanujam et al. 2000; Chance 2005). The generated heat was justified by using cooling fans for the halogen lamps. Recently, trans-abdominal oxygen saturation (S_pO_2) in animal (Nioka et al. 2005) and human fetuses were successfully obtained in the laboratory (Vintzileos et al. 2005). However, the proposed techniques require high power (a total of 80 W optical power) and a relatively expensive detection unit (photo-multiplier).

In this project, we propose to design and develop a low-power optical FHR monitor. The signal of interest is the photoplethysmogram (PPG), which is generated when a beam of light is modulated by blood pulsations. PPG is a noninvasive technique for detecting blood volume changes in living tissue by optical means consisting of a light emitting diode (LED) and a photo-detector. One of the potential applications of the PPG technology is non-invasive fetal heart rate detection through the maternal abdomen. In this application, the light intensity is modulated by the mother as well as fetal blood circulation, producing a combined signal which needs to be separated via digital signal processing (DSP) techniques. The design of a fixed filter would not be adequate as the frequency spectrum of the noise (maternal PPG) overlaps with the desired signal (fetal PPG). The adaptive filter will automatically adjust its coefficients therefore achieve the high degree of noise rejection.

Such an approach - based on adaptive noise cancellation (ANC) - has been evaluated for extraction of the fetal heart-rate using PPG signals from the maternal abdomen. A simple optical model has been proposed in which the maternal and fetal blood pulsations result in emulated signals where the lower SNR limit (fetal to maternal) is -25 dB (Zahedi & Beng, 2008). It is shown that the RLS algorithm is capable to extract the peaks of the fetal PPG from these signals, corresponding to typical values of maternal and fetal tissues.

Subsequently, an optical fetal heart rate detection (OFHR) system has been designed and developed using low-cost, low-power IR light (890 nm with optical power < 68 mW) and a commercially available silicon photo-detector (Gan et al. 2009). Previous literature (Ramanujam et al. 2000; Chance 2005; Zourabian et al. 2000; Nioka et al. 2005; Vintzileos et al. 2005; Choe et al. 2003) shows that the Source-Detector separations depends on the type of sources and the photo-detectors implemented in their studies. Since in our work the developed instrument utilizes low optical power, the source-to-detector separation plays an important role as it affects the detectivity of the photo-detector. This chapter discusses the selection of S-D separation for the OFHR system based on the ANC limit and photo-detector's noise. The implementation of the ANC algorithm in OFHR system is also discussed and the clinical trial results are also reported.

2. Materials and methods

2.1 Adaptive noise cancellation

Conventional digital signal processing techniques do exist to extract a desired biomedical signal from a mixed signal which is usually contaminated by unwanted noises. Adaptive filters are used for non-stationary signals where a sample-by-sample adaptation process is required (Vaseghi, 2000; Widrow et al., 1975). Applications of adaptive filtering include multi-channel noise reduction, radar or sonar signal processing, channel equalization for cellular mobile phones, echo cancellation and low delay speech coding. This section

discussed the concept of the adaptive filtering, adaptive algorithm and the Recursive Least Square (RLS) algorithm.

2.1.1 Concept of adaptive filtering

Adaptive filters consist of two distinct parts: a digital filter and the corresponding adaptive algorithm, used to adjust the coefficients of the filter (Figure 1). In these algorithms, the error signal $e(n)$ defined as the difference between the output of the filter ($y(n)$) and a primary input signal ($d(n)$), is minimized according to a least squares error criterion (Ifeachor & Jervis, 2002).

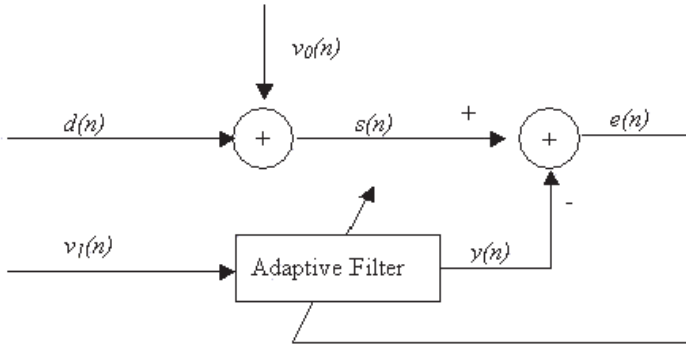


Fig. 1. ANC system

The desired signal $d(n)$ (Figure 1) is contaminated by an uncorrelated noise signal $v_0(n)$, where n is the running time index. The result $d(n) + v_0(n)$ is the primary measurement signal $s(n)$. The reference input, $v_1(n)$ is only correlated with $v_0(n)$ and fed to an adaptive FIR filter. The output of the FIR adaptive filter $y(n)$ is subtracted from the primary input $s(n)$ to produce the error signal $e(n)$:

$$e(n) = d(n) + v_0(n) - y(n) \quad (1)$$

The adaptive filter uses $e(n)$ to adjust its own impulse response to produce an output $y(n)$ as close a replica as possible to $v_0(n)$. Squaring and applying the expectation operator to both sides of Equation 1:

$$E\{e^2(n)\} = E\{d^2(n)\} + E\{(v_0(n) - y(n))^2\} - 2E\{d(n)(v_0(n) - y(n))\} \quad (2)$$

$d(n)$ being uncorrelated with $v_0(n)$ and $v_1(n)$, $E\{d(n)(v_0(n) - y(n))\} = 0$. Therefore Equation 2 can be simplified:

$$E\{e^2(n)\} = E\{d^2(n)\} + E\{(v_0(n) - y(n))^2\} \quad (3)$$

An iterative procedure minimizes $E\{e^2(n)\}$, which will occur when $y(n) = v_0(n)$ (ideal situation) producing $e(n) = d(n)$.

2.1.2 Adaptive algorithm

In most adaptive systems, the digital filter in Figure 2 is realized using a transversal or finite impulse response (FIR) structure. The FIR structure is the most widely used because of its simplicity and stability.

A m th-order adaptive transversal filter is a linear time varying discrete-time system that can be represented by:

$$y(n) = \sum_{i=0}^{m-1} w_i(n) v_1(n-i) \quad (4)$$

where $w_i(n)$ is the adjustable weight and $v_1(n)$ and $y(n)$ are the input and output of the filter. The filter output is a time varying linear combination of the past input (Figure 2).

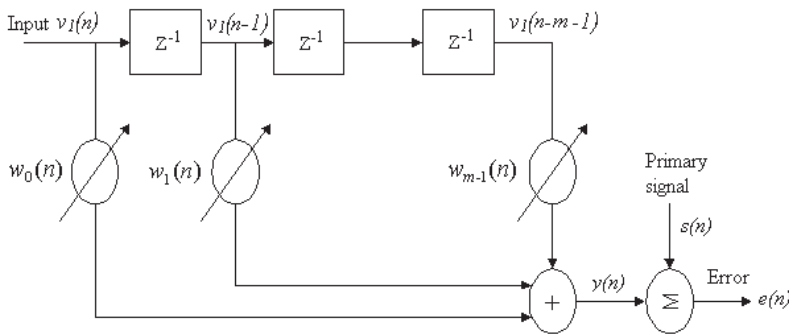


Fig. 2. Illustration of the configuration of an adaptive filter

Adaptive algorithm are used to adjust the coefficient of the digital filter (Figure 2) such that the error signal $e(n)$, is minimized according to the mean square error and least squares error criterion (Ifeachor & Jervis, 2002). Common adaptation algorithms are least mean square (LMS) and the RLS. The RLS algorithm minimizes the sum of the square of the error whereas the LMS algorithm minimizes the mean square error. In terms of the computational and storage requirements, the LMS algorithm is the most efficient and does not suffer from the numerical instability problem (Ifeachor & Jervis, 2002). However, the recursive least square (RLS) algorithm has superior convergence properties (Ifeachor & Jervis, 2002). It is suitable for offline processing where computational requirement is not an issue.

2.1.3 Linear least-square error estimation

The principle of least-squares (LS) was introduced by the German mathematician Carl Friedrich Gauss, who used it to determine the orbit of the asteroid ceres in 1821 by formulating the estimation problem as an optimization problem (Manolakis et al., 2005).

The least-square approach provides a mechanism for designing fixed filters when the properties of the signal source are known. More importantly, it provides a vehicle for adaptive filter design that can operate in an environment of changing signal properties. The source signal is modeled as the output of a linear discrete-time system with parameters which are either known for the fixed algorithm or unknown in the adaptive case. Noise added to the observations completes the signal description. The least-square algorithm is then required to

do the “best” filtering of the signal, employing as much of the priori signal and noise models as is known. If these priori properties are unknown, then the LS algorithm is required to identify the changed conditions and to adapt its parameters to the new signal environment. The basic idea of the LS method is shown in Figure 3. An output signal, $s(n)$ measured at the discrete time, n in response to a set of input signal, $v_1(n)$. The input and output signals are related by the simple regression model.

$$s(n) = \sum_{i=0}^{m-1} w_i(n)v_1(n) + e(n) \quad (5)$$

where $e(n)$ is the measurement errors and $w_i(n)$ is the adjustable weight with m th order.

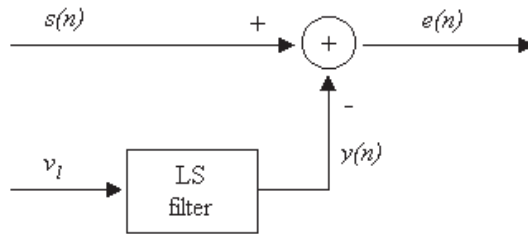


Fig. 3. An illustration of the basic idea of the LS method

The estimation error is defined as

$$\begin{aligned} e(n) &= s(n) - \sum_{i=0}^{m-1} w_i(n)v_1(n) \\ &= s(n) - \mathbf{w}^T \mathbf{v}_1 \end{aligned} \quad (6)$$

where $\mathbf{v}_1 = [v_1(n), v_1(n-1), \dots, v_1(n-m+1)]^T$ and $\mathbf{w} = [w_0(n), w_1(n), \dots, w_{m-1}(n)]^T$. The filter weight, $w_i(n)$ are determined by minimizing the sum of the squared errors

$$E \Delta \sum_{n=0}^{N-1} |e(n)|^2 \quad (7)$$

that is, the energy of the signal.

To explore the relation between the filter coefficient, w , and the error signal, $e(n)$, Equation 6 can be written in matrix form for N samples measurement of the signals $[s(0), s(1), \dots, s(N-1)]$ and signals $[v_1(0), v_1(1), \dots, v_1(N-1)]$ as

$$\begin{pmatrix} e(0) \\ e(1) \\ e(2) \\ \vdots \\ e(N-1) \end{pmatrix} = \begin{pmatrix} s(0) \\ s(1) \\ s(2) \\ \vdots \\ s(N-1) \end{pmatrix} - \begin{pmatrix} v_{10}(0) & v_{11}(0) & v_{12}(0) & \cdots & v_{1m-1}(0) \\ v_{10}(1) & v_{11}(1) & v_{12}(1) & \cdots & v_{1m-1}(1) \\ v_{10}(2) & v_{11}(2) & v_{12}(2) & \cdots & v_{1m-1}(2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{10}(N-1) & v_{11}(N-1) & v_{12}(N-1) & \cdots & v_{1m-1}(N-1) \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{m-1} \end{pmatrix} \quad (8)$$

or more compactly as

$$\mathbf{e} = \mathbf{s} - \mathbf{V}\mathbf{w} \quad (9)$$

where

$$\begin{aligned} \mathbf{e} &\triangleq [e(0), e(1), \dots, e(N-1)]^T && \text{error data vector } (N \times 1) \\ \mathbf{s} &\triangleq [s(0), s(1), \dots, s(N-1)]^T && \text{primary data vector } (N \times 1) \\ \mathbf{V} &\triangleq [\mathbf{v}_1(0), \mathbf{v}_1(1), \dots, \mathbf{v}_1(N-1)]^T && \text{input data matrix } (N \times m) \\ \mathbf{w} &\triangleq [w_0, w_1, \dots, w_{m-1}]^T && \text{weight vector } (m \times 1) \end{aligned} \quad (10)$$

where $\mathbf{v}_1(n) \triangleq [v_{10}(n), v_{11}(n), \dots, v_{1m-1}(n)]$. The energy of the error vector, that is the sum of squared elements of the squared error vector, is given by the inner vector product as:

$$\begin{aligned} \mathbf{e}^T \mathbf{e} &= (\mathbf{s} - \mathbf{V}\mathbf{w})^T (\mathbf{s} - \mathbf{V}\mathbf{w}) \\ &= \mathbf{s}^T \mathbf{s} - \mathbf{s}^T \mathbf{V}\mathbf{w} - \mathbf{V}^T \mathbf{w}^T \mathbf{s} + \mathbf{V}^T \mathbf{w}^T \mathbf{V}\mathbf{w} \end{aligned} \quad (11)$$

The gradient of the squared error function with respect to the filter coefficients is obtained by differentiating Equation 11 with respect to $\mathbf{w}$ as:

$$\frac{\partial \mathbf{e}^T \mathbf{e}}{\partial \mathbf{w}} = -2\mathbf{s}^T \mathbf{V} + 2\mathbf{w}^T \mathbf{V}^T \mathbf{V} \quad (12)$$

The filter coefficients are obtained by setting the gradient of the squared error function of Equation 12 to zero and yield:

$$(\mathbf{V}^T \mathbf{V})\mathbf{w} = \mathbf{V}^T \mathbf{s} \quad (13)$$

or

$$\mathbf{w} = (\mathbf{V}^T \mathbf{V})^{-1} \mathbf{V}^T \mathbf{s} \quad (14)$$

Note that the matrix $\mathbf{V}^T \mathbf{V}$ is a time-averaged estimate of the autocorrelation matrix of the input signal, $\mathbf{R}_{yy}$ and the vector $\mathbf{V}^T \mathbf{s}$ is a time-averaged estimate of the cross-correlation vector of the input and the primary signals, $\mathbf{r}_{yx}$

2.1.4 Recursive least square algorithm

The RLS algorithm is based on the least-square method (Ifeachor & Jervis, 2002; Haykin, 2002). In recursive implementations of the method of least squares, the computation is started with known initial conditions and use the information contained in new data samples to update the old estimates. The RLS adaptive filters are designed so that the updating of the coefficients is always achieved the minimization of the sum of the squared errors. The RLS adaptive algorithm for updating the coefficients of the FIR filter is superior to the LMS algorithm in convergence properties, eigen value sensitivity, and excess MSE. The price paid for this improvement is additional computational complexity.

The computation of $\mathbf{w}$ in Equation 14 requires time-consuming computation of the inverse matrix. With the RLS algorithm the estimate of $\mathbf{w}$ can be updated for each new set of data

acquired without repeatedly solving the time-consuming matrix inversion directly. A suitable RLS algorithm can be obtained by exponential weighting the data to remove gradually the effect of old data on w and to allow the tracking of slowly varying signal characteristic.

The derivation of the RLS algorithm can be found in the report (Gan, 2009) and the RLS algorithm can be summarized as follows:

Input signals: $v_I(n)$ and $d(n)$

Initial values:

$$\Phi_{yy}(n) = \delta^{-1} I$$

$$w(0) = w_I$$

For $n = 1, 2, \dots$, compute

1. Filter gain vector update :

$$k(n) = \frac{\lambda_f^{-1} \Phi_{yy}(n-1) v_I(n)}{1 + \lambda_f^{-1} v_I^T(n) \Phi_{yy}(n-1) v_I(n)} \quad (15)$$

2. Error signal equation:

$$e(n) = d(n) - w^T(n-1) v_I(n) \quad (16)$$

3. Filter coefficients adaptation:

$$w(n) = w(n-1) - k(n) e(n) \quad (17)$$

4. Inverse correlation matrix update:

$$\Phi_{yy}(n) = \lambda_f^{-1} \Phi_{yy}(n-1) - \lambda_f^{-1} k(n) v_I^T(n) \Phi_{yy}(n-1) \quad (18)$$

2.2 Photoplethysmography

Photoplethysmograph is an optoelectronic method for measuring and recording changes in the volume of body parts such as finger and ear lobes caused by the changes in volume of the arterial oxygenated blood, associated with cardiac contraction (Bronzino 2000). A sample of few normal periodic PPG pulse waves is shown in Figure 4, where the steep rise and dicrotic notch on the falling slope are clearly visible. When light travels through a biological tissue (earlobe or finger), it is absorbed by different absorbing substances. Primary absorbers are the skin pigmentation, bones and the arterial and venous blood. The characteristics of the PPG pulse are influenced by arterial ageing and arterial disease (Allen & Murray 2000).

The emitted light either red or infrared light emitting diode is detected by a photo-detector. The time varying signals of the detected signal is called PPG. The PPG signal contains AC and DC components: the AC component is mainly due to the arterial blood pulsation and the DC component comes from the non-pulsating arterial blood, venous blood and other tissues.

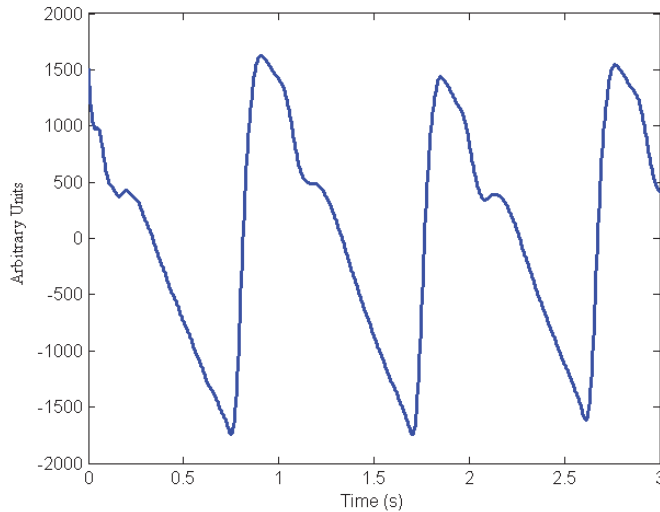


Fig. 4. Typical PPG pulse wave signal acquired in our laboratory

The probes can be of two types, transmission or reflection. A transmission probe measures the amount of light that passes through the tissue as in a finger clip probe. The photodiode is located on the opposite side of the LED and the tissue is located between them. A reflectance probe measures the amount of light reflected to the probe. However, the detected light intensity of a reflectance probe is weaker than the transmission probe with the same source to detector separation.

In this application, transmission probes are not suitable due to the very long optical path that the light would have to travel to the photo-detector which is located opposite sides of the maternal abdomen (Zahedi & Beng, 2008). The reflectance probe becomes the method of choice where the photo-detector is placed on the same body surface (abdomen) making the measurement of abdominal PPG signal possible (Zahedi & Beng, 2008).

2.3 Photo-detector noise

When designing an optical instrument, the photo-detector is an essential component. Selection of an appropriate photo-detector resulted in better signal quality of the acquired signals. The noise floor of the photo-detector will determine the maximum S-D separation which is useful in the optical instruments.

Currently, the low noise photo-detector (from Edmund Optics Inc.) with noise equivalent power as low as $1.8 \times 10^{-14} \text{ W/Hz}^{1/2}$ (0.051 cm^2) (W57-522, Edmund Optics, Inc.) and $8.6 \times 10^{-14} \text{ W/Hz}^{1/2}$ (1.00 cm^2) (W57-513, Edmund Optics, Inc.). Noise equivalent power is the incident optical power required to produce a signal on the photo-detector that is equal to the noise when the SNR is equal to one. These silicon photo-detectors are then utilized in the following analysis.

Photo-detector area (cm ²)	R (A/W)	R_{sh} min (M Ω)	Bandwidth (Hz)	I_{PN} (A)
0.051	0.62	600	100	8.29×10^{-14}
			1000	2.63×10^{-13}
			10000	8.29×10^{-13}
			100000	2.63×10^{-12}
			100	3.71×10^{-13}
1.00	0.62	30	1000	1.17×10^{-12}
			10000	3.71×10^{-12}
			100000	$\times 10^{-11}$

Table 1. P_{Noise} during photovoltaic operation at various bandwidths

The photo-detector can either operate in photovoltaic or photo-conductance condition. Photovoltaic operation offered a low noise system compared to the photo-conductance operation. Shot noise (due to the dark current) is the dominant noise component during photo-conductance operation. Small photo-detector's active area resulted in lower noise level compared to the large photo-detector's active area. Since strong scattering process for the human tissue dispersed the light in random fashion (Bronzino, 2000), large photo-detector's active area increases the probability of detecting photons that exit from the maternal layer. Therefore, photo-detector with 1 cm² area is proposed for the optical fetal heart rate instrument. This value has thus been used in the rest of this work. Table 1 showed the proposed silicon photo-detector's noise, P_{Noise} during photovoltaic operation at various bandwidths. It shows that photo-detector's noise increases with its bandwidth.

3. Results and discussions

This section discusses the determination of S-D separation based on the limit of ANC operation. Results obtained in previous work (Zahedi & Beng, 2008) encouraged us to take one step forward via practical implementation of the circuitry whereas digital synchronous detection is utilized to further enhance the SNR. The design and development of the OFHR system is described and results of the clinical trial are also reported.

3.1 Adaptive noise cancellation and the limit of the photo-detector

Since the adaptive noise canceling limit is -34.7 dB, the photo-detector used in the optical fetal heart rate instrument must be able to detect fetal signal at this limit. By using Equation 19, the expected fetal optical power, P_F at -34.7 dB is estimated and tabulated in Table 2.

$$10\log_{10}\left[\frac{P_F}{P_{M+am}}\right] = -34.7 \text{ dB} \quad (19)$$

where P_F is the estimated fetal optical power, P_{M+am} is the optical power at photo-detector using Monte Carlo simulation and -34.7 dB is the limit of the ANC operation. These values were obtained through Monte-Carlo simulation using a three-layered tissue model (maternal, amniotic, and fetal) (Zahedi & Beng, 2008). Optical properties (scattering and

absorption coefficients) of the tissue model as well as respective thicknesses were obtained from previous studies (Ramanujam et al. 2000; Gan, 2009), and simulation results were based on the launching of two million photons with 1 mW optical power. The detailed discussion of the Monte-Carlo simulation can be found in the previous report (Gan, 2009) and will not be further discussed here.

From Figure 5, when S-D separation larger than 4 cm (6 cm, 8 cm and 10 cm), the expected optical power is below the photo-detector noise level. At 2 cm and 4 cm source to detector separation, the expected fetal optical powers, 2293.99×10^{-12} W/cm² and 5.94×10^{-12} W/cm² respectively, are higher than the photo-detector's noise (1.17×10^{-12} W/cm²) level. The photo-detector is assumed to be operated at the photovoltaic condition with 1000 Hz bandwidth and 1 cm² active area. Therefore, source to detector separation of 4 cm, which results in 70% of optical power from fetal layer, is suitable to use with this low noise photo-detector. At 890 nm and 4 cm source-detector separation, the receiver sensitivity is optimized by considering the limitation of the adaptive filter in FHR detection.

Source to detector separation (cm)	Expected signal level, P_{M+am} ($\times 10^{-9}$)	Expected P_F signal level of -34.7 dB ($\times 10^{-12}$)
2	6767.09	2293.99
4	17.53	5.94
6	0.31	0.11
8	0.37	0.13
10	0.09	0.03

Table 2. Expected P_F signal level (-34.7 dB) at various source to detector separation

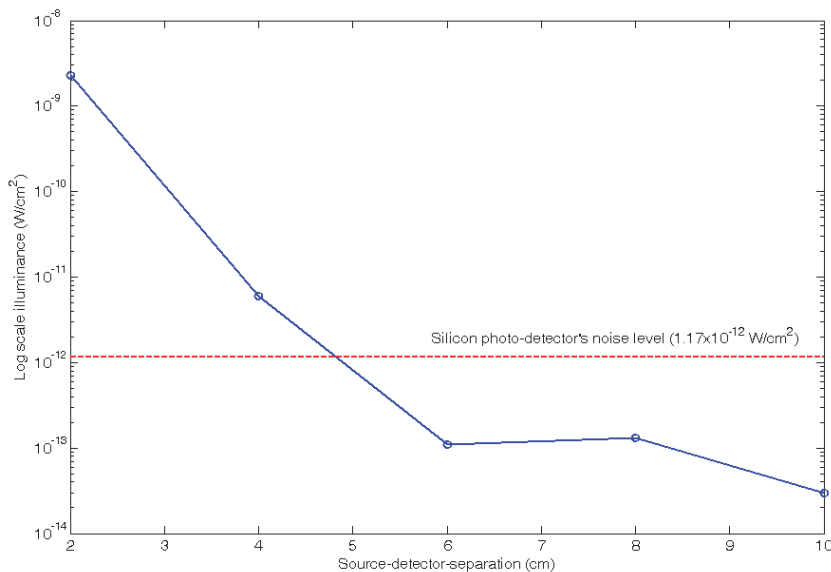


Fig. 5. Estimated P_F (-34.7 dB) at 2.5 cm fetal depth

3.2 Implementation of ANC in transabdominal fetal Heart rate detection using PPG

In our work (Gan et al. 2009), a low-power optical technique is proposed based on the PPG to non-invasively estimate the FHR. A beam of LED light (<68 mW) is shone to the maternal abdomen and therefore modulated by the blood circulation of both mother and fetus whereas maximum penetration is achieved at a wavelength of 890 nm. This mixed signal is then processed by an adaptive filter with the maternal index finger PPG as reference input. Figure 6 shows the optical fetal heart rate detection (OFHR) system block diagram whereas the implementation by using National Instrument hardware and LabVIEW software are illustrated in Figure 7 and Figure 8. In the OFHR system, the fetal probe (primary signal) is attached to the maternal abdomen using a Velcro belt to hold the IR-LED and photo-detector, separated by 4 cm. The reference probe is attached to the mother's index finger as generally practiced in pulse oximetry. As the selected IR-LED could only emit a maximum optical power of 68 mW, the OFHR system operates with an optical power less than the limit of 87 mW specified by the International Commission on Non-Ionizing Radiation Protection (ICNIRP) (International Commission on Non-Ionizing Radiation Protection, 2000). In order to modulate the IR-LED, the modulation signal is generated at a frequency of 725 Hz using software subroutine through a counter port (NI-USB 9474) to the LED driver (Figure 6).

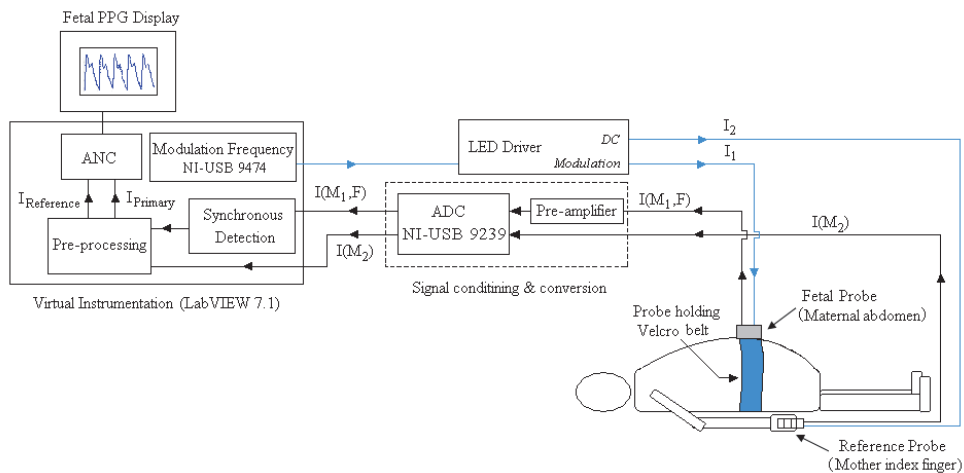


Fig. 6. OFHR system block diagram showing the hardware modules have been implemented in LabVIEW.

The diffused reflected light from the maternal abdomen, detected by the low-noise photo-detector, is denoted as $I(M_1, F)$, where M_1 and F denote the contribution to the signal from the mother abdomen and fetus, respectively. A low-noise ($6 \text{ nV/Hz}^{1/2}$) transimpedance amplifier is utilized to convert the detected current to a voltage level. The reference probe (mother's index finger) consists of an IR-LED and a solid-state photodiode with an integrated preamplifier. The signal from this probe is denoted as $I(M_2)$, where M_2 refers to the maternal contribution. Synchronous detection is not required at this channel as the finger photoplethysmogram has a high signal to noise ratio (SNR).

Detected signals from both probes are simultaneously digitized with a 24-bit resolution data acquisition card (NI-USB 9239, National Instruments, Inc.) at a rate of 5.5 kHz. The

demodulation, digital filtering, and signal estimation are all performed in the digital domain. Software implementation consists of generating a modulation signal, a synchronous detection algorithm, down-sampling, high-pass filtering and ANC algorithm (Zahedi & Beng, 2008). The entire algorithm and part of the instrument have been implemented using Laboratory Virtual Instrumentation Engineering Workbench (LabVIEW 7.1, from National Instruments, Inc.). After pre-processing and applying the ANC algorithm (Figure 9), the fetal signal and the fetal heart rate are displayed. The FHR is found by estimating the prominent peak of the power spectral density using the Yule-Walker autoregressive (AR) method (order of 20).

Figure 7 shows the laboratory prototype and the graphical user interface of the OFHR system (in Figure 7, left) where the maternal index finger PPG (top), the abdominal PPG (middle) and the estimated fetal PPG (bottom) are presented. There are three types of

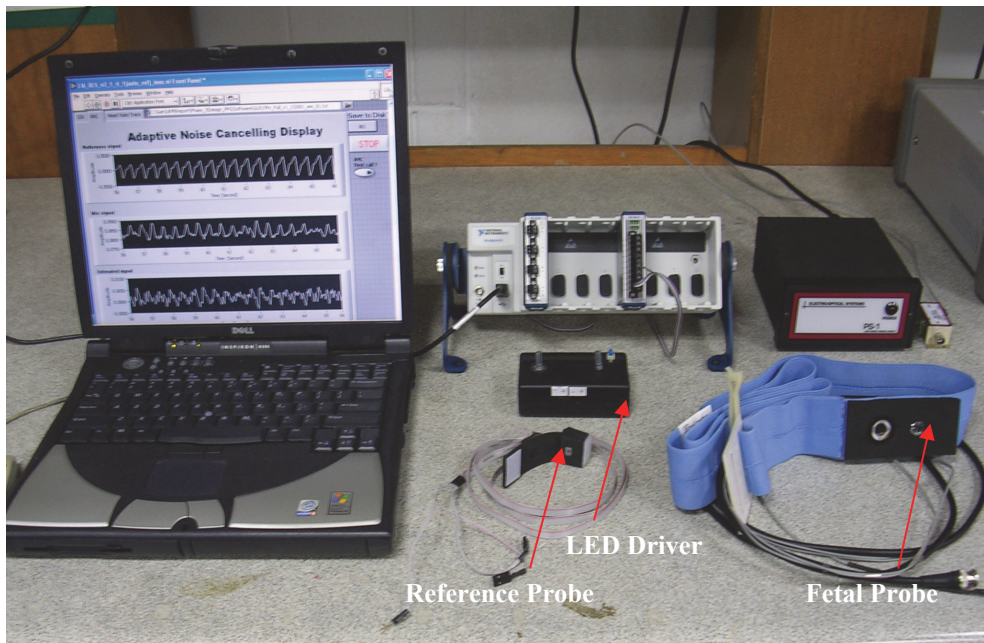


Fig. 7. OFHR prototype

selectable displays (Figure 8) namely digital synchronous or lock-in amplifier (LIA), ANC and heart rate trace. The purpose of the first two displays is to assist development and the third one (Figure 8) indicates FHR values versus time (clinical application). The user can either save the data to the personal computer for further analysis or just display it online.

Finally, a total of 24 data sets were acquired from six subjects at 37 ± 2 gestational weeks from the Universiti Kebangsaan Malaysia Medical Centre. This study was reviewed and approved by the University Ethical Committee and written consent was obtained from all patients who participated in this study after the procedure was clearly explained to them. The process for subject recruitment and data acquisition are complied with the rules and regulation as stated in Good Clinical Practice.

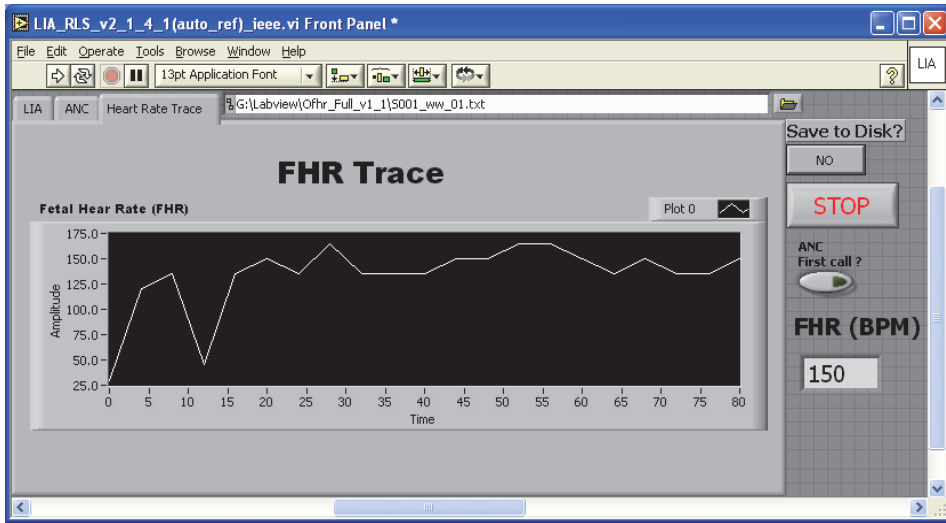


Fig. 8. Graphical User Interface of OFHR system. FHR trace menu

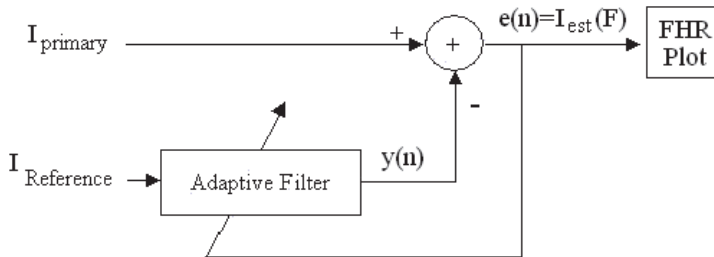


Fig. 9. ANC block diagram

In this study, all fetuses were singleton with gestation weeks from 30 week to 40 week. Subjects with twin pregnancies, anterior placed placenta, obesity ($BMI \geq 30$), gestational diabetes mellitus (GDM) and hypertension were excluded from this study. In addition, all fetuses in this study were found to be healthy by the obstetrician and born naturally (vaginally) without any complication.

During the data acquisition, the fetal probe is fixed to a maternal abdomen and the reference probe on her index finger in semi-Fowler position. The data analysis shows a correlation coefficient of 0.97 (p -value < 0.001) between optical and ultrasound FHR with a maximum error of 4%. Clinical results indicate that positioning the probe over the nearest fetal tissues (not restricted to head or buttocks) improves signal quality and therefore detection accuracy.

4. Conclusions

A low power OFHR detection system has been designed and developed using low cost, very low power (< 68 mW) IR light and a commercially available silicon photo-detector. The digital synchronous detection and adaptive filtering techniques have been successfully

implemented using LabVIEW 7.1. By applying digital synchronous detection and adaptive filtering techniques the FHR was determined with acceptable accuracy (maximum error of 4%) when compared to Doppler ultrasound. Attested by clinical results the probe positioning influences the acquired signal's quality and therefore affects the FHR results. Locating the nearest fetal tissues (not restricted to head or buttocks) to the probe will help to increase the signal quality and FHR determination accuracy.

The limitations of the optical technique are due to the presence of motion artifacts and sensitivity to the probe placement. The presence of motion artifacts may cause loss of correlation between the reference signal and the noise source (maternal PPG) in the mixed signal recorded from the maternal abdomen. The performance of the adaptive filtering scheme will suffer as a consequence, making the probe placement and stability an important criterion. Besides that, finding a proper location is needed in order to get signals with good SNR.

For the future development, by using an array of sensors to automatically select the channel with the highest SNR will eliminate the positioning problem. The topology of the sources and the photo-detector has to be determined. For the cost effective design, it is recommends that more light sources are used instead of photo-detectors. A wearable system will make the device more convenient for clinical applications in the near future. To ensure real-time and low-power function, the whole system can be implemented using embedded processor. The FHR will be wirelessly transmitted to another computing platform (PC or PDA) for further analysis, storage and transmission (to the nursing entity at a clinic). The main performance factors which will be considered are robustness, battery life, weight, dimensions and ergonomics. It is thought that the using of the selected platform (ARM) implementation will lead to a sufficiently low-cost bill-of material for the final product. During development phase, EMC directives will be taken into account so that the system's operation does not affect nor will be affected by other electronic devices. As a by-product of the project and contribution to the scientific community, it is proposed that acquired data during the project to be made available to a public data-base of biological signals (www.physionet.org) maintained by MIT in the USA.

5. Acknowledgment

This work has been partially supported by research university grant UKM-AP-TKP-07-2009. The authors would like to express their gratitude to Prof. Dr. M. A. J. M. Yassin and Associate Prof. Dr. S. Ahmad for their assistance in collecting the clinical data, and the staff at the Universiti Kebangsaan Malaysia Medical Centre, especially N. F. Mujamil for her assistance in determining the fetal position through ultrasound scan.

6. References

- Vaseghi, S.V. (2000). *Advanced digital signal processing and noise reduction*, Baffins Lane: John Wiley & Sons Ltd
- Widrow, B.; Glover Jr, J.R.; McCool, J.M.; Kaunitz, J.; Williams, C.S.; Hearn, R.H.; Zeidler, J.R.; Eugene, D. & Goodlin, R.C. (1975). Adaptive noise cancelling: principles and applications, *Proceedings of the IEEE*, Vol. 63, pp. 1692-1716
- Ifeachor, E.C. & Jervis B.W. (2002). *Digital signal processing: A practical approach*, England: Prentice Hall

- Freeman, R.K.; Garite, T.J. & Nageotte, M.P. (2003). *Fetal heart rate monitoring*, Lippincott William & Wilkins
- Philip, J.S. (2002). Fetal distress. *Current Obstetrics & Gynaecology*, 12(1):5-21
- Hershkovitz, R.; Sheiner, E. & Mazor, M. (2002). Ultrasound in obstetrics: a review of safety, *European Journal of Obstetrics & Gynecology and Reproductive Biology*, Vol. 101, pp. 15-18
- Karlsson, B.; Berson, M.; Helgason, T.; Geirsson, R.T. & Pourcelot, L. (2000). Effects of fetal and maternal breathing on the ultrasonic Doppler signal due to fetal heart movement, *European Journal of Ultrasound*, pp. 47-52
- Khandpur R. S. (2004). *Biomedical Instrumentation: Technology and Applications*, McGraw-Hill Professional
- Najafabadi, F. S.; Zahedi, E. and Mohd Ali, M.A. (2006). Fetal heart rate monitoring based on independent component analysis, *Computers in Biology and Medicine*, Vol. 36, No. 3, pp. 241-252
- Ramanujam, N.; Vishnoi, G.; Hielscher, A.H.; Rode, M.E.; Forouzan, I. & Chance, B. (2000) Photon migration through the fetal head in utero using continuous wave, near infrared spectroscopy: clinical and experimental model studies, *Journal of Biomedical Optics*, pp. 163-172
- Chance, B. (2005). Transabdominal Examination Monitoring and Imaging of Tissue. U.S. Patent 2005/0038344A1
- Zourabian, A.; Chance, B.; Ramanujam, N.; Martha, R. & David A.B. (2000). Trans-abdominal monitoring of fetal arterial blood oxygenation using pulse oximetry, *Journal of Biomedical Optics*, Vol. 5, pp. 391-405
- Nioka, S.; Izzetoglu, M.; Mawn, T.; Nijland, M.J.; Boas, D.A. & Chance, B. (2005). Fetal transabdominal pulse oximeter studies using a hypoxic sheep model, *The Journal of Maternal-Fetal and Neonatal Medicine*, Vol. 17, No. 6, pp. 393-399
- Vintzileos, A.M.; Nioka, S. & Lake, M. (2005). Transabdominal fetal pulse oximetry using near-infrared spectroscopy, *American Journal of Obstetric & Gynaecology*, Vol. 192, pp. 129-133
- Zahedi, E. & Beng, G.K. (2008). Applicability of adaptive noise cancellation to fetal heart rate detection using photoplethysmography. *Computers in Biology and Medicine*, Vol. 38, No. 1, pp. 31-41
- Choe, R.; Durduran, T.; Yu, G.; Nijland, M.J.M.; Chance, B.; Yodh, A.G. & Ramanujam, N. (2003). Transabdominal near infrared oximetry of hypoxic stress in fetal sheep brain in utero, *Proceedings of the National Academy of Sciences*, vol. 100, No. 22, pp. 12950-12954
- Gan, K.B.; Zahedi, E. & Mohd. Ali, M.A. (2009). Trans-abdominal fetal heart rate detection using NIR photoplethysmography: instrumentation and clinical results, *IEEE Transactions on Biomedical Engineering*, Vol. 56, No. 8, pp. 2075-2082.
- Manolakis, D.G.; Ingle, V.K. & Kogon, S.M. (2005). *Statistical and adaptive signal processing*. Norwood:Artech House, Inc.
- Haykin, S. (2002). *Adaptive filter theory*. Prentice Hall.
- Gan K.B. (2009). Non-invasive fetal heart rate detection using near infrared and adaptive filtering. Available online from: (<http://ptsldigital.ukm.my>)

- International Commission on Non-Ionizing Radiation Protection. (2000). ICNIRP statement on light-emitting diodes (LEDs) and laser diodes: Implications for hazard assessment, *Health Phys.*, Vol. 78, No. 6, pp. 744–752
- Bronzino, J.D. (2000). *The biomedical engineering handbook: volume 1*, Florida: CRC Press LLC
- Allen, J. & Murray, A. (2000). Variability of photoplethysmography peripheral pulse measurements at the ears, thumbs and toes, *IEEE Proceeding Science and Technology*, Vol. 147, No. 6, pp. 403–407

Adaptive Filtering by Non-Invasive Vital Signals Monitoring and Diseases Diagnosis

Omar Abdallah^{1,2} and Armin Bolz¹

¹*Institute for Biomedical Engineering, Karlsruhe Institute of Technology*

²*Biomedatronic, Karlsruhe
Germany*

1. Introduction

The reliability, reproducibility and accuracy of in-vivo measurements are of great importance and have to be thoroughly studied and to a great extent achieved. Reproducibility problems may result from the electronic components of the applied devices and the variability of measured variables as well as noise sources. The inaccuracy is caused by the approximation in the calculations or the used methods and by diverse sources of errors resulting from the subject under considerations and its surroundings. In sensible measurement like blood components, the positioning of the measuring sensor as well as the variation in the applied pressure and the characteristics of contact area between sensor and skin have a great effect on the accuracy and reproducibility of the measurements. The ambient noise like high frequency and line frequency (50 or 60 Hz) noise can be filtered by the detected biosignals like Photoplethysmogram (PPG) using the conventional analog or digital filters without a great effort. The motion artifact of the subject caused by him as well as by physical motion of body parts or by the surrounding has a varying frequency which may have the same range of the signal frequency. It is difficult to filter noise from these signals, and errors resulting from filtering can distort them. Usually physicians are misled by these noisy signals and the analysis can go wrong. An adaptive filter is essential by bio-signal and bio-image processing for noise cancellation without destroying or manipulating the valuable detected information.

Biomedical signals such as photoplethysmogram (PPG) (Figure 1), electrocardiogram (ECG), electroencephalogram (EEG), electromyogram (EMG) and impedance cardiogram (ICG) are very important in the diagnosis of different pathological variations. By the detection of these bio-signals as well as by the further derived parameters like oxygen saturation by pulse oximeter, the motion artifact is a great challenge, which may lead to erroneous results or even no results can be delivered [Lee].

The effectiveness of ECG monitors can be significantly impaired by motion artifact, which can cause misdiagnoses, lead to inappropriate treatment decisions or trigger false alarms. However, it is difficult to separate the noise from bio-signal due to its frequency spectrum overlapping that of the ECG. A portable ECG recorder using accelerometer based on motion artifact removal technique will be a great help for patients for tele-homecare or ambulatory ECG monitoring.

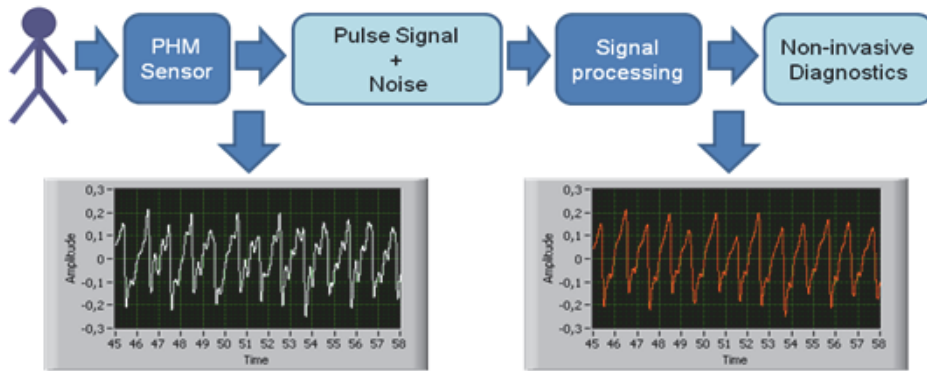


Fig. 1. Signal detection and processing by noninvasive diagnosis

A maternal electrocardiogram (mECG) and abdominal noise in abdominal maternal recordings (especially by cardiotocography) can be orders of magnitude stronger than the fetal electrocardiogram (fECG) signal. An adaptive filter using frequency-domain or time-domain electrocardiogram features can be applied by the automatic extraction of a beat-to-beat fECG from mECG using surface electrodes placed on the maternal abdomen [Rik] [Prasad]. This will allow early diagnosis and monitoring treatment of certain fetal cardiac disorders.

By non-contact ECG monitoring, cardiac activity and movements (that may be seen in part in cardioballistogram CBG) may cause also disturbance to the detected signals, which can be eliminated by applying an appropriate adaptive filter

High-quality EEG recording is crucial for diagnosis of different pathological variations. EEG has biological artifacts and external artifacts. Biological artifacts can be EMG-, EOG- (Electrooculograph) CBG or ECG-signal [Rasheed]. These artifacts appear as noise in the recorded EEG signal individually or in a combined manner. These noise sources increase the difficulty in analyzing the EEG and to obtaining clinical information. For this reason, it is necessary to design specific filters to decrease such artifacts in EEG records. EEG quality in the MR scanner is compromised by artifacts caused by interaction between the subject, EEG electrode assemblies, and the scanner's magnetic fields [Rasheed2009]. The three most significant causes of EEG artifacts in the scanner are the large movements in the static field like swallowing; the cardioballistogram and blood flow effects in the field associated with the subject's pulse; and the changing fields applied during fMRI image acquisition. Pulse artifact is potentially a significant problem as it is normally large amplitude, widespread on the scalp, and continuous. Using a cascade of adaptive filters based on a least mean squares (LMS) algorithm can eliminate the undesired signals or interferences.

2. Photoplethysmogram

The photoplethysmogram (PPG) waveform comprises a pulsatile physiological waveform superimposed on a slowly varying baseline with various lower frequency components. The pulsatile one is attributed to cardiac synchronous changes in the blood volume with each heart beat, and the second is attributed to respiration, sympathetic nervous system activity and thermoregulation. Figure 2 shows a typical PPG signal without motion artifact. The

PPG technology has been used in a wide range of commercially available medical devices for measuring oxygen saturation, blood pressure and arterial stiffness, cardiac output, assessing autonomic function and detecting peripheral vascular diseases. Although the origins of the components of the PPG signal are not fully understood, there is no doubt that they can provide valuable information about the cardiovascular system and autonomic nervous system. Hence, there is a great interest in the technique in recent years, driven by the demand for low cost, very compact size, simple and portable technology for the primary care and community based clinical settings, non-invasive technology without side effects or risks as well as online monitoring capability and the advancement of computer-based pulse wave analysis techniques and diagnosis [Allen, Abicht]. A computer aided analysis tool for the hemodynamic diagnosis using PPG can be very helpful in clinical applications. Automatic assessment of the reliability of reference heart rates from patient vital-signs monitors incorporating both ECG and PPG based pulse measurements has been proposed by Yu et al. They expressed reliability as a quality index for each reference heart rate. The physiological waveforms were assessed using a support vector machine classifier and the independent computation of heart rate made by an adaptive peak identification technique that filtered out motion-induced noise [Allen].

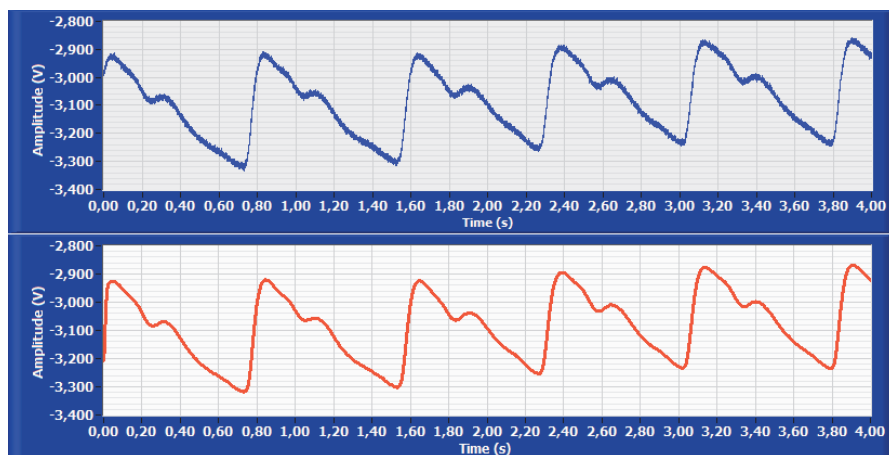


Fig. 2. Photoplethysmogram PPG; top: detected raw signal, bottom: filtered signal

Also, due to demographic change, especially in the industrial countries, the personal health care of old people is of great importance for prevention and rehabilitation. Continuous monitoring of vital parameters is essential for that aim. By long term as well as by emergency, a monitoring without interruption is crucial for the diagnosis of the case under consideration. In many cases, a motion artifact caused by patient as well as by physical motion of body parts or by the surrounding may have the same range of the signal frequency. It is difficult to filter noise from these signals, and errors resulting from filtering can distort them. Pulse oximeter for measuring oxygen saturation (SrO_2) using more than one PPG signal is a valuable device for monitoring patients in critical conditions. PPG and the derived oxygen saturation are susceptible for motion artifact.

Pulse oximetry sensors use two Light Emitting Diodes (LEDs) which emit red and infrared light that shine through a reasonably translucent part of the patient's body. In pulse

oximetry, it is called red light to the light band whose wavelength is comprised between 600-750 nm, while infrared light's wavelength varies between 850 and 1000 nm. These two wavelengths values are chosen because light absorption coefficient varies with the oxygen concentration of in both the red and the infrared light. Figure 3 shows the two principles of pulse oximetry: transmission and reflection pulse oximetry. By transmission Pulse oximetry, the light sensitive photodetector (Photodiode PD), which acts as a receiver picking up the light that passes through the measuring site, is opposite to the light emitter (light emitting diode LED). By reflection pulse oximetry the PD and the LED's lie at the same side of the irradiated body portion.

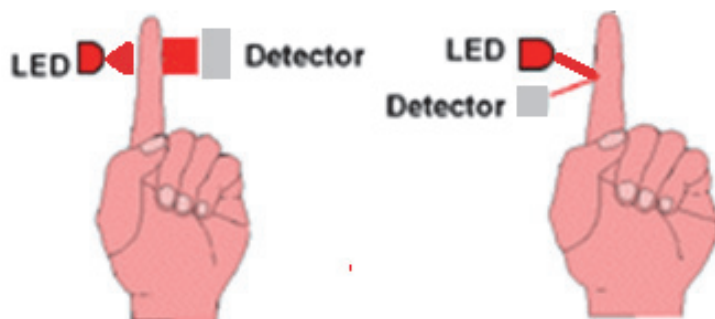


Fig. 3. Operation of the pulse oximeter sensor, left: transmission pulse oximetry, right: reflection pulse oximetry

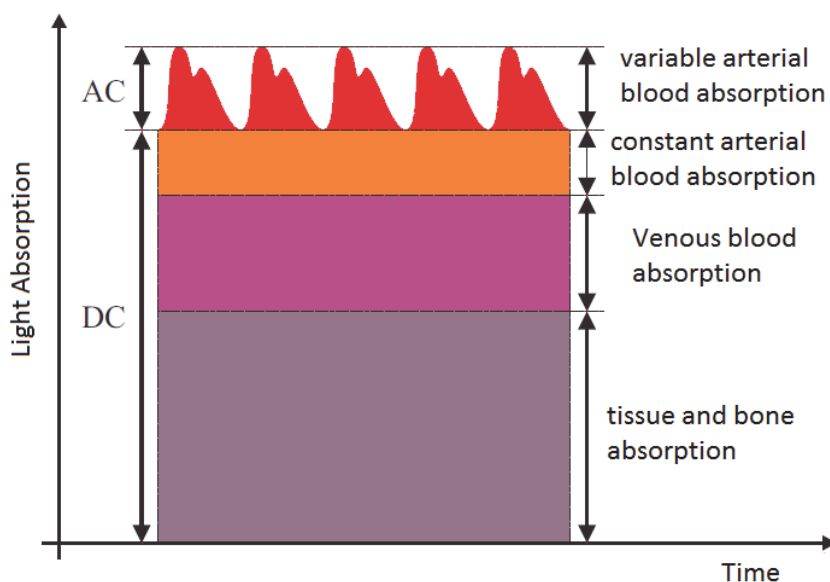


Fig. 4. Light absorption by different tissue, at the top we see the plethysmogram generated by the arterial pulsationen

Pulse oximeter works according to two physical principles: first, the presence of a pulse wave generated by changes in blood volume (plethysmography) in the arteries and capillaries (Figure 4) and second, the fact that oxyhemoglobin (O₂Hb) and reduced hemoglobin (Hb) have different absorption spectra (spectroscopy). Oxygenated hemoglobin absorbs more infrared light and allows more red light to pass through. Deoxygenated (or reduced) hemoglobin absorbs more red light and allows more infrared light to pass through.

3. Adaptive filtering of photoplethysmogram

We emphasize here on the use of the adaptive filter by PPG, because of the importance of this signal by detecting further parameters like pulse transit time (PPT); blood pressure monitoring, Pulse rate variability and the application of it for the risk estimation and diagnosis of cardiovascular diseases. Also the non-invasive calculation of concentration, fractional oxygen saturation and further blood components like glucose may require also the PPG signal analyzing. AC component of PPG signal carries important information for diagnosis, but it may be affected by noise, which is sharing the same bandwidth. An important application for the PPG is the calculation of oxygen saturation in emergency and in intensive care, where the oxygen supplement of tissue has to be measured continuously. The problem will be greater for example by detecting the PPG by low perfusion for the monitoring of oxygen saturation, where a low signal to noise ratio is the result. An adaptive filter will be the solution for this problem. Conventional filtering cannot be applied to eliminate those types of artifacts because signal and artifacts have overlapping spectra. For long term monitoring an adaptive filter is essential [Com 2007].

By pulse oximetry, Masimo adaptive filter is well known to the people working in this area. The principle is easy and shortly described here: all detected samples of PPG's by red and infrared causing oxygen saturation below a certain value (e.g. 80%) are coming from venous blood signals caused by motion artifact and has to be filtered. All signals causing saturation higher than a threshold value (e.g. 90%) are the arterial signal. An intelligent algorithm is designed according to this principle for the robust detection of oxygen saturation. By using one PPG signal we cannot apply this algorithm. We used another algorithm by Filtering and generation of reference noise depending on the detected signal.

Motion artefacts are one of the most important handicaps of photoplethysmography and pulse oximetry, as they suppose a big limitation and often become an insurmountable obstacle on the utilization of this technology, since they are quite hard to cancel mainly due to spectral characteristics of both, pulse signals and motion artifacts. In order to improve the quality of Photoplethysmograms and pulse oximetry, some signal processing must be implemented. Our research proposes, as viable solution, an Adaptive Filter in Noise Cancellation configuration, working with a Least Mean Square Algorithm. At the end of the system, we have carried out a reconstruction of the Photoplethysmogram and the signal that we recover has a high enough quality for measuring fractional oxygen saturation of hemoglobin in blood and for further diagnosis purposes.

An Adaptive Noise Cancellation (ANC) System has two inputs. This fact can be seen in the Figure 5 presented below, more specifically in the diagram on the top. One is the Input Signal, i.e., the signal corrupted by noise, coming from the sensor output, and the other one is the Noise Reference, coming from the Synthesizer output. Both, the graphic of the Input Signal and the generated plot of the Noise Reference appear in the Front Panel of the corresponding LabVIEW program. Given that the Least Mean Square Algorithm provides

adaptive filtering, the Noise Reference is adjusted to the real noise measured with the sensor and, as a result, the output, Filtered Signal, naturally will be the filtered signal. In the diagram below from the Figure 5 the main blocks of the Least Mean Square Algorithm (LMS) implementation are presented. It is worth mentioning the fact that this algorithm is recursive: the weights of the filter are calculated recursively to minimize the Mean Square Error [Abdallah].

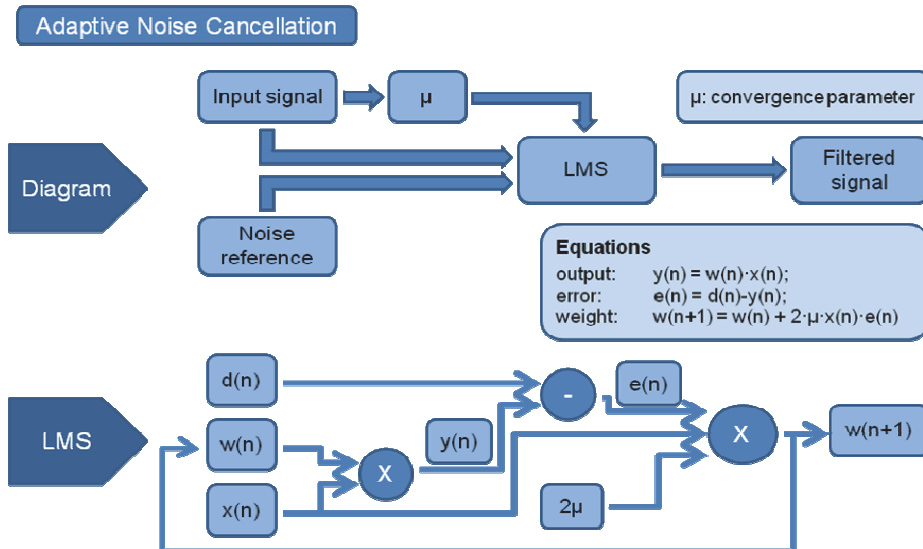


Fig. 5. Block diagram of the Adaptive Noise Cancellation (ANC)

4. Method and results by adaptive filtering of photoplethysmogram

Adaptive filters have been used to enable the measurement of photoplethysmogram PPG under conditions, where movement of the body parts where the sensor is applied causes a high noise to the signal. In this adaptive filter a noise reference and a signal reference are used. We use the least mean square (LMS) method to extract the actual signal from the noisy one.

For the first approximation to generate the reference signal a lowpass filter is used. Using the resulting signal from this lowpass, an appropriate reference signal is generated. This reference signal is in turn subtracted from the detected signal to generate a noise signal. The generated noise signal is modified to synthesize the noise reference signal. The synthesized reference noise is adjusted by the adaptive algorithm to the real one contained in the measurement, and then subtracted from the detected noisy signal. The resulting signal is modified to fulfill certain requirements (Figure 6)

The applied method discussed above can be used for the detection of a photoplethysmogram signal without the need for further signals of the same type or requiring a further sensor. Figure 7 shows an example of the results obtained using this method. The algorithm was tested for the calculation of oxygen saturation and accurate results are delivered under artificial motion artifacts.

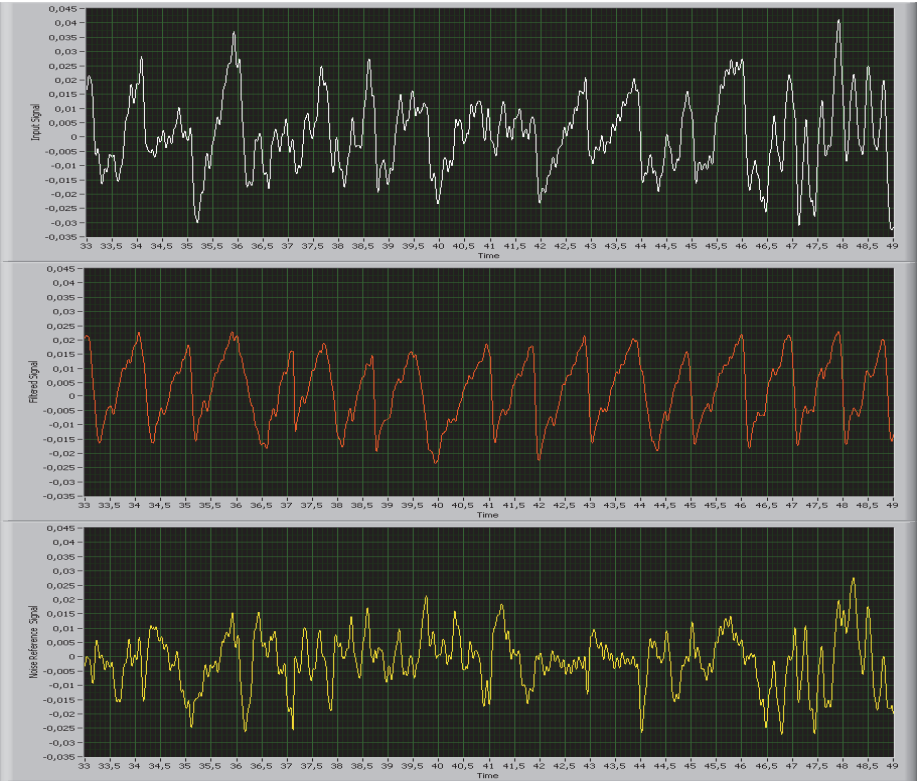


Fig. 6. Detected signal (top), generated reference signal (middle) and generated reference noise for PPG filtering

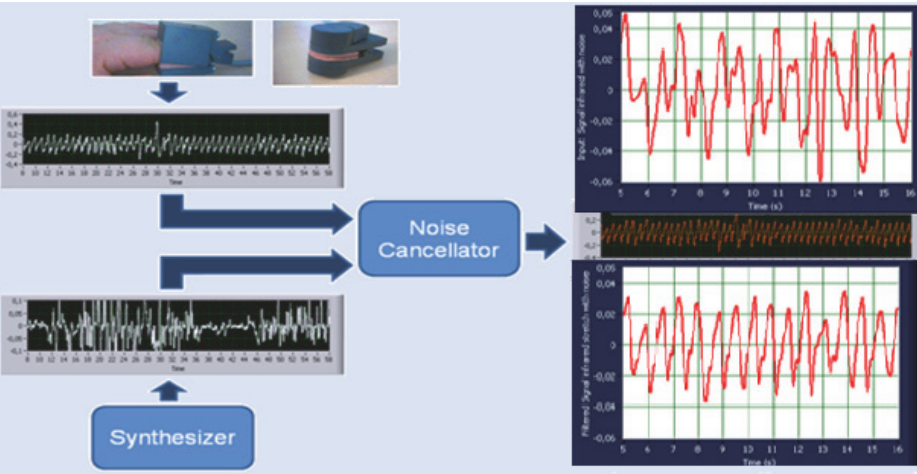


Fig. 7. Schematic of the PPG filtering

Each measurement from the applied PHM sensor contains seven signals of LEDs having different wavelengths. Besides, a LED (which is off) acts as zero reference level. Since we need two of them, first we have to separate them. Once these signals are presented separately, we select the two of them that have been measured with the proper wavelengths value for the calculation of oxygen saturation (LED having the wavelength 970 nm, representing infrared light and a LED having the wavelength 660 nm, representing red light). Then they are already adapted for being filtered by our system, which remove the motion artifact from them. Finally, the filtered signals obtained after the program execution can already be used to compute ratios regarding the SpO₂, such as the so-called Ω ratio:

$$\Omega = \frac{\ln \frac{I(\lambda_1, t_1)}{I(\lambda_1, t_2)}}{\ln \frac{I(\lambda_2, t_1)}{I(\lambda_2, t_2)}} \quad (1)$$

Where:

$I(\lambda_1, t_1)$, $I(\lambda_1, t_2)$, $I(\lambda_2, t_1)$ and $I(\lambda_2, t_2)$ are the light intensities measured at the instants t_1 , t_2 with the wavelengths λ_1, λ_2 respectively.

As results, examples of each step of the process described here are presented. First of all, examples of the appearance of PHM measurements (and therefore, multiwavelength measurements) are shown, both the whole measurement and a zoom of it (figure 8).

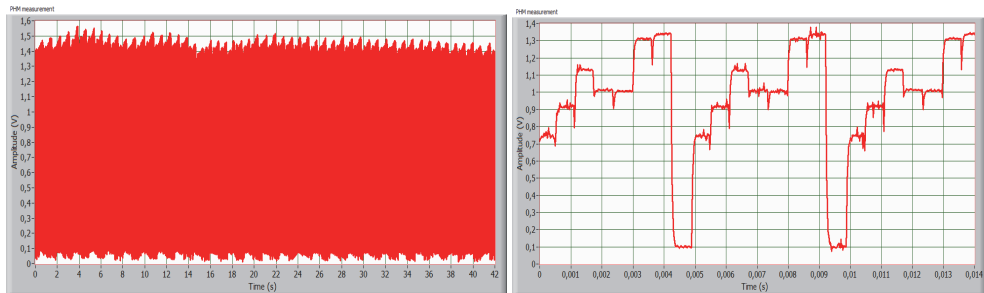


Fig. 8. Measurements of photoplethysmogram signals of PHM (right) and a zoom of it (left)

Next, the output given by the recovery of each signal is also presented. To demonstrate the ability of the system presented here to make possible a precise enough computation of the SpO₂, we have calculated the value of the above-named Ω ratio for several measurements. In order to make sure that the adaptive filter works well enough to get accurate SpO₂ readings, the main goals are: first, to prove that the ratios obtained are included in an acceptable range (bearing in mind that the values of this ratio allow us to estimate the calibration that has to be applied later to the exact calculation of the SpO₂). Next, it must be proved that the values for the ratio when the signal is affected by motion artifacts keep quite unchanging compared to those derived from the same signals without motion artifacts [Figure 10]. The pulse amplitudes of the red and infrared signals are detected by the pulse oximeter and measured to produce a certain ratio value, which is intrinsically related to the functional oxygen saturation of (SpO₂).

The signals shown in Figure 8 are measured by a Pulshemometer (PHM) sensor for the aim of calculation of concentration and fractional oxygen saturation SaO₂, which based on the

Principle of plethysmography (here volume change of arterial blood due to pulsation generated from the heart) and optical spectroscopy. Also by our Project for the non-invasive monitoring of glucose concentration in blood an adaptive filter for this aim is essential. For in vivo measurement of blood components, the adaptive filter is necessary to get rid of the noise and disturbances to the signal without any distortion of the detected useful signal that may cause erroneous additive signals or that may reduce the information contents in the detected signal. The Pulshemometer PHM sensor with a compact hardware circuit for driving the LED's and programmable digital potentiometer for adaptive programmable gain amplification is shown in Figure 9.

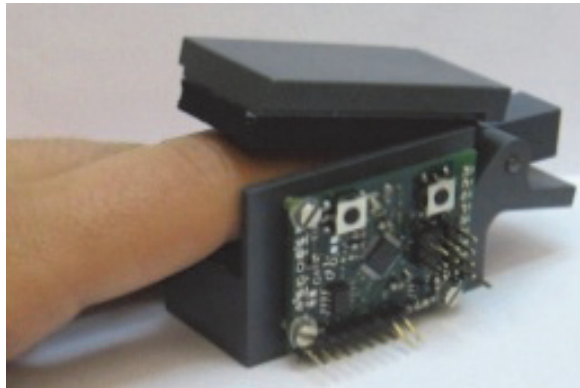


Fig. 9. Pulshemometer PHM sensor for hemoglobin concentration and fractional and fractional

Seven separated filtered PHM signals for and fractional oxygen saturation measurement von PHM are shown in Figure 11 after normalization. For this sensor an adaptive filter is essential for reliable and high accuracy results.

Wavelet transformation in combination with fuzzy and neuronal Networks (in some cases cascaded) adaptive filtering is applied by different research groups. An energy ratio-based method and a wavelet-based cascaded adaptive filter (CAF) can be applied for detecting and removing baseline drift from pulse waveforms. This CAF outperforms traditional filters both in removing baseline drift and in preserving the diagnostic information of pulse waveforms [Lisheng]. Daubechies wavelet adaptive filter based on Adaptive Linear Neuron networks is used to extract the signal of the pulse wave. Wavelet transform is a powerful tool to disclose transient information in signals. The wavelet used is adaptive because the parameters are variable, and the neural network based adaptive matched filtering has the capability to "learn" and to become time-varying. So this filter estimates the deterministic signal and removes the uncorrelated noises with the deterministic signal. This filter is found to be very effective in detection of symptoms from pulsatile part of the entire optical signal [Xiaoxia]. Fuzzy logic and Neuro-fuzzy can be used by adaptive filtering.

The method that has to be applied depends on the sensor applications and the case under consideration, because intensive computation time, a high speed processor and a large saving space may be needed, which may cause a delay time that disables an online monitoring. In applications by multi-monitoring it will be possible to use other detected signals for the purpose of filtering of a certain signal as will be discussed on the following section.

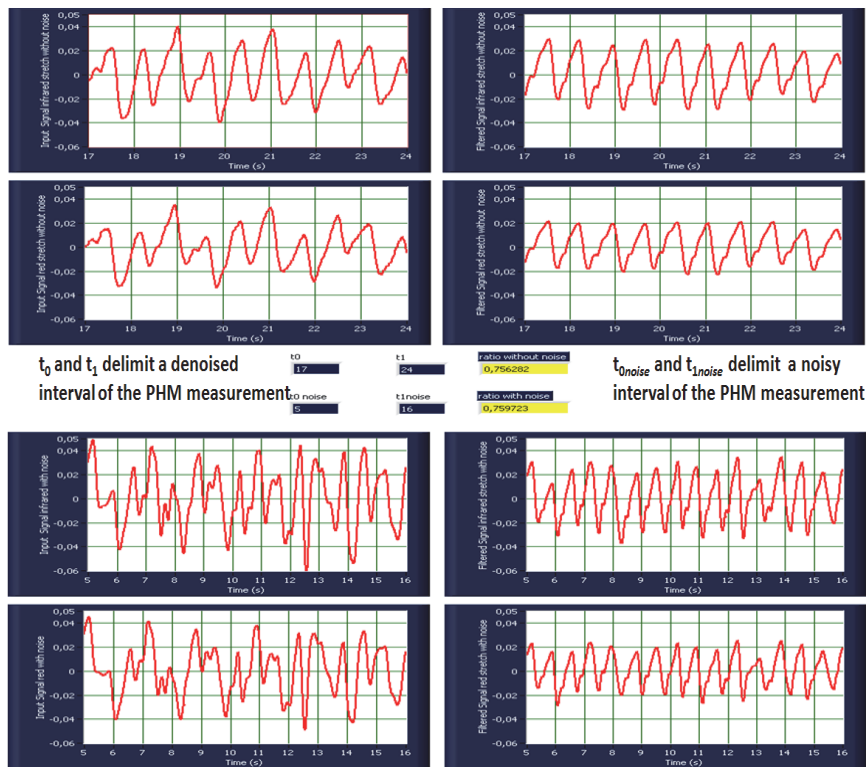


Fig. 10. LabVIEW printing for filtering a PHM measurement and computing Ω

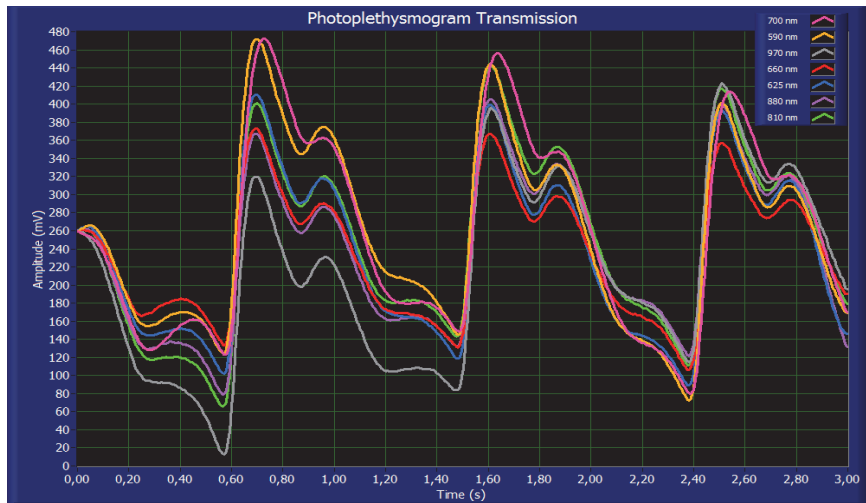


Fig. 11. LabVIEW printing for 7 filtered PHM signals for and fractional oxygen saturation measurement

5. Application of further signals for adaptive filter of photplethysmogram

In one method is to use a simple acceleration sensor of a piezoelectric element [Lit. Han2009]. In a current work we use also a method based on adaptive filtering by taking the advantage of piezoelectric sensor signal to get information about the desired signal or motion to estimate the motion artifact noise or to generate the reference signal in order to get the filtered one. Hence the piezoelectric signal describes the velocity in the blood volume change; numerical integration of the piezoelectric element signal may be helpful by synthesizing of reference signal or the noise that has to be subtracted from the noisy detected signal by generating artificial motion artifacts after signal processing. Figure 12 shows the simultaneously detected optical signal detected on fingertip and a piezoelectric signal detected on the wrist with artificial motion artifact in the time period between 11 and 15 second. The piezoelectric signal is still clean and can be used for the generating of the optical signal.

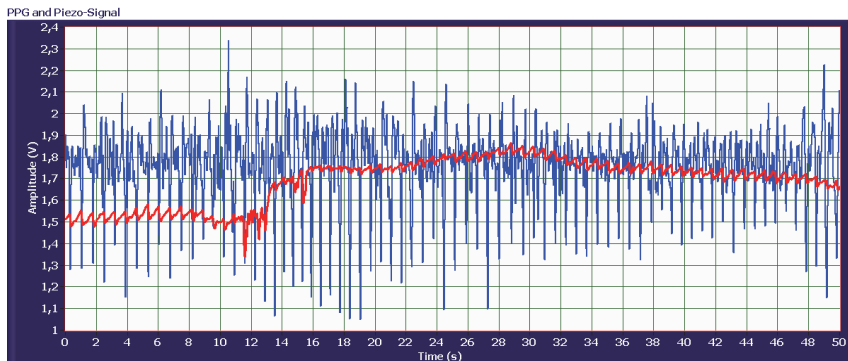


Fig. 12. PPG and piezosignal by motion artifact

Figure 13 shows the simultaneously detected optical signal detected on fingertip and a piezoelectrical signal in the case of extrasystole also detected on the wrist without motion artifact.

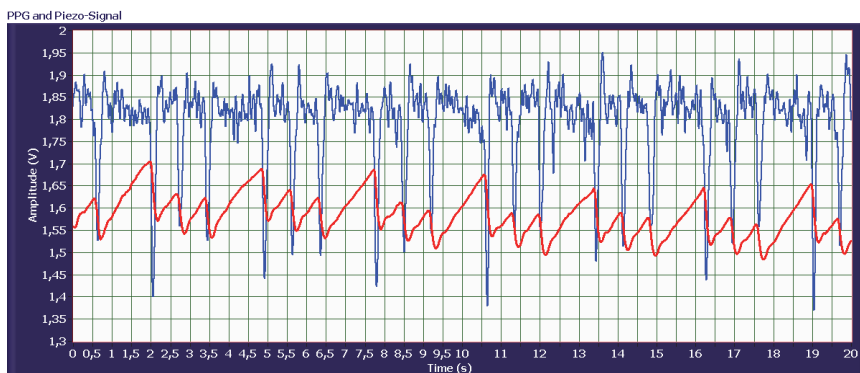


Fig. 13. PPG and piezosignal by extrasystole

Figure 14 shows the simultaneously detected optical signal detected on fingertip and the piezoelectrical signal detected on the wrist with artificial motion artifact. The piezosignal is still good enough for the generating of the PPG signal.

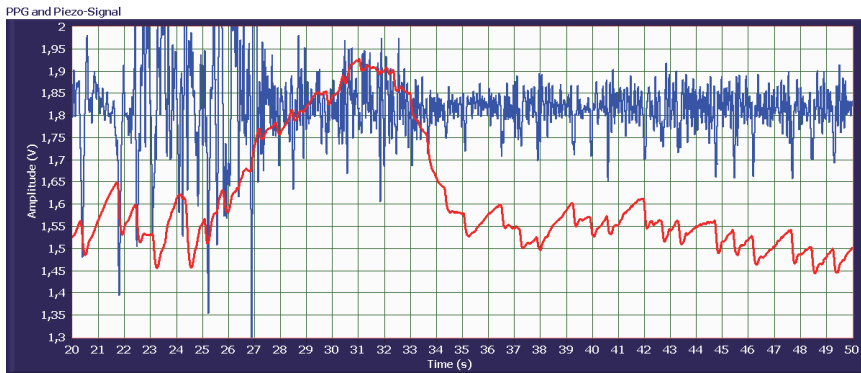


Fig. 14. PPG and piezosignal by extrasystole under motion artifact

The piezoelectric sensor may be also used for the measurement of other physiological parameters when PPG or ECG is also available. Pulse transit time PTT, which has shown its potential in cardiovascular studies, can be calculated from ECG and piezoelectric signal detected on the radial artery instead of PPG signal by low perfusion [Abdiel]. Under normal conditions the PTT can be calculated using a simple and compact piezoelectric sensor and a PPG one or by using only two simple, compact, low power piezoelectric elements. Piezoelectric elements have also different applications in biomedical engineering as sensors and actuators [Fannin].

Also a light source having the wave length around 590 nm can be used to generate signals that are larger than that of other wavelengths due to the high light absorption of hemoglobin at this wavelength.

Motion artifacts effects on the signal quality of photoplethysmographic signals are also discussed by different research groups. Using 3-D acceleration sensors are applied and showed very good results for adaptive filtering. The focus by [Volmer] lies on a performance estimation for a reconstruction method based on adaptive filtering with help of acceleration signals acquired at the fingertip. The acceleration in the direction of finger is an optimal setup for a continuous long-term application with low cost and low calculation complexity.

6. Adaptive filtering by further biosignals

The previous discussed method for the removal of the motion artifact from a PPG signal can be applied by other biosignals like ECG or EEG. We discussed in this chapter briefly the adaptive filtering by other biosignals described by different authors as examples to emphasize on the importance of adaptive filtering. Figure 15 shows a 3-chanel ECG, subjected at the end portion to motion artifact. Low signal to noise ratio can make the task of an adaptive filter more difficult.

An automated system for efficient detection of brain tumors in EEG signals using artificial neural networks (ANNs) is described by Murugesan and Sukanesh [Murugesan]. Generally, the EEG signals are bound to contain an assortment of artifacts from both subject and equipment interferences along with essential information regarding abnormalities and brain activity. Adaptive filtering has to be applied to remove the artifacts present in the EEG signal.



Fig. 15. ECG with motion artifact by the last three seconds

Artifacts in EEG records are caused by various factors, like line interference, EOG and ECG. These noise sources increase the difficulty in analyzing the EEG and to obtaining clinical information. For this reason, it is necessary to design specific filters to decrease such artifacts in EEG. A cascade of three adaptive filters based on a least mean squares (LMS) algorithm will be helpful. The first one eliminates line interference, the second adaptive filter removes the ECG artifacts and the last one cancels EOG spikes [Correa].

The MEG signal can also be used to control rehabilitation systems like prostheses and artificial neuromuscular electrical stimulation toward restoring movement to spinal cord injured patients. These mobile systems are usually used in different environments and thus are being exposed to different noise levels with characteristics not completely known. Different techniques for noise reduction have been compared for that aim: wavelet transform (WT), adaptive digital filters, and non-adaptive digital filters [Ortolan].

By non-invasive hemodynamic measurements like stroke volume and cardiac output, impedance cardiography (ICG) can be applied for continuously measurements. The impedance cardiography is designed for assessment and management of congestive heart failure, hypertension, and pacemaker patients. It is also appropriate for select ICU patients. The ICG waveform is generated by thoracic electrical bioimpedance (TEB) technology, which measures the level of change in impedance in the thoracic fluid [Philips]. Adaptive Filtering is in this case essential for reliable measurement at different conditions. Impedance plethysmography is a well known and extensively used noninvasive method for physical parameters monitoring. A major problem with this technique is its sensitivity to body movement. More than one frequency impedance plethysmograph can be used to measure different body parts movements. Adaptive filtering is heir very essential to detect and reduce motion artifacts [Rosell]. Wavelet based cancellation of artifacts in impedance cardiography is discussed by Pandey [Pandey].

Phonocardiography, Ballistocardiography may be also important for the extracting more information of the cardiovascular system as well as the low frequency signals caused by sympathetic and parasympathic autonomic nervous system for the diagnostic of cardiovascular diseases and its autoregulations [Xinsheng].

Contactless measurement for physiological parameter like ECG, EEG [Oehler] and respiration rate is research area by different groups, where adaptive filtering is of great importance. Cardiac pulse based on the information contained in the thermal signal emitted from major superficial vessels is discussed by Pavlidis et. al. [Garbey]. Non-contact monitoring of breathing function using infrared imaging is also discussed by R. Murthy and I. Pavlidis.

At the end of this chapter we want only to indicate that by image processing in biomedical engineering the adaptive filtering is very essential and discussed in many literatures [Sudha].

In summary for both, long-term and short-term monitoring (for example in case of an emergency or for prevention) of biosignals, the use of an adaptive filter is essential for reliable results.

7. Conclusions

The motion artifact caused by subject motion as well as by physical motion of body parts has a varying frequency which may lie in the same range of the signal frequency. It is difficult to filter noise from these signals using traditional filters, and errors resulting from filtering can distort them and physicians may be misled by these noisy signals and this may make the diagnosis not possible or an erroneous diagnosis is the result. By intensive care and emergency the continuous supplement of tissue with the oxygen is crucial (especially brain tissue, where after few minutes oxygen deficiency causes irreversible damage of tissue). The calculation of oxygen saturation by pulse oximetry is based on photoplethysmography and photospectroscopy. Hence the photoplethysmogram plays a great role by the calculation of oxygen saturation and heart rate as well as many further diagnostic parameters like stiffness, blood pressure and more cardiovascular pathologies can be calculated using other parameters like impedance cardiography and ECG. For a reliable detection of these parameters, an adaptive filtering is essential. By oxygen saturation detection, since many years an adaptive filter from the company Masimo that accepts all measurement results for the calculation of oxygen saturation above a threshold value and rejects all measurements yielding results below another value, is robust and in clinical use. By our method using a tiepass with a low cut off frequency a robust result was obtained. The potential of this method is that only one signal is required, which enables its application by compact sensors for different applications. Also a high velocity in the ascending portion of a signal followed by exponential or polynomial decay may be considered as a PPG signal. In multi-monitoring or using multisensory for diagnosis, another signal can be used by adaptive filtering. As an example ECG or piezoelectric signals can be applied for the robust detection of PPG. For its simplicity and cost effectiveness as well as its quick application without great disturbance of the patient we use piezoelectric signals. Acceleration sensors find their application by the adaptive filtering for PPG or ECG and further physiological parameters. By multisensors some signals can be used to generate the reference signal that has to be extracted from the measured signal or the reference noise that has to be removed from the measured data. Kalman Filter, Neuronal Fuzzy with optimization methods like swarm algorithms and Wavelet Transformation may be in some cases a good choice for

adaptive filtering of biosignals. The method which has to be applied depends on the case under consideration and the availability of other sensors. For emergency, intensive care, home care and long term monitoring and over all, where non-invasive measurement are applied, the use of adaptive filter is of a great importance and in many cases is compulsory to get the required results. It will also radically reduce the disturbances (alarm) for patient and medical care stuff, reduce costs and enhance the medical systems.

8. References

- Abdallah, O.; Piera Tarazona, A., Martínez Roca, T., Boutahir, H., Abo Alam, K. & Bolz, A. (2006). Photoplethysmogram Signal Conditioning by Monitoring of Oxygen Saturation and Diagnostic of Cardiovascular Diseases, *4th European Congress for Medical and Biological Engineering*, ISBN 978-3540892076, pp. (303-306), Antwerp, September 2008,
- Abdiel Foo Jong Yong & Sing Lim Chu. (2006). Pulse Transit Time based on Piezoelectric Technique at the radial Artery. *Journal of clinical monitoring and computing*, (May 2006) Vol. 20, Nr. 3, pp. 185-192
- Abicht Jan-Michael. (2003) Computerunterstuetzte Analyse photoplethysmographischer Signale, *Dissertation zum Erwerb des Doktorgrades der Medizin an der Medizinischen Fakultat der Ludwig-Maximilians Universitaet zu Muenchen*, October 2003, available from: http://edoc.ub.uni-muenchen.de/1793/1/Abicht_Jan_Michael.pdf
- Allen John. (2007) Photoplethysmography and its application in clinical physiological Measurement, *Physiological Measurement* 28, , (February 2007), pp. R1-R39
- Comtois, G.; Mendelson, Y., Ramuka, P. (2007). A Comparative Evaluation of Adaptive Noise Cancellation Algorithms for Minimizing Motion Artifacts in a Forehead-Mounted Wearable Pulse Oximeter, *Conf proceeding IEEE Eng Medicine Biology Soc EMBS*, ISBN: 978-1-4244-0787-3, pp. 1528 - 1531, Lyon, August 2007
- Fannin Christopher A. (1997). Design of an Analog Adaptive Piezoelectric Sensoriactuator, *Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE*, 1997
From : <http://schoar.lib.vt.edu/thesis/etd-8897-171952/unrestrictd/Cfannin.pdf>
- Garbey Marc Sun Nanfei, Merla Arcangelo, & Pavlidis Ioannis. (2007). Contact-Free Measurement of Cardiac Pulse Based on the Analysis of Thermal Imagery, *IEEE Transactions On Biomedical Engineering*, Vol. 54, No. 8, August 2007, pp. 1418-1426
- Han D. K., Hong J. H., Shin J. Y. & Lee T. S. (2009). Accelerometer based motion noise analysis of ECG signal, *World Congress on Medical Physics and Biomedical Engineering*, , IFMBE Proceedings, Vol. 25/5, pp. 198-201, Munich, Germany, September 2009
- Lee, Ju-Won and Lee, Gun-Ki (2005) Design of an Adaptive Filter with a Dynamic Structure for ECG Signal Processing, *International Journal of Control, Automation, and Systems*, vol. 3, no. 1, (March 2005), pp. 137-142,
- Lisheng, X., David Z., & Kuanquan W. (2005). Wavelet-Based Cascaded Adaptive Filter for Removing Baseline Drift in Pulse Waveforms, *IEEE Transactions on Biomedical Engineering*, Vol. 52, No. 11, (November 2005), pp. 1973-1975, ISSN 0018-9294
- Murugesan M. & Sukanesh R. (2009) Towards Detection of Brain Tumor in Electroencephalogram Signals Using Support Vector Machines, *International Journal of Computer Theory and Engineering*, Vol. 1, No. 5, (December, 2009), pp. 1793-8201
- Murugesan, M. & Sukanesh, R. (2009). Automated Detection of Brain Tumor in EEG Signals Using Artificial Neural Networks, *Int. Conf. on Advances in Computing, Control, and Telecommunication Technologies*, pp. 284 - 288, Trivandrum, India, December 2009

- Oehler Martin Johannes (2009) Kapazitive Elektroden zur Messung bioelektrischer Signale, *Technischen Universität Carolo-Wilhelmina zu Braunschweig*, Dissertation 2009
Available from: http://rzbl04.biblio.etc.tu-bs.de:8080/docportal/receiv/DocPortal_document_00031116
- Ortolan, RL., Mori, RN. Pereira, RR., Cabral, CM., Pereira, JC. & Cliquet, AJ. (2003) Evaluation of adaptive/nonadaptive filtering and wavelet transform techniques for noise reduction in EMG mobile acquisition equipment, *Neural Systems and Rehabilitation Engineering, IEEE Transactions Vol. 11, No. 1, (March 2003)* pp. 60 – 69, ISSN 1534-4320
- Pandey Vinod, K. & Pandey Prem, C. (2007). Wavelet based cancellation of respiratory artifacts in impedance cardiography, *IEEE Intl. Conf. on Digital Signal Processing*, Cardiff, Wales, UK, July 2007
- Philips Healthcare: ICG Impedance Cardiography, Non-invasive hemodynamic measurements, http://www.healthcare.philips.com/main/products/patient_monitoring/products/icg/
- Prasad D.V. & Swarnalatha R. (2009) A New Method of Extraction of FECG from Abdominal Signal, *Int. Conf. On Biomedical Engineering, IFMBE Proceedings*, Vol. 23, pp. 98-100, Singapore, December 2008
- Rasheed Tahir, Ho In Myung, Lee, Young-Koo, Lee Sungyoung, Lee Soo Yeol & Kim Tae-Seong (2006). Constrained ICA Based Ballistocardiogram and Electro-Oculogram Artifacts Removal from Visual Evoked Potential, *EEG Signals Measured Inside MRI, Lecture Notes in Computer Science*, Vol. 4232, 2006, pp. 1088-1097,
- Rik Vullings, Chris Peters, Massimo Mischi, Rob Sluijter, Guid Oei, & Jan Bergmans (2007) Artifact reduction in maternal abdominal ECG recordings for fetal ECG estimation, *Proceedings of the 29th Annual International Conference of the IEEE EMBS*, Lyon, France, August 2007
- Rosell, Javier., Cohen Kevin P. & Webster John G. (1995) Reduction of motion artifacts using a two-frequency impedance plethysmograph and adaptive filtering, *IEEE Transactions On Biomedical, Engineering*, Vol. 42, No. 10, (October 1995), pp.1044-148, ISSN 0018-9294
- Sudha S., Suresh G. R. & Sukanesh R. (2009) Speckle Noise Reduction in Ultrasound Images by Wavelet Thresholding based on Weighted Variance, *International Journal of Computer Theory and Engineering*, Vol. 1, No. 1, April 2009, pp. 1793-8201,
- Rasheed T., Young-Koo L., Soo LY. and Kim TS. (2009). Attenuation of artifacts in EEG signals measured inside an MRI scanner using constrained independent component analysis, *Physiol. Meas.*, Vol. 30, No. 4, April 2009, pp. 387-404
- Volmer Achim, Orglmeister Reinhold & Feese Sebastian. (2010). Motion Artifact Compensation for Photoplethysmographic Signals by Help of Adaptive Noise Cancellation Motion, *Automatisierungstechnik*, Vol. 58, No. 5, May 2010, pp. 269-276
- Xiaoxia Li, Gang Li, Ling Lin, Yuliang Liu, Yan Wang & Yunfeng Zhang. (2004). Application of a Wavelet Adaptive Filter Based on Neural Network to Minimize Distortion of the Pulsatile Spectrum, *Advances in Neural Networks Lecture Notes in Computer Science*, Vol. 3174, 2004, pp. 279-301, ISBN 2004
- Xinsheng Yu, Don Dent & Colin Osborn. (1996). Classification of Ballistocardiography using Wavelet Transform and Neural Networks, *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 3, October 1996, pp. 937 – 938, ISBN 07803-38111
- Yan, YS., Poon CC., & Zhang, YT. (2005). Reduction of motion artifact in pulse oximetry by smoothed pseudo Wigner-Ville distribution, *Journal of NeuroEngineering and Rehabilitation*, 2 : 3, March 2005,

Noise Removal from EEG Signals in Polisomnographic Records Applying Adaptive Filters in Cascade

M. Agustina Garcés Correa and Eric Laciár Leber

*Gabinete de Tecnología Médica, Facultad de Ingeniería, Universidad Nacional de San Juan
Argentina*

1. Introduction

Polisomnography (PSG) is the standard technique used to study the sleep dynamic and to identify sleep disorders. In order to obtain an integrated knowledge of different corporal functions during sleep, a PSG study must perform the acquisition of several biological signals during one or more nights in a sleep laboratory. The signals usually acquired in a PSG study include the electroencephalogram (EEG), the electrocardiogram (ECG), the electromiogram (EMG), the electro oculogram (EOG), the abdominal and thoracic breathings, the blood pressure, the oxygen saturation, the oro-nasal airflow and others biomedical records (Collop et. al., 2007).

Particularly, the EEG is usually analyzed by physicians in order to detect neural rhythms during sleep. However, it is generally contaminated with different noise sources and mixed with other biological signals. Their common artifacts sources are the power line interference (50 or 60 Hz), the ECG and EOG signals. Figure 1 shows an example of real EEG ECG and EOG signals recorded simultaneously in a PSG study. It can be seen that EEG signal is contaminated by the QRS cardiac complexes which appear as spikes at the same time in ECG record. Likewise, the low frequencies present in the contaminated EEG correspond to the opening, closing or movements of the eyes recorded in EOG signal. These noise sources increase the difficulty in analyzing the EEG and obtaining clinical information.

To correct, or remove the artifacts from the EEG signal, many techniques have been developed in both, time and frequency domains (Delorme et. al., 2007; Sadasivan & Narayana, 1995). More recently, component-based techniques, such as principal component analysis (PCA) and independent component analysis (ICA); (Akhtar et. al., 2010; Astolfi et. al., 2010; Jung et. al., 2000), have also been proposed to remove the ocular artifacts from the EEG. The use of Blind Source Separation (BSS) (De Clercq et. al., 2005) and Parallel Factor Analysis (PFA) methods to remove artifacts from the EEG have been used in this area too (Cichocki & Amari, 2002; Makeig et. al., 2004). Wavelet Transform (WT) (Senthil Kumar et. al., 2009), WT combined with ICA (Ghandeharion et. al., 2009) and Autoregressive Moving Average Exogenous (ARMAX) (Hass et. al., 2003; Park et. al., 1998), have been applied too, to remove artifacts from EEG.

In this chapter, it is described a cascade of three adaptive filters based on a Least Mean Squares (LMS) algorithm to remove the common noise components present in the EEG signal recorded in polysomnographic studies.

Adaptive filters method has been used, among other applications, in external electroenterogram records (Mejia-García et. al., 2003) and in impedance cardiography (Pandey et. al., 2005). Other applications of this filtering technique in biomedical signals include, for example, removal of maternal ECG in fetal ECG records (Soria et. al., 1999) detection of ventricular fibrillation and tachycardia (Tompkins, 1993), cancellation of heart sound interference in tracheal sounds (Cortés, 2006), for pulse wave filter (Shen et. al., 2010), for tumor motion prediction (Huang et. al., 2010), detection of single sweep event related potential in EEG records (Decostre et. al., 2005), detection of SSVEP in EEG signals (Diez et. al., 2011) and for motor imagery (Jeyabalan et. al., 2007).

In the particular case of artifacts removal in EEG records, He et. al. (2007) studied the accuracy of adaptive filtering method quantitatively using simulated data and compared it with the accuracy of the domain regression for filtering ocular artifacts from EEG records. Their results show that the adaptive filtering method is more accurate in recovering the true EEG signals. Kumar et. al. (2009) shows that adaptive filtering can be applied to remove ocular artifacts from EEG with good results. Adaptive filters have been used to remove biological artifacts from EEG by others authors (Chan et. al., 1998; Karjalainen et. al., 1999; Kong et. al., 2001).

In order to improve the signal to noise ratio of EEG signals in PSG studies, we had proposed in a previous work a cascade of three adaptive filters based on a LMS algorithm (Garcés et. al., 2007). The first filter in the cascade eliminates line interference, the second adaptive filter removes the ECG complexes and the last one cancels EOG artifacts. Each stage uses a Finite Impulse Response (FIR) filter, which adjusts its coefficients to produce an output similar to the artifacts present in the EEG. In this chapter, we explain in detail the operation of the cascade of adaptive filters including novel tests to determinate the optimal order of FIR filter for each stage. Finally, we describe the results of the proposed filtering scheme in 18 real EEG records acquired in PSG studies.

2. Materials

Eighteen PSG records belonging to sixteen subjects were selected from the MIT-BIH Polysomnographic Database. All subjects are aged 44 +/- 12 years. This database contains

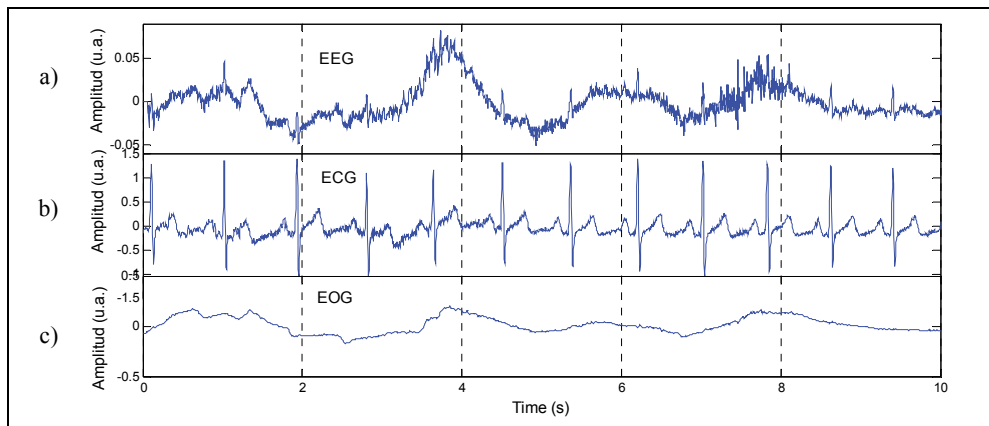


Fig. 1. Some biological signals acquired in a PSG study a) EEG recording (corresponding to Patient 41) corrupted with ECG and EOG artifacts, b) Real ECG signal, and c) Real EOG signal.

over 80 hours of four-, six-, and seven-channel PSG recordings. All of them contain EEG, ECG and Blood Pressure (BP) signals, some of them have Nasal or Plethysmograph Respiratory signals, five of them have O₂ Saturation signal, EOG and EMG signals. All the subjects have ECG signals annotated beat-by-beat, and EEG and respiration signals annotated by an expert with respect to sleep stages and apnea (Goldberger et. al., 2000). In this work were used only the EEG, ECG and EOG signals, all of them were sampled at 250 Hz.

3. Common artifacts in EEG records

By artifacts it is understood all signals that appear in the EEG record which don't come from the brain. The most common artifacts in the EEG signal appear during the acquisition due to different causes, like as bad electrodes location, not clean hairy leather, electrodes impedance, etc. There is also a finding of physiological artifacts, that is, bioelectrical signals from other parts of the body (heart and muscle activity, eye blink and eyeball movement) that are registered in the EEG (Sörnmo & Laguna, 2005).

The problem of those artifacts is that they can made a mistake in the analysis of a EEG record, either in automatic method or in visual inspection by specialist (Wang et. al., 2008).

3.1 Power line interference

Biological records, especially EEG signals, are often contaminated with the 50 or 60 Hz line frequency interference from wires, light fluorescents and other equipments which are captured by the electrodes and acquisition system. The ignition of light of fluorescents usually causes artificial spikes in the EEG. They are distributed in several channels of EEG and can made a mistake in the analysis of the record (Sanei & Chambers, 2007)

3.2 Ocular artifacts

The human eye generates an electrical dipole caused by a positive cornea and negative retina. Eye movements and blinks change the dipole causing an electrical signal known as an EOG. The shape of the EOG waveform depends on factors such as the direction of eye movements. A fraction of the EOG spreads across the scalp and it is superimposed on the EEG (Vigon et. al., 2000).

Two kinds of ocular artifacts can be observed in EEG records, eye blinks and eye movements. Eye blinks are represented by a low frequency signal (< 4 Hz) with high amplitude. It is a symmetrical activity mainly located on the front electrodes (FP1, FP2) with low propagation. Eye movements are also represented by a low frequency signal (< 4 Hz) but with higher propagation, (Crespel et. al., 2006). In order for the EEG to be interpreted for clinical use, those artifacts need to be removed or filtered from the EEG.

3.3 Cardiac artifacts

Cardiac activity may have pronounced effects on the electroencephalogram (EEG) because of its relatively high electrical energy, especially upon the no-cephalic reference recordings of EEG. The QRS complexes appear in the EEG signal like regular spikes (Sörnmo & Laguna, 2005). In figure 1 it can be observed the QRS complex present in a segment of EEG record. The QRS amplitudes in the ECG are of the order of mV, but in the external EEG they have been reduced. These artifacts in the EEG records could be clinically misleading.

3.4 Other artifacts

The muscle disturbances are introduced in the EEG by involuntary muscle contractions of the patient, thus generating an electromyogram (EMG) signal present in the EEG record. The EMG and other biological artifacts have not been analyzing in the present work.

4. Methodology

Herein, we propose the use of adaptive filters to remove artifacts from EEG signal acquired in PSG studies. Usually, biological signals (ECG, EOG and others) have overlapped spectra with the EEG signal. For that, conventional filtering (band-pass, lower-pass or high-pass filters) cannot be applied to eliminate or attenuate the artifacts without losing significant frequency components of EEG signal.

Due to this reason, it is necessary to design specific filters to attenuate artifacts of EEG signals in PSG studies. The adaptive interference cancellation scheme is a very efficient method to solve the problem when signals and interferences have overlapping spectra.

Since the PSG recordings usually contain the ECG, EOG and EEG signals it is very convenient to apply this method to filter this kind of records.

4.1 Adaptive filter

Adaptive filters are based on the optimization theory and they have the capability of modifying their properties according to selected features of the signals being analyzed (Haykin, 2005). Figure 2 illustrates the structure of an adaptive filter. There is a primary signal $d(n)$ and a secondary signal $x(n)$. The linear filter $H(z)$ produces an output $y(n)$, which is subtracted from $d(n)$ to compute an error $e(n)$.

The objective of an adaptive filter is to change (adapt) the coefficients of the linear filter, and hence its frequency response, to generate a signal similar to the noise present in the signal to be filtered. The adaptive process involves minimization of a cost function, which is used to determine the filter coefficients. Initially, the adaptive filter adjusts its coefficients to minimize the squared error between its output and a primary signal. In stationary conditions, the filter should converge to the Wiener solution. Conversely, in non-stationary circumstances, the coefficients will change with time, according to the signal variation, thus converging to an optimum filter (Decostre & Arslan, 2005).

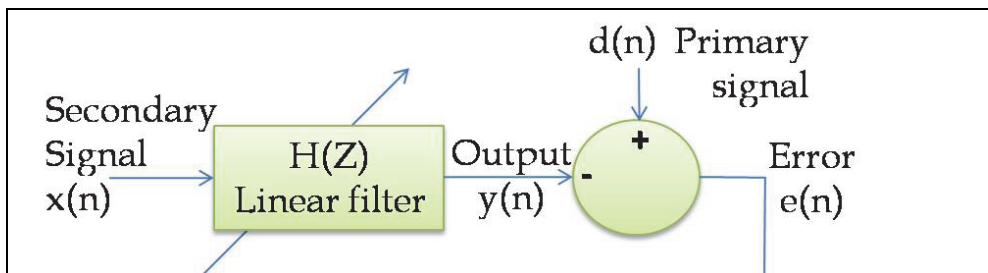


Fig. 2. Structure of an adaptive filter.

In an adaptive filter, there are basically two processes:

- a. A filtering process, in which an output signal is the response of a digital filter. Usually, FIR filters are used in this process because they are linear, simple and stable.

- b. An adaptive process, in which the transfer function $H(z)$ is adjusted according to an optimizing algorithm. The adaptation is directed by the error signal between the primary signal and the filter output. The most used optimizing criterion is the Least Mean Square (LMS) algorithm.

The structure of the FIR can be represented as,

$$y(n) = \sum_{k=0}^L w_k x(n-k) \quad (1)$$

where L is the order of the filter, $x(n)$ is the secondary input signal, w_k are the filter coefficients and $y(n)$ is the filter output.

The error signal $e(n)$ is defined as the difference between the primary signal $d(n)$ and the filter output $y(n)$, that is,

$$e(n) = d(n) - y(n) \quad (2)$$

where,

$$e(n) = d(n) - \sum_{k=0}^L w_k x(n-k) \quad (3)$$

The squared error is,

$$e^2(n) = d^2(n) - 2d(n) \sum_{k=0}^L w_k x(n-k) + \left[\sum_{k=0}^L w_k x(n-k) \right]^2 \quad (4)$$

The squared error expectation for N samples is given by

$$\zeta = E[e^2(n)] = \sum_{k=0}^N e^2(n) \quad (5)$$

$$\zeta = \sum_{n=1}^N \left[d^2(n) \right] - 2 \sum_{k=0}^L w_k r_{dx}(n) + \sum_{k=0}^L \sum_{l=0}^L w_k w_l r_{xx}(k-l) \quad (6)$$

where $r_{dx}(n)$ and $r_{xx}(n)$ are, respectively, the cross-correlation function between the primary and secondary input signals, and the autocorrelation function of the secondary input, that is

$$r_{dx}(n) = \sum_{n=1}^N d(n)x(n-k) \quad (7)$$

$$r_{xx}(n) = \sum_{n=1}^N x(n)x(n-k) \quad (8)$$

The objective of the adaptation process is to minimize the squared error, which describes a performance surface. To get this goal there are different optimization techniques. In this work, we used the method of steepest descent (Semmlow, 2004). With this, it is possible to calculate the filter coefficient vector for each iteration k having information about the previous coefficients and gradient, multiplied by a constant, that is,

$$w_k(n+1) = w_k(n) + \mu(-\nabla_k) \quad (9)$$

where μ is a coefficient that controls the rate of adaptation.

The gradient is defined as,

$$\nabla_k = \frac{\partial \{e^2(n)\}}{\partial w_k(n)} \quad (10)$$

Substituting (10) in (9) leads to,

$$w_k(n+1) = w_k(n) - \mu \frac{\partial \{e^2(n)\}}{\partial w_k(n)} \quad (11)$$

Deriving with respect to w_k and replacing leads to,

$$w_k(n+1) = w_k(n) - 2\mu e(n) \frac{\partial \{e(n)\}}{\partial w_k(n)} \quad (12)$$

$$w_k(n+1) = w_k(n) - 2\mu e(n) \frac{\partial \left\{ d(n) - \sum_{k=0}^L w_k x(n-k) \right\}}{\partial w_k(n)} \quad (13)$$

Since $d(n)$ and $x(n)$ are independent with respect to w_k , then,

$$w_k(n+1) = w_k(n) - 2\mu e(n)x(n-k) \quad (14)$$

Equation (14) is the final description of the algorithm to compute the filter coefficients as function of the signal error $e(n)$ and the reference input signal $x(n)$. The coefficient μ is a constant that must be chosen for quick adaptation without losing stability. The filter is stable if μ satisfies the following condition, (Sanei & Chambers, 2007).

$$0 < \mu < \frac{1}{10L.P_{xx}}; P_{xx} \approx \frac{1}{M+1} \sum_{n=0}^{M-1} x^2(n) \quad (15)$$

where L is the filter order and P_{xx} is the total power of the input signal.

4.2 Artifacts removal from EEG

As it is mentioned above, the adaptive interference cancellation is a very efficient method to solve the problem when signals and interferences have overlap spectra.

The adaptive noise canceller scheme is arranged on the basic structure showed in Figure 2, where the primary and secondary inputs are called as "corrupted signal" and "reference signal", respectively.

In this scheme, it is assumed that the corrupted signal $d(n)$ is composed of the desired $s(n)$ and noise $n_0(n)$, which is additive and not correlated with $s(n)$. Likewise, it is supposed that the reference $x(n)$ is uncorrelated with $s(n)$ and correlated with $n_0(n)$. The reference $x(n)$ feeds the filter to produce an output $y(n)$ that is a close estimate of $n_0(n)$ (Tompkins, 1993).

To remove the main artifacts of the EEG signal, we propose a cascade of three adaptive filters (see Figure 3). The input $d_1(n)$ in the first stage is the EEG corrupted with artifacts (EEG + line-frequency + ECG + EOG). The reference $x_1(n)$ in the first stage is an artificial sine function generated with 50 Hz (or 60 Hz, depends on line frequency). The output of $H_1(z)$ is $y_1(n)$, which is an estimation of the line artifacts present in the EEG. This signal $y_1(n)$ is subtracted from the corrupted $d_1(n)$ to produce the error $e_1(n)$, which is the EEG without line-interference. The $e_1(n)$ error is forwarded as the corrupted input signal $d_2(n)$ to the second stage. The reference input $x_2(n)$ of the second stage can be either a real or artificial ECG. The output of $H_2(z)$ is $y_2(n)$, representing a good estimate of the ECG artifacts present in the EEG record. Signal $y_2(n)$ is subtracted from $d_2(n)$; its result produces error $e_2(n)$. Thus, we have obtained the EEG without line and ECG artifacts. Then, $e_2(n)$ enters into the third stage as the signal $d_3(n)$. The reference input $x_3(n)$ of filter $H_3(z)$ is also a real or artificial EOG and its output is $y_3(n)$, which is a replica of the EOG artifacts present in the EEG record. Such $y_3(n)$, subtracted from $d_3(n)$, gives error $e_3(n)$. It is the final output of the cascade filter, that is, the clean EEG without artifacts.

The reference signals ECG and EOG and the corrupted EEG were acquired simultaneously in polysomnographic studies. EEG, ECG and EOG records belonged to adult patients and were downloaded from the MIT-BIH Polysomnographic Databas-Physiobank (Goldberger et. al., 2000).

In section 4.3 there are present the tests that were carried out to determine the optimum order of $H_1(z)$, $H_2(z)$ and $H_3(z)$.

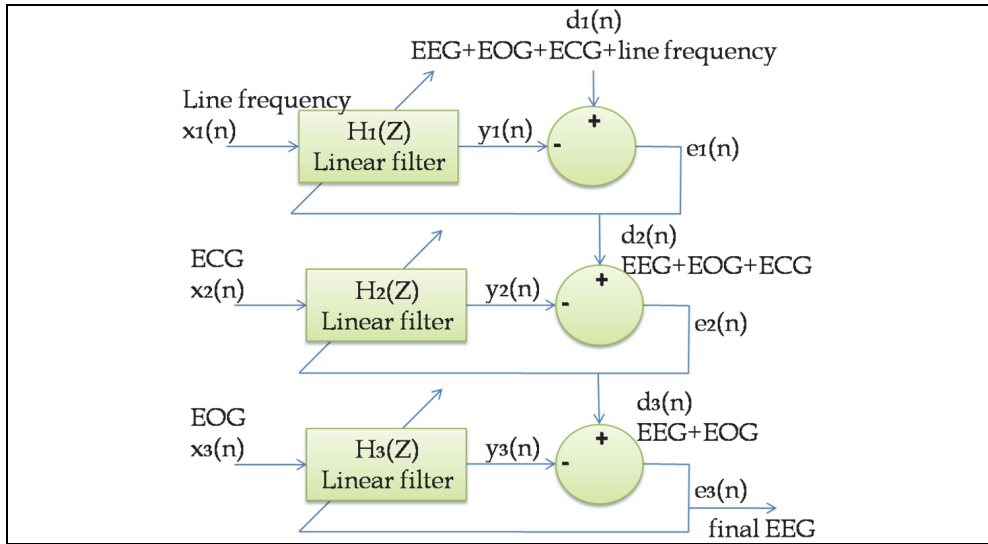


Fig. 3. Structure of adaptive filters cascade for artifacts removal on EEG signal acquired in PSG studies.

4.3 Optimal order of FIR filters

To determine the optimum values of the orders L_1 , L_2 and L_3 of $H_1(z)$, $H_2(z)$ and $H_3(z)$ filters the EEG signal were artificially contaminated with different coloured noises. The test to

determinate the optimum values of the orders L_1 , L_2 and L_3 was done with a coefficient convergence rates μ fixed in 0.001. As soon as the optimum value of the L of each stage was obtained the coefficient convergence rates μ of each stage was recalculated with Eq. (15) to assure an adequate adaptation. If μ is too big, the filter becomes unstable, and if it is too small, the adaptation may turn out too slow.

The tests were done using one stage of adaptive filter per time without using the cascade of three filters.

4.3.1 Optimal estimation of order L_1 for filter $H_1(z)$.

The first stage filter attenuates the line frequency and was used to determinate the optimum value L_1 of $H_1(z)$. To determinate L_1 , the EEG was artificially contaminated with a sinusoidal signal of 50 Hz which amplitude is adjusted in 30%, 50%, 80% and 100% of the Root Mean Square (RMS) value of original EEG signal. Then, the filter order L_1 was adjusted with different values of 8, 16, 32, 64 and 128.

In order to study the filter performance, we estimated the Power Spectral Density (PSD) of the original real EEG signal, the contaminated EEG and the different filtered versions of the EEG signal. PSD was computed using the Burg method with a model order equal to 12. Those graphics for one patient are presented in Figure 4 as an example.

Then, we estimated the normalized area below the frequency coherence function and the maximum of temporal cross-correlation normalized function between the filtered EEG signals and the contaminated EEG. If the signals are identical these parameters must be equal to 1. This test was done for each patient.

Table 1 show the averaged values of two parameters for all EEG records of the database.

Contamination of line frequency	L_1	Coherence	Cross- correlation
30%	8	0.9943	0.9760
	16	0.9940	0.9727
	32	0.9947	0.9657
	64	0.9939	0.9497
	128	0.9912	0.9062
50%	8	0.9936	0.9426
	16	0.9932	0.9393
	32	0.9938	0.9326
	64	0.9930	0.9171
	128	0.9902	0.8751
80%	8	0.9918	0.8739
	16	0.9914	0.8706
	32	0.9919	0.8643
	64	0.9909	0.8500
	128	0.9879	0.8111
100%	8	0.9903	0.8223
	16	0.9898	0.8191
	32	0.9901	0.8131
	64	0.9890	0.7996
	128	0.9859	0.7631

Table 1. Average values of the normalized parameters between filtered EEG signal and contaminated EEG signal with line interference for different values of L_1 .

Figure 4 is an example of PSD graphics for a EEG recording (corresponding to Patient 48) but all records of the database have a similar behaviour in the test. In this figure it could be observed that as L_1 increases, the attenuation of the 50 Hz interference is more significant. However, if L_1 is higher than 32, it can be seen than other frequencies of spectrum are modified.

For this reason, there is a tradeoff between the 50 Hz interference attenuation and the modification of the main frequency components of EEG signal.

In table 1 it can be observed that the best option between $L_1=8$, $L_1=16$ and $L_1=32$ is $L_1=16$, because it have the minimum area of coherence and similar values of maximum in cross-correlation with $L_1=32$. Chosen this value of the order L_1 there is a loss of information of original signal and there is not a modification in the rest of the spectrum.

It is concluded that the optimum value of L_1 for the first filter is $L_1=16$ (for a sampling frequency of 250 Hz). For this order, the optimum value of the coefficient convergence rates μ calculated with Eq. (15) must be positive and lower than 0.047

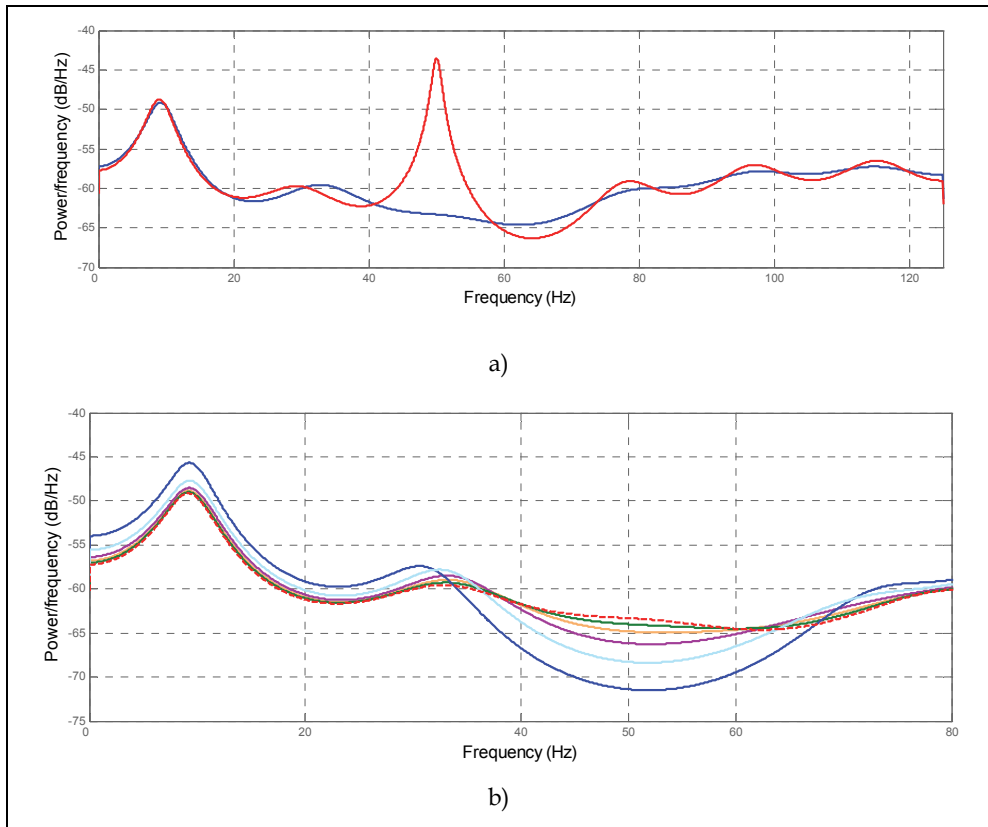


Fig. 4. Power Spectral Density (PSD) of a EEG signal before and after the first adaptive filter $H_1(z)$. a) In blue: PSD of original EEG, in red: PSD of EEG signal contaminated with an artificial line interference. b) PSD of filtered EEG signal for different values of the order L_1 . Red: original EEG, Green: $L_1=8$, Orange: $L_1=16$, Purple: $L_1=32$, Light Blue: $L_1=64$, Blue: $L_1=128$.

4.3.2 Optimal estimation of order L_2 for filter $H_2(z)$.

The second stage filter attenuates ECG artifacts (mainly QRS complexes) present in EEG signal, and was used to determinate the optimum value of the order L_2 of $H_2(z)$. To determinate L_2 , the EEG was artificially contaminated with a coloured noise, with a -3dB bandwidth between 5 Hz and 40 Hz. This bandwidth was selected considering that QRS complexes have almost their total energy in this frequency band (Thakor, 1984). Then, the filter order L_2 was adjusted with the different values of 16, 32, 64, 128, 256 and 512.

As a similar way to optimum value estimation of L_1 , we estimated the PSD of the original real EEG signal, the contaminated EEG and the different filtered versions of the EEG signal. Figure 5 shows the PSD graphics for an EEG recording before and after the second adaptive filter. In this figure it could be observed that the possible optimum values of L_2 to filter the cardiac frequencies between 5Hz and 40Hz are $L_2=16$, $L_2=32$ or $L_2=64$, because the rest of the values of L_2 modify the frequencies of the entire spectrum.

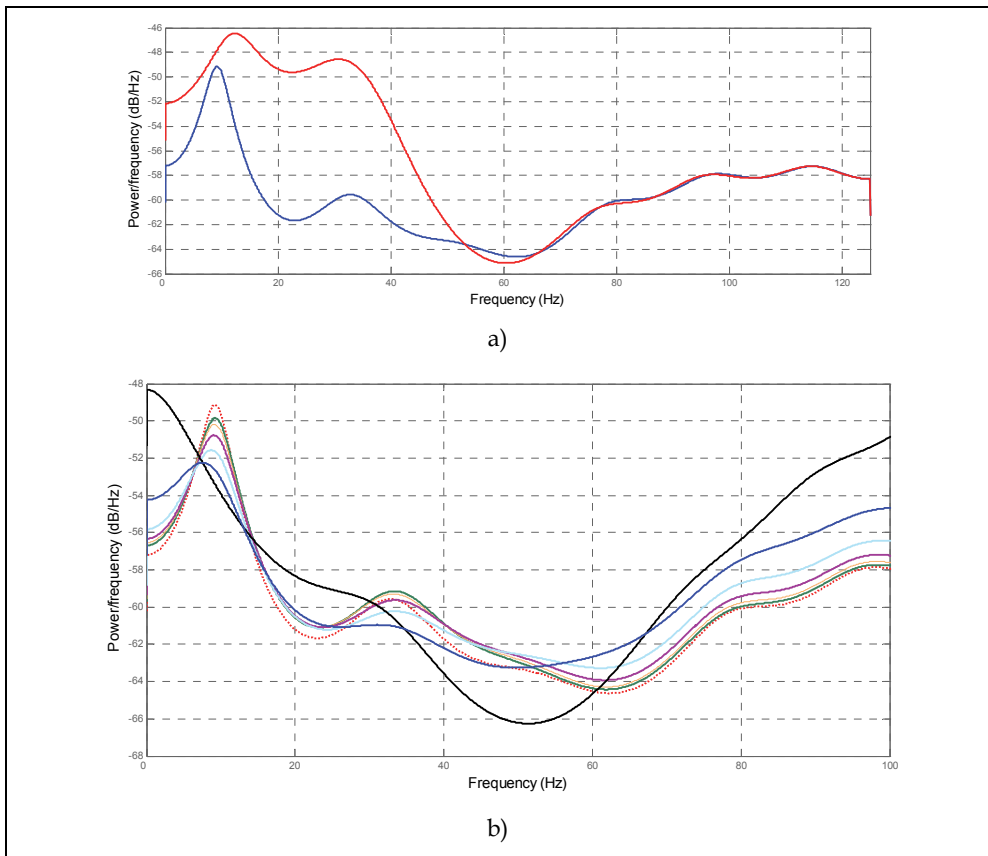


Fig. 5. Power Spectral Density (PSD) of a EEG signal before and after the second adaptive filter $H_2(z)$. a) In blue: PSD of original EEG, in red: PSD of EEG signal contaminated with coloured noise (5Hz to 40 Hz). b) PSD of filtered EEG for different values of the order L_2 . Red: original EEG, Green: $L_2=16$, Orange: $L_2=32$, Purple: $L_2=64$, Light Blue: $L_2=128$, Blue: $L_2=256$, Black: $L_2=512$.

Table 2 shows the average of the normalized area below the frequency coherence function and the maximum of temporal cross-correlation normalized function (between the filtered EEG signals and the contaminated EEG) for all recordings analyzed and for different values of L_2 .

L_2	Coherence	Cross-correlation
16	0.2588	0.5686
32	0.2596	0.5595
64	0.2927	0.5406
128	0.2641	0.5087
256	0.1756	0.4576
512	0.1579	0.3463

Table 2. Average values of the normalized parameters between filtered EEG signal and contaminated EEG signal for different values of L_2 .

In table 2 it can be observed that the best option between $L_2=16$, $L_2=32$ or $L_2=64$ is $L_2=32$, because it have the minimum value of the normalized area below the frequency coherence function and the lower values of maximum of cross- correlation normalized function without losing information and not modifying the spectrum of the original EEG signal. It is concluded that the optimum value of L_2 for second filter is $L_2=32$. For this order, the optimum value of the coefficient convergence rates μ calculated with (15) must be positive and lower than 0.02367.

4.3.3 Optimal estimation of order L_3 for filter $H_3(z)$.

As it is mentioned above, the third and last stage filter attenuates EOG artifacts present in EEG. In this section, we determinate the optimum value of the order L_3 of $H_3(z)$. To determinate it, the EEG was artificially contaminated with a coloured noise with a -3dB bandwidth between 0.5 Hz and 10 Hz. This bandwidth includes the main frequency components of EOG artifacts. Then, we evaluated the filter performance with different L_3 values (4, 8, 16 and 32).

As a similar way to optimum value estimation of L_1 and L_2 , we estimated the PSD of the original real EEG signal, the contaminated EEG and the different filtered versions of the EEG signal.

L_3	Coherence	Cross-correlation
4	0.8773	0.8014
8	0.8586	0.7979
16	0.8579	0.7937
32	0.8584	0.7863
64	0.8586	0.7842

Table 3. Average values of the normalized parameters between filtered EEG signal and contaminated EEG signal for different values of L_3 .

Figures 6 and 7 show the PSD graphics for an EEG recording before and after the third adaptive filter. It can be observed that all the values of the order L_3 chosen have good result

to filter the frequencies lower than 10 Hz (see Figure 6). No one introduce interferences in other frequencies. But with values bigger than 256 it could be observed a distortion in high frequencies and a loss of information of the original signal in low frequencies (see Figure 7). The modification of the high frequencies and the losing of information in low frequencies are shown in figure 7, where there have been filtered the contaminated EEG with values of $L_3=256$ and $L_3=512$

Table 3 shows the averaged values of the normalized area below the frequency coherence function and temporal cross-correlation normalized function (between the filtered EEG signals and the contaminated EEG) for all recordings analyzed and for different values of L_3 .

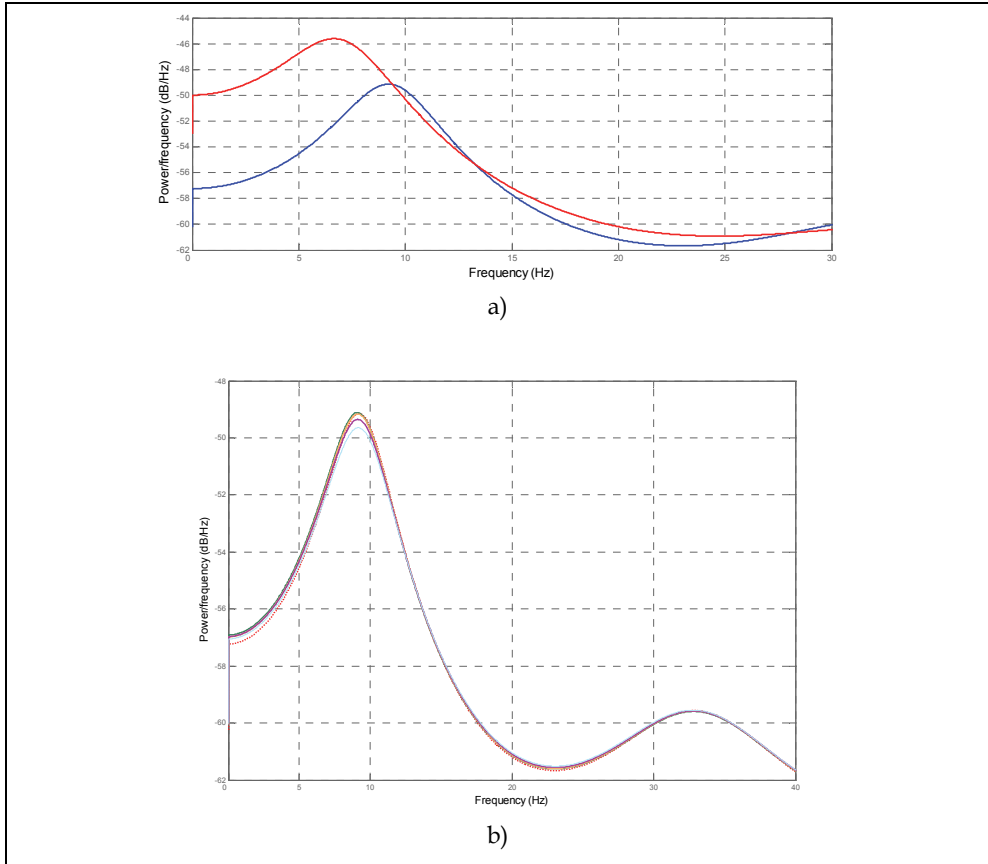


Fig. 6. Power Spectral Density (PSD) of a EEG signal before and after the third adaptive filter $H_3(z)$. a) In blue: PSD of original EEG, in red: PSD of EEG signal contaminated with coloured noise (0.5 Hz to 10 Hz). b) PSD of EEG signal filtered for different values of the order L_3 , Red: original EEG, Green: $L_3=4$, Orange: $L_3=8$, Purple: $L_3=16$, Light Blue: $L_3=32$,

In Table 3 it can be observed that the best option of the value of the order L_3 for the third filter is $L_3=16$, because it have the minimum value of the normalized area below the frequency coherence function and the lower values of maximum of cross- correlation

normalized function without losing information of original signal and not modifying the spectrum of the original EEG. The results of the test using values of L_3 bigger than $L_3=256$ have not been included in Table 3.

It is concluded that the optimum value of L_3 for the third filter is $L_3=32$. For this value, the optimum value of the coefficient convergence rates μ calculated with (15) must be positive and lower than 0.02367.

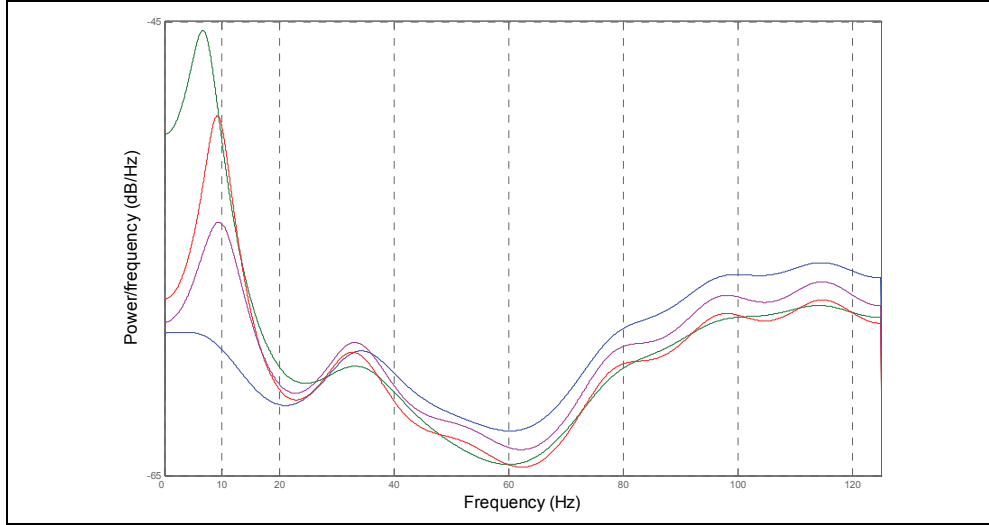


Fig. 7. Power Spectral Density of a EEG signal before and after the third adaptive filter $H_3(z)$. In Red: PSD of the original EEG signal. In Green: PSD of the EEG signal contaminated with coloured noise (0.5 Hz to 10 Hz). In Purple: PSD of the EEG filtered for order $L_3=256$. In Blue: PSD of the EEG filtered for $L_3=512$. Note the modification in high frequencies and losing of information in low frequencies.

5. Results

Eighteen real EEG records acquired in PSG studies were processed with the cascade of adaptive filters. According to the previous tests, the values of the orders L_1 , L_2 and L_3 were adjusted as $L_1=16$, $L_2=32$ and $L_3=32$.

As it was mentioned in section 2, only five subjects from the entire database have EOG signals. So, the EEG signals of these five patients have been filtered with the entire cascade shown in Figure 3. The others thirteen EEG (belonging to the rest of the patients) have not been filtered with the last third stage.

The input $d_1(n)$ in the first stage is the EEG corrupted with artifacts (EEG + line-frequency + ECG + EOG). The reference $x_1(n)$ in the first stage is an artificial sine function generated with 50 Hz with the same RMS of the EEG signal. The $e_1(n)$, which is the EEG without line-interference, is forwarded as the corrupted input signal $d_2(n)$ to the second stage. The reference input $x_2(n)$ of the second stage is the real ECG. The error $e_2(n)$ is the EEG without line and ECG artifacts and enters into the third stage as the signal $d_3(n)$. The reference input

$x_3(n)$ of filter $H_3(z)$ is a real EOG. The error $e_3(n)$ is the final output of the cascade filter, that is, the clean EEG without artifacts.

In order to study the filter performance we estimated the normalized area below the frequency coherence function and the maximum of temporal cross-correlation normalized function between the filtered EEG signals of each stage and the original EEG for the entire data base.

Table 4 shows the results obtained for each record of the database processed by the first stage of the propose filter. In this table, it is presented the values of the normalized area of frequency coherence function and the normalized maximum of temporal cross-correlation between the contaminated signal $d_1(n)$ and the error signal $e_1(n)$. Those values show that the first stage attenuates the line interference.

Patient	Coherence %	Cross-correlation
1a	0.8690	0.6730
1b	0.8901	0.6349
2a	0.9833	0.4724
2b	0.9507	0.5417
3	0.9279	0.4044
4	0.9776	0.3615
14	0.9807	0.4698
16	0.9816	0.4452
32	0.9879	0.8309
37	0.9881	0.9293
41	0.9963	0.9857
45	0.9928	0.7017
48	0.9983	0.9413
59	0.9839	0.3970
60	0.9747	0.2807
61	0.9663	0.4281
66	0.9783	0.4213
67	0.9734	0.5504
average	0.9667	0.5816

Table 4. Normalized area of frequency coherence function and maximum of temporal cross - correlation function between the signals $d_1(n)$ and $e_1(n)$ of the first stage of proposed filter.

Figure 8 illustrates a temporal segment of 10s of the original EEG record (corresponding to Patient 41) and its filtered version after the first stage of adaptive filter. In this figure it can be observed that the 50 Hz power line component is significantly filtered.

Figure 9 shows the PSD function of the same original and filtered EEG signals shown in Figure 8. The PSD of the filtered signal shows that the first stage attenuates the line-frequency artifacts. The $H_1(z)$ filter adapts the amplitude and the phase of the artificial

sinusoidal signal $x_1(n)$ (50Hz) in order to have as output a replica, $y_1(n)$, of the line-frequency artifacts present in the EEG.

After 50 Hz filtering, the EEG is forwarded to the second stage in order to remove ECG artifacts (see Figure 3).

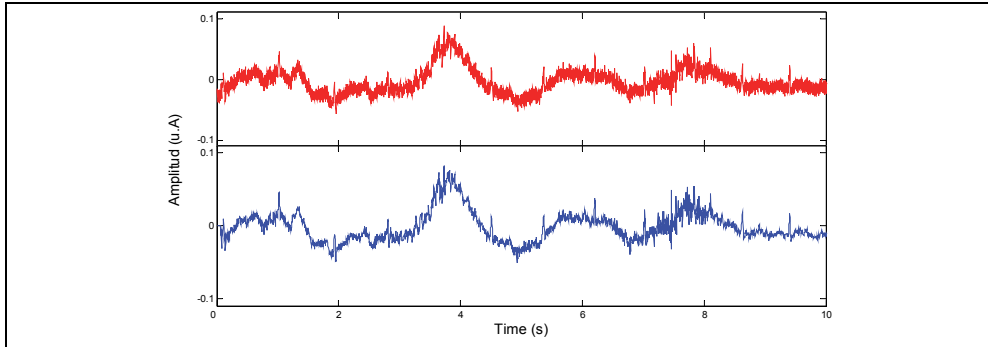


Fig. 8. Example of a temporal segment of EEG filtered with stage 1 for patient 41. a) Red: Original EEG contaminated with 50 Hz power line interference, $d_1(n)$. b) Blue: EEG without line interference, $e_1(n)$.

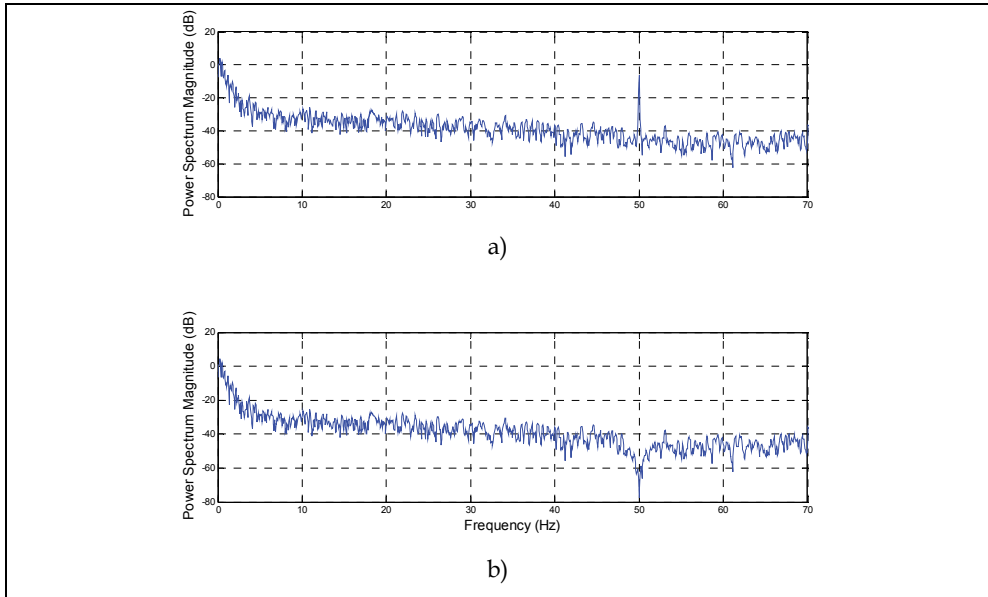


Fig. 9. Example of first stage of the proposed filter. a) PSD of original EEG with artifacts. b) PSD of first stage output $e_1(n)$, where the 50 Hz component is attenuated.

Table 5 shows the results obtained for each record of the database processed by the second stage. In this table, it is presented the values of the normalized area of frequency coherence function and the normalized maximum of temporal cross-correlation between the contaminated signal $d_2(n)$ and the error signal $e_2(n)$. Those values show that the second stage attenuates QRS complexes artifacts introduced by ECG signal.

Patient	Coherence	Cross-correlation
1a	0.8528	0.7514
1b	0.8801	0.5180
2a	0.9709	0.9467
2b	0.9946	0.9845
3	0.9107	0.9460
4	0.9120	0.7910
14	0.9276	0.8768
16	0.9070	0.8757
32	0.8364	0.3333
37	0.8550	0.6725
41	0.8204	0.7826
45	0.7985	0.7981
48	0.9096	0.6893
59	0.9106	0.5431
60	0.8224	0.3027
61	0.8979	0.2482
66	0.8097	0.5319
67	0.8464	0.8209
average	0.8342	0.6439

Table 5. Normalized area of frequency coherence function and maximum of temporal cross - correlation function between the signals $d_2(n)$ and $e_2(n)$ of the second stage of proposed filter.

Figure 10 shows an example of 10s of EEG signal (corresponding to patient 41) processed by the second filter. The contaminated signal $d_2(n)$ is shown in red. It could be observed the presence and morphology similarity of QRS complexes of the ECG (in green) in the EEG record. The output signal $y_2(n)$ of $H_2(z)$ is drawn in black colour, this signal is an estimation of the ECG artifacts present in the EEG. The $H_2(z)$ filter adapts the amplitude and the phase of the reference signal $x_2(n)$ (ECG signal) in order to have as output a replica of the artifacts present in the EEG

After 50 Hz and ECG filtering, the EEG is forwarded to the third stage in order to remove EOG artifacts.

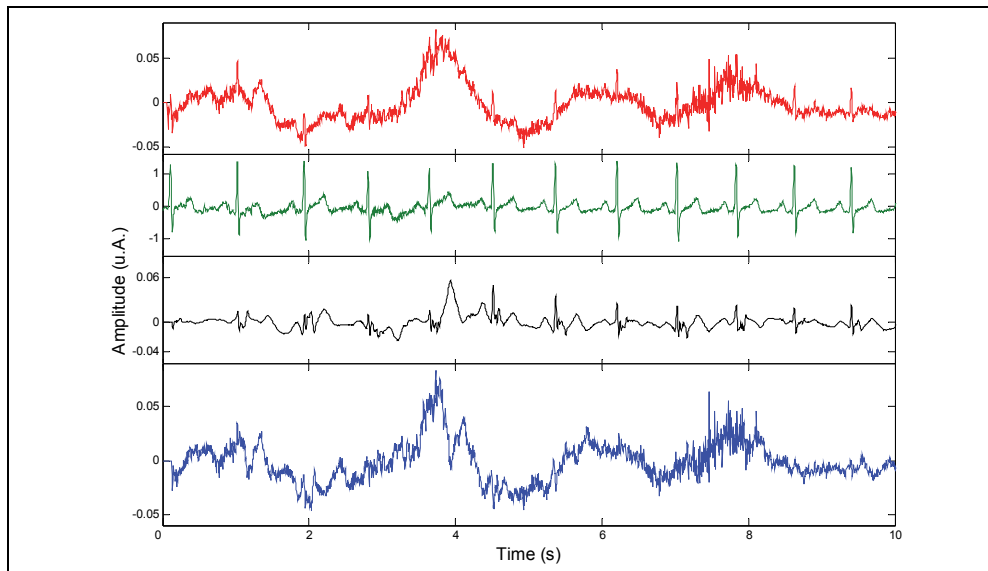


Fig. 10. Example of a temporal segment of EEG filtered with stage 2 for patient 41. In Red: Contaminated EEG, $d_2(n)$. In Green: ECG signal. In Black: output signal from $H_2(z)$, that is $y_2(n)$. In Blue: EEG without ECG artifacts, $e_2(n)$.

Table 6 shows the results obtained for five records of the database processed by the third stage. In this table, it is presented the values of the normalized area of frequency coherence function and the normalized maximum of temporal cross-correlation between the contaminated signal $d_3(n)$ and the error signal $e_3(n)$, which is the final output of the proposed filter. As it has been mentioned before only five patients have been filtered with the third stage, the rest of them do not have the reference signal $x_3(n)$. Those values show that this last stage attenuates artifacts introduced by the EOG.

Patient	Coherence	Cross-correlation
32	0.9985	0.9907
37	0.9912	0.7949
41	0.9859	0.6052
45	0.9990	0.9500
48	0.9527	0.7943
average	0.9855	0.8270

Table 6. Normalized area of frequency coherence function and maximum of temporal cross - correlation function between the signals $d_3(n)$ and $e_3(n)$ of the third stage of proposed filter.* Patient without available EOG signal.

Figure 11 shows the same 10s of temporal EEG signal of patient 41. There it can be observed all signals of third stage. The contaminated signal $d_3(n)$ is drawn in red colour. It can be observed the presence and morphology similarity of the EOG signal in the EEG record. The output signal $y_3(n)$ of $H_2(z)$ is in black colour in the figure, this signal is an estimation of the EOG signal present in the EEG. The $H_3(z)$ filter adapts the amplitude and the phase of the reference signal $x_3(n)$ (EOG signal) in order to have as output a replica of the EOG artifacts present in the EEG.

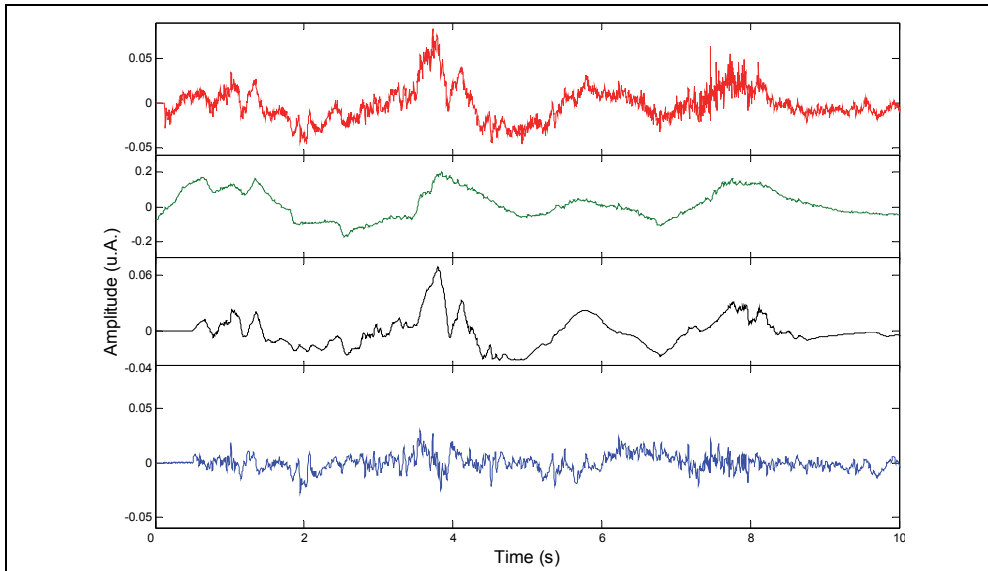


Fig. 11. Example of temporal segment of EEG filtered with stage 3 for patient 41. In Red: Contaminated EEG, $d_3(n)$. In Green: EOG signal. In Black: output signal from $H_3(z)$, that is $y_3(n)$. In Blue: EEG without EOG artifacts, $e_3(n)$.

Figure 12 show the PSD of the contaminated EEG of third stage, $d_3(n)$, of the reference signal $x_3(n)$, EOG, and of the filtered EEG signals illustrated in Fig. 11. Note that the low frequencies of the EOG present in the contaminated EEG are attenuated in the filtered EEG signal.

Figure 13 is shown temporal temporal segments of 10s of EEG. In this figure it could be observed the attenuation of line frequency and biological artifacts without losing important information of the EEG signal. Results show that the proposed adaptive filter cancels correctly the line frequency interference and attenuate very well the biological artifacts introduced by the ECG and the EOG.

6. Discussion and conclusion

In this chapter, a novel filtering method based on three adaptive filters in cascade has been proposed to cancel common artifacts (line interference, ECG and EOG) present in EEG signals recorded in PSG studies.

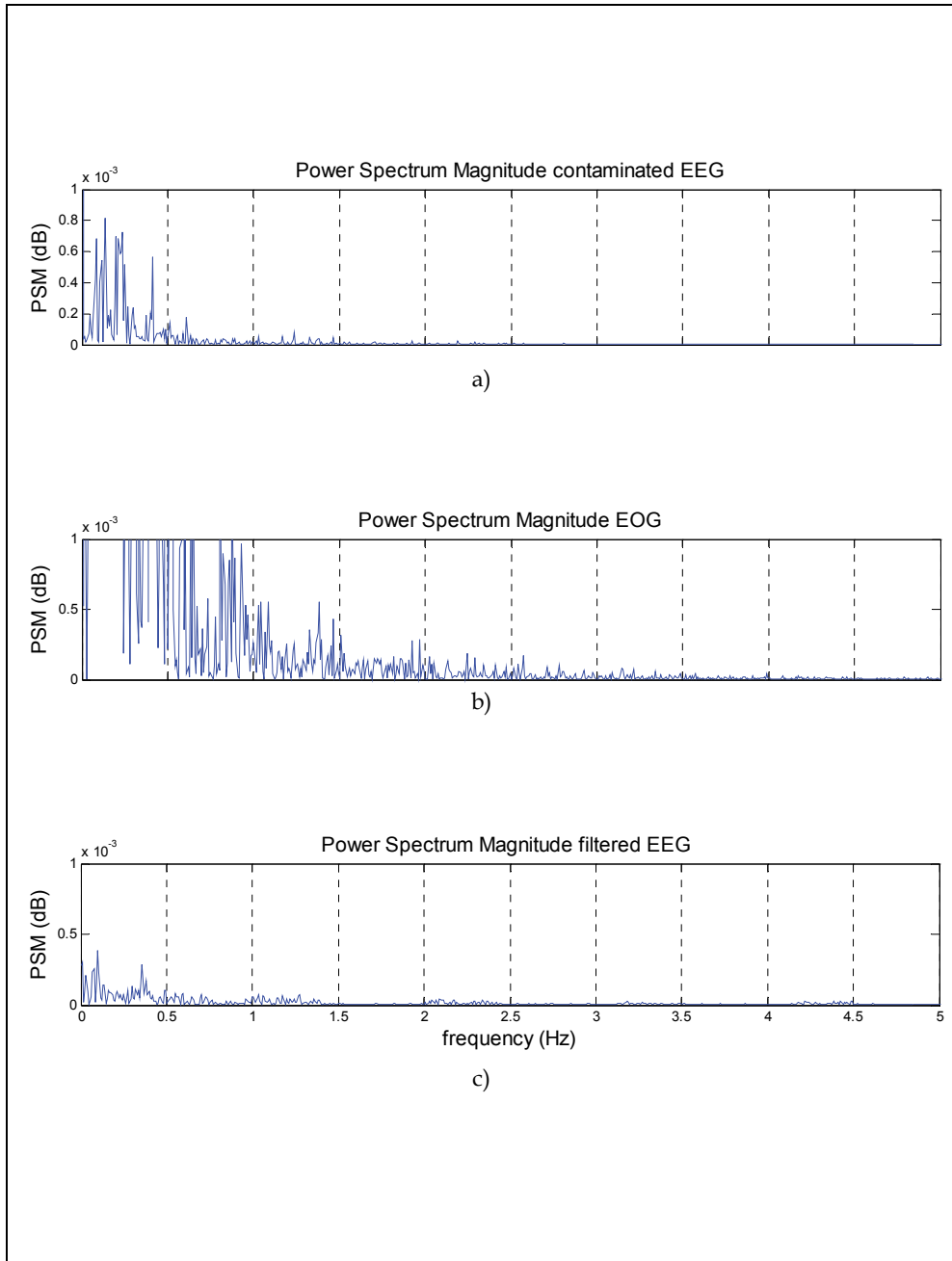


Fig. 12. Example of third stage of the proposed filter a) PSD of the contaminated EEG, $d_3(n)$, b) PSD of the reference signal $x_3(n)$, EOG, c) PSD of the filtered EEG signal.

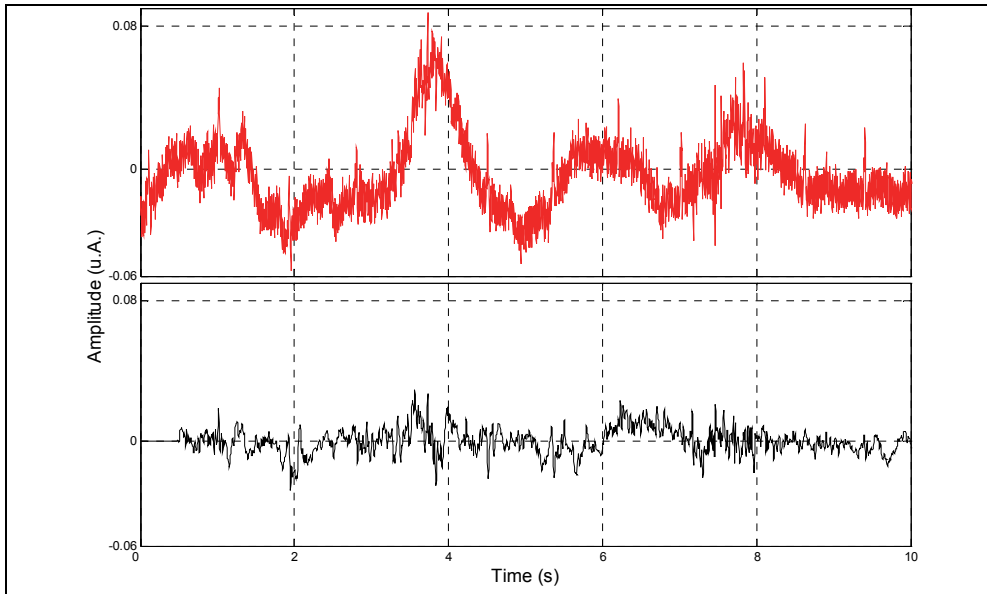


Fig. 13. Example of temporal segments of contaminated EEG and EEG filtered with the entire cascade for patient 41. In Red: Contaminated EEG, $d_1(n)$, In Black: final filtered EEG without line interference, ECG and EOG artifacts, $e_3(n)$.

Other methods (like PCA, ICA, BSS or WT) have been described in the bibliography to cancel these artifacts in the EEG signals. However, those methods have some restrictions. For example, the properties of WT make it has an advantage in processing short-time instantaneous signal, but it needs that the frequency range of the EEG signal was not overlap with the bandwidth of noise sources and in this case the frequencies bands of the ECG and EOG signal are overlap with the frequencies of the EEG. ICA is a developed method for transforming an observed multidimensional vector into components that are statistically as independent from each other as possible. This method needs that the dimension of the signals were larger than that of original signals, and every original signal must be non-Gaussian. With more observed signals ICA will get better filtering result, which limits the application of this technique in few channels EEG recordings.

The main advantages of the proposed adaptive filtering method can be summarized as:

- The method does not have restrictions about the signal to be filtered.
- The implementation of adaptive filtering is very simple and fast and the results can be obtained without complex calculations.
- The filter coefficients can be adapted to variations in heart frequency, abrupt changes in the line frequency (caused, say, by ignition of electric devices) or modifications due to eye movements.
- At each stage output, the error signals $e_i(n)$, EEG with one of the three attenuated artifacts are present; such separation (by artifacts) may be useful in some applications where such output might be enough.
- The filters have a linear phase response so no phase distortion is made. This is particularly important for the analysis of neurological rhythms in EEG signals

As soon as the optimal orders of the three filters were determinate, the method was tested in 18 real EEG records acquired in PSG studies. Figure 13 is a good example of an EEG record corrupted by three types of artifacts and its corresponding filtered version. It can be seen that all artifacts have been eliminated or attenuated, improving the quality of EEG record. The remaining records analyzed in the work had obtained similar results and their filtered EEGs don't have large artifacts.

It has been concluded that proposed adaptive filtering scheme with the appropriate values of order L_i , attenuate correctly ECG, EOG and line interference without removing significant information embedded in EEG signals registered in PSG studies. Due to the fact that these studies usually have the ECG, EOG and EEG signals, the proposed cascade of adaptive filters is very useful and appropriate for the analysis of PSG recordings in sleep laboratories. The cascade could be used in others biomedical applications and in BCI applications.

7. Acknowledgment

This work has been supported by grants from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) and Universidad Nacional de San Juan (UNSJ), both Argentinian institutions.

8. References

- Akhtar, M.T.; James, C.J & Mitsuhashi W. (2010). Modifying the Spatially-Constrained ICA for Efficient Removal of Artifacts from EEG Data. *Proceedings of 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE)*, pp. 1-4. ISBN: 978-1-4244-4713-8, Chengdu, China, June 18-20, 2010.
- Astolfi, L.; Cincotti, F.; Mattia, D.; Babiloni, F.; Marciani, M.G.; De Vico Fallani, F.; Mattiocco, M.; Miwakeichi, F.; Yamaguchi, Y.; Martinez, P.; Salinari, S.; Tocci, A.; Bakardjian, H.; Vialatte, F.B. & Cichocki, A. (2006). Removal of ocular artifacts for high resolution EEG studies: a simulation study, *Proceedings of 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 976 – 979, ISBN 14244-0033-3, New York City, USA, Aug 30-Sept 3, 2006.
- Chan, F.H.Y.; Qiu, W.; lam, F.K. & Poon, P.W.F. (1998). Evoke potential estimation using modified time-sequence adaptive filter, *Medical & Biological Engineering & Computing*, Vol. 36, No. 4, (July 1998) pp. 407-414
- Cichocki, A. & Amari, S. (2002). *Adaptive Blind Signal and Image Processing*, John Wiley & Sons. ISBN 0-471-60791-6.
- Collop, N.A.; Anderson, W.M.; Boehlecke, B.; Claman, D.; Goldberg, R. & Gottlieb, D.J. (2007). Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine. *Journal Clinical Sleep Medicine*. Vol. 3, No. 7, (December 2007), pp.737-47.
- Cortés, S.; Jane, R.; Torres, A.; Fiz, J.A. & Morera, J. (2006). Detection and Adaptive Cancellation of Heart Sound Interference in Tracheal Sounds, *Proceedings of the 28th IEEE EMBS Annual International Conference*, pp. 2860-2863, ISBN 1-4244-0033-3, New York City, USA, Aug 30-Sept 3, 2006.

- Crespel, A.; Gélisse, P.; Bureau, M. & Genton, P. (2005). *Atlas of Electroencephalography* (1st ed), Vol. 1 & 2, J. Libbey Eurotext, ISBN 2-7420-0600-1, Paris.
- De Clercq, W.; Vergult, A.; Vanrumste, B.; Van Hees, J.; Palmini, A.; Van Paesschen, W. & Van Huffel, S. (2005). A new muscle artifact removal technique to improve the interpretation of the ictal scalp electroencephalogram, *Proceedings of the 27th Annual Conference IEEE Engineering in Medicine and Biology*, pp. 944 - 947 ,ISBN: 0-7803-8741-4, Shanghai, China, September 1-4, 2005.
- Decostre, A. & Arslan, B. (2005). An Adaptive Filtering Approach to the Processing of Single Sweep Event Related Potentials Data, *Proceedings 5th International. Workshop Biosignal Interpretation*, pp. 1-3 , Tokyo, Japan, September 6-8, 2010.
- Decostre, A. & Arslan, B. (2005). An Adaptive Filtering Approach to the Processing of Single Sweep Event Related Potentials Data, *Proceeding 5th International. Workshop Biosignal Interpretation*, pp. 1-3 ,Tokyo- Japan, September 2005
- Delorme, A.; Sejnowski T. & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher order statistics and independent component analysis, *NeuroImage*, Vol. 34, (2007), pp. 1443-1449, ISSN: 10538119.
- Diez, P.; Garcés Correa, M.A. & Laciár, E. (2010). SSVEP Detection using Adaptive filters, *V Congreso Latinoamericano de Ingeniería Biomédica (CLAIB2011)*, Habana, Cuba, May 16-21, 2011, In press.
- Garcés Correa, A.; Laciár, E.; Patiño, H.D. & Valentinuzzi, M.E. (2007). Artifact removal from EEG signals using adaptive filters in cascade, *Journal of Physics*, Vol.90, (September 2007), pp.1-10,
- Ghandeharion H. & Ahmadi-Noubari, H. (2009). Detection and Removal of Ocular Artifacts using Independent Component Analysis and Wavelets, *Proceedings of the 4th International IEEE EMBS Conference on Neural Engineering*, pp.653-656, ISBN 978-1-4244-2073-5, Antalya, Turkey, April 29 - May 2, 2009.
- Goldberger, A.L.; Amaral, L.A.; Glass, N L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K. & Stanley H.E. (2000). PhysioBank, PhysioToolkit, andPhysioNet: *Components of a New Research Resource for Complex Physiologic Signals*, *Circulation*, Vol. 101, No.23, pp. 215-220, June 2000
- Hass, S.H.; Frei, M.G.; Osorio, I.; Pasik-Duncan, B. & Radel, J. (2003). EEG ocular artifact removal through ARMAX model system identification using extended least squares, *Communication in formation and system*, Vol. 3, No. 1, (June 2003), pp. 19-40.
- Haykin, S. (2005). *Neural Network* (2nd), Pearson Prentice Hall, ISBN 81-7808-300-0, Delhi, India.
- He, P.; Kahle, M.; Wilson, G. & Russell, C. (2005). Removal of Ocular Artifacts from EEG: A Comparison of Adaptive Filtering Method and Regression Method Using Simulated Data, *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 1110-1113, ISBN 0-7803-8740-6, Shanghai, China, September 1-4, 2005
- Huang, K.; Buzurovic, I.; Yu, Y. & Podder, T.K. (2010). A Comparative Study of a Novel AE-nLMS Filter and Two Traditional Filters inPredicting Respiration Induced Motion

- of the Tumor, *2010 IEEE International Conference on Bioinformatics and Bioengineering*, pp.281-282, ISBN 978-0-7695-4083-2, Philadelphia, May 31-Jun3, 2010.
- Jeyabalan, V.; Samraj, A. & Chu-Kiong, L. (2007). Motor Imaginary Signal Classification Using Adaptive Recursive Bandpass Filter and Adaptive Autoregressive Models for Brain Machine Interface Designs *International Journal of Biological and Life Sciences*, Vol. 3, No.2 (March 2007), pp. 116-123. eISSN 2010-3832.
- Jung, T. P.; Makeig, S.; Westerfield, M.; Townsend, J.; Courchesne, E. & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, Vol.111, No.10, (October 2000), pp. 1745-1758. ISSN 1388-2457.
- Karjalainen, P.; Kaipio, J.; Koistinen, A. & Vauhkonen, M. (1999) Subspace regularization method for the single trial estimation of evoked potentials, *IEEE Transactions on Biomedical Engineering*, Vol. 46, No. 7,(July 1999), pp. 849-860, ISBN 0018-9294.
- Kong, X. & Qiu, T. (2001) Latency change estimation for evoked potentials a comparison of algorithms, *Medical & Biological Engineering & Computing*, Vol. 39, No. 2, (2001), pp. 208-224.
- Makeig S.; Debener S.; Onton J. & Delorme A. (2004). A. Mining event-related brain dynamics. *Trends Cogn Sci.*, pp.204-10., May 2004; Vol.8, No.5.
- Mejia-Garcia, J. H.; Martinez-De-Juan, J. L.; Saiz, J.; Garcia-Casado, J. & Ponce, J.L. (2003) Adaptive cancellation of the ECG interference in external electroenterogram, *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2003*, Vol. 3, pp. 2639-2642, ISBN 0-7803-7789-3, Cancun, Mexico September 17-21, 2003.
- Pandey, V.K. & Pandey,P.C. (2005). Cancellation of Respiratory Artifact in Impedance Cardiography, *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 5503-5506, ISBN 0-7803-8740-6, Shanghai, China, September 1-4, 2005
- Park, H.-J.; Jeong, D.-U. & Park, K.-S. (2002).Automated Detection and Elimination of Periodic ECG Artifacts in EEG Using the Energy Interval Histogram Method. *IEEE transactions on Biomedical Engineering*. Vol. 49, No. 12, (December 2002), pp.1526-1533, ISSN 0018-9294
- Sadasivan, P.K & Narayana, D. (1995). Line interference cancellation from corrupted EEG signals using Modified linear phase FIR digital filters, *Engineering in Medicine and Biology Society, 1995 and 14th Conference of the Biomedical Engineering Society of India. An International Meeting, Proceedings of the First Regional Conference*, pp. 3.35-3.36. India, February 15-18, 1995.
- Sanei, S. & Chambers, J. (2007). *EEG Signal Processing* (1st ed.), Jhon Wiley & Sons, ISBN 13-978-0-470-02581-9, England.
- Semmlow, J.L. (2004). *Biosignal and Medical Image Processing*, Marcel Dekker, ISBN 0-8247-4803-4, United State of America.
- Senthil Kumar, P.; Arumuganathan, R. & Vimal, C. (2009). An Adaptive method to remove ocular artifacts from EEG signals using Wavelet Transform, *Journal of Applied Sciences Research*, Vol. 5, No. 7, (2009), pp. 741-745, ISBN 741-745.

- Shen, Z.; Hu, C. & Meng M.Q.-H. (2010). A Pulse Wave Filter Method Based on Wavelet Transform Soft-threshold and Adaptive Algorithm, *Proceedings of the 8th World Congress on Intelligent Control and Automation*, pp. 1947-1952, ISBN 978-1-4244-6712-9, Jinan, China, July 6-9, 2010.
- Soria, E.; Martínez, M.; Calpe, J.; Guerrero, J.F. & Serrano, A.J. (1999) A new recursive algorithm for extracting the fetal ECG, *Revista Brasileira de Engenharia Biomédica*, Vol. 15, (1999), pp. 135-139. ISBN 1517-3151.
- Sörnmo, L. & Laguna, P. (2005). *Bioelectrical signal processing in cardiac and neurological applications* (1st ed.), Elseiver, ISBN 0-12-437552-9, USA.
- Thakor N.V., Webster J.G., Tompkins W.J. (1984). Estimation of QRS complex power spectrum for design of a QRS filter, *IEEE Transtaction Biomededical. Engenier.*, Vol. 31, No. 11, (1984), pp. 702-706.
- Tompkins W. J. (1993). *Biomedical digital Signal Processing*, Prentice Hall, New Jersey,
- Vigon, L.; Saatchi, M.R.; Mayhew, J.E.W. & Fernandes, R. (2000). Quantitative evaluation of techniques for ocular artifact filtering of EEG waveforms, *IEE Processing Science Meas. Technology*, Vol. 147, No. 5, (September 2000), pp. 219-228.
- Wang, Y.L.; Liu, J.H. & Liu, Y.C. (2008) Automatic Removal of Ocular Artifacts from Electroencephalogram Using Hilbert-Huang Transform, *The 2nd International Conference on Bioinformatics and Biomedical Engineering, ICBBE 2008*, pp. 2138 - 2141, ISBN 978-1-4244-1748-3, Shanghai, China, May 16-18, 2008.

Fast Extraction of Somatosensory Evoked Potential Based on Robust Adaptive Filtering

Yuxian Zou¹, Yong Hu² and Zhiguo Zhang²

¹*Peking University Shenzhen Graduate School,*

²*The University of Hong Kong
China*

1. Introduction

Somatosensory Evoked Potentials (*SEPs*) are brain electrical physiological signals elicited by the direct electrical stimulation of peripheral nerves. In other words, *SEP* is viewed as the nerve electric response produced by spinal cord sending or receiving sensory information in response to a stimulus (Turner et al., 2003). *SEP* has been widely used during the clinical testing and monitoring of the spinal cord and the central nervous system with the surface electrical stimulation. It can be said that the *SEP* is the most popular technique for intraoperative spinal cord monitoring in the operating room over 30 years (Nash et al., 1977; El-Hawary et al., 2006). However, in practice, the *SEP* signals recorded in the operating theaters are always contaminated by severe background noises (Krieger & Sciabassi 2001). The factors which cause noises may be electrical, physiological, anesthetic, surgical or abrupt event such as cough, body movement or adverse response to the stimulus of the patients. Generally, the recorded *SEP* signal is of a very poor signal-to-noise ratio (*SNR*) nature of the typical values between -20 dB to 0 dB (McGillem et al., 1981).

Literature review of *SEP* extraction techniques showed that the Ensemble Averaging (*EA*) is the most commonly used practical technique for *SEP* extraction (MacLennan & Lovely 1995). Research studies reveal that the *EA-SEP* approach is a kind of stimulus-locked signal averaging method, which is able to enhance the *SNR* in evoked potential recordings when a huge number of independent stimulus trails are used (such as hundreds or more than one thousands stimuli). This means that the *EA-SEP* extraction may lengthen the surgical time and hinder the surgical procedures (El-Hawary et al., 2006). Furthermore, *EA-SEP* approach is lack of ability to provide the timely warning of the eminent danger of cord injury in spine surgeon monitoring. In conclusion, the major drawbacks of *EA-SEP* approach are: First, the assumption that the captured *SEP* signals are truly deterministic and invariant between ensembles is dubious. Actually, a number of studies showed that *SEPs* are nonstationary and time-varying across stimulus trails (Nishida et al., 1993; Woody, 1967). Second, the procedure is very time-consuming, requiring up to 2000 ensembles to identify the *SEP* signal, which causes the discomfort to the subjects, and brings larger opportunity for the interference to degrade the *SEP* extraction. Moreover, careful evaluation of the working principle of the *EA-SEP* method reveals that the averaging process may merge the details carrying the information of certain neurological function in *SEP*. With the analysis above, we can conclude that *EA-SEP* method may fail to track trial-to-trial variations both in

latency and amplitude. A more effective and reliable technique is expected to minimize the number of trials for *SEP* extraction, and the single trail *SEP* extraction is desired.

A lot of researches have been carried out for *SEP* extraction and various signal processing techniques have been investigated, including parametric modelling, nonlinear filtering, wavelet transform, adaptive filtering and independent component analysis (ICA) (Lange & Inbar 1996; Wei et al., 2002). It can be seen that a large number of records are still required to obtain a qualitative estimation in parametric model, and the study by Lange and Inbar suggested that it may not be able to provide adequate estimation of *SEP* (Lange & Inbar 1996).

In recent years, the present authors and some other researchers intensively investigated on the *SEP* extraction using adaptive filtering technique (*AF-SEP*) (Lin et al., 2004) (Lin et al., 2004; Hu et al., 2005; Lam et al., 2005). Research results showed that *AF-SEP* performs better compared with the *EA-SEP* or other parameter estimation approaches in either stationary or non-stationary situation, and *AF-SEP* was recommended as the most appropriate method to improve SNR of *SEP* (Lam et al., 2005). Specifically, *AF-SEP* under investigation usually employed the conventional linear transversal adaptive filter. There are two different structures have been proposed: one is the adaptive noise canceller (ANC) *SEP* extraction method (*ANC-SEP*) (Hu et al., 2005; Ren et al., 2009), another is the multi-filter *SEP* extraction method (*MAF-SEP*) for low SNR *SEP* estimation, where the ANC is used to remove the correlated noise in a primary signal and the uncorrelated noise, while the *SEP* components enhancement is carried out by the adaptive signal enhancer (ASE). Experimental results have shown that *MAF-SEP* method can greatly reduce the number of input trials for *SEP* extraction (Lam et al., 2005). Adaptive filter theory tells that the different adaptive algorithms provide different filtering performances (Haykin, 2001). The least mean squares (*LMS*) based adaptive noise canceller *SEP* method (*LMS-ANC-SEP*) was found to be a fast, simple, and reliable *SEP* extraction method for intraoperative spinal cord monitoring (Lam et al., 2005). The *LMS* algorithm is famous for its simplicity at the price of having a relatively slow convergence rate and sensitive to the noise disturbance. To speed up the convergence, a Recursive Least Squares (*RLS*) based *ANC-SEP* (*RLS-ANC-SEP*) extraction algorithm was developed and studied in (Ren et al., 2009), where the Least Square cost function has been employed. *RLS* is a stable and accurate adaptive filtering algorithm (Haykin, 2001) since it updates the estimate using all the past available information, instead of the instantaneous measurement and error values in *LMS*. Intensive experimental results demonstrate that the *RLS-ANC-SEP* extraction outperforms the *EA-SEP* and the *LMS-ANC-SEP*. It also showed that the *RLS-ANC-SEP* is much less sensitive to noise disturbance over its counterpart algorithms, but at the expense of a heavier computational load.

Some research has shown that the conventional adaptive filters minimizing least squares (*LS*) or mean square error (*MSE*) are very sensitive to non-Gaussian or impulsive noise (Chan & Zou, 2004; Hazarika et al., 1997; Kong & Qiu, 1999). This is of increasing importance in biomedical signal processing field. Kong and Qiu (Kong & Qiu 1999) have done some preliminary research on a latency change detection and estimation algorithm under α -stable noise condition. They showed that the adaptive time delay estimation (*TDE*) algorithms based on the least mean square criterion failed to give an accurate estimation of the latency changes in the *EP* signal, and they employed the direct least mean square (*DLMS*) adaptive *TDE* algorithm derived based on the direct least mean p-norm criterion proposed by Etter and Stearn (Etter & Stearn, 1981). Theoretical analysis and simulation studies concluded that the *DLMS* algorithm is robust to the noises in *EP* signals with both Gaussian and non-Gaussian distributions.

SEP signals recorded in the operating room have illustrated the impulsive characteristics under certain circumstance, such as some orthopedic manipulations using saw, drill, bone

taps or bone bits. Based on our knowledge, there is no research carried out for the *SEP* extraction under the impulsive noise environment.

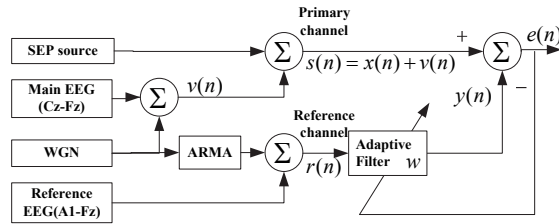
In this research, we will investigate the incorporation of robust M-estimator in the adaptive noise canceller structure for the *SEP* extraction. A recursive least M-estimate *SEP* extraction algorithm named as *RLM-ANC-SEP* has been developed by minimizing a robust M-estimator cost function. The performance of the *RLM-ANC-SEP*, *RLS-ANC-SEP*, *LMS-ANC-SEP*, and *EA* methods regarding to *SEP* extraction will be evaluated and compared quantitatively.

2. Materials and methods

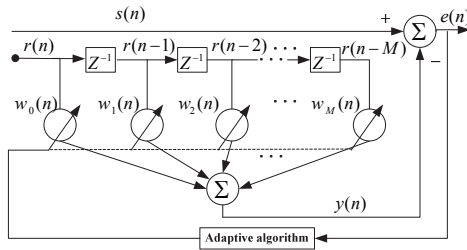
In this section, the framework and the working principle of the adaptive noise canceller (ANC) using the finite impulse response (FIR) filter for *SEP* extraction is introduced. The *SEP* extraction system setup and data generation is presented accordingly. The *SEP* extraction methods using least mean square algorithm (*LMS-ANC-SEP*) and recursive least square algorithm (*RLS-ANC-SEP*) are provided for completeness and comparison purpose. The *SEP* extraction using the recursive least M-estimate (*RLM*) algorithm is derived and discussed at last.

2.1 Adaptive noise canceller (ANC) for SEP extraction

In Figure 1 (a), a block diagram of ANC for *SEP* extraction is illustrated, which mainly consists of a primary channel and a reference channel. The primary channel receives the source signal which refers to the raw *SEP* recording and can be modelled as



(a)



(b)

Fig. 1. (a) A block diagram of adaptive noise canceller for *SEP* extraction, (b) Diagram of the *M* order FIR adaptive filter in ANC

$$s(n) = x(n) + v(n) \quad (1)$$

where $x(n)$ is the true SEP signal and $v(n)$ represents the background noise and interferences. In Figure 1, the reference channel represents a noise source denoted as $r(n)$, and $e(n)$ is the output of the ANC system, which is considered as the estimated version of the true SEP signal which can be formulated as

$$e(n) = s(n) - y(n) = x(n) + v(n) - \mathbf{w}^T(n)\mathbf{r}(n) \approx \hat{x}(n) \quad (2)$$

where the output of the adaptive filter is denoted as $y(n) = \mathbf{w}^T(n)\mathbf{r}(n)$. $\mathbf{r}(n) = [r(n), \dots, r(n-M)]^T$ and $\mathbf{w}(n) = [w_0(n), w_1(n), \dots, w_M(n)]^T$ are the output, input data vector and weight vector of the adaptive FIR filter (AF), respectively. The derivation of the adaptive filtering algorithm is governed by a meaningful cost function. As the result, after the convergence of the AF, the difference between the filter output and the desired response will be minimized.

It is worthy to note that for the ANC approach for SEP extraction, there are some important assumptions for achieving global convergence of the adaptive filter and the unbiased estimation of the desired signal. Firstly, the desired signal ($x(n)$) is corrupted by an additive interference (or noise) ($v(n)$) to form the primary signal $s(n)$; Secondly, if the reference signal ($r(n)$) is a correlated version of the interference signal ($v(n)$), then a FIR filter can be applied to transform $r(n)$ to approximate $v(n)$ and then suppress $v(n)$ from $s(n)$, which is illustrated in Figure 1(b); Thirdly, the reference signal ($r(n)$) must not contain a correlated component of $x(n)$, otherwise, the SEP signal component may also be cancelled at the output of the ANC. Therefore, it can be concluded that in the study of SEP extraction under ANC framework, the SEP recording and the reference signal generation must be designed carefully to satisfy the above requirements. Some discussion of the SEP extraction system setup will be presented in the next section.

2.2 SEP Extraction system setup and data generation

2.2.1 SEP extraction system setup and signals

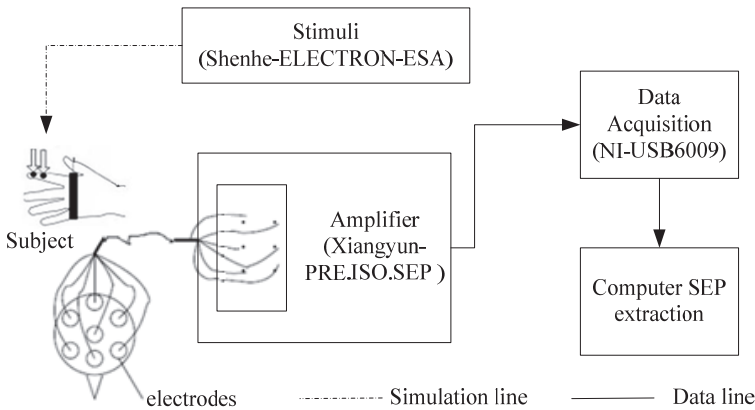


Fig. 2. A typical setup of the SEP extraction system

In our SEP extraction study, the SEP extraction system setup is illustrated in Figure 2. The SEP signals were collected over Cz' (2 cm posterior to Cz, 10-20 international system for EEG

electrode placement) versus the Fz of the 10-20 system. The stimulation for *SEP* recording was applied on the posterior tibial nerve with the duration of 0.3 ms, the rate of 5.1 Hz and the constant current of 10 to 30 mA. The signals were amplified one hundred thousand times, bandpass filtered at 20-3000Hz. The *SEP* signals were acquired and recorded to a computer with 12-bit resolution and the sampling rate of 5 kHz. We collected 500 trials for one subject and then the average of these trials is taken as a standard *SEP* template in our study, which is shown in the first row of Figure 3 as xn .

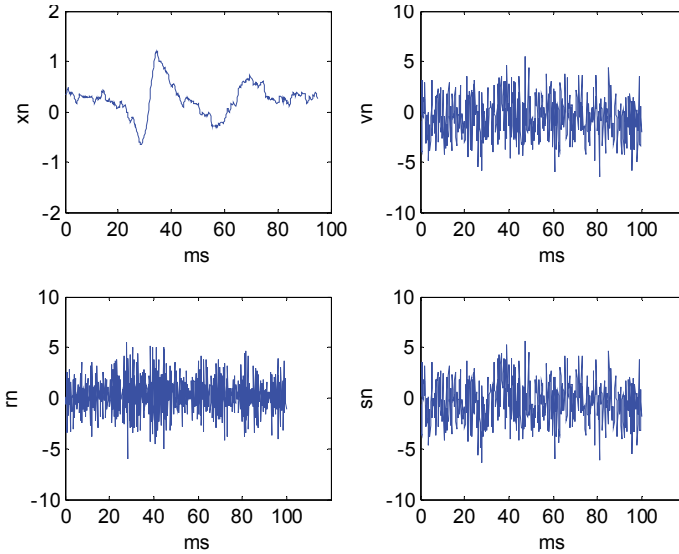


Fig. 3. *SEP* signals. (1) xn : An example of a *SEP* template obtained from ensemble averaging of 500 trials; (2) vn : one example of recorded $A1-Fz$ used as *EEG* together with *WGN* for primary channel; (3) rn : one example of recorded $Cz-Fz$ recording used as *EEG* together with correlated *WGN* as the reference channel signal; (4) sn : one example of the primary channel signal ($EEG + SEP + WGN$) at $SNR = -15$ dB.

2.2.2 The primary channel signal and reference channel signal generation

It is observed that continuous *EEG* is the major source of noise found in the primary *SEP* recording channel and its variation version in the reference channel (Lam et al., 2005). In our study, *EEG* is recorded over Cz and $A1$ (auricular) versus Fz , respectively from the awoken subjects at a sitting position in a quiet environment. This is because the $A1-Fz$ recording signal has much less *SEP* component, and $A1-Fz$ recording is suitable used for the generation of the reference channel signal $r(n)$ (MacDonald et al., 2005). Meanwhile, the $Cz-Fz$ recording was used to superimpose onto the *SEP* template to generate the *EEG*-contaminated *SEP* signal, that is, continuous *EEG* signals of different SNR levels were added to each *SEP* template to generate the *EEG*-contaminated *SEP* trials (the primary channel signal $s(n)$ at different SNR level) for algorithm performance evaluation purpose. An autoregressive moving average (*ARMA*) filter was employed to simulate the correlated *WGN* noise for the reference channel. One example of the $A1-Fz$ *EEG* signal and $Cz-Fz$ *EEG* signal recorded from scalp is plotted in Figure 3 as vn and rn , respectively. The SNR of *EEG*-

contaminated *SEP* was set to -10, -15 and -20 dB. The last row of Figure 3 (*sn*) shows the simulated primary channel *SEP* signal at -15dB with *EEG* and *WGN*.

2.3 Adaptive noise canceller for SEP extraction using mean square estimation

The *SEP* extraction method under the adaptive noise canceller (ANC) framework derived from the Mean-Square-Error was firstly introduced and evaluated by some of the present authors in (Lam et al., 2005). From the adaptive filter theory (Haykin, 2001) and the configuration of the ANC shown in Figure 1, the *SEP* extraction problem can be solved as a linear ANC problem since the *FIR* adaptive filter is used (named as ANC-*SEP* approach in this study). The commonly used error measure is the Mean-Square-Error (*MSE* defined as $J_{MSE} = E[e^2(n)]$, where $E[\cdot]$ represents the ensemble operator). The minimisation of the *MSE* results in the Wiener normal equation under some statistically independent and signal wide-sense stationary (*WSS*) assumptions. The optimal solution of Wiener normal equation can be denoted as:

$$\mathbf{w}_{opt}(n) = \mathbf{R}_{mse}^{-1}(n) \mathbf{P}_{mse}(n) \quad (3)$$

where $\mathbf{P}_{mse}$ is the cross-correlation vector between $s(n)$ and $\mathbf{r}(n)$, and $\mathbf{R}_{mse}$ is the autocorrelation matrix of $\mathbf{r}(n)$, which can be written as

$$\begin{aligned} \mathbf{P}_{mse} &= E[s(n)\mathbf{r}(n)] \\ \mathbf{R}_{mse} &= E[\mathbf{r}(k)\mathbf{r}^T(k)] \end{aligned} \quad (4)$$

As discussed in many literatures, the well-known least mean squares (*LMS*) algorithm is a stochastic gradient based adaptive algorithm to obtain the optimal solution of J_{MSE} . The updating of the adaptive filter coefficient vector can be denoted as (Haykin, 2001)

$$\mathbf{w}(n) = \mathbf{w}(n-1) + 2\mu_{lms}e(n)\mathbf{r}(n) \quad (4)$$

where μ_{lms} is the stepsize which is one of the most important factors that controls the initial convergence rate and steady state error of the *LMS*-ANC for *SEP* extraction. Generally, a big stepsize yields rapid convergence but larger steady-state misadjustment error. A small stepsize yields slow convergence but a corresponding smaller steady-state misadjustment error. There exists a theoretical lower and upper bound of the choice of μ_{lms} (details can be referred to (Haykin, 2001). Usually, the choice of μ_{lms} is suggested by the following condition in the *LMS* algorithm (Haykin, 2001)

$$0 < \mu < 1 / (MP_{in}) \quad (5)$$

where P_{in} and M is the input power and the order of the adaptive *FIR* filter, respectively. In principle, the selection of the stepsize not only depends on the desired steady-state error level but also the statistical properties of the input signal of the adaptive filter. In other words, the convergence rate of the *LMS* algorithm is greatly affected by the dynamic range of the eigenvalues of the autocorrelation matrix $\mathbf{R}_{mse}$. Considering this essential limitation, it is not difficult to understand that the performance of the ANC-*SEP* approach using *LMS* algorithm may suffer from the conflict to the *WSS* assumption for $s(n)$ and $\mathbf{r}(n)$ and the nonstationary property of the $\mathbf{r}(n)$.

2.4 Adaptive noise canceller for SEP extraction using least square estimation

Motivated by the performance enhancement of the *ANC-SEP* method using *LMS* algorithm compared with *EA-SEP* (Hu et al., 2005; Lam et al., 2005; Cui et al., 2008), some investigations of the *ANC-SEP* using *RLS* algorithm have been carried out and presented in (Ren et al., 2009). Instead of using *MSE* cost function, a conventional least square (*LS*) cost function is employed and the optimal solution of J_{LS} is described as follows (Haykin 2001)

$$J_{LS}(n) = \sum_{k=1}^n \lambda^{n-k} |e(k)|^2, \text{ and } \mathbf{w}_{opt}(n) = \mathbf{R}_{LS}^{-1}(n) \mathbf{P}_{LS}(n) \quad (6)$$

where, λ is the forgetting factor with the value between 0 and 1, which controls the effective amount of data used in the averaging and hence the degree to which the *RLS* algorithm can track the signal variation. The closer the value of λ goes to one, the lower will be the steady-state misadjustment error of the *RLS* algorithm. Its tracking ability, however, will also be slower. $\mathbf{R}_{LS}(n)$ is the autocorrelation matrix of the input vector at time index n and $\mathbf{P}_{LS}(n)$ is the cross-correlation vector between the input vector and the reference signal at time index n . Generally, they can be estimated as

$$\mathbf{R}_{LS}(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{r}(i) \mathbf{r}^T(i) = \lambda \mathbf{R}_{LS}(n-1) + \mathbf{r}(n) \mathbf{r}^T(n) \quad (7)$$

$$\mathbf{P}_{LS}(n) = \sum_{i=1}^n \lambda^{n-i} s(i) \mathbf{r}(i) = \lambda \mathbf{P}_{LS}(n-1) + s(n) \mathbf{r}(n) \quad (8)$$

By applying the matrix inversion lemma to the optimal solution in (6), the famous recursive least square (*RLS*) algorithm can be derived, and it is summarized in Table 1 for the completeness (Interested readers can refer to (Haykin, 2001)). From Table 1, it is noted that the computational complexity of the *RLS* algorithm is of order M^2 .

- | |
|---|
| <p>1) Initialization: $\mathbf{w}_{RLS}(-1) = 0$, $\mathbf{P}_{RLS}(n-1) = \delta^{-1} \mathbf{I}_M$, $n=0$, where M is the filter order of the adaptive filter using in <i>ANC</i>, δ can be the inverse of an estimation of the input signal power.</p> <p>2) Calculation of the adaptive filter output: $y(n) = \mathbf{r}^T(n) \mathbf{w}_{RLS}(n-1)$</p> <p>3) Estimation error: $e(n) = s(n) - y(n)$</p> <p>4) Calculation of the Kalman gain vector:</p> $\mathbf{K}_{RLS}(n) = \frac{\mathbf{P}_{RLS}(n-1) \mathbf{r}^*(n)}{\lambda + \mathbf{r}^T(n) \mathbf{P}_{RLS}(n-1) \mathbf{r}^*(n)}$ <p>5) Update of the inverse correlation matrix:</p> $\mathbf{P}_{RLS}(n) = \frac{1}{\lambda} \left[\mathbf{P}_{RLS}(n-1) - \mathbf{K}_{RLS}(n) \mathbf{r}^T(n) \mathbf{P}_{RLS}(n-1) \right]$ <p>6) Update of the filter weights: $\mathbf{w}_{RLS}(n) = \mathbf{w}_{RLS}(n-1) + \mathbf{K}_{RLS}(n) e(n)$</p> <p>$n=n+1$, back to step 2)</p> |
|---|

Table 1. *RLS-ANC-SEP* algorithm

2.5 Adaptive noise canceller for SEP extraction using robust estimation

Carefully evaluating the properties of the recording *SEP* signals in the operating room, it noted that these *SEP* signals may have some nonstationary and impulsive like properties when the trial patients happen to the eye movement, cough and stimulus etc, which commonly exist. Under these kind of circumstances, the performance of the *ANC-SEP* methods using *LMS* or *RLS* will degrade or fail to extract *SEP* signal due to the adverse effect of the noise. The new method is desired. Motivated by the research work done by Chan and Zou (Chan & Zou, 2004), a new error measure method based on the *M*-estimate has been introduced and the corresponding cost function instead of J_{MSE} or J_{LS} is used for providing the robustness in the algorithm, which is given as

$$J_R(n) = \sum_{i=1}^n \lambda^{n-i} \rho(e(i)) = \sum_{i=1}^n \lambda^{n-i} \rho(s(i) - \mathbf{w}^T(n) \mathbf{r}(i)) \quad (10)$$

where λ is the positive forgetting factor and ρ is an *M*-estimate function, which provides certain ability to suppress the adverse effect of impulsive noise on the cost function when the error signal becomes very large. In our study, the Huber *M*-estimate function and the related weighting function are used, which can be denoted as

$$\rho(e) = \begin{cases} e^2 / 2, & 0 < |e| < \xi \\ \xi^2 / 2 & \text{otherwise} \end{cases} \quad (11)$$

where ξ is the threshold parameter. The optimal solution $\mathbf{w}(n)$ for minimizing $J_R(n)$ can be obtained by differentiating (10) with respect to $\mathbf{w}(n)$ and setting the derivatives to zero. This yields the following *M*-estimate normal equation

$$\mathbf{R}_R(n) \mathbf{w}(n) = \mathbf{P}_R(n) \quad (12)$$

where

$$\mathbf{R}_R(n) = \lambda \mathbf{R}_R(n-1) + q(e(n)) \mathbf{r}(n) \mathbf{r}^T(n) \quad (13)$$

$$\mathbf{P}_R(n) = \lambda \mathbf{P}_R(n-1) + q(e(n)) s(n) \mathbf{r}(n) \quad (14)$$

$$q(e) = \frac{d\rho(e)/de}{e} = \begin{cases} 1, & 0 < |e| < \xi \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

where, $\mathbf{R}_R(n)$ and $\mathbf{P}_R(n)$ are called the *M*-estimate correlation matrix of $\mathbf{r}(n)$ and the *M*-estimate cross-correlation vector of $\mathbf{r}(n)$ and $s(n)$, respectively. The adaptive algorithm for solving the normal equation (12) can be obtained in the same way as developing *RLS* algorithm, and the resulting algorithm is called recursive least *M*-estimate algorithm (*RLM*) and it is summarized in Table 2. From Table 1 and Table 2, it can be seen that the computational complexity of *RLM* and *RLS* is similar except the cost to determine the weighting function $q(e)$ in (15). It is also noted that when the signal is Gaussian distributed, *RLS* and *RLM* are identical. The contribution of the weight function $q(e(n))$ lies at the suppression of the adverse effects of the large estimation error due to the undesired impulsive interference on the adaptive filter weight vector $\mathbf{w}(n)$. The degree of this suppression is controlled by the parameter ξ , in our study, a recursive estimation approach

is adopted which directly connects to the variance of the estimation error under the assumption of the interference is with contaminated Gaussian (CG) or alpha-stable distributions. The parameter ξ has been determined (shown in Table 2) when there is 95% confidence to detect and reject the impulses (Chan & Zou, 2004).

1) Initialization: $\mathbf{w}_{RLM}(-1) = 0$, $\mathbf{P}_{RLM}(n-1) = \delta^{-1} \mathbf{I}_M$, $n=0$, where M is the filter order of the ANC, δ can be the inverse of an estimation of the input signal power.

2) Calculation of the adaptive filter output: $y(n) = \mathbf{r}^T(n) \mathbf{w}_{RLM}(n-1)$

3) Estimation error calculation: $e(n) = s(n) - y(n)$

4) Estimate the variance of the estimation error, determine the parameter ξ and determine the weighting function (Chan & Zou, 2004) :

$$\sigma^2(n) = \lambda_\sigma \sigma^2(n-1) + (1 - \lambda_\sigma) c_1 \text{med}(A_e(n)), \quad \xi = 2.24\sigma(n), \quad q(e) = \begin{cases} 1, & 0 < |e| < \xi \\ 0, & \text{otherwise} \end{cases}$$

where λ_σ is the forgetting factor, $A_e(n) = \{e^2(n), \dots, e^2(n - N_w + 1)\}$, N_w is the length of the estimation window, and $c_1 = 1.483(1 + 5 / (N_w - 1))$ is the finite sample correction factor

5) Calculation of the Kalman gain vector: $\mathbf{K}_R(n) = \frac{q(e(n)) \mathbf{P}_{RLM}(n-1) \mathbf{r}(n)}{\lambda + q(e(n)) \mathbf{r}^T(n) \mathbf{P}_{RLM}(n-1) \mathbf{r}(n)}$

6) Update of the inverse correlation matrix:

$$\mathbf{P}_{RLM}(n) = \frac{1}{\lambda} [\mathbf{P}_{RLM}(n-1) - \mathbf{K}_{RLM}(n) \mathbf{r}^T(n) \mathbf{P}_{RLM}(n-1)]$$

7) Update of the filter weights: $\mathbf{w}_{RLM}(n) = \mathbf{w}_{RLM}(n-1) + \mathbf{K}_R(n) e(n)$

8) Calculate the estimation error: $e(n) = \begin{cases} e(n), & q(e) = 1 \\ e(n-1), & q(e) = 0 \end{cases}$

$n=n+1$, back to step 2)

Table 2. RLM algorithm

3. Simulation study and discussion

As discussed above, we have introduced the SEP extraction approaches under the ANC framework by using different adaptive filtering algorithms. Specifically, employing LMS, RLS and RLM algorithms to update the weighting vector of the adaptive FIR filter in ANC results in the LMS-ANC-SEP method, RLS-ANC-SEP method and RLM-ANC-SEP method, respectively. In this section, the performance of these adaptive filtering methods for SEP extraction under Gaussian and impulsive noise environment has been evaluated and compared by intensive simulation experiments.

3.1 Experiment 1: SEP extraction under Gaussian noise

In this section, we aim to visually illustrate the SEP extraction performance of the algorithms discussed above under Gaussian noise condition. The detailed performance

comparison between *EA-SEP*, *LMS-ANC-SEP*, *RLS-ANC-SEP* and *RLM-ANC-SEP* methods under *EEG* and *WGN* contamination can be referred to the work presented in papers (Lam et al., 2005) and (Ren et al., 2009). Here, we only illustrate one set of the *SEP* extraction results for reader's favorite review. Figure 4 shows the *SEP* extraction results from 50 *SEP* trails by different algorithms. In this experiment, the *SEP* template (xn), simulated primary signals (sn , vn) and reference signal (rn) are the same as those shown in Figure 3 at $SNR=-15dB$. The order of the adaptive filter M is set to be 10, the step size μ of the *LMS-ANC-SEP* is chosen as 2×10^{-4} , the forgetting factor of the *RLS-ANC-SEP* and *RLM-ANC-SEP* algorithms is set to be 0.99. The parameters for *RLM-ANC-SEP* in Table 2 are set as $\lambda_\sigma=0.9$ and $N_w=7$. From Figure 4, it is clear to see that the signals extracted from 50 trials by *EA-SEP* and *LMS-ANC-SEP* are difficult to detect the positive and negative peaks required for quantitative analysis and diagnosis of the *SEP* signal. More precisely, the positive peak around 35ms and the negative peak around 40ms, which are two most commonly-used criteria for the online monitoring during the spinal surgery, are still buried in the heavy background noise, so that their latencies and amplitudes cannot be measured accurately. On the other hand, we can see that the performance of *RLM-ANC-SEP* is almost the same as that of *RLS-ANC-SEP*, which outperforms than other two algorithms. It is apparent that two peaks around 35ms and 40 ms can be easily observed and their latencies and amplitudes can be precisely measured in the results using *RLS-ANC-SEP* and *RLM-ANC-SEP* methods. All these findings in practice can be well explained in theory. That is, the *RLS/RLM*-based algorithms have a fast convergence rate than *LMS*-based algorithm. Furthermore, the *RLM-ANC-SEP* algorithm is comparable to *RLS-ANC-SEP* algorithm under *EEG* and *WGN* environment. We next test and compare their performances when few *SEP* trials are contaminated with impulsive noises.

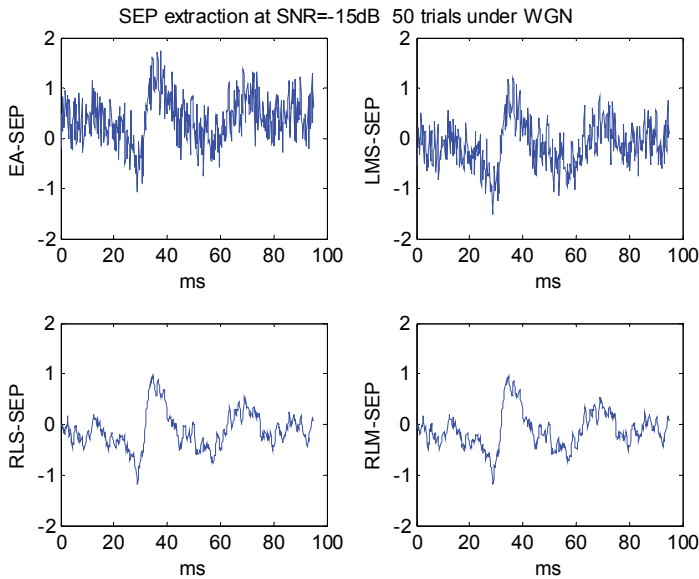


Fig. 4. 50-trial *SEP* extraction results obtained by *EA-SEP*, *LMS-ANC-SEP*, *RLS-ANC-SEP*, and *RLM-ANC-SEP* method, respectively ($SNR=-15dB$)

3.2 Experiment 2: SEP extraction under impulsive noise

This simulation is set up to compare the *SEP* extraction performance of the *EA-SEP*, *LMS-ANC-SEP*, *RLS-ANC-SEP* and *RLM-ANC-SEP* under *EEG* and individual impulse contaminated noise environment. Generally, the impulsive noise can be generated by a contaminated Gaussian (CG) model proposed in (Haweel & Clarkson, 1992). The impulses are generated individually with arrival probability $P_{ar}=2\times 10^{-3}$ and the variance is chosen as 200. In our study, only for performance illustration purpose, the positions of the impulses are assumed to occur at 19ms, 28ms, 35ms, 44ms, and 78ms, respectively (which is not necessary to fix the position of the impulses, but here it is for us to gain the better performance visualization for different algorithms). The *SEP* template (xn), one sample primary interference (vn) with impulses, one sample of the reference signal (m) and the resultant primary signal (sn) at -15dB are shown in Figure 5. The difference between Figure 3 and Figure 5 only lies at several impulses added in the primary interference signal (vn). In this case, the primary signal is composed of a *SEP* template, an *A1-Fz EEG* component, and a contaminated Gaussian noise.

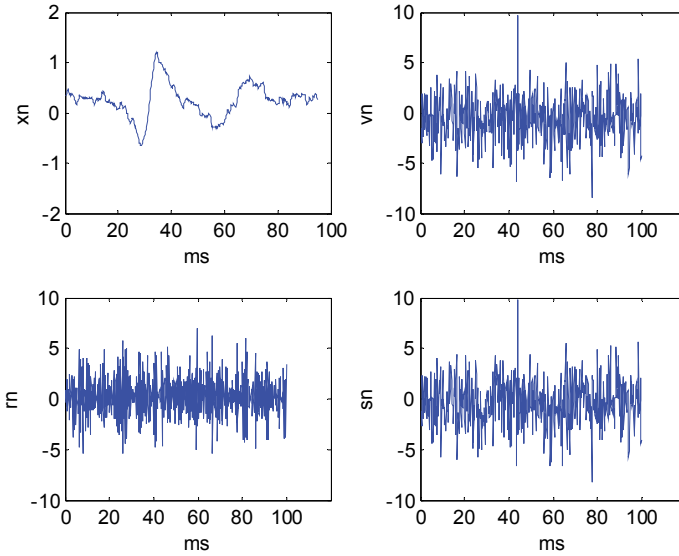


Fig. 5. *SEP* signals with impulsive noise, (1) xn and m are the same as those in Figure 3. (2) vn : one example of recorded *A1-Fz* used as *EEG* together with CG noise for primary channel; (3) sn : One example of the primary channel signal (*EEG* + *SEP* + CGN) at -15 dB.

For this simulation, all parameter settings are the same as those used in Experiment 1. The *SEP* extraction results from 50 *SEP* trails under impulsive noise by different algorithms are shown in Figure 6. If no impulsive noise occurs, the extraction results of four different methods should be approximately identical to their counterparts in Figure 4. As a result, Figure 5 can be regarded as a standard to evaluate the robustness of these methods when impulsive noises are added. From Figure 6, it is clear to see that the adverse impact of the impulses on the *SEP* extraction for *EA-SEP*, *LMS-ANC-SEP* and *RLS-ANC-SEP* algorithms compared with their counterpart algorithms under *WGN* shown in Figure 4. More specifically, for the *EA-SEP*

method, since the amplitudes of the impulsive noise are rather large compared to that of WGN, they cannot be averaged out completely using finite number of trials. As for the *LMS-ANC-SEP* and *RLS-ANC-SEP* methods, which employ an *LS* criterion for the updating of the filter coefficients in *ANC*, their performances are degraded severely because the coefficient estimates in *ANC* are unstable and may be greatly deviated from the reasonable values when impulsive noise occurs. The performance degradation can be more easily observed in the result of *RLS-ANC-SEP* in Figure 6, where the adverse impacts of impulsive noises around 35ms and 44ms are distinct and its difference with *RLS-ANC-SEP* of Figure 4 is obvious. Unlike those methods based on averaging or *LS* criterion, *RLM-ANC-SEP* employs an *M*-estimation function in *ANC* so that the impulsive noise can be detected and suppressed effectively. As the result, its harmful impact on *SEP* extraction is reduced considerably. The simulation results illustrate the advantage of *RLM-ANC-SEP*, and we can see that *RLM-ANC-SEP* shows its robustness to the impulsive interferences and its performance is close to that under WGN condition. In Figure 6, we can hardly find the traces of impulsive noise in the *RLM-ANC-SEP* result and peaks were clearly seen and measurable. In a simple word, impulsive noise which degrades the outputs of *EA-SEP*, *LMS-ANC-SEP* and other *LS*-based *SEP* extraction methods will do little harm to *RLM-ANC-SEP*.

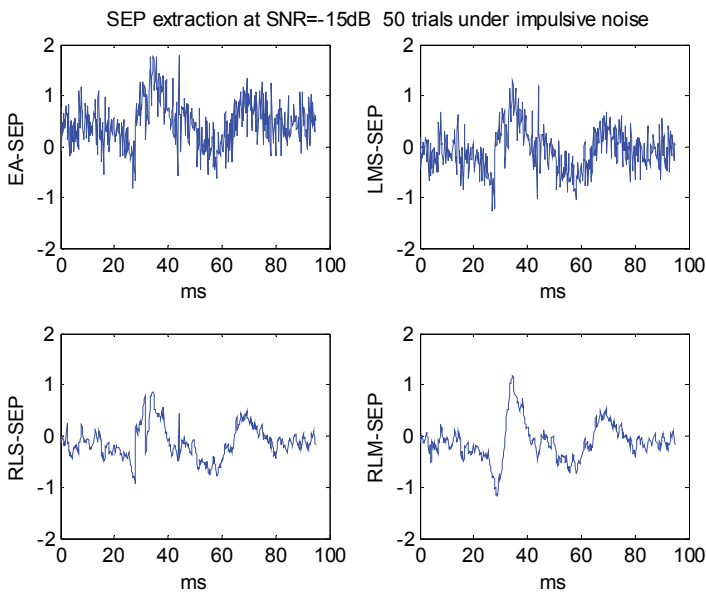


Fig. 6. 50-trial *SEP* extraction results obtained by *EA-SEP*, *LMS-ANC-SEP*, *RLS-ANC-SEP*, and *RLM-ANC-SEP* method, respectively (SNR=-15dB)

As mentioned before, impulsive noise often occurs during spinal surgery in operating theatres and it will greatly decrease the quality of *SEP* recording. Current *SEP* recording technique works in this way when some *SEP* trials are contaminated with impulsive noise, they will be discarded. However, these trials with impulsive noise also contain useful *SEP* information, and the rejection of these trials will increase the time to record a useful *SEP*

signal, and make the recording and monitoring discontinuous, which is undesirable. Therefore, making use of *SEP* trials contaminated with impulsive noise is necessary and robust *SEP* extraction method, such as the proposed *RLM-ANC-SEP* method, is advantageous. Our preliminary study and experimental results show that the *RLM-ANC-SEP* method has an excellent performance in impulsive noise environment, it may be taken as a good solution to achieve reliable and continuous *SEP* recording for monitoring under Gaussian and impulsive noise environment.

4. Conclusion

Aiming at developing the efficient *SEP* recording system, we have introduced the *SEP* extraction methods under the ANC framework using adaptive *FIR* filter. A new *SEP* extraction method called *RLM-ANC-SEP* was developed to obtain the fast and robust performance under Gaussian and Contaminated Gaussian noise environment. *RLM-ANC-SEP* minimizes the modified Huber *M*-estimator based cost function instead of the conventional mean square error and least squares error based cost functions, which provides the robust ability when impulses occurring in the primary channel, and maintains the fast convergence as the *RLS-ANC-SEP* algorithm. Simulation study proved that either *RLM-ANC-SEP* or *RLS-ANC-SEP* has better and more robust convergence performance than *LMS-ANC-SEP*. The performances of *RLM-ANC-SEP* and *RLS-ANC-SEP* showed equivalent under *WGN* condition, but *RLM-ANC-SEP* presented its robustness to the impulsive interferences. Clinical application and validation study could be our future work on this proposed *SEP* signal extraction approach.

5. Acknowledgment

This work was partially supported by Shenzhen Science and Technology Program (No. 08CXY-01), Hong Kong ITF Tier 3 (ITS/149/08), and Research Grants Council of the Hong Kong SAR (GRF HKU7130/06E).

6. References

- Chan, S. C. & Y. X. Zou (2004). A Recursive Least M - Estimate Algorithm for Robust Adaptive Filtering in Impulse Noise: Fast Algorithm and Convergence Performance Analysis, *IEEE Transactions on Signal Processing*, Vol. 52, No.4, pp. 975-991.
- Cui, H., C. Shen, et al. (2008). Study on Adaptive Noise Canceller Based on Fixed-Point Algorithm for Real-Time Somatosensory Evoked Potential Monitoring, *Bioinformatics and Biomedical Engineering*, Vol. 3, pp.2213 - 2216.
- El-Hawary, R., S. Sparagana, et al. (2006). Spinal Cord Monitoring in Patients with Spinal Deformity and Neural Axis Abnormalities: A Comparison with Adolescent Idiopathic Scoliosis Patients, *SPINE*, Vol. 31, No. 19, pp.E698-E706.
- Etter, D. M. & S. D. Stearn (1981). Adaptive Estimation of Time Delays in Sampled Data Systems, *IEEE Transactions on Acoustic, Speech, Signal Processing*, Vol. 29, pp. 582-587.
- Haweel, T. I. & P. M. Clarkson (1992). A Class of Order Statistic LMS Algorithms, *IEEE Transactions on Signal Processing*, Vol. 40, No. 1, pp. 44-53.

- Haykin, S. (2001). *Adaptive Filter Theory* (4th Edition), Prentice Hall.
- Hazarika, N., A. Tsoi, et al. (1997). Nonlinear Considerations in EEG Signal Classification, *IEEE Transactions on Signal Processing*, Vol. 45, pp. 829-836.
- Hu, Y., B. Lam, et al. (2005). Adaptive Signal Enhancement of Somatosensory Evoked Potential for Spinal Cord Compression Detection: An Experimental Study, *Computers in Biology and Medicine*, Vol. 35, pp. 814-828.
- Kong, X. & T. S. Qiu (1999). Adaptive Estimation of Latency Change in Evoked Potentials by Direct Least Mean p-Norm Time-Delay Estimation, *IEEE Transactions On Biomedical Engineering*, Vol. 46, No.8, pp. 994-1003.
- Krieger, D. & R. Sclabassi (2001). Real-Time Intraoperative Neurophysiological Monitoring, *Methods*, Vol. 25, pp. 272-287.
- Lam, B., Y. Hu, et al. (2005). Multi-adaptive Filtering Technique for Surface Somatosensory Evoked Potentials Processing, *Medical Engineering & Physics*, Vol. 27, pp. 257-266.
- Lange, D. H. & G. F. Inbar (1996). A Robust Parametric Estimator for Single-Trial Movement Related Brain Potentials, *IEEE Transactions on Biomedical Engineering*, Vol. 43, pp. 341-347.
- Lin, B. S., B.-S. Lin, et al. (2004). Adaptive Interference Cancel Filter for Evoked Potential Using High-Order Cumulants. *IEEE Engineering in Medicine and Biology Society*, Vol. 1, pp.396-398.
- MacDonald, D. B. & A. Z. Zayed (2005). Tibial Somatosensory Evoked Potential Intraoperative Monitoring: Recommendations Based on Signal to Noise Ratio Analysis of Popliteal Fossa, *Optimized P37, Standard P37, and P31 Potentials*, *Journal of Clinical Neurophysiology*, Vol. 116, No.8, pp. 1858-1869.
- MacLennan, A. R. & D. F. Lovely (1995). Reduction of Evoked Potential Measurement Time by a TMS320 Based Adaptive Matched Filter, *Medical Engineering & Physics*, Vol. 17, pp. 248-256.
- McGillem, C. & J. Aunon (1981). Signal Processing in Evoked Potential Research: Applications of Filtering and Pattern Recognition, *Critical Reviews in Bioengineering*, CRC, Vol. 9, pp. 225-265.
- Nash, C. L., R. A. Lorig, et al. (1977). Spinal Cord Monitoring During Operative Treatment of the Spine, *Clinical Orthopaedics and Related Research*, Vol. 126, pp. 100-105.
- Nishida, S. & M. Nakamura (1993). Method for Single-Trial Recording of Somatosensory Evoked Potentials, *Journal of Biomedical Engineering*, Vol. 15, pp. 257-262.
- Ren, Z. L., Y. X. Zou, et al. (2009). Fast Extraction of Somatosensory Evoked Potential using RLS Adaptive Filter Algorithms, *The 2nd International Congress on Image and Signal Processing (CISP'09)*, pp. 4444-4447.
- Turner, S., P. Picton, et al. (2003). Extraction of Short-Latency Evoked Potentials Using a combination of Wavelets and Evolutionary Algorithms, *Medical Engineering & Physics*, Vol. 25, pp. 407-412.
- Wei, Q. & K. Fung (2002). Adaptive Filtering of Evoked Potentials with Radial-Basis-Function Neural Network Prefilter, *IEEE Transactions on Biomedical Engineering*, Vol. 49, pp. 225-232.
- Woody, C. D. (1967). Characterization of an Adaptive Filter for the Analysis of Variable Latency Neuroelectric Signals, *Medical Biology Engineering*, Vol. 5, pp. 539-553.

Part 3

Communication Systems

A LEO Nano-Satellite Mission for the Detection of Lightning VHF Sferics

Ghulam Jaffer^{1,2}, Hans U. Eichelberger²,
Konrad Schwingenschuh² and Otto Koudelka¹

¹*Institute of Communication Networks and Satellite Communications
Graz University of Technology, Graz*

²*Space Research Institute, Austrian Academy of Sciences, Graz
Austria*

1. Introduction

The LiNSAT is a proposed project for the detection of electromagnetic signatures produced by lightning strokes (Sferics) in very high frequency (VHF) range in low-earth-orbit (LEO) around 800 km. The satellite is 20 cm cube and weighs ~ 5 kg. The main scientific objective of the planned LiNSAT is the investigation of impulsive electromagnetic signals generated by electrical discharges in terrestrial thunderstorms (lightning), blizzards, volcanic eruptions, earthquakes and dust devils. These electromagnetic phenomena called Sferics cover the frequency range from a few Hertz (Schumann resonances) up to several GHz. Depending on the source mechanism, the wave power peaks at different frequencies, e.g. terrestrial lightning has a maximum power in the VLF and HF range, also trans-ionospheric pulses reaching at LEO and possibly to satellites in geostationary-Earth-orbit (GEO) peak at VHF. The global terrestrial lightning rate is in the order of 100 lightning flashes per second with an average energy per flash of about 10^9 Joule (Rakov and Uman 2003). Only a small percentage of the total energy is converted to electromagnetic radiation. Other forms are acoustic (thunder), optical and thermal, so the whole power of lightning flash is distributed into many chunks of energies. The signal strength received by a satellite radio experiment depends on the distance and the energy of a lightning stroke as well as on the orientation of the discharge channel.

The nano-satellite project under study emphasizes on the investigation of the global distribution and temporal variation of lightning phenomena using electromagnetic signals. In contrast to optical satellite observations the Sferics produced by lightning can be observed on the day and night side but with a smaller spatial resolution. We know from the Fast On-orbit Recording of Transient Events (FORTE) satellite mission (Jacobson, Knox et al. 1999) that at an altitude of about 1000 km the impulsive events produced by lightning can reach amplitudes up to 1 mV/m in a 1 MHz band around 40 MHz.

The LiNSAT is based on the design and the bus similar to the Austrian first astronomical nano-satellite TUGSat-1/ BRITE-Austria (Koudelka, Egger et al. 2009) which is scheduled to launch in April 2011. The LiNSAT will carry a broadband radio-frequency receiver payload for the investigation of Sferics. Special emphasis is on the investigation of transient

electromagnetic waves in the frequency range of 20 – 40 MHz, well above plasma frequency to avoid ionospheric attenuations. The on-board RF lightning triggering system is a special capability of the LiNSAT. The lightning experiment will also observe signals of ionospheric and magnetospheric origin. To avoid false signals detection (false alarm), pre-selectors on-board LiNSAT are part of the Sferics detector. Adaptive filters will be developed to differentiate terrestrial electromagnetic impulsive signals from ionospheric or magnetospheric signals.

One of the major challenges of using a nano-satellite for such a scientific payload is to integrate the lightning experiment antenna, receiver and data acquisition unit into the small nano-satellite structure. The optimization in this mission is to use one of the lightning antennas integrated into gravity gradient boom (GGB) that increases the sensitivity and directional capability of the satellite toward nadir direction. The section 5.1 and section 5.4 describe the space segment and modes of operation. Electromagnetic compatibility (EMC) issues are specially treated. Results of the payload in a simulated environment are presented in section 9.

The lightning emissions are the transient electrical activity of thunderstorms (primarily RS and IC activity) generates broadband electromagnetic radiations with spectrum range from ULF to UHF and also visible-light. A typical RS radiation peaks at ~10 kHz and an IC stroke produces radiations peaking at a slightly higher frequency at 40 kHz with 2 orders of magnitude less energy than a typical RS (Volland 1995). Electromagnetic radiations at these frequencies propagate through the earth - ionosphere waveguide, so can be observed at large distances, thousands of km from the source.

The lightning electromagnetic pulse (LEMP) is time-varying electromagnetic field that varies rapidly around 10 ns, reaches its maxima and then on its descending way is less fast around a few tens of μ s and goes to a negligible value. The LEMP is very dangerous due to its ability to damage unprotected electronic devices. LEMPs are powerful radio emissions that radiate across a broad spectrum of frequencies from tens of kHz or lower to at least several hundred MHz as indicated by the inverse of LEMP rise time. As mentioned earlier, these broadband emissions reach LEO and possibly GEO, so the payload on-board LiNSAT will be designed as a broadband receiver.

The LiNSAT will operate in the VHF portion of the electromagnetic spectrum because lower frequency radio emissions (HF and below) often cannot penetrate through the earth's ionosphere and thus, do not reach LEO. Also, Sferics in higher band from VHF are less powerful, so, more difficult to detect. It complicates its detection and time tagging in the case of broadband VHF signals from LEO by the dispersive and refractive effects of the ionosphere. These effects become increasingly severe at lower frequencies in proportion to wavelength squared.

The LiNSAT radio receiver will record waveforms using a fixed-rate 200 MS/s, 12-bit digitizer that takes its input from either of a 3-antennas wideband sub-resonant monopole and a VHF receiver. The instrument utilizes a coarse trigger based on preset amplitude level to detect transient events.

As the radio emissions from natural lightning produce broadband transients in the VHF spectrum, so a potential source of false alarms for space based detection of other phenomena in the same band. One of the main objectives of the LiNSAT payload development on-board LEO nano-satellite is the need to characterize the Earth's radio background. The characterization is necessary for both transient signals, like those produced by lightning and continuous wave (CW) signals emitted by commercial broadcasting radio and television

stations. Even if a receiver is well-matched to the detection of broadband transients, CW signals can still degrade its sensitivity when many, powerful carriers exist within its bandwidth. Extensive experiments have been performed by the detection of natural and artificial lightning discharges in urban environment to visualize and verify the detectability of transient signals by LiNSAT payload in carrier-dominated radio environments and are discussed by (Jaffer and Schwingenschuh 2006a; Jaffer 2006b; Jaffer, Koudelka et al. 2008; Jaffer, Eichelberger et al. 2010d; Jaffer, Koudelka et al. 2010e; Jaffer 2011c; Jaffer and Koudelka 2011d)

2. Space heritage

2.1 CASSINI-HUYGENS

The LiNSAT research team in Graz is experienced in conducting field and particle experiments for planetary and interplanetary missions (Schwingenschuh, Molina-Cuberos et al. 2001). A milestone was the participation in an electric field experiment aboard ESA's HUYGENS mission, which for the first time explored the atmosphere of the Saturnian moon Titan (Fulchignoni, Ferri et al. 2005). After a 7 years cruise to the Saturnian system and two close Titan encounters NASA's CASSINI orbiter released the HUYGENS probe on 25 December 2004. On 14 January 2005 the atmosphere of Titan was first detected by the HUYGENS Atmospheric Structure Instrument (HASI) accelerometers at an altitude of about 1500 km. About 5 minutes later at an altitude of 155 km the main parachute was deployed and the probe started to transmit data of the fully operational payload. About 2.5 h later the probe landed near the equator of Titan and continued to collect data for about one hour.

The orbit of the HUYGENS probe has been reconstructed using the data of the entry phase and of the descent under the parachute. The electric field sensor of HASI carried out measurements during the descent (2 hours and 27 minutes) and on the surface (32 minutes) about 3200 spectra in two frequency ranges from DC - 100 Hz and from DC - 11 kHz. The major emphasis of the data analysis is on the detection of electric and acoustic phenomena related to lightning (Fulchignoni, Ferri et al. 2005; Schwingenschuh, Hofe et al. 2006a; Schwingenschuh, Hofe et al. 2006b; Schwingenschuh, Besser et al. 2007; Schwingenschuh, Lichtenegger et al. 2008b; Schwingenschuh, Tokano et al. 2010).

Three methods are used to identify lightning in the atmosphere of Titan:

- Measurements of the low frequency electric field fluctuations produced by lightning strokes
- Detection of resonance frequencies on the Titan surface - ionosphere cavity
- Determination of the DC fair weather field of the global circuitry driven by lightning

Several impulsive events have been detected by the HASI lightning channel. The events were found to be similar to terrestrial Sferics and are most likely produced by lightning. Large convective clouds have been observed near the South Pole during the summer season and lightning generated low frequency electromagnetic waves can easily propagate by ionospheric reflection to the equatorial region. The existence of lightning would also be consistent with the detection of signals in the Schumann range and a very small fair weather field, but there is yet no confirmation by the CASSINI orbiter.

Contrary to the HUYGENS VLF lightning detector, the LiNSAT radio receiver is planned to operate in the VHF range which is less affected by the terrestrial ionosphere.

2.2 TUGSat-1/ BRITE

The predecessor of LiNSAT is the first Austrian nano-satellite TUGSat-1/BRITE-Austria, being developed by the Graz University of Technology with University of Vienna, Vienna University of Technology and the Space Flight Laboratory of University of Toronto (Canada) as partners. The scientific objective is the investigation of the brightness variation of massive luminous stars of magnitude +3.5. The satellite has a size of 20 x 20 x 20 cm with a mass of about 6 kg and carries a differential photometer as the science instrument. It will fly in a sun-synchronous orbit.

Figure 1 shows a mock-up of the TUGSat-1/ BRITRE nano-satellite. Power is generated by multiple body-mounted strings of triple-junction solar cells. The available power is about 6 W on average. Energy is stored in a 5.3 Ah Lithium-Ion battery. The power subsystem has been designed for direct energy transfer.

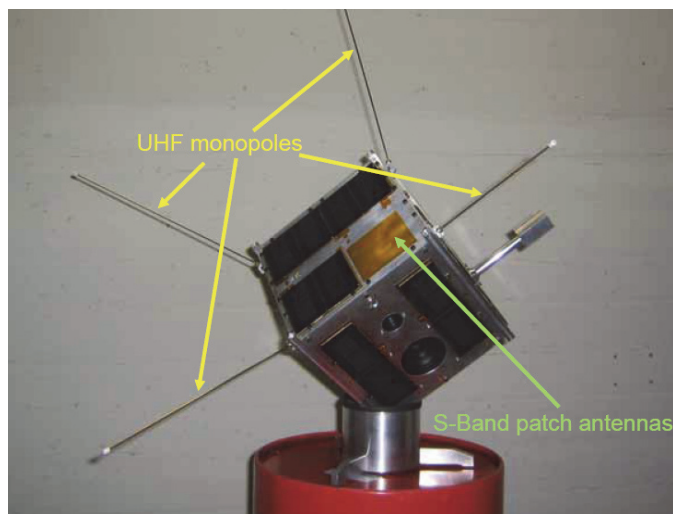


Fig. 1. TUGSat1 ADCS and antennas configuration.

A novelty is an advanced attitude determination and control system (ADCS) with three very small momentum wheels and a Star Tracker providing a pointing accuracy at the arc-minute level. Three on-board computers are installed in the spacecraft, one for housekeeping and telemetry, one for the science instrument and one for the ADCS.

The telemetry system comprises an S-band transmitter of about 0.5 W. It is capable for transmitting data with a minimum rate of 32 kbit/s. Data rates of up to 512 kbit/s are feasible with existing ground stations. The uplink is in the 70 cm band. A beacon in the 2 m band is also implemented (Koudelka, Egger et al. 2009).

Other objectives of TUGSat1 and LiNSAT are training of students, hands-on experience in conducting of a challenging space project and synergies between several scientific fields. The investigation of massive luminous stars with a precise star camera opens up new dimension for astronomers as observation of stars without interference by earth atmosphere can be carried out in LEO with such a small and low-cost spacecraft. Moreover, LiNSAT is a pure student satellite and will contribute as a low-cost atmospheric research platform.

3. Lightning discharges and classification

Lightning is a hazard, and sometimes a killer, as observed with severe storms and lightning strikes. Worldwide thousands of individual lightning discharges, including dramatic bolts, occur each day. The map in Figure 2 shows the geographic distribution of the frequency of strikes averaged over 8 years (1995 - 2003) of data collecting by NASA's Optical Transient Detector (OTD) and by the Lightning Imaging Sensor (LIS) on Tropical Rainfall Measurement (TRMM) (NASA 2011b). As indicated by the figure, the central part of Africa has been an area with the most lightning strikes; almost all of South America is prone to frequent electrical storm activity. A different perspective by looking at the distribution of lightning strikes worldwide average lightning strikes per square km per year can be seen in the same figure.

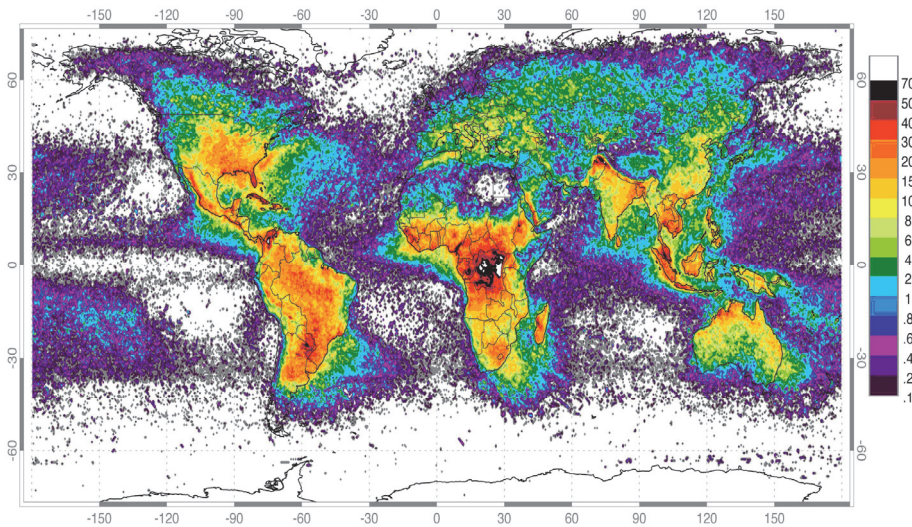


Fig. 2. Global distribution of lightning April 1995 – February 2003 from the combined observation of the NASA OTD (4/95-3/0) and LIS (1/98-2/03) instruments. Courtesy NASA TRMM team (NASA 2011a).

Each lightning strike exhibit unique signature and streak lightning being the most-commonly observed in the world. That is actually a return stroke (RS) that is the visible part of the lightning stroke. The majority of strokes occur within a cloud or clouds (IC), as indicated by Figure 3. Fair weather field is ~ 150 V/m close to the surface of earth. this field changes significantly the cloud electric field underneath (~ 10 kV/m) and within (~ 100 kV/m) the cloud (Uman 2001).

4. Lightning detection

It is well known from the very beginning of radio technology that lightning is a source of interference in amplitude-modulated (AM) radio reception. In fact before the radio use in transmissions/ broadcasting had shown that lightning causes distinctive noise in a radio channel, so that in this sense lightning detection can be said to pre-date other uses of the radio. Radio measurements of lightning were made extensively until the 1960's, although

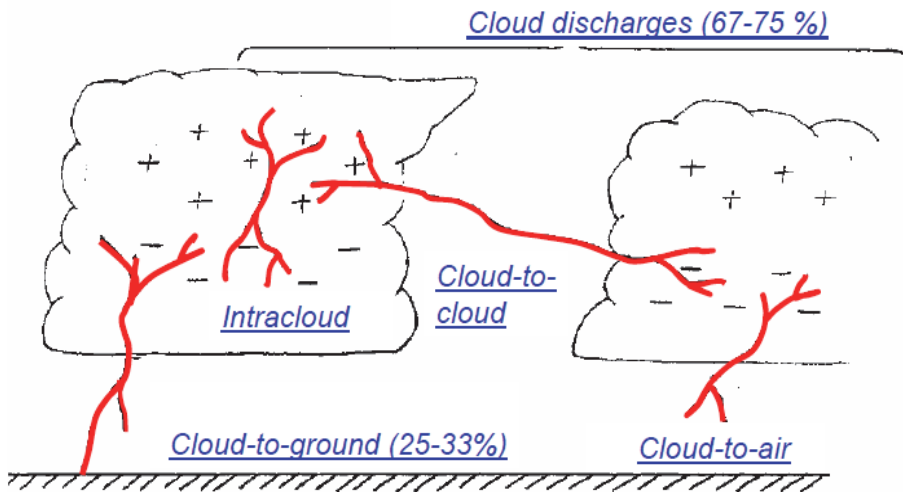


Fig. 3. The lightning classification indicates that almost two third of the lightning discharges occur within and inter- cloud that has direct impact on air traffic. Adapted from (Rakov and Uman 2003).

mainly with the purpose of improving radio transmissions. Some differences between optical and RF detection are elaborated in Figure 4.

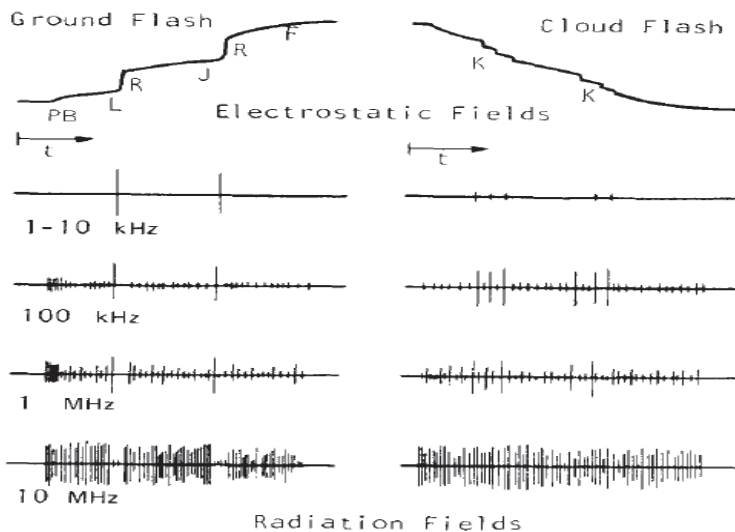


Fig. 4. Electromagnetic fields in lightning associated different channels, Preliminary Breakdown (PB), Leader (L), Return Stroke (R), Changes in (J, F and K). Adapted from ((Le Vine 1987)).

The major advantage of RF detection of lightning over optical is that optical detection is unable to distinguish many signatures like CG versus IC, RS, Leader and TIPPs etc. Also, the atmosphere has least effect on electromagnetic waves (Suszcynsky, Kirkland et al. 2000).

5. LiNSAT constellation, space and ground segments

5.1 Space segment

Space segment consists of constellation of three identical nano-satellites and each satellite is comprised of following units

- Attitude control through gravity gradient boom (GGB) Figure 5.
- Three orthogonal lightning antennas (GGB-LA, LA2, LA3), one antenna (GGB-LA) is integrated into GGB at nadir direction, Figure 5.
- Power subsystem (Solar panels, battery charge and discharge regulator (BCDR) and battery).
- Thermal subsystem
- VHF electronics
- Data processing unit (DPU)
- On-board event detector together with Adaptive Filtering
- Housekeeping
- Communication UHF/VHF monopole antenna (U-MP/V-MP): 2m and 70 cm frequency bands and S-band patch antenna

All subsystems are shown in Figure 6. The components selection criteria for on-board memory, telemetry volume, and power budget are the “cost effectiveness” therefore, commercial off the shelf components (COTS) will be used.

As a heritage from TUGSat1, the same Generic Nano-satellite Bus (GNB) (de Carufel 2009) will be used on LiNSAT with data rate from 32 – 256 kbps. Data volume per day will be ~ 15 MB/day. For LiNSAT, the actual amount of data is mode-dependent. The requirement of on-board memory with optimum data volume ~ 150 MB/day (3 GS) with 256 kbps data rate is possible. Based on studies done by the SPOT team (Barillot and Calvel 2002), around 8 events upset per year occur in LEO 800 km orbit. Countermeasures for the memory are necessary and the cold redundancy is considered. Additionally, a current limiter is foreseen to reset the experiment and the sub-systems on-board LiNSAT.

5.2 Constellation: local and global coverage

After launch and commissioning phase, all three satellites will be close together (local small scale coverage) and would be used for combined investigations (TOA). Due to natural orbital variations they separate from each other and, in the long run, the satellites will no longer remain in the constellation (global intermediate and large coverage). At this point they will be treated as individual entity (LiNSAT).

5.3 Ground segment

The global coverage emphasizes on the main purpose of ground segment as distributed GSs network (DGSN) to track the satellite and receive housekeeping and scientific data on global scale in real-time. DGSN consists of three GS

- Automated remote GS at Graz University of Technology, Austria (AR-TUG), (Jaffer and Koudelka 2011e)
- I-2-O gateway (Hermes-A) in Ecuador
- One proposed GS at Lahore, Pakistan (LiNSAT-GS)

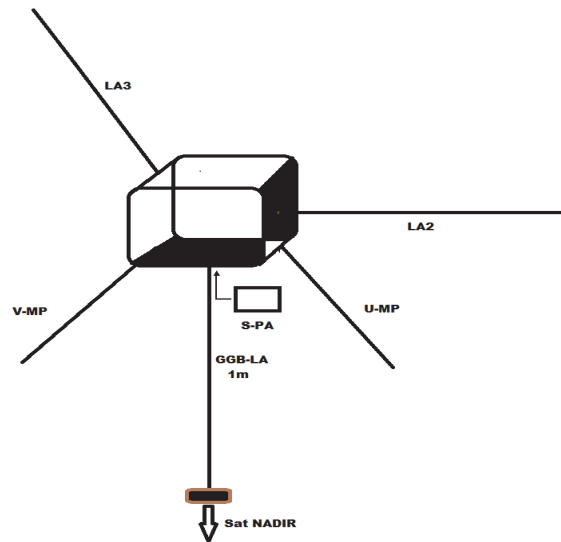


Fig. 5. LiNSAT structure with antenna configuration of three orthogonal lightning antennas (GGB-LA, LA2 and LA3). GGB is a passive attitude control sub-system to nadir direction. Additionally the multi-purpose boom is integrated as lightning antenna.

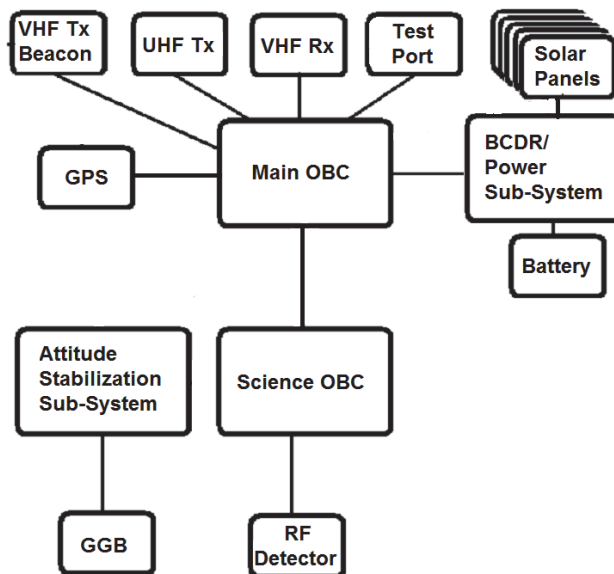


Fig. 6. Block diagram of all subsystems of LiNSAT. The science on-board computer (OBC) and ADCS OBC are connected to main OBC. Other subsystems like Communication and power are interfaced with main OBC.

First two GS are already functional and the third is proposed and we will pursue its development in near future.

From the scientific data reception point of view, DGSN opens up the mission to a wider range. Therefore, DGSN would eventually make difference as compared by the amount of data collected manually with standalone GS. Moreover, transformation of DGSN as autonomously operating network is foreseen. This would ultimately support future nano-satellite experiments effectively. Additionally, its scheduling capability which we pursue in near future will enhance its functionalities. The I-2-O gateway and virtual GS are detailed in (Jaffer, Klesh et al. 2010a; Nader, Carrion et al. 2010a; Jaffer, Nader et al. 2010b; Nader, Salazar et al. 2010b; Jaffer, Nader et al. 2010f; Jaffer, Nader et al. 2011a; Jaffer, Nader et al. 2011h; Jaffer, Nader et al. 2011i). The setup is shown in Figure 7.

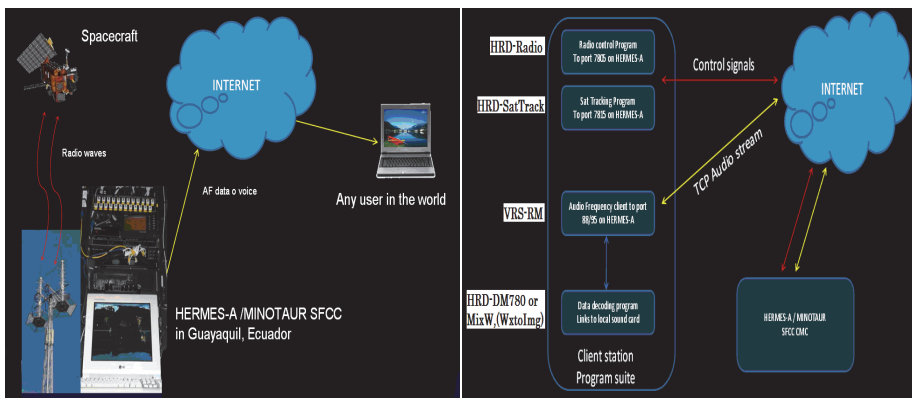


Fig. 7. LiNSAT ground segment, *Left*: HERMES-A set-up and working scenario, *Right*: virtual ground station at remote user end

5.4 LiNSAT modes of operation

LiNSAT consists of several mission modes (MO) of operation after successful deployment (Table 1). The LiNSAT is planned to perform functions in various modes depending upon task/ idle situations.

Mission Operation Mode No.	Tasks, Annotation
1	Deployment, <i>Commissioning phase</i>
2	Stabilization, <i>Commissioning phase</i>
3	Lightning Experiment, <i>Scientific Payload</i>
4	Ground Communication, <i>Telemetry</i>
5	Conserve Power/ Recharge, <i>in Eclipse</i>
6	Standby

Table 1. LiNSAT mission operational (MO) modes.

Modes 1 and 2 apply to initial deployment and stabilization of LiNSAT. Mode 3 is the key mode with five experiment mode options (Table 2). Switching to mode 4 occurs for telemetry operations. Mode 5 is applicable when the power subsystem is no longer capable to support normal operations, e.g. during eclipse. Mode 6 is the default mode when no other operations are going on. Either telemetry or lightning experiment will be carried out at a time for efficient use of on-board power.

For LiNSAT, EMC is a vital part to avoid intra- and inter-system disturbances from electromagnetic radiations and coupling with special considerations in the 20 - 40 MHz VHF range. The printed circuit board (PCB) layout tools considering routing and grounding concepts of the highly integrated electronics together with shielding and harness strategies. Verification on board level occurs with the aid of EMC pre-compliance measurements and validation of the system in an anechoic chamber. Fine tuning of the components and adaptive filtering (AF) of unwanted signals results in a high signal to noise ratio (SNR).

Electromagnetic pulses from DC converters could produce spurious signals similar to lightning spikes even in the same frequency range. Laboratory tests are currently carried out in order to optimize antenna, receiver and adaptive filter design.

The scientific payload on-board LiNSAT will perform detection of lightning events, measurement of time series and transmit to one of the GS within communication window. This mode is further sub-categorized and elaborated in (Table 2).

Experiment Mode	Tasks, Annotation
Survey Mode	Statistical; Number of lightning events above threshold level (coarse trigger)
Event Mode	Sferics time series data dumped into on-board memory
Background / CW Mode	All events/ signals, galactic, interference detected from ionosphere between two lightning events
Test Mode	Software test/ Polar lightning (If any)
EMC Mode	Artifacts from satellite itself

Table 2. LiNSAT experiment modes closely associated with mission modes (Table 1)

5.5 Payload instrumentation

5.5.1 Antennas and gravity gradient boom

The voltage received at LiNSAT

$$V = h_{\text{eff}} * E, \quad (1)$$

and depends on the lightning electrical field E and the effective length h_{eff} of the antenna. Also, $h_{\text{eff}} \sim h_m/2$ for $h_m \ll \lambda$, where h_m is the mechanical length of the antenna and λ the wavelength.

No stringent pointing requirements for LiNSAT are foreseen as compared with TUGSAT-1/BRITe, so a simple and inexpensive gravity gradient stabilization (GGS) technique already proven on many missions (NASA 2011b) and detailed in (Wertz and Larson 1999); (Wertz 1978) which points to the nadir of the satellites envisaged for satellite attitude control. The GGB fulfills the requirements and is selected due to its economical features like least power consumption (once during deployment) and the cheapest of all other stabilization mechanisms of the same breed.

The GGB, acting as an antenna for lightning detection, is a deployable with 10 % of the satellite mass (tip mass). A three-antenna system has multifold advantages, like redundant directional capability and simultaneous back up. In-orbit characteristics depend on several factors, e.g. mechanical forces, non-conservative forces and induced pendulum motions. The boom torque needs to overcome environmental torque for a maximized stabilization capability. The antenna works in non-resonant mode so lightning investigation is performed in different frequency ranges but with reduced efficiency.

5.5.2 Data Processing Unit (DPU)

The signal from the antenna is fed to the data processing unit (DPU) through a pre-amplifier prior to analog/digital conversion and further processing. As mentioned earlier, typical lightning electric fields at 1000 km altitude are 1mV/m at 40 MHz, 1 MHz bandwidth (Jacobson, Knox et al. 1999). After filtering and amplification, the 200 MS/s ADC will sample the received signals and the digitized signals are dumped into cyclic memory with the help of two levels of event detection (coarse and fine). Data acquisition captures a waveform record. A triggering unit helps in elimination of unwanted signals stemming from galactic and magnetospheric origins.

5.5.3 Data acquisition system (DAQ)

The data acquisition system (DAQ) is capable of retriggering a new record within microseconds of the end of the previous record. Data of many lightning events will be stored in cyclic memory/ buffer. The memory is capable to be overwritten all the times. A significant number of events can be stored in a solid-state mass memory before downloading via telemetry to the ground segment. The minimum telemetry transfer rate for science data is 180 kbytes per day. 256 MB of flash memory for long-term storage of measurement data are foreseen. The Sferics data will be identified with minimum pulse width $\sim 50 \mu\text{s}$ and sharp rising amplitude with pulse rise time $\sim 10 \text{ ns}$. The records will then be analyzed on ground to investigate VHF signatures in time and frequency domains. The payload configuration is shown in Figure 8.

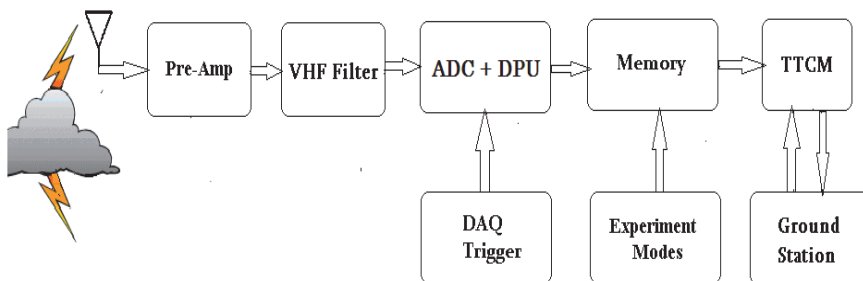


Fig. 8. Block diagram of LiNSAT scientific payload.

Event detector is an important part of the detection system; a signal processing subsystem performing various signal processing functions to classify the signals into distinct categories.

6. Adaptive filtering

The adaptive filtering (AF) structure shown in Figure 9 is based on and draws heritage from adaptive noise cancellation (ANC) (Haykin 1996). The input signal $d[k]$ is a lightning transient pulse that is contaminated with artifacts $a_d[k]$ from LiNSAT. The co-efficients for the reference signal $x[k]$ can be derived from two sources, either by ground-based EMC investigations as a preliminary estimation as pre-selector co-efficients for AF (GPC) or the output from the sensors on-board nano-satellite. The goal is to detect natural lightning spikes, so finally we'll have to rely on on-board sensor as an ultimate for updating the co-efficients of AF. The reason behind could be

- slightly shifting of subsystems emissions curve within scientific payload measurement range
- cancellation of disturbances generated by LiNSAT or subsystems, $a_x[k]$.

Electric field emissions in the measurement range 20-40 MHz, from on-board subsystems has got an emission curve as a proxy of the artifacts generated by LiNSAT. Being a narrowband disturbance, it will be eliminated using a notch filter. It can happen that the frequency of the emission curve shifts within the measurement range, therefore, we need to update the filter co-efficients to track the movement in a robust manner.

We are using coarse and fine triggering mechanism for the lightning detector on-board LiNSAT and the adaptive filtering approach will be used in coincidence with the fine trigger on-board the satellite (DAQ Trigger, Figure 8) for redundancy.

The orbital footprint of the LiNSAT will be scanning the whole Earth during several orbits but as we know the lightning flash rate varies along the orbit from the poles to equator as shown in Figure 2. This is valuable statistical environmental input for the $d[k]$ channel for precise triggering and to avoid false alarms.

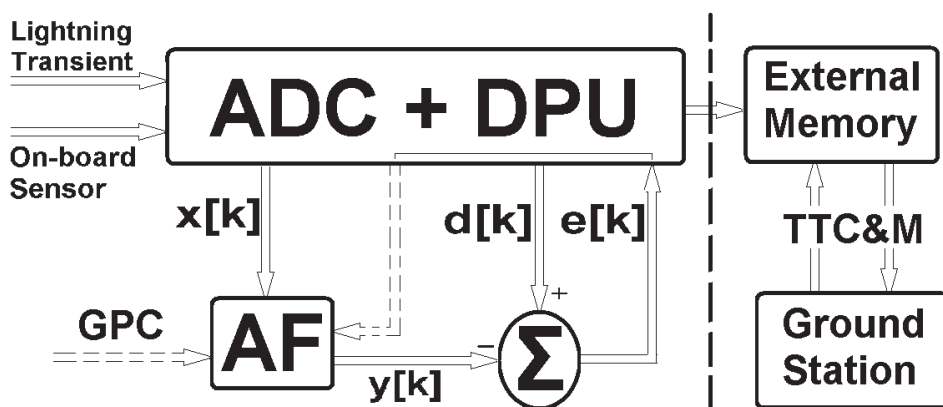


Fig. 9. DAQ trigger for lightning detector on-board LiNSAT based on adaptive filters

$x[k]$ is a digitized output of the sensor on-board which detects artifacts in terms of electric field emissions for LiNSAT itself. This could be broadband or narrowband, e.g. a clock signal or harmonics of the digital electronics

$$x[k] = a_x[k]. \quad (2)$$

$d[k]$ is digitized signal of the lightning transient after passing through antenna, VHF filter and pre-amplification including noise picked up in this receiving channel,

$$d[k] = s[k] + a_d[k], \quad (3)$$

where $a_d[k]$ and $a_x[k]$ are correlated noise sources.

AF is an FIR filter while the initial co-efficients for the filter are derived from ground based investigations (GPC) during development phase.

$y[k]$ is the output of the AF that sums up with $d[k]$ to produce error signal that ultimately is used for

1. Trigger purpose for the on-board lightning detector to dump the ring buffer/ cyclic memory contents into external memory for future download using telemetry, tracking, command and monitoring (TTC&M) by one of the ground stations through visibility/ communication window
2. To modify the coefficients of the AF accordingly using LMS algorithm.

Among several requirements, e.g. robustness, tracking speed and stability of the AF, the on-board computational power of the DPU is one of the constraints in space.

The filter is in the digital domain and is application dependent. The output is noise-removed lightning signal which will trigger the ring buffer to store the transients for future download through all three GS.

To determine the capability of the filter, we tested it with real life signals i.e. artificial discharges in high voltage chamber (section 9.2) and natural signals (section 9.3) in particular TIPP event recorded by ALEXIS satellite (Massey, Holden et al. 1998; Jacobson, Knox et al. 1999). The outputs were found to be as close as noise-free real lightning transients. The major advantage of the output (error signal) is that it would trigger the memory (on-board LiNSAT) with lower threshold level. Therefore, lightning pulses with even small peaks will be captured using adaptive filter fine trigger, otherwise would be hard to capture due to higher noise floor. The out puts are shown in Figure 10 and Figure 11.

The code to generate through such signals using AF through Matlab function is elaborated in Table 3.

```
function
[E,Y,noise,sig_plus_filterednoise]=anc_lightning(signal,filterorder,cutoff_frequency)
% [E,Y,noise,sig_plus_filterednoise]=anc_lightning(b23(1.18*10^6:1.28*10^6,1),4,0.4);
noise=0.1*randn(length(signal),1); .....% noise modeling
nfilt=fir1(filterorder,cutoff_frequency); .....% filter order LP
fnoise=filter(nfilt,1,noise); .....% correlated noise data
sig_plus_filterednoise=signal+fnoise; .....% input plus noise
coeffs=nfilt-0.01; .....% filter initial conditions
mu=0.05; .....% step size for algorithm updating
S=initlms(coeffs,mu); .....% Init LMS FIR filter
[Y,E,S]=adaptlms(noise,sig_plus_filterednoise,S);...% Y=filtered data, E=Prediction Error
```

Table 3. Matlab code to test Adaptive filter algorithm

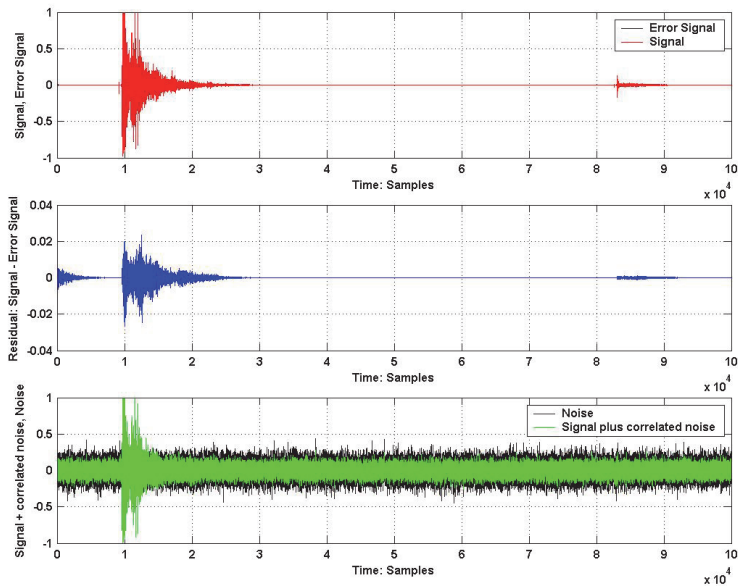


Fig. 10. Acoustic signal generated in high voltage chamber and received using Adaptive filter.

The other signal of interest to test AF is natural lightning (i.e. TIPP) captured with ALEXIS satellite shown in Figure 11. Again, the noise-removed signal can be observed.

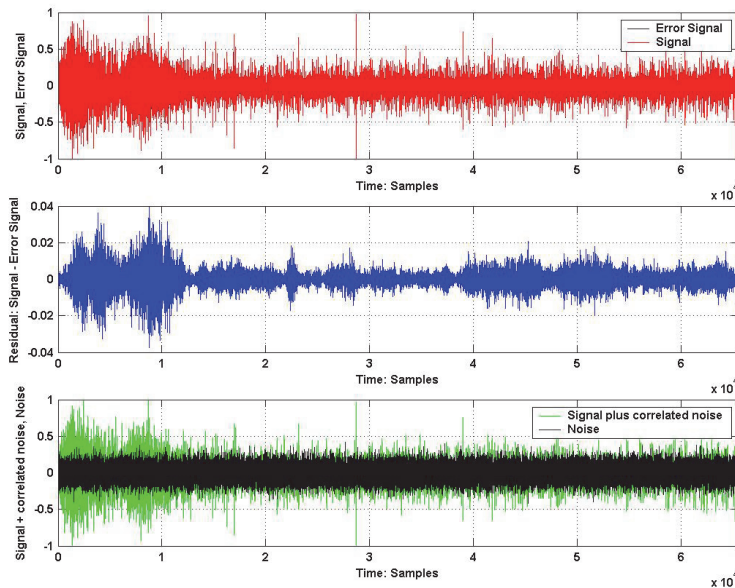


Fig. 11. TIPP event captured with Adaptive filter. The error signal is close to real input, an advantage of using AF to trigger on-board memory for saving noise-free transients.

Based on previous experiments using engineering models and TUGSat-1 study, the DAQ power estimations are ~ 3 W on average up to 5 W (peaks) for short time. Power consumption is mainly determined by the consumption of digital electronics. Among several requirements, e.g. robustness, tracking speed and stability of the AF, the on-board computational power of the DPU is one of the constraints in space. A technique to overcome this deficiency on-board is power aware computing (Graybill and Melhem 2002) and one case-study for FORTE mission is discussed in (Shriver, Gokhale et al. 2002).

7. Attitude control system

Unlike TUGSat-1/ BRITE-Austria which is three-axis stabilized, LiNSAT will use the widely used and inexpensive gravity gradient stabilization (GGS) technique for its attitude control. The deployable GGB has a length of 1 m. The vector of the boom and its counterweight will be rotating around a vector pointing directly towards the center of the Earth. If this oscillation can be dampened, it is possible to control the attitude of the satellite such that the nadir surface points towards the Earth within limits of $\pm 10^\circ$, (Taylor-University 2011). This is sufficient for antenna pointing. The setup from Figure 5 benefits from the RF noise survey (Burr, Jacobson et al. 2004; Burr, Jacobson et al. 2005) and enables investigations of major types of lightning, including TIPPs (Holden, Munson et al. 1995; Massey and Holden 1995; Massey, Holden et al. 1998; Tierney, Jacobson et al. 2002).

A redundant approach (passive magnetic and GGB) is being used by Quakesat (Quakesat 2011). It used GGB back in 2003 and still operating. The ADCS of this satellite is primarily a magnet, but it extends a magnetotorquer out on a boom providing more than the average effects from gravitational gradient. As without a large tip mass, the gravity gradient effect is reduced, so magnets are needed to reduce damping effects and ultimately stabilizing pointing accuracy.

The tip mass is comparable to total mass of the satellite (10%). In principle the GGB works for imaging payload with a pointing accuracy $\sim \pm 5^\circ$, but roll of the satellite must be taken into account. There are only a few small satellites with "camera" as optical payload using GGB and/ or passive magnets (no 3-axis ADCS) and manage the pointing requirements. Although for communication link, GGB technique works well as long as we assume the roll rate is slow, we can probably take a picture without smear, but only if the assumption holds or the camera is designed to counterbalance this rolling effect.

8. Experimental setup and results

8.1 Lightning detection simulation

In Austria, lightning flash density is between 0.5 and 4 flashes per square km per year, depending on terrain (Diendorfer, Schulz et al. ; Schulz and Diendorfer 1999; Diendorfer, Schulz et al. 2002; Schulz and Diendorfer 2004; Schulz, Cummins et al. 2005). Lightning flash occurrence duration is about half a second. The simulated signals are evaluated with coarse and fine triggered detection using "lightning detector" program written in Matlab. Also, execution and processing time of 10,000 data points are checked in Matlab on standard PC and will be compared to Microprocessor time after transformation into machine language. Execution time in Matlab found to be 5 s on the average. As the algorithm contains many loops, so the computation time is higher.

The blocks of the receiving chain (Figure 8) are simulated individually and as a whole using Matlab functions. To verify the simulations, lab measurements (RF and Acoustic) are performed. The objective of these measurements/ simulations was to find algorithm which could implemented on the on-board lightning detector electronics (requirements: must be fast, effective and running with limited resources).

One second lightning strokes data for simulation using the program is analyzed. The program computes the pulse width, pulse rise time, number of detected pulses, frequency and pulse-amplitude and stores all the information in a resultant matrix; emulation of cyclic memory. The algorithm is mode dependent and basically captures lightning signals above threshold level. It classifies and characterizes the pulses. A standard class of lightning signals was used to see e.g. total electron contents (TEC) effects on lightning transients passing through ionosphere. The algorithm will work along with experiment modes (Table 2).

For coarse detection, the whole window was divided in one thousand 1 ms sub-windows for computing mean and standard deviation (power) in each sub-window above threshold level of noise floor. This procedure will help us in pre-selection of the data and to change the experimental modes. Telemetry using coarse trigger will be downloaded for ground based analysis. The resultant matrix of the program is shown in Figure 12. Other consideration is mode-switching while above Equator, Poles etc with the lightning detector on-board LiNSAT.

PULSE=>No.	Start-Time	Widthth	thLevel	AmplitudeRMS	Rise-Time	Max-Frequency
resultantMatrix =						
1.0000	4.0000	2.0000	16.0000	17.3449	17.8759	0.0196
2.0000	7.0000	2.0000	16.0000	17.9203	21.3362	0.0164
3.0000	13.0000	2.0000	16.0000	19.1966	28.3573	0.0123
4.0000	19.0000	3.0000	16.0000	18.0037	33.4029	0.0105
5.0000	27.0000	1.0000	16.0000	17.6707	41.1366	0.0085
6.0000	31.0000	1.0000	16.0000	22.0093	48.6074	0.0072
7.0000	34.0000	1.0000	16.0000	21.3789	51.1032	0.0068
8.0000	37.0000	2.0000	16.0000	21.6722	54.3378	0.0064
9.0000	40.0000	1.0000	16.0000	16.7201	53.3761	0.0066
10.0000	47.0000	1.0000	16.0000	22.8012	65.2410	0.0054
11.0000	50.0000	12.0000	16.0000	134.4105	157.5284	0.0022
12.0000	64.0000	2.0000	16.0000	19.4476	79.5581	0.0044
13.0000	70.0000	1.0000	16.0000	21.2800	87.0240	0.0040
14.0000	74.0000	2.0000	16.0000	18.5694	88.8555	0.0039
15.0000	78.0000	2.0000	16.0000	17.2528	91.8023	0.0038
16.0000	81.0000	1.0000	16.0000	19.4400	96.5520	0.0036
17.0000	84.0000	1.0000	16.0000	21.9625	101.5700	0.0034
18.0000	88.0000	1.0000	16.0000	20.0368	104.0295	0.0034
19.0000	92.0000	1.0000	16.0000	21.6809	109.3447	0.0032
20.0000	98.0000	1.0000	16.0000	18.6548	112.9238	0.0031

Fig. 12. Matlab™ simulated results of the program "lightning detector". In this computation, threshold level for coarse detection is set to 16 mV to extract all pulses above noise floor. Twenty pulses were detected along with their information about pulse start time, pulse rise/ fall time, amplitude and maximum frequency.

8.2 Artificial lightning measurement campaigns

8.2.1 Electrical discharges in high voltage chamber

In these measurement campaigns, artificial lightning was produced in laboratory at Space Research Institute, Austrian Academy of Sciences, Graz and Institute of High voltage Technology and System Management of the Graz University of Technology high voltage chamber. The artificial lightning discharge was from 700 kV to ~ 1.8 MV under normal pressure and temperature conditions. The system has selectable +ve and -ve polarity and electrodes inter-distance. Two discharges using different models are shown in Figure 13. The lightning measurement setup consisted of electric field probe (60 Hz to 100 MHz) and digital oscilloscope (bandwidth = 200 MHz) with external and manual trigger (pre-, post-trigger) option.

We focus mainly on the received time series including noisy features to extract characteristic parameters. We determined the chamber inter-walls distance by considering reflections in the measurement. Round-trip-time (direct and reflected) waves indicate many reflections from all walls of the chamber. The captured image shows similarities with natural lightning signal shown in Figure 20. All experimental evidences are shown in figures below.



Fig. 13. *Left:* Lightning event occurrence on a model. Also branched lightning (to ground and air) was observed. *Right:* Color changes (intensity) due to discharges along the rope. The variations indicate current interaction with fibers of the rope under impact of lightning discharge of ~ 1.8 MV. Post-event inspection revealed no damage in the rope, macroscopically.

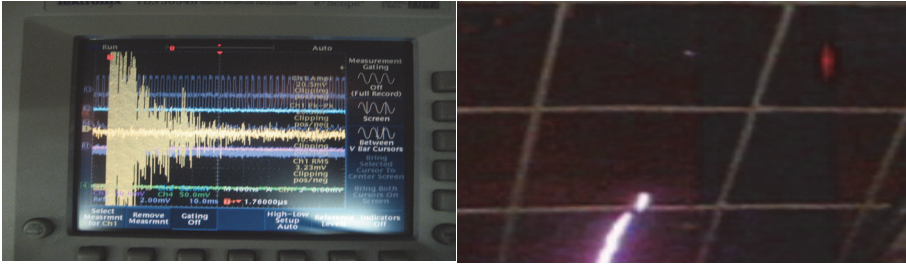


Fig. 14. *Left*: Lightning electric field captured with oscilloscope. A round-trip-time (direct and reflected) wave indicates many reflections from all walls of the chamber. The captured image shows similarities with natural lightning signal shown in Figure 20, *Right*: The striking phenomenon, upward lightning initiated by sharp objects in the vicinity of chamber ground was observed. In principle, this phenomenon occurs with sharp endings tall buildings like church etc. It has similarity with return stroke (RS) phenomenon.



Fig. 15. *Left*: Basic structure of the tent as a model for the Faraday cage. There can be arcing if some metal is present inside the tent (no lightning safety), *Mid*: Lightning event occurs near the object due to slight moment of the electrode. *Right*: The measured electric field due to arcing (impact) of the lightning discharge on the model of Faraday cage.

8.2.2 Acoustic measurement

We performed acoustic measurements in parallel to RF detection of artificial lightning in high voltage chamber (Eichelberger, Prattes et al. 2010; Eichelberger, Prattes et al. 2011). The outcomes helped to correlate RF signatures.

8.2.3 Electrical discharges in a lab

Besides the measurements from the previous section, we wanted to determine all associated phenomena with natural lightning, e.g. corona discharges, glow and sparks. Moreover, these phenomena were easily produced in the lab due to controlled setup. But it was not necessary to perform these measurements in the high voltage chamber. This setup helps to investigate distinct features of all three mentioned discharge types and the transition among all these phenomena related to voltage in a pre-defined setup.

The setup comprises of

- High voltage power supply unit (HV-PSU)
- Adjustable electrodes: a metal plate (Anode) and screw (Cathode) combination with inter-distance of ~ 1 inch (2.54 cm)
- Electric field probe (60 Hz to 100 MHz)
- Digital Oscilloscope (eTek, 200 MHz bandwidth)

8.2.3.1 Corona discharges

The very first broadband phenomenon occurs with a hissy sound due to increase in the voltage. The maximum corona acoustic along with glow was observed at 11 – 12 kV.

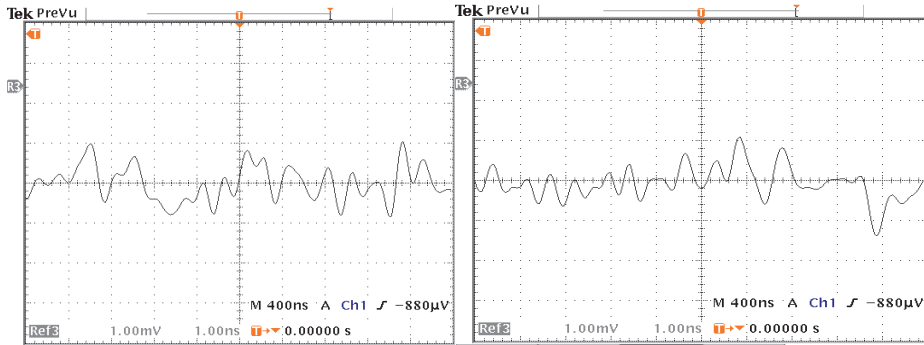


Fig. 16. eTek Oscilloscope display of Corona discharges.

8.2.3.2 Glow discharges

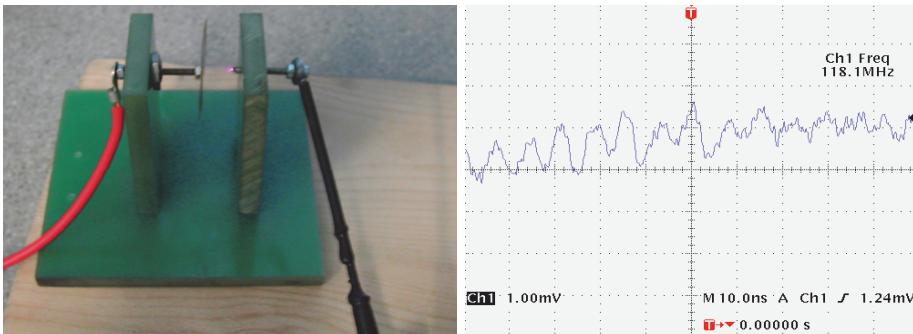


Fig. 17. Left: Artificial lightning: indication of corona and glow discharge, Right: Glow discharge measurement. Frequency ~ 118 MHz. Applied Voltage is 11 kV. The maximum corona sound with glow was observed at 11 - 12 kV and decaying afterwards.

8.2.3.3 Spark discharges

The HF and VHF radiations associated with initial breakdown pulse in cloud flashes were studied by (Krider, Weidman et al. 1979) and they found that radiations at 3 MHz, 69 MHz, 139 MHz and 259 MHz tend to peak during the initial half cycle of the pulse. We have also observed the feature in the lab at 139.6 MHz shown in Figure 18.

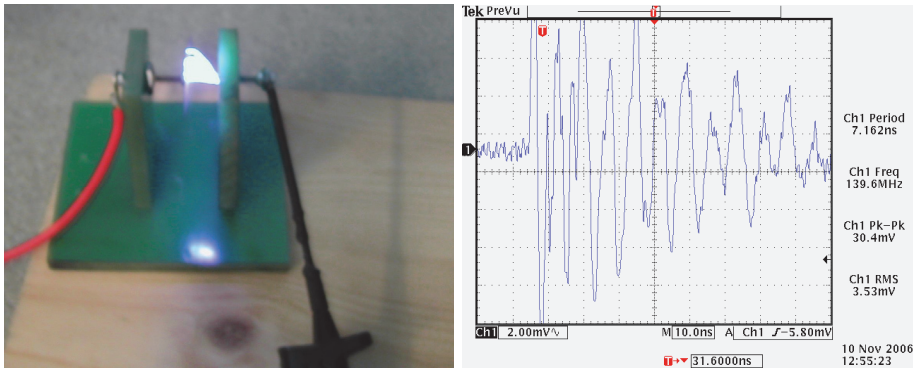


Fig. 18. *Left:* Artificial lightning, spark discharges on cathode. The maximum frequency observed for spark was 140 MHz. The spark was produced at 13 kV and higher voltages, *Right:* Spark discharge measurement, the maximum frequency observed for Spark is 140 MHz. The spark produced at 13 kV and higher, also measured on oscilloscope.

8.3 Natural lightning measurement

During intense thunderstorm activity on June 30, 2010, in urban area of Graz, Austria, natural lightning measurements were performed using broadband discone antenna, 15 m shielded cable and digital oscilloscope (Bandwidth = 200 MHz) to correlate with artificial lightning discharges measured in high voltage chamber. The radiation patterns of such antenna are shown in Figure 19.

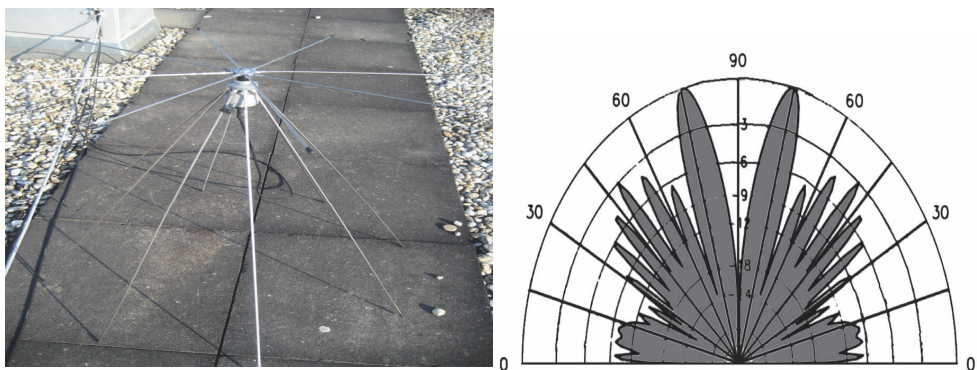


Fig. 19. *Left:* The broadband discone antenna used for natural lightning measurements. The antenna was put on roof of the Graz University of Technology building for better reception and to avoid interferences within the campus, *Right:* Radiation patterns of discone antenna (DA-RP 2011).

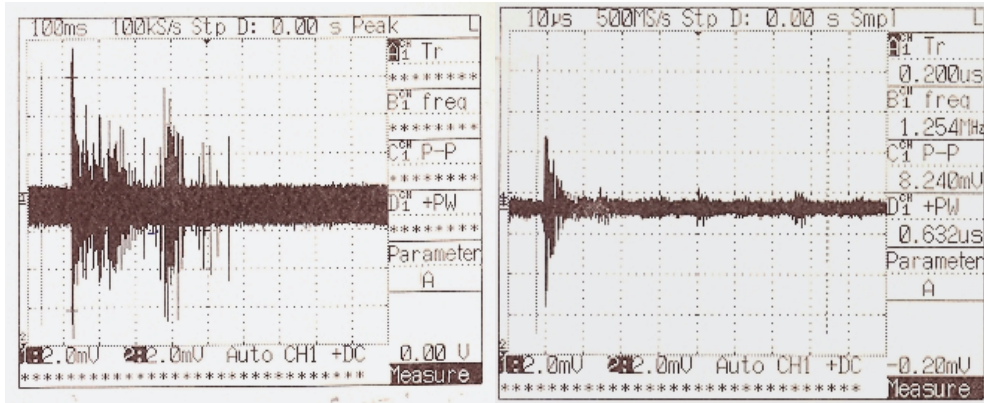


Fig. 20. *Left*: Natural lightning measurement with digital oscilloscope (Bandwidth = 200 MHz), with sampling rate 100 kS/s. It shows two individual strokes within a lightning flash, *Right*: Natural lightning measurement with digital oscilloscope (Bandwidth = 200 MHz) with sampling rate 500 MS/s indicates a single stroke with a few reflections.

No.	f_{sampling}	$V_{\text{p-p}}$	V_{noise}	t_{rise}	t_{fall}	$t_{\text{inter-pulse}}$
Figure 20 (Left)	100 kS/s	18 mV	2 mV	10 ms	200 ms	250 ms
Figure 20(Right)	500 MS/s	6 mV	1 mV	1 μ s	5 μ s	15 μ s

f_{sampling}	Sampling frequency of the oscilloscope
$V_{\text{p-p}}$	Peak-to-peak voltage
V_{noise}	Noise floor
t_{rise}	Pulse rise time (10-90% of the peak voltage)
t_{fall}	Pulse fall time (90-10% of the peak voltage)
$t_{\text{inter-pulse}}$	Time between two pulses (reflections, TIPP etc)

Table 4. Natural lightning: setup and obtained resultant parameters

9. Data analysis conclusions

The measurements from the HV chamber and natural environment have been evaluated in the time domain. We also determined statistically that how the rise/ fall time for each stroke is different and relevant to indicate unique signature of each sub-process of lightning event. The envelope of the signal is analyzed

- Events: by coinciding the size of the HV chamber (reflections) with the signal trace
- The ambient noise (and carrier) properties in these measurements
- Out of these results we have deduced the requirements for the lightning electronics of the LiNSAT (sample rate, buffer size, telemetry rate)
- The Fourier transform of the signals (frequency domain) helped in indicating the bandwidth of the lightning detector on-board LiNSAT.

10. Summary and conclusions

We presented a feasibility study of LiNSAT for lightning detection and characterization as part of climate research with low-cost scientific mission, carried out in the frame of university-class nano-satellite mission. In order to overcome the mass, volume and power constraints of the nano-satellite, it is planned to use the gravity gradient boom as a receiving antenna for lightning Sferics and to enhance the satellite's directional capability.

We described an architecture of a lightning detector on-board LiNSAT in LEO. The LiNSAT will be a follow-up mission of TUGSat1/BRITE and use the same generic bus and mechanical structure. As the scientific payload is lightning detector and it has no stringent requirement of ADCS to be three axis stabilization, so GGS technique is more suitable for this mission.

In this chapter we elaborated results of two measurement campaigns; one for artificial lightning produced in high voltage chamber and lab, and the second for natural lightning recorded at urban environment. We focused mainly on the received time series including noisy features and narrowband carriers to extract characteristic parameters. We determined the chamber inter-walls distance by considering reflections in the first measurements to correlate with special lightning event (TIPPs) detected by ALEXIS satellite.

The algorithm for the instruments on-board electronics has been developed and verified in Matlab™. The time and frequency domain analysis helped in deducing all the required parameters of the scientific payload on-board LiNSAT.

To avoid false signals detection (false alarm), pre-selectors on-board LiNSAT are part of the Sferics detector. Adaptive filters are formulated and tested with Matlab functions using artificial and real signals as inputs. The filters will be developed to differentiate terrestrial electromagnetic impulsive signals from ionospheric or magnetospheric signals on-board LiNSAT.

11. Acknowledgements

Authors wish to thank Prof. Stephan Pack for RF measurements in high voltage chamber. We are grateful to Ecuadorian Civilian Space Agency (EXA) and Cmdr. Ronnie Nader for providing access to the Hermes-A. Many thanks to Prof. Klaus Torkar for valuable discussions and comments. This work is funded by Higher Education Commission (HEC) of Pakistan.

12. References

- Barillot, C. and P. Calvel (2002). "Review of commercial spacecraft anomalies and single-event-effect occurrences." *Nuclear Science, IEEE Transactions on* 43(2): 453-460.
- Burr, T., A. Jacobson, et al. (2004). "A global radio frequency noise survey as observed by the FORTE satellite at 800 km altitude." *Radio Science* 39(4).
- Burr, T., A. Jacobson, et al. (2005). "A dynamic global radio frequency noise survey as observed by the FORTE satellite at 800 km altitude." *Radio Science* 40(6).
- DA-RP (2011). <http://www.moonraker.com.au/techni/discs&cones.htm>.

- de Carufel, G. (2009). Assembly, Integration and Thermal Testing of the Generic Nanosatellite Bus, University of Toronto.
- Diendorfer, G., W. Schulz, et al. Comparison of correlated data from the Austrian lightning location system and measured lightning currents at the Peissenberg tower.
- Diendorfer, G., W. Schulz, et al. (2002). "Lightning characteristics based on data from the Austrian lightning locating system." *Electromagnetic Compatibility, IEEE Transactions on* 40(4): 452-464.
- Eichelberger, H., G. Prattes, et al. (2010). Acoustic measurements of atmospheric electrical discharges for planetary probes. European Geosciences Union (EGU). Vienna, Austria.
- Eichelberger, H., G. Prattes, et al. (2011). Acoustic outdoor measurements with a multi-microphone instrument for planetary atmospheres and surface. European Geosciences Union (EGU). Vienna, Austria.
- Fulchignoni, M., F. Ferri, et al. (2005). "In situ measurements of the physical characteristics of Titan's environment." *Nature* 438(7069): 785-791.
- Graybill, R. and R. Melhem (2002). Power aware computing, Plenum Pub Corp.
- Haykin, S. (1996). Adaptive Filter Theory, Prentice Hall.
- Holden, D., C. Munson, et al. (1995). "Satellite Observations of Transionospheric Pulse Pairs." *Geophys. Res. Lett.* 22(8): 889-892.
- Jacobson, A., S. Knox, et al. (1999). "FORTE observations of lightning radio-frequency signatures: Capabilities and basic results." *Radio Sci.* 34(2): 337-354.
- Jacobson, A. R., S. O. Knox, et al. (1999). "FORTE observations of lightning radio-frequency signatures: Capabilities and basic results." *Radio Sci.* 34(2): 337 - 354.
- Jaffer, G. (2006b). Measurements of electromagnetic corona discharges. Educational workshop: "Lehrerfortbildung des Pädagogischen Institutes des Bundes in Steiermark: von Monden, Kometen und Planeten im Sonnensystem" Space Research Institute, Austrian Academy of Sciences, Graz, Austria.
- Jaffer, G. (2011c). Austrian Lightning Nanosatellite (LiNSAT): Space and Ground Segments. 3rd International Conference on Advances in Satellite and Space Communications (SPACOMM 2011). Budapest, Hungary. (accepted).
- Jaffer, G., H. U. Eichelberger, et al. (2010d). LiNSAT: Austrian lightning nano-satellite. UN-OOSA/Austria/ESA Symposium on Small Satellite Programmes for Sustainable Development: Payloads for Small Satellite Programmes. Graz, Austria.
- Jaffer, G., A. Klesh, et al. (2010a). Using a virtual ground station as a tool for supporting higher education. 61st International Astronautical Congress (IAC). Prague, Czech Republic.
- Jaffer, G. and O. Koudelka (2011d). Lightning detection onboard nano-satellite (LiNSAT). International Conference on Atmospheric Electricity (ICAE). Rio de Janeiro, Brazil. (accepted).
- Jaffer, G. and O. Koudelka (2011e). Automated remote ground station for Austrian lightning nanosatellite (LiNSAT). 8th IAA Symposium on Small Satellites for Earth Observation. Berlin, Germany,, (accepted). (accepted).
- Jaffer, G., O. Koudelka, et al. (2008). The detection of sferics by a nano-satellite. 59th International Astronautical Congress (IAC). Glasgow, UK 8.

- Jaffer, G., O. Koudelka, et al. (2010e). A Lightning Detector Onboard Austrian Nanosatellite (LiNSAT). American Geophysical Union, Fall Meeting`.
- Jaffer, G., R. Nader, et al. (2010b). Project Agora: Simultaneously downloading a satellite signal around the world. 61st International Astronautical Congress (IAC). Prague, Czech Republic: 8.
- Jaffer, G., R. Nader, et al. (2011a). "Using a virtual ground station as a tool for supporting space research and scientific outreach." *Acta Astronautica* (in press).
- Jaffer, G., R. Nader, et al. (2011h). Online and real-time space operations using Hermes-A I-2-O gateway. 1st IAA Conference On University Satellites Missions. Rome, Italy: -.
- Jaffer, G., R. Nader, et al. (2011i). "Online and real-time space operations using Hermes-A I-2-O gateway." *Actual Problems of Aviation and Aerospace Systems* (submitted).
- Jaffer, G., R. Nader, et al. (2010f). An online and real-time virtual ground station for small satellites, UN-OOSA/ Austria/ESA Symposium on Small Satellite Programmes for Sustainable Development: Payloads for Small Satellite Programmes.
- Jaffer, G. and K. Schwingenschuh (2006a). Lab experiments of corona discharges. Graz, Austria, Space Research Institute (IWF), Austrian Academy of Sciences. 37.
- Koudelka, O., G. Egger, et al. (2009). "TUGSAT-1/BRITE-Austria-The first Austrian nanosatellite." *Acta Astronautica* 64(11-12): 1144-1149.
- Krider, E., C. Weidman, et al. (1979). "The Temporal Structure of the HF and VHF Radiation Produced by Intracloud Lightning Discharges." *J. Geophys. Res.* 84(C9): 5760-5762.
- Le Vine, D. M. (1987). "Review of measurements of the RF spectrum of radiation from lightning." *Meteorology and Atmospheric Physics* 37(3): 195-204.
- Massey, R. and D. Holden (1995). "Phenomenology of transionospheric pulse pairs." *Radio Sci.* 30(5): 1645-1659.
- Massey, R., D. Holden, et al. (1998). "Phenomenology of transionospheric pulse pairs: Further observations." *Radio Sci.* 33(6): 1755-1761.
- Massey, R. S., D. N. Holden, et al. (1998). "Phenomenology of transionospheric pulse pairs: Further observations." *Radio Science* 33(6): 1755-1761.
- Nader, R., H. Carrion, et al. (2010a). HERMES Delta: The use of the DELTA operation mode of the HERMESA/ MINOTAUR Internet-to-Orbit gateway to turn a laptop in to a virtual EO ground station. 61st International Astronautical Congress (IAC). Prague, Czech Republic
- Nader, R., P. Salazar, et al. (2010b). The Ecuadorian Civilian Space Program: Near-future manned research missions in a low cost, entry level space program. 61st International Astronautical Congress (IAC). Prague, Czech Republic
- NASA (2011a). <http://thunder.msfc.nasa.gov/data/query/mission.png>.
- NASA (2011b). <http://science.nasa.gov/missions/ats/>.
- Price, C. and D. Rind (1993). "What determines the cloud to ground lightning fraction in thunderstorms?" *Geophysical Research Letters* 20(6): 463-466.
- Quakesat (2011). <http://www.quakefinder.com/joomla15/index.php>.
- Rakov, V. A. and M. A. Uman (2003). *Lightning: physics and effects*, Cambridge Univ Pr.

- Schulz, W., K. Cummins, et al. (2005). "Cloud-to-ground lightning in Austria: A 10-year study using data from a lightning location system." *J. Geophys. Res.* 110(D9): 1-20.
- Schulz, W. and G. Diendorfer (1999). Lightning Characteristics as a function of altitude evaluated from lightning location network data, SOC AUTOMATIVE ENGINEERS INC.
- Schulz, W. and G. Diendorfer (2004). Lightning peak currents measured on tall towers and measured with lightning location systems.
- Schwingsenschuh, K., B. P. Besser, et al. (2007). HUYGENS in-situ observations of Titan's atmospheric electricity. European Geosciences Union (EGU) General Assembly. Vienna, Austria.
- Schwingsenschuh, K., R. Hofe, et al. (2006a). In-situ observations of electric field fluctuations and impulsive events during the descent of the HUYGENS probe in the atmosphere of Titan. European Planetary Science Congress. Berlin, Germany. 37: 2793.
- Schwingsenschuh, K., R. Hofe, et al. (2006b). Electric field observations during the descent of the HUYGENS probe: evidence of lightning in the atmosphere of Titan. 36th COSPAR Scientific Assembly. Beijing, China. 37: 2793.
- Schwingsenschuh, K., H. Lichtenegger, et al. (2008b). Electric discharges in the lower atmosphere of Titan: HUYGENS acoustic and electric observations. 37th COSPAR Scientific Assembly. Montral, Canada. 37: 2793.
- Schwingsenschuh, K., G. J. Molina-Cuberos, et al. (2001). "Propagation of electromagnetic waves in the lower ionosphere of Titan." *Io, Europa, Titan and Cratering of Icy Surfaces, Advances of Space Research* 28(10): 1505-1510.
- Schwingsenschuh, K., T. Tokano, et al. (2010). Electric field transients observed by the HUYGENS probe in the atmosphere of Titan: Atmospheric electricity phenomena or artefacts? 7th International Workshop on Planetary, Solar and Heliospheric Radio Emissions (PRE VII) Graz, Austria
- Shriver, P. M., M. B. Gokhale, et al. (2002). A power-aware, satellite-based parallel signal processing scheme, Power aware computing, Kluwer Academic Publishers, Norwell, MA.
- Suszcynsky, D. M., M. W. Kirkland, et al. (2000). "FORTE observations of simultaneous VHF and optical emissions from lightning: Basic phenomenology." *Journal of Geophysical Research-Atmospheres* 105(D2): 2191-2201.
- Taylor-University (2011). GGB Design Document.
- Tierney, H. E., A. R. Jacobson, et al. (2002). "Transionospheric pulse pairs originating in maritime, continental, and coastal thunderstorms: Pulse energy ratios." *Radio Sci.* 37(3).
- Uman, M. A. (2001). The lightning discharge, Dover Pubns.
- Volland, H. (1995). Handbook of atmospheric electrodynamics, CRC.
- Wertz, J. R. (1978). Spacecraft attitude determination and control, Kluwer Academic Pub.
- Wertz, J. R. and W. J. Larson (1999). "Space mission analysis and design."

Williams, E. R., M. E. Weber, et al. (1989). "The Relationship between Lightning Type and Convective State of Thunderclouds." *Journal of Geophysical Research-Atmospheres* 94(D11): 13213-13220.

Adaptive MIMO Channel Estimation Utilizing Modern Channel Codes

Patric Beinschob and Udo Zölzer

*Helmut-Schmidt-Universität / Universität der Bundeswehr Hamburg
Germany*

1. Introduction

For the ever increasing demand in high data rates the spectrum from 300 MHz to 3500 MHz gets crowded with radio, smartphones, and tablets and their competition for bandwidth. Regulators cannot realistically reduce demand, nor can they expand the overall supply.

A solution is seen in the uprising of Multiple-Input Multiple-Output (MIMO) communications. The sometimes poor spectral efficiency of established radio systems can be increased dramatically without expanding bandwidth and at reasonable signal power levels.

The term MIMO pays tribute to the fact that multiple antennas at sender and receiver are used in order to have spatially distributed access to the channel thus establishing additional degrees of freedom also referred to as spatial diversity. Spatial diversity can be used for solely transmit redundant symbols, e.g. Space-Time Block Codes, as well as the transmission of independent data streams via the spatial layers known as Spatial Multiplexing (SM). This mode is preferred over pure diversity usage as recently discussed by Lozano & Jindal (2010). However, the benefit comes at the price of increasing RF hardware expenses and geometry in case of many installed antennas which are the main reasons for reluctant implementations in the industry in former times. Additional algorithmic complexity at one point in the communication system is another reason. For SM mode, this is mainly in the receiver, where the independent data streams have to be separated in the detection process, leaving open questions in implementation issues of MIMO technologies in handheld devices.

For high data rate communications, MIMO in conjunction with Orthogonal Frequency Division Multiplexing (OFDM) offers the opportunity of exploiting broadband channels within reasonable algorithmic complexity measures (Bölcskei et al., 2002).

OFDM used as a standard technique in broadband modulation eases the equalization issue in MIMO broadband channels. For a given system with n_R receive antennas and n_T transmit antennas the MIMO channel is described by the $n_R \cdot n_T$ Single-Input Single-Output (SISO) spatial subchannels established between each transmit-receive antenna pair. For the sake of notation they are arranged in a so called channel matrix.

MIMO-OFDM modulation technique allows to consider the MIMO problem for each OFDM subcarrier separately. Thus, complexity is reduced by turning a $K \cdot n_R \times K \cdot n_T$ matrix inversion into K inversions of $n_R \times n_T$ matrices in the case of linear MIMO detection algorithms (Beinschob & Zölzer, 2010b).

For coherent receivers channel estimation is necessary. Recent advances in channel coding theory and feasibility of “turbo” principles and techniques led to new receiver designs, (Akhtman & Hanzo, 2007b; Hagenauer et al., 1996; Liu et al., 2003), optimal Detectors (Hochwald & ten Brink, 2003) and optimized codes for MIMO transmission (ten Brink et al., 2004) with the help of EXIT chart analysis (ten Brink, 2001) on LDPC Codes (Gallager, 1962; 1963), which were in turn rediscovered and revised by MacKay (1999).

Iterative decoding to approximate a posteriori probability (APP) information on the received data enhances the possibilities of classical adaptive signal processing approaches. On the other hand, MIMO Spatial Multiplexing APP detectors are very complex and only slowly convergent.

However, in practical systems large gaps between theoretically calculated capacity and realized data rates can be observed. The negative impact of imperfect channel knowledge on detection performance is significant (Dall’Anese et al., 2009). Those errors are especially high in mobile scenarios. Constraints on the amount of reference symbols that use exclusive bandwidth is natural. So, as a solution decision-directed techniques in adaptive channel estimators are considered that utilize information from the obligatory forward error correction in order to increase the channel estimation accuracy.

Our approach focuses on a minimization of pilot symbols. Therefore, only a small initial training preamble is send followed by data symbols only as shown in Fig. 2. The use of distributed pilot symbols, a common approach for slow fading channels – also employed in LTE, is avoided that way. The application of adaptive filtering in combination with decision-directed techniques is shown here to provide the necessary update of the channel state information in time varying scenarios like mobile receivers.

The discussed channel estimation techniques aim to add only reasonable complexity, so non-iterative approaches are considered. It is non-iterative in the sense that no a priori feedback is given to the detector. Hence it is suited for low latency applications, too. Channel estimates are readily available at OFDM symbol rate as well as the decoded data bits.

The chapter is organized as follows. The basic system model is presented in the next section, with a discussion of channel characterization and used pilot symbols for minimum training length in Section 2.3. Common approaches to channel estimation with minimum training length are reviewed in Section 3. The receiver structure we focus on is presented in Section 4. Results of conducted numerical experiments are discussed in Section 5.

Notation is used as follows. Bold face capital letters denote matrices, column vectors are typed in bold small letters. The operator $(\cdot)^H$ applies complex-conjugate transposition to a vector or matrix. Time domain signals carry the check accent, e. g. $\check{x}$, in order to distinguish them from their frequency domain counterpart.

2. System model

2.1 Bit-interleaved coded MIMO-OFDM

A multiple antenna systems is represented as a time discrete model in a multi-path channel in the following fashion: The vector of received values $\check{\mathbf{r}}$ at the time sample m of a MIMO system is the superposition of $L \cdot n_T$ previously sent samples and the current n_T samples, where $L + 1$ is the length of the sampled channel impulse response. It is given by

$$\check{\mathbf{r}}[m] = \sqrt{E_s} \sum_{l=0}^L \check{\mathbf{H}}[l, m] \cdot \check{\mathbf{s}}[m - l] + \sigma_w \check{\mathbf{w}}[m], \quad (1)$$

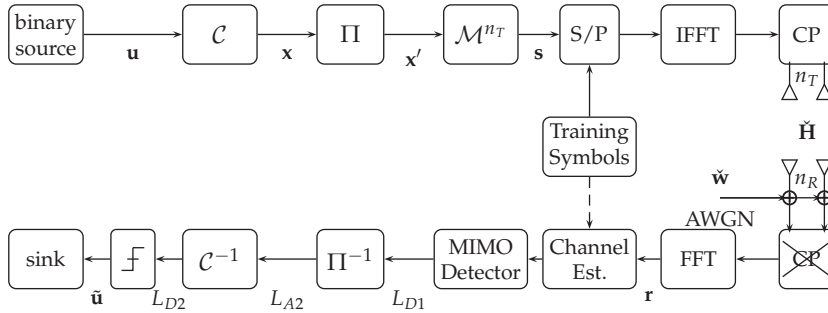


Fig. 1. MIMO-OFDM system with standard receiver processing.

where $\mathbf{s}[m]$ denotes the current vector of symbols of each of the transmit antenna, $\mathbf{w}$ is an identically, independently distributed (iid) additive white Gaussian noise term and $\mathbf{H}[l, m]$ is the MIMO channel matrix in delay and time domain, indexed with l respectively m . It is therefore the MIMO Channel Impulse Response per time sample m . The past sent samples are denoted by $\mathbf{s}[m-l]$, for $l \neq 0, l \leq L$. The data symbols of the K subcarriers are modulated by an inverse Fast Fourier Transform (IFFT). In simulations every value corresponding to a transmit antenna of the resulting vectors is transmitted using the formula above.

The MIMO-OFDM system model in frequency domain is described by

$$\mathbf{r}[n, k] = \sqrt{E_s} \mathbf{H}[n, k] \cdot \mathbf{s}[n, k] + \sigma_w \mathbf{w}[n, k], \quad (2)$$

where n denotes the time index of an OFDM symbol and k its subcarrier index, where K is the total number of subcarriers. As a Signal-to-noise measure E_b/N_0 is defined with noise variance given by $\sigma_w^2 = N_0$, where N_0 is the spectral noise power density in equivalent base band domain and with the energy per (QAM) symbol

$$E_s = R_c \cdot \kappa \cdot E_b, \quad (3)$$

where R_c is the code rate and κ bits per QAM symbol.

The receive vector $\mathbf{r}[n, k]$ and noise vector $\mathbf{w}[n, k]$ are of dimension $n_R \times 1$, the send vector $\mathbf{s}[n, k]$ of $n_T \times 1$ and the matrix $\mathbf{H}[n, k]$ of $n_R \times n_T$, at which n_R is the number of transmit antennas. The entries of $\mathbf{w}[n, k]$ are complex circular-symmetric Gaussian distributed random variables where $w_r[n, k] \sim \mathcal{CN}(0, 1), r = 1, \dots, n_R$ holds.

A perfect synchronization and total avoidance of block interference is assumed, so the OFDM cyclic prefix L_{cp} is longer than the discrete maximum path delay denoted by the channel order L , hence $L_{cp} > L$. The system overview is depicted in Fig. 1.

The MIMO-OFDM sent symbols are separately bit-interleaved LDPC codewords, where the EXIT chart of the employed LDPC code is shown in Fig. 4. The sender limits the codeword and interleaver length to the number of available bits in a MIMO-OFDM symbol n that is $n_T \cdot K \cdot \kappa$. The data symbols are drawn from an M -order QAM modulation alphabet $\mathcal{S}$. The mapping, denoted by $\mathcal{M}\{\cdot\}$, modulates $\kappa = \log_2 M$ bits to a QAM symbol. This is done consecutively for all n_T sent streams/layers hence the notation $\mathcal{M}^{n_T}\{\cdot\}$. The QAM constellations are considered unit power-normalized to simplify notation. At the receiver, the Log-Likelihood Ratios (LLRs) can be de-interleaved and LDPC decoded at once after reception, FFT and MIMO detection, which yields the approximated a-posteriori LLRs $L_{D2}[n]$

out of the received symbols:

$$L_{D2}[n] = \mathcal{C}^{-1} \left\{ \Pi^{-1} \{ L_{D1}[n] \} \right\}. \quad (4)$$

Scrutinizing the sign of $L_{D2}[n]$ yields the most probable sent codeword $\mathbf{y}[n]$. Finally, the transmitted information bits $\hat{\mathbf{u}}[n]$ are recovered by discarding the redundancy bits in $\mathbf{y}[n]$.

2.2 MIMO channel model

Typically a (static) MIMO channel realization can be modeled by drawing the coefficients $\check{H}_{r,t}$ independently from a complex circular-symmetric Gaussian distribution.

$$\check{H}_{r,t}[l] \sim \mathcal{CN} \left(0, \frac{1}{(L+1)} \right), r = 1, \dots, n_R, t = 1, \dots, n_T. \quad (5)$$

Doing so for all $L+1$ -multi-path time-domain MIMO channel coefficients implies a constant power delay profile for all spatial subchannels.

Of course, in mobile communication time-variant channel behaviour is expected. For multiple antennas systems in urban environments we have array size limitations thus small distances between the colocated antennas which renders the assumption of i.i.d. channel coefficients unrealistic. In order to conduct realistic simulations the 3GPP developed a Spatial Channel Model (SCM) suitable to test algorithms supporting mobile MIMO systems in macro- or micro urban scenarios (*Spatial channel model for Multiple Input Multiple Output (MIMO) simulations*, 2008).

Mobile receivers experiences velocity-dependent Doppler frequency shifts in components of the superposed received signal. For an OFDM system the consequence might be a gradually loss of orthogonality of the subcarriers which results in Intercarrier Interference (ICI).

Considering a wireless OFDM system at carrier frequency f_0 with OFDM symbol duration $T_{\text{OFDM}} = (K + L_{\text{cp}})/f_s$ in seconds (f_s being the sampling rate), a maximum Doppler frequency in Hertz for a given mobile station's relative radial velocity of v_{MS} is given by

$$f_D = \frac{v_{\text{MS}}}{c} \cdot f_0, \quad (6)$$

with c being the speed of light. As a measure in OFDM systems the normalized Doppler frequency is of more interest because of its independence of the system parameters K and L_{cp} :

$$f_{D,n} = f_D \cdot T_{\text{OFDM}}. \quad (7)$$

As a rule of thumb, significant ICI appears if $f_{D,n} > 5 \times 10^{-3}$. Associated with f_D a coherence time interval T_{coh} can be defined as by Proakis & Salehi (1994)

$$T_{\text{coh}} = \frac{1}{2f_D}. \quad (8)$$

2.3 Training symbol design

Training symbols must be carefully chosen in order to maximize the signal-to-noise ratio during estimation. In OFDM systems, it is important to design training symbols that have low peak-to-average-power ratio (PAPR) in time-domain. Spatial orthogonality should be preserved in frequency-domain for the different transmit antennas. As basic construction of

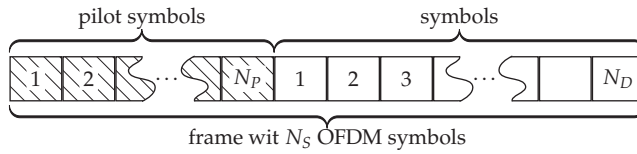


Fig. 2. Frame structure for proposed MIMO-OFDM RLS-DDCE with preamble length N_P symbols with $N_P \geq n_T$

orthogonal sequences Frank-Zadoff-Chu (FZC) sequences $\mathbf{p}_t$, $t = 1, \dots, n_T$ are chosen (Chu, 1972):

$$\mathbf{p}_i^H \mathbf{p}_j = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j. \end{cases} \quad (9)$$

It is a special property of FZC sequences that the sequence $\mathbf{p}_j$ is yielded by cyclic shifting of $\mathbf{p}_i$ by $j - i$ positions. The sequences are inserted over time and a subcarrier-specific phase to lower the PAPR is added, e.g. for the first sequence

$$p_1[n, k] = e^{j\pi M'(n-1)^2/n_T + \varphi[k]}, \quad n = 1, \dots, n_T. \quad (10)$$

The phases are taken from another FZC sequence of length K :

$$\varphi[k] = \pi M''(k-1)^2/K, \quad k = 1, \dots, K. \quad (11)$$

M' and M'' are prime numbers less or equal to $n_T + 2$ and $K + 2$, respectively. Each antenna t sends its $p_t[n, k]$ as training preamble at the beginning of the frame. Construction is possible for all length of n_T and K and leads to time and frequency domain signals with minimum PAPR.

The underlying frame structure provides a training sequence at the beginning of each frame as shown in Fig. 2.

3. Decision-directed channel estimation techniques

From Eq. (2) it is clear that estimating the channel matrix $\mathbf{H}$ is difficult even if the send vector is known due to the rank-deficit of the problem. Therefore, for the estimate it needs a scheme that efficiently exploits all given diversities: time, frequency and space. A promising approach is given by Akhtman & Hanzo (2007a), that proposed an adaptive channel estimation structure. In the first step, a spatial auto- and crosscorrelation estimator is employed for each subcarrier individually. Originally, a further stage for dimension reduction – using the PAST scheme – is employed. It is not considered here in order to eliminate further influence of parameters and to separate the effects. However, in order to exploit the correlation of adjacent subcarriers, LDPC codewords are interleaved over spatial streams and subcarriers. So the structure is enhanced by the usage of short yet powerful LDPC codes, employing the belief propagation decoder to approximate posteriori information on the send symbol which are used in the decision-feedback processing. Deep fading occurring occasionally on individual subcarriers would result in low LLRs, which are less trusted in belief propagation decoding. But through message-passing their information is recovered from the other connected nodes. By simple parity or syndrome check – a property which LDPC codes inherit from the family

of linear block codes –, a reliable and readily available criterion is given to control the overall decision feedback of the channel estimator.

3.1 Recursive least squares estimation

Due to the unknown error distribution the channel estimation is often formulated as a *Least Squares* problem: Find a channel matrix estimate $\tilde{\mathbf{H}}[n]$ at the symbol n that projects the send vector $\mathbf{s}[n]$ in the receive vector space, such that the euclidean distance to the actual received vector $\mathbf{r}[n]$ is be minimized:

$$J_{RLS}[n] = \sum_{m=1}^n \zeta^{n-m} \mathbf{e}^H[m, n] \mathbf{e}[m, n], \quad (12)$$

with the error signal

$$\mathbf{e}[m, n] = \mathbf{r}[m] - \tilde{\mathbf{H}}[n] \cdot \mathbf{s}[m]. \quad (13)$$

This classic approach yields good results with increasing samples if the unknown channel matrix $\mathbf{H}$ is constant. For time-variant channels old samples will increase the estimation error as the channel coefficients keep changing slowly. To gain adaptivity a “forgetting” factor $0 < \zeta \leq 1$ is introduced, that applies a weighting depending on the sample index such that newer sample have stronger influence on the estimate than older ones. An exponential decreasing weighting has some implementation qualities that will be pointed out in the following.

A LS channel estimate of the channel matrix $\mathbf{H}$ is yielded by

$$\tilde{\mathbf{H}}[n] = (\tilde{\Phi}^{-1}[n] \tilde{\Theta}[n])^H. \quad (14)$$

with the estimated spatial auto- and cross-correlation matrices based on

$$\tilde{\Phi}[n] = \sum_{m=1}^n \zeta^{n-m} \mathbf{s}[m] \mathbf{s}^H[m] = \zeta \tilde{\Phi}[n-1] + \mathbf{s}[n] \mathbf{s}^H[n], \quad (15)$$

and

$$\tilde{\Theta}[n] = \sum_{m=1}^n \zeta^{n-m} \mathbf{s}[m] \mathbf{r}^H[m] = \zeta \tilde{\Theta}[n-1] + \mathbf{s}[n] \mathbf{r}^H[n]. \quad (16)$$

The known pilot symbols are used as substitutes for the sent vectors $\mathbf{s}[n, k]$ if $n \leq N_p$, otherwise the decision-feedback is used. A $\zeta := 1$ is optimal if a static channel is considered because the estimation error keeps decreasing with increasing n as long as there are no false decisions in the feedback.

If only pilot symbols are utilized, no further information is available beyond the training and the channel estimates need to be used for the rest of the frame. For $\zeta = 1.0$, this technique is referred as ordinary Least-Squares (LS) Channel Estimation in the following.

Due to the orthogonal designed pilot symbols, the matrices $\tilde{\mathbf{H}}[n, k]$ have full condition at $n = n_T$ yet they are superposed by noise.

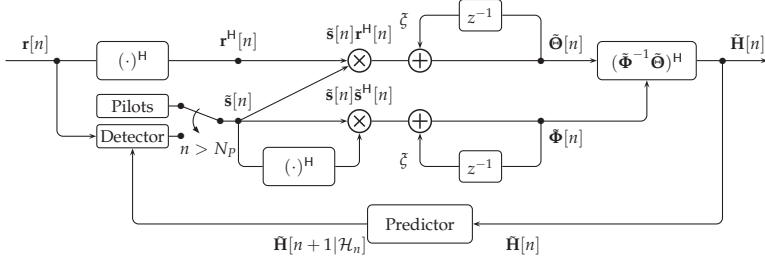


Fig. 3. Signal Flow diagram RLS algorithm.

3.2 Algorithmic structure

In Fig. 3 the algorithmic structure is shown. After N_p pilots of constant-energy type (Chu, 1972), the output of the MIMO detector is used instead of the known pilots in order to estimate the spatial auto- and cross-correlation matrices $\tilde{\Phi}$ resp. $\tilde{\Theta}$ independently per subcarrier. Averaging is performed in the recursive part of the structure and weighting with a forgetting factor ζ is applied to suppress older samples and adapt to newer ones. The detector uses the channel estimate for detection. To mitigate the effect of outdated CSI, a predictor is employed that tracks the time-variant MIMO channel $\mathbf{H}$ and calculates an prediction $\hat{\mathbf{H}}[n+1|\mathcal{H}_n]$. Through the immediately detection of data this algorithm is in principle suited to low delay applications as pointed out by Beinschob & Zölzer (2010a).

3.3 Decision feedback

3.3.1 Hard decision feedback

Further information on the channel can be acquired by using the detection output in Eq. (15) and (16), i. e. estimated sent vectors as proposed in Akhtman & Hanzo (2007a),

$$\tilde{\mathbf{s}}[n] = \mathcal{M}^{n_T} \{ \text{sgn} \{ L_{D1}[n] \} \}, \quad \forall n > N_p. \quad (17)$$

The algorithm is illustrated in Fig. 3. This is referred to as decision-directed channel estimation. It is prone to error propagation since incorrect decisions increases the channel estimation error, which in return increases the probability of incorrect decisions. Feedback with incorrect symbols in an early stage of the frame renders the channel estimate for the rest completely useless.

3.3.2 Soft decision feedback

In contrast to Eq. (17) hard decision, the sent MIMO-OFDM symbols can be estimated by evaluating the symbol expectation values (Glavieux et al., 1997) based on the detection probabilities p associated with $L_{D1}[n]$:

$$\tilde{s}_t[n] = E \{ s_t \} = \sum_{c \in \mathcal{S}} c \cdot p(\tilde{s}_t[n] = c), \quad \forall t. \quad (18)$$

The reconstructed sent vectors can be applied in Eq. (15) and (16). The *soft* symbol value is determined by the reliability of LLRs, i. e. magnitude. If low LLRs occur Eq. (18) evaluates to near zero, which can lead to stability problems in Eq. (14) for $\zeta < 1$ due to exponentially decreasing values in $\tilde{\Theta}$. This scheme is referred to as soft-decision RLS (RLSsd).

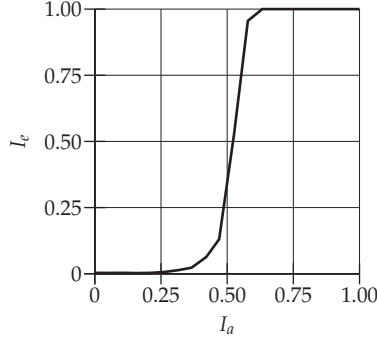


Fig. 4. Extrinsic mutual information transfer (EXIT) chart of used LDPC code 1024 bit length constructed with edge distributions from Richardson et al. (2001).

4. Proposed receiver structure

4.1 Conditioned a posteriori decision-feedback

Following the decision-feedback strategy of Eq. (17) the estimator has no information about the certainty of the feedback. Using soft-decision feedback the estimated sent vectors still contain the noise from detection stage. The developing of the channel estimation error $\varepsilon[n]$ can be very roughly modeled as follows:

$$\varepsilon[n] = \frac{\sigma_e^2}{n} + N_e \sigma_c^2 \quad (19)$$

On one hand the channel estimation error decreases with growing n through the averaging effect, where the error energy per gained channel estimate sample due to noise and spatial interference is denoted σ_e^2 . On the other hand an incorrect decision directly adds the error energy σ_c^2 , which contains the energy of a false channel estimate plus noise and interference. The cost of incorrect decisions are unequally higher as the benefit of another correct sample for averaging, especially for higher n . Hence an incorrect decision should be avoided even at the price of low number of samples to average.

By utilizing channel decoder L_{D2} information at the decision stage, the feedback of incorrect decisions can be avoided. For time-invariant channels the weighting factor is set to $\xi := 1.0$ because an update via Eq. (15) and (16) is applied only if the codeword associated with the MIMO-OFDM symbol is successfully decoded. For linear block codes this is indicated by the syndrome vector

$$\gamma[n] = \mathbf{A} \cdot \mathbf{y}[n]. \quad (20)$$

The parity check matrix is denoted $\mathbf{A}$, the received codeword is

$$\mathbf{y}[n] = \text{sgn}\{L_{D2}[n]\}. \quad (21)$$

Note that Eq. (20) and (21) are evaluated in most LDPC decoder implementations as break criterion for the iteration process. So, if the decoder signals $\|\gamma[n]\|_H = 0$, where $\|\cdot\|_H$ denotes the Hamming distance, the sent vectors in Eq. (15) and (16) are substituted by

$$\{\tilde{\mathbf{s}}[k, n]\}_{k=1}^K = \mathcal{M}^{n_T} \{\Pi\{\mathbf{y}[n]\}\}, \quad (22)$$

using the property of the systematic linear block code which renders additional re-encoding unnecessary. Because of the high codeword distance it is unlikely that the codeword $\mathbf{y}[n]$ contains undetected errors or being another valid codeword. However, this is a source of the algorithm's residual errors. In the case $\|\gamma[n]\|_H \neq 0$, no update is performed instead the former channel estimate is assumed to be still valid:

$$\tilde{\mathbf{H}}[n, k] := \tilde{\mathbf{H}}[n - 1, k], \quad (23)$$

which can be seen as a simple zero order hold prediction.

As pointed out with the rough model in Eq. (19), feedback of erroneous data increases the estimation error over-proportional compared to the benefit of another sample to average. However, for valid codewords this approach leads to an extension of the training length due to the usage of data symbols as virtual pilots. The benefit is especially large at the beginning of a frame, directly after the end of pilots. There, the channel estimate error is still high and therefore the probability of false detection is high. The proposed scheme is referred in the following as Conditional Feedback (RLSCF).

The estimated channel matrix $\tilde{\mathbf{H}}$ is used in the MIMO detection, where LLR channel values L_{D1} are determined per OFDM symbol and subcarrier on a Maximum Likelihood (ML) basis:

$$L_{D1}(t, \nu) = \frac{1}{\sigma_w^2} \left(\min_{\mathbf{s} \in \mathcal{S}^{\nu, +1}} \|\mathbf{r} - \mathbf{H}\mathbf{s}\| - \min_{\mathbf{s} \in \mathcal{S}^{\nu, -1}} \|\mathbf{r} - \mathbf{H}\mathbf{s}\| \right), \quad (24)$$

where $\mathcal{S}^{\nu, +1}$ denotes the set of all possible send vectors that ν -th bit is $+1$, $|\mathcal{S}| = |\mathcal{M}|^{n_T}$ and considering unit-power constellations. Of course, ML-approximating detection algorithms, i. e. List Sphere Decoder (Hochwald & ten Brink, 2003) can be applied here as well. σ_w^2 is the channel noise power which is often assumed to be known at the receiver.

4.2 Complexity discussion

Most of the proposed processing is straight-forward: OFDM demodulation, subcarrierwise MIMO detection, de-interleaving and LDPC decoding has to be done in any coherent MIMO receiver coping with multipath channels. The additional complexity comes from an interleaver and a MIMO symbol mapper in the channel estimation feedback chain. In the soft decision case there is more computing power necessary with only marginal performance gains compared to the hard decision technique. However, the main contribution to complexity raises in the channel estimation itself for possibly updating the channel estimation at each OFDM symbol if the SNR is sufficient high and the channel decoder is successful decoding the codeword. In this case the LDPC decoder that is aware of a valid codeword and stops the iteration, needs less internal iterations so there is a shift in complexity from the decoder to the channel estimator.

The estimation procedure can be further simplified by help of the matrix inversion lemma to avoid explicit inversion of the autocorrelation matrix. The resulting algorithm is described in Beinschob et al. (2009).

5. Numerical experiments

5.1 Channel estimation accuracy & system performance

In a MIMO system, the detection depends on the received vector $\mathbf{r}$ and channel matrix $\mathbf{H}$ (Hochwald & ten Brink, 2003). In real scenarios the channel matrix is not available and an

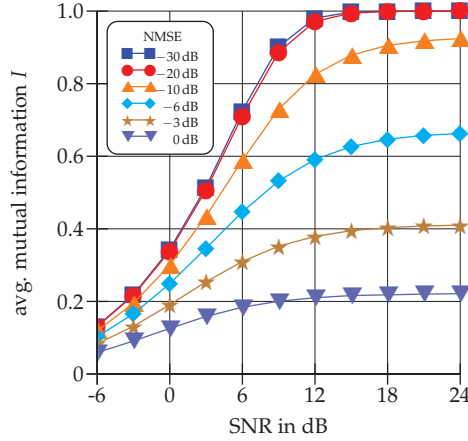


Fig. 5. Average Mutual Information of a MIMO 4×4 Spatial Multiplexing with 4-QAM modulation and different levels of Channel Estimation Errors in terms of NMSE.

estimate $\tilde{\mathbf{H}}$ is used instead. To evaluate the quality of the detector output, which is of most importance for the decision-directed scheme, the mutual information of the gained (uncoded) LLRs is used (ten Brink, 1999). The performance degradation in presence of estimation errors is assessed by modelling a channel estimate as follows:

$$\tilde{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}, \quad (25)$$

where $H_{r,t} \in \mathcal{CN}(0, 1)$ are the i.i.d. MIMO Rayleigh Fading channel coefficients and elements of $\Delta\mathbf{H}$ are i.i.d. circular symmetric complex Gaussian random variables with error variance σ_e^2 . This also covers to some extent the case of mobility induced interference because the ICI leads to additional noise in the Frequency Domain as pointed out by Russell & Stuber (1995). As a measure for the estimation error the normalized mean squared error is defined as

$$\text{NMSE} = \frac{\sum_{t=1}^{n_T} \sum_{r=1}^{n_R} |H_{r,t} - \tilde{H}_{r,t}|^2}{\sum_{t=1}^{n_T} \sum_{r=1}^{n_R} |H_{r,t}|^2}. \quad (26)$$

So, the channel capacity for discrete input alphabet (4-QAM) and continuous channel output is shown in Fig. 5.

In general, the capacity increases with SNR and with decreasing channel estimation NMSE. However, a saturation of I over SNR can be observed: With increasing estimation error the maximum level of mutual information I decreases.

At minimum, a NMSE of -6 dB to -10 dB is necessary for a half-rate coded system to work properly. Below -20 dB NMSE there is no significant difference in the system's performance using either the channel estimate or the actual channel matrix.

5.2 System level simulations for time-invariant channels

In Tab. 1 the system parameters are given for the experiments described in this section. For the

$n_T \times n_R$	4×4
Detector	Maximum Likelihood
OFDM Subcarriers	128
Modulation	4-QAM
MIMO-OFDM symbols per frame	516
bandwidth efficiency	$\approx 3.8 \text{ bit/s/Hz}$
Pilot symbols per antenna	4 (0.8%)
Channel Model	$\mathcal{CN}(0,1)$
channel multi-path order L	6
OFDM cyclic prefix L_{cp}	7
channel coherence time T_c	$> \text{frame length}$
LDPC code design rate R_c	1/2
codeword & interleaver length	1024 bit

Table 1. Simulation parameters for time-invariant channel

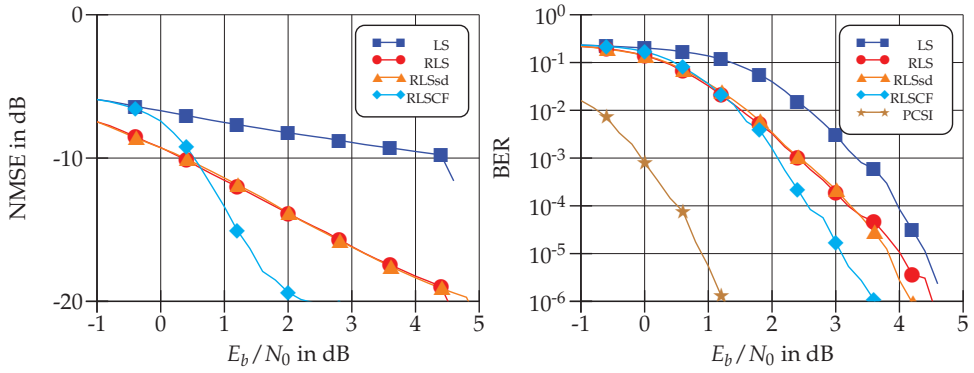
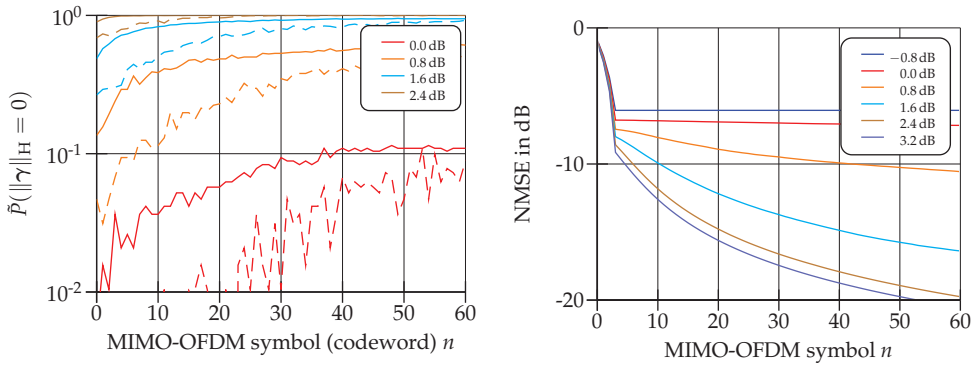


Fig. 6. BER and NMSE for the discussed schemes evaluated by Monte Carlo simulations using time-invariant channels with system parameters defined in Tab. 1.

BER/NMSE system level simulations a Maximum Likelihood bit detector with LLR output, combined with a short LDPC code of 1024 bit length, inserted in an MIMO-OFDM symbol, were used. Simulations were conducted for the case of minimum pilot symbol length $N_P = n_T$. The proposed channel estimation (RLSCF) was compared to the ordinary LS approach, the hard/soft-decision feedback approach by Eq. (17) (RLS resp. RLSsd) and detection with perfect channel state information (PCSI). Independent block-fading MIMO channel realization were generated which were frequency-selective but time-invariant during the frame, see Eq. (5).

The channel estimation error is assessed in terms of normalized mean squared error (NMSE), shown in Fig. 6. Least Squares channel estimation using only the available short training symbol sequence was inferior to the discussed decision-directed schemes. An improvement of RLS/RLSsd over LS can be observed. With increasing E_b/N_0 more correct feedback was available and decreased the NMSE, as reflected by the stronger slope of the RLS/RLSsd curve. RLS and RLSsd performed equally on average. The proposed method, RLSCF, performed worse than RLS/RLSsd for low E_b/N_0 . This is because the feedback is limited to valid codewords only, which occurred seldom. So the averaging effect is almost as limited as in



(a) Rate of successful decoding as indicated by $\|\gamma\|_H = 0$, for different channel E_b/N_0 , comparing proposed RLSCF (solid) and RLS (dashed lines). (b) NMSE per MIMO-OFDM symbol (time) index n and selected E_b/N_0 .

Fig. 7. Developing of NMSE over time resp. OFDM symbol index for RLSCF scheme (multiple frames and channel realizations averaged).

the LS case. Beyond 1 dB E_b/N_0 RLSCF performed better than RLS/RLSsd because it can be deduced that there were plenty of virtual pilots, as depicted in Fig. 7(a).

There the rate of indicated successful decoding over MIMO-OFDM symbol index n is shown. Especially in the beginning of the frame a higher rate of successful decoding could be achieved with the proposed channel estimator RLSCF. This led to more virtual pilots and improved NMSE depicted in Fig. 7(b).

The developing channel estimation error in terms of normalized mean square error (NMSE) can be seen for different E_b/N_0 in Fig. 7(b). Due to the short training the channel estimation error is rather high at the beginning of the frame and decreases with $\propto 1/n$ if the SNR is sufficient so that the decoder could deliver successful decoded feedback. This the case for E_b/N_0 above ≈ 1 dB, below only insignificant improvements could be realized. This threshold is a property of the LDPC code, which is visualized in the EXIT chart in Fig. 4. Only if the input information exceeds the threshold successful decoding is possible. This means for the proposed scheme, for $E_b/N_0 < 1$ dB only marginal gains could to realized compared to LS, where as for higher values significant improvements in terms of NMSE were achieved.

For completeness the resultant system Bit Error Rate (BER) of the conducted simulations are shown in Fig. 6. For reference a curve for detection with perfect channel state information (PCSI) is shown, too. On the right hand side the detection based on Least Squares channel estimation using only the available training symbols is shown. The performances of the discussed decision-directed schemes were in between, reflecting the same tendency as the NMSE plot in Fig. 6. Here for higher E_b/N_0 RLSsd performed marginal superior. By using detection-based decision-directed strategies it was possible to decrease the BER but due to erroneous feedback it was rather small. For the proposed conditional a-posteriori feedback (RLSCF) a gain of 1 dB could be achieved. The gap of 2 dB to the decoding with perfect channel knowledge can be explained with the limited codeword distance of the short LDPC code used.

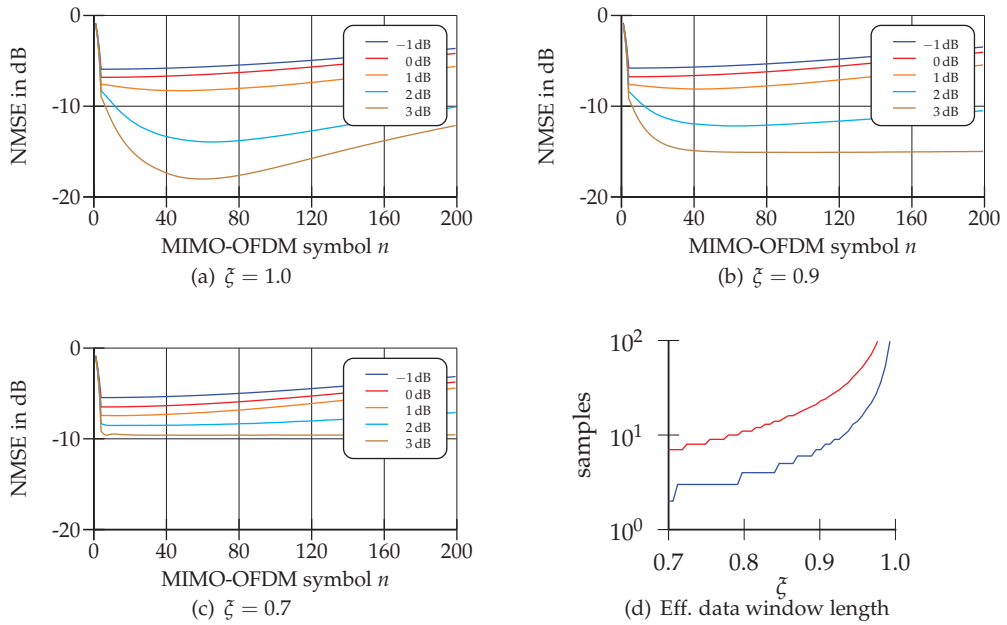


Fig. 8. Comparison of adaptivity on time variant channels (10 m/s) of channel code constraint feedback (RLSCF) for different forgetting factors ζ , also shown the effective data window length of forgetting factor in samples for coherence time (3 dB - blue line) and 10 dB (red line).

5.3 Simulations for 3GPP channels

To evaluate algorithms for real application a realistic channel model is needed. In particular the mobility component for the multipath MIMO case must be modelled very well. For this purpose the 3GPP Spatial Channel Model (*Spatial channel model for Multiple Input Multiple Output (MIMO) simulations*, 2008) is applied here.

5.3.1 Time adaptivity analysis

Fig. 8 illustrates the adaptivity of the DDCE structure depicted in Fig. 3. Clearly visible is the balance of the forgetting factor ζ which was too inflexible if set to $\zeta = 1.0$ then the algorithm failed to track the channel well, although the NMSE became very low after 60 symbols. On the other hand, if set to $\zeta = 0.7$, even for 3 dB E_b/N_0 the NMSE was insufficient for reliable transmissions. So a higher E_b/N_0 is needed for lower forgetting factors and thus higher adaptivity to achieve the same system performance for mobility scenarios.

5.3.2 System level results

For the 3GPP channel simulations RLSCF-based DDCE schemes were investigated through Monte Carlo simulations due to the lack of analytical tools to this kind of detection and channel estimation feedback structure. A Point-to-Point link is considered with Signal-to-noise-based assessment instead of range-dependence. Independent channel realization were generated for each frame. A frame consisted of 516 MIMO-OFDM symbols

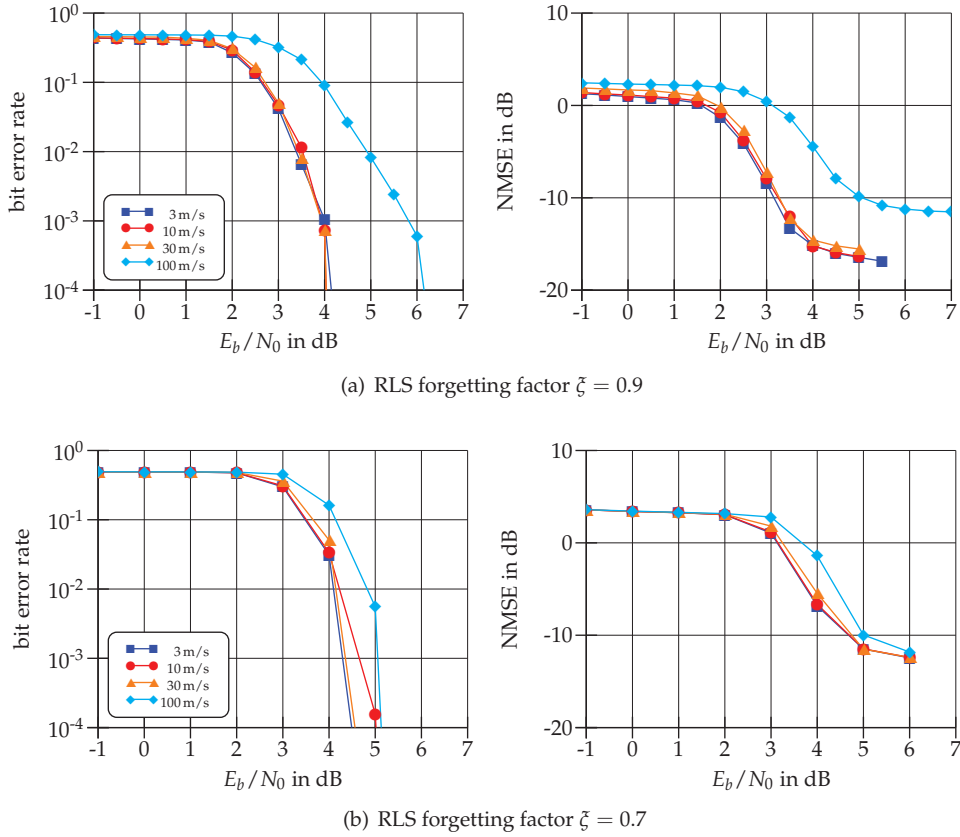


Fig. 9. On the left BER vs. E_b/N_0 for burst transmissions in time variant channels with minimal pilot length $N_p = n_T$ for 4×4 MIMO, on the right channel estimation error in NMSE (same legend as left hand).

constructed by modulating uncorrelated random bits. Transmissions were simulated until either 3×10^5 bit were transmitted or at least 200 frames errors were encountered per E_b/N_0 and velocity setting. The E_b/N_0 level was not further increased if in 3×10^5 bit no bit errors occurred after the channel decoder.

As seen from the results for time-invariant channels, RLSCF clearly gave best results in terms of BER/NMSE. As already mentioned, in order to maximize the bandwidth efficiency, we focus on having minimal number of pilot symbols, that is when $N_p = n_T$. So for the 4×4 MIMO system we employed 4 pilot tones per subcarrier and spatial stream, having a pilot rate of 0.8%.

As pointed out in the previous section, it is obvious that the forgetting factor has influence on the NMSE performance due to the averaging effect. Left of Fig. 9 shows the BER of RLSCF for increasing velocities of the mobile terminal. But it is difficult to predict whether an improved adaptivity on time-varying channels pays off the lost samples for the averaging. Simulation results for comparative study can be seen in Fig. 9. For a forgetting factor of $\zeta = 0.9$ (see

Fig. 9(a)), it can be seen that performance degradation started somewhere above 30 m/s. For lower velocities virtually error-free reception was observed for E_b/N_0 above 4.5 dB. In the case of very high velocities, strong variations in the channel occurred which led to an increased channel estimation error in terms of NMSE.

In comparison with the forgetting factor setting of $\xi = 0.7$ (see Fig. 9(b)), we observed a general rise in NMSE due to the smaller effective data window length, which in turn leads to a worse BER. An exception is the highest velocity, which seemed unaffected by the shorter data window size in terms of NMSE. The same NMSE performance for $\xi = 0.7$ and $\xi = 0.9$ could be achieved. We can deduce that for 100 m/s during the longer data window size the channel changed significantly and therefore the additional samples were of no use for the averaging process. While the NMSE seemed unaffected there was influence visible in the BER results. A shorter window length led to a earlier E_b/N_0 level of virtually error-free transmission, about 1 dB.

6. Conclusion

A promising structure to perform channel estimation for a multitude of channel scenarios using minimum amount of pilots enhanced by decision-directed techniques is presented. Through the employment of modern channel coding in the feedback the algorithm is capable to improve the channel estimate thus decreasing the estimation error beyond the pilot sequence only by using data symbols as virtual pilots. Optimal tracking of time-variant channels is still an open problem, although good results can be achieved by regarding influence length resp. coherence time of the channel and choosing the forgetting factor ξ appropriately. However, in many practical cases lower velocities are encountered which can be exploited quite well with a high forgetting factor like $\xi = 0.9$ as shown in the results. For highest velocities we loose about 1 dB in E_b/N_0 but gain in NMSE and BER for the lower velocities. There is a general trade-off between good system performance in low E_b/N_0 or high velocities scenarios where the usable data window size gets too small for significant averaging gains in the channel estimation.

7. References

- Akhtman, J. & Hanzo, L. (2007a). Advanced channel estimation for MIMO-OFDM in realistic channel conditions, *IEEE International Conference on Communications, ICC '07* pp. 2528–2533.
- Akhtman, J. & Hanzo, L. (2007b). Iterative receiver architectures for MIMO-OFDM, *IEEE Wireless Communications and Networking Conference, WCNC*, pp. 825–829.
- Beinschob, P., Lieberei, M. & Zölzer, U. (2009). An error propagation prevention method for MIMO-OFDM RLS-DDCE algorithms, *Proc. IEEE International Symposium On Wireless Communication Systems 2009 (ISWCS'09), Siena* pp. 31–35.
- Beinschob, P. & Zölzer, U. (2010a). Improving Low-Delay MIMO-OFDM channel estimation, *International ITG Workshop on Smart Antennas 2010, Bremen, Germany*.
- Beinschob, P. & Zölzer, U. (2010b). MIMO-OFDM equaliser for spatial multiplexing transmission modes, *Advances in Radio Science* 8: 81–85.
URL: <http://www.adv-radio-sci.net/8/81/2010/>
- Bölcskei, H., Gesbert, D. & Paulraj, A. (2002). On the capacity of OFDM-based spatial multiplexing systems, *IEEE Transactions on Communications* 50(2): 225–234.

- Chu, D. C. (1972). Polyphase codes with good periodic correlation properties, *IEEE Transactions on Information Theory* 18(4): 531–532.
- Dall’Anese, E., Assalini, A. & Pupolin, S. (2009). On the effect of imperfect channel estimation upon the capacity of correlated MIMO fading channels, *IEEE 69th Vehicular Technology Conference, VTC Spring 2009*, pp. 1–5.
- Gallager, R. G. (1962). Low-density parity-check codes, *IRE Transaction on Information Theory* 8(1): 21–28.
- Gallager, R. G. (1963). *Low Density Parity-Check Codes*, MIT Press, Cambridge, MA.
- Glavieux, A., Loat, C. & Labat, J. (1997). Turbo equalization over a frequency selective channel, *Proceedings of the International Symposium of Turbo Codes, Brest, France* pp. 96–102.
- Hagenauer, J., Offer, E. & Papke, L. (1996). Iterative decoding of binary block and convolutional codes, *IEEE Transactions on Information Theory* 42(2): 429–445.
- Hochwald, B. & ten Brink, S. (2003). Achieving near-capacity on a multiple-antenna channel, *IEEE Transactions on Communications* 51(3): 389–399.
- Liu, H., Wang, X. & Xiong, Z. (2003). Iterative receivers for OFDM coded broadband MIMO fading channels, *IEEE Workshop on Statistical Signal Processing* pp. 355–358.
- Lozano, A. & Jindal, N. (2010). Transmit diversity vs. spatial multiplexing in modern MIMO systems, *IEEE Transactions on Wireless Communications* 9(1): 186–197.
- MacKay, D. (1999). Good error-correcting codes based on very sparse matrices, *IEEE Transactions on Information Theory* 45(2): 399–431.
- Proakis, J. G. & Salehi, M. (1994). *Communication Systems Engineering*, Prentice Hall International, Eaglewood Cliffs, New Jersey 07632.
- Richardson, T. J., Shokrollahi, M. A. & Urbanke, R. L. (2001). Design of capacity-approaching irregular low-density parity-check codes, *IEEE Transactions on Information Theory* 47(2): 619–637.
- Russell, M. & Stuber, G. (1995). Interchannel interference analysis of OFDM in a mobile environment, *IEEE 45th Vehicular Technology Conference, VTC 1995*, Vol. 2, pp. 820–824.
- Spatial channel model for Multiple Input Multiple Output (MIMO) simulations* (2008). *Technical Report 25.996*, 3rd Generation Partnership Project (3GPP), Technical Specification Group Radio Access Network, 06921 Sophia-Antipolis Cedex, France.
- ten Brink, S. (1999). Convergence of iterative decoding, *Electronics Letters* 35(10): 806–808.
- ten Brink, S. (2001). Convergence behavior of iteratively decoded parallel concatenated codes, *IEEE Transactions on Communications* 49(10): 1727–1737.
- ten Brink, S., Kramer, G. & Ashikhmin, A. (2004). Design of low-density parity-check codes for modulation and detection, *IEEE Transactions on Communications* 52(4): 670–678.

An Introduction to ANFIS Based Channel Equalizers for Mobile Cellular Channels

K. C. Raveendranathan

*Government Engineering College Bartonhill Vanchiyoor P.O.
India*

1. Introduction

The purpose of a communication system is to transfer information between two separate points over some medium in the presence of disturbances or distortions such as noise and dispersion. This distortion is manifested in the time domain as pulse dispersion and is labeled as Inter-Symbol Interference (ISI). As data rates increase in modern digital communication systems, ISI becomes an inevitable consequence of the dispersive nature of band-limited propagation channels. The receiver must include an equalizer to mitigate the effects of ISI. The function of the equalizer is to combat the ISI and to utilize the available spectrum most efficiently.

Equalizers are cascaded to almost all kinds of channels, right from telephone lines to radio and optical fiber channels, to make the channel performance optimal. Ideally, an equalizer, when cascaded to the end of a channel, will make it behave like an ideal channel, the one which will not distort the signals in any manner. In the case of mobile cellular channels, which are generally considered to be Non-Linear and Time Variant (NLTV), the design of equalizers is *not a trivial problem*. Moreover, the above said channel has certain uncertainties in its behaviour, which need to be tackled in the equalizer design. The Co-Channel Interference (CCI) due to frequency reuse and Adjacent Channel Interference (ACI) due to spectral leakage, both contribute to the reduction in overall Signal-to-Interference-Noise-Ratio (SINR) in mobile cellular channels.

In applications in which the Channel Impulse Response (CIR) is unknown and no training sequence is available, the equalizer must be computed/ updated blindly from the received signal and knowledge of the statistics of the data source alone. A common approach in continuous transmission systems is to blindly update a Linear Equalizer (LE) using the Constant Modulus Algorithm (CMA), and then switch to a Decision Directed (DD) mode when the Symbol Error Rate (SER) is low enough (Widrow, 1973).

2. Fading characteristics of mobile channels

In mobile cellular radio transmission between a base station and a mobile telephone, the signal transmitted from the base station to the mobile receiver is usually reflected from surrounding buildings, hills, and other obstructions. As a consequence, we observe multiple propagation paths arriving at the receiver at different delays. Hence the received signal has characteristics similar to those for ionospheric propagation. The same is true for

transmission from the mobile telephone to the base station. Moreover, the speed that the mobile (automobile, train, etc.) is traveling results in frequency offsets, called *Doppler shifts*, of the various frequency components of the signal (Proakis & Salehi, 2002). As the intervening medium changes its characteristics with respect to time, the mobile radio channel is time varying.

An important requirement for assessing technology for mobile radio applications is to have an accurate description of the wireless channel. Channel models are heavily dependent upon the radio architecture. For example, in the first generation systems, a super-cell or "single-stick" architecture is used where the Base Transmitting Station (BTS) and the subscriber station are in Line-Of-Sight (LOS) condition and the system uses a single cell with no co-channel interference. For the second generation systems, a scalable multi-cell architecture with Non-Line-Of-Sight (NLOS) conditions becomes necessary.

The mobile cellular channel can be modeled as one of the following:

- The Free Space model.
- Tapped Delay Line (TDL) Channel model.
- Rayleigh Fading Channel model.
- Ricean Fading Channel model.
- The Hata-Okumura model.
- Cost 231 Walfish-Ikegami (W-I) model.
- Modified Stanford University Interim (SUI) Channel model.
- The FCC model.
- The ITU-R model.

The free space model is the simplest model, but it does not model the terrestrial environment reliably. A general model for a time-variant multipath channel is a TDL structure. It consists of a tapped delay line with uniformly spaced taps. The tap spacing between adjacent taps is $1/W$, where W is the bandwidth of the signal transmitted through the channel. The tap coefficients, denoted as $\{c_n(t) \equiv a_n(t) e^{jqm(t)}\}$ are usually modeled as complex valued, Gaussian random processes which are mutually uncorrelated (Proakis & Salehi, 2002). The most widely used path loss model for signal strength prediction and simulation in macrocellular environments is the *Hata-Okumura model* (Okumura et.al., 1968, Hata, 1980). This model is valid for the 500-1500 MHz frequency range, receiver distances greater than 1km from the base station, and base station antenna heights greater than 30m. It is shown that the Cost 231 Walfish-Ikegami (W-I) model matches extensive experimental data for flat suburban and urban areas with uniform building height (Smith & Dalley, 2000). It has also been found that the suburban path loss model for the Category C (flat terrain, light tree density) is in a good agreement with the Cost 231 W-I model for suburban areas, providing continuity between the two proposed models. The Modified Stanford University Interim channel model can be used for simulations, design, and development and testing of technologies suitable for fixed broadband wireless applications. The FCC propagation model has three basic elements that affect the predicted field strength at the receiver, viz. Line-Of-Sight (LOS) mode, using basic free-space path loss; Non-Line-Of-Sight (NLOS) mode, using multiple wedge diffraction; and Partial first Fresnel zone obstruction losses applicable to either mode. The ITU-R model has the following salient features:

- It takes into account various physical phenomena such as line-of-sight, diffraction, tropospheric scatter, surface ducting, elevated layer reflection and refraction and hydrometeor scatter.

- It uses the Deygout method, for multiple diffraction.
- Path loss is calculated for clear line-of-sight, line-of-sight with sub-path obstruction and trans-horizon cases.
- In analyses, which include terrain information, the FCC or ITU-R models are recommended.
- In analyses, which do not include terrain information, the FCC or ITU-R models can be used, provided that the model for the terrain profiles can be justified.

The “two ray” model is recommended for simple analyses, in which the propagation conditions are clearly line-of sight. To keep it simple, we can choose one or two models that will be the most conclusive and will cover most common cases. We have used the Rayleigh Fading and Ricean Fading Channel models in our discussion.

3. Modelling and simulation of mobile channels

The modeling and simulation of mobile cellular channels have been successfully carried out by several researchers. Various interference patterns including Ricean/Rayleigh fading, co-channel and adjacent-channel interferences can be found in literatures. In this Chapter, we intend to model the mobile cellular channel used in an indoor environment, where the channel can be taken to be of slow fading type. The study is focused to consider the noise contributions from various sources, when they fall within the spectrum of the frequencies used in cellular telephony, and then to design an equalizer which will mitigate the noise present, due to CCI and ACI. When the channel over which data is sent is unknown, which is common, one must employ adaptive equalization. The Decision Feedback Equalizer (DFE) is one such adaptive equalizer. It is known that DFE generally outperforms Linear Equalizer (LE) for the same hardware complexity. Further, as it is indicated earlier, when the channel characteristics show Rayleigh/Ricean fading (due to the presence of multipath), Adjacent Channel Interference (ACI) and Co-Channel Interference (CCI), realization of equalizers based on neuro-fuzzy techniques seems to be most appropriate option for the mobile cellular channel.

Linear space-time equalization is shown to be effective in coping with the complicated propagation conditions for wireless broadband communication in an industrial, indoor environment. This is demonstrated by realistic simulations that use real channel sounder for modeling the influence of the radio channel. Industrial indoor environments like large factory halls show typically a complicated radio channel because of the presence of many reflecting objects. This results in wide delay spreads and a considerably changing channel for a moving mobile unit. There exist a number of options to overcome the difficulties of heavy multipath propagation.

The issues in the design of the neuro-fuzzy channel equalizer to null the effects of fading are investigated. One of the objectives of this article is, to establish the fact that, within an acceptable bound, the mobile cellular channel is Linear Time Variant (LTV). *Another major objective of this article is to investigate the suitability of neuro-fuzzy models as applicable to the analysis and design of mobile cellular channel equalizers.* Three solutions to the channel equalizer problem are investigated in this article. First, a type-2 Fuzzy Adaptive Filter (FAF) for the above purpose is considered. Simulations show that it performs better than a type-1 FAF or Neural Network Classifier (NNC) equalizer. Then the use of Adaptive Network Based Fuzzy Inference System (ANFIS) based equalizer is investigated. Lastly, a Compensatory Neuro-Fuzzy Filter (CNFF) for channel equalization is considered (Lin & Ho, 2004).

Subsequently, an attempt is made to bring the various equalizer realizations in the study under the generic framework of radial basis function (RBF) neural network. Further to this, a novel modular approach for the simulation and design of equalizers for Non-Linear Time-Variant (NLTV) channels is aimed. A suitable model for Ultra-Wide Band (UWB) channel and its equalization is the last goal. It may be noted that the mobile channels are considered to be Linear, and Time Variant (LTV), as a limiting case. However, when the transmitter stages are driven to their non-linear regions, the channel need to be modeled as Non-Linear (to account for the non-linearities thus introduced to the transmitted signal) and Time-Variant (NLTV). The modular approach in combating ISI is to cascade an adaptive non-linear preprocessor filter and linear adaptive equalizer, which simplifies the equalizer design. It is also shown that the ANFIS model can be successfully adapted to equalization of UWB channels.

4. Channel equalization of mobile channels

Channel equalizers are used in digital communication receivers to mitigate the effects of Inter-Symbol Interference (ISI) and inter user interference in the form of Co-Channel Interference (CCI) and adjacent channel interference (ACI) in the presence of Additive White Gaussian Noise (AWGN). Linear equalizers based on adaptive filtering techniques have long been used for this application. Recently, use of nonlinear signal processing techniques like Artificial Neural Networks (ANN) and Radial Basis Functions (RBF) has shown encouraging results in this application. In order to facilitate digital data communication with bandwidth efficiency, we need to use adaptive equalizers. There have been many solutions to the problem of active noise cancellation in mobile indoor channels. Several investigators have studied various active noise cancellers in detail. The main objectives of this article are modeling of a mobile broadband communication channel and designing of a Neuro-Fuzzy Adaptive Equalizer for it. In digital transmission, the problem of Inter-Symbol Interference (ISI) is most often mitigated by the receiving-end equalization. The two important issues in the design and implementation of equalizers are its complexity and its training.

The advancements in Digital Signal Processing (DSP) provide scope for very large scale implementation of many complex algorithms in a lucid manner. The programming capability of the Digital Signal Processors (DSP) makes them very suitable for fast realization. This is definitely an advantage while developing a new system based on an unproven technology, as modifying the design is a matter of re-programming the DSP chip. Thus the product development cycle gets simplified. And, modifications can be incorporated quite easily. Due to the reasons mentioned above, non-linear equalizers are being investigated by many researchers. They include Artificial Neural Networks (ANN), Radial Basis Functions (RBF), recurrent networks and neuro-fuzzy systems.

4.1 Nomenclature of channel equalizers

Recent advances in nonlinear signal processing techniques have provided a rich variety of nonlinear equalizers. Some of the equalizers developed with these processing techniques are based on Volterra filters, ANN, perceptrons, Multi Layer Perceptrons (MLP), RBF networks, fuzzy filters and fuzzy basis functions. All of these nonlinear equalizers, during their training period, optimize some form of a cost function like the *mean square error* (MSE) or probability of error and have the capability of providing the optimum Bayesian equalizer performance in terms of BER.

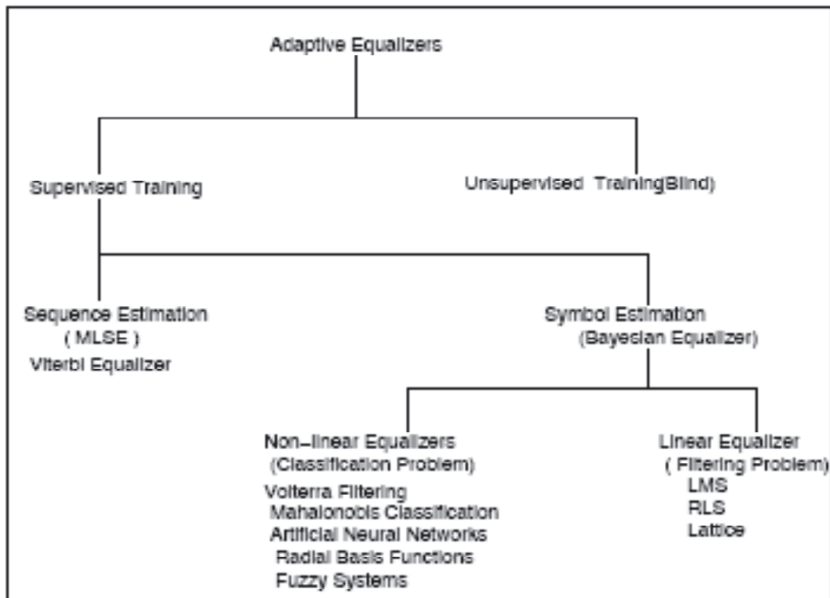


Fig. 1. Classification of Equalizers.

The nonlinear equalizers treat equalization as a pattern classification process where the equalizer attempts to classify the input vector into a number of transmitted symbols. The fuzzy equalizers investigated in this Chapter fall into this category. The Figure 1 shows a typical classification of equalizers. In general the family of adaptive equalizers can be classified as *supervised equalizers* and *unsupervised equalizers*.

5. Neuro fuzzy adaptive equalizers for cellular channels

Fuzzy logic and neural networks (with genetic algorithms) are complementary technologies in the design of intelligent systems (Wang & Mendel, 1993). Fuzzy logic is based on the way the brain deals with inexact information, while neural networks are modeled after the physical architecture of the brain. Fuzzy systems and neural networks are both numerical *model-free estimators* and dynamic systems. They share the common ability to improve the intelligence of systems working in an uncertain, imprecise, and noisy environment. Neural networks provide fuzzy systems with learning abilities, and fuzzy systems provide neural networks with a structured framework with high-level fuzzy *IF-THEN* rule thinking and reasoning. The fuzzy adaptive filter is constructed from a set of fuzzy *IF-THEN* rules that change adaptively to minimize some criterion function as new information becomes available. The concept can be generalized to include complex parameters and complex signals. The fuzzy filter as adaptive equalizer is applied to Quadrature Amplitude Modulation (QAM) digital communication with linear complex channel characteristics.

The fuzzy adaptive filter has drawn a great deal of attention because of its universal approximation ability in nonlinear problems. These fuzzy rules come either from human experts or by matching input-output pairs through an adaptation procedure. Some

application examples of the fuzzy filter to signal processing include classification and signal prediction, communications channel equalization, and nonlinear systems modeling and identification. Most fuzzy filters available to us, are real-valued and are suitable for signal processing in real multidimensional space. In some applications, however, signals are complex valued and processing is done in complex multidimensional space. An example is the equalization of digital communication channels with complex signaling schemes such as QAM. For complex signal processing problems, many existing fuzzy filters cannot directly be applied. It may be emphasized that the cellular, mobile communication channel is essentially a non-linear, time varying, fading channel. Therefore, the design of the equalizer is, it follows that, primarily a non-linear problem. Thus it becomes imperative that we can design appropriate equalizers for wireless fading channels by employing non-linear techniques. Several researchers have come up varying solutions to this daunting problem. In fact, the adaptive equalization is a highly challenging problem, a fact which is evident from the large research literature available and forthcoming.

5.1 The type-2 fuzzy adaptive filter

The type-2 fuzzy adaptive filter for channel equalization is obtained by generalizing the unnormalized output type-1 Takagi-Sugeno-Kang (TSK) fuzzy logic system to a type-2 TSK fuzzy logic system (Liang & Mendel, 2000). For equalization, the antecedents of type-1 TSK FLS are generalized to type-2 fuzzy sets, whereas the consequent is unchanged (i.e., it is a crisp number). In a type-2 FAF with a rule base of M rules, where each rule has p antecedents, the i^{th} rule R_i is denoted as:

$$R_i : \text{IF } x_1^i \text{ is } F_1^i \text{ and } x_2^i \text{ is } F_2^i \text{ and } \dots x_p^i \text{ is } F_p^i \text{ THEN } y^i = c_0^i + c_1^i x_1^i + c_2^i x_2^i + \dots + c_p^i x_p^i \quad (1)$$

where $i = 1, 2, \dots, M$; $c_j^i (j = 0, 1, \dots, p)$ are the consequent parameters that are crisp numbers; y^i is an output from the IF-THEN rule, which is a crisp number and the $F_i^k (k = 1, 2, \dots, p)$ are type-2 fuzzy sets.

6. Performance of various equalizers

We will now examine the performance of typical equalizers employed in wireless, mobile, cellular channels. The block diagram of a typical digital communication system is shown in Figure 2.. It clearly illustrates where the equalizer fits in as a part of the receiver sub system.

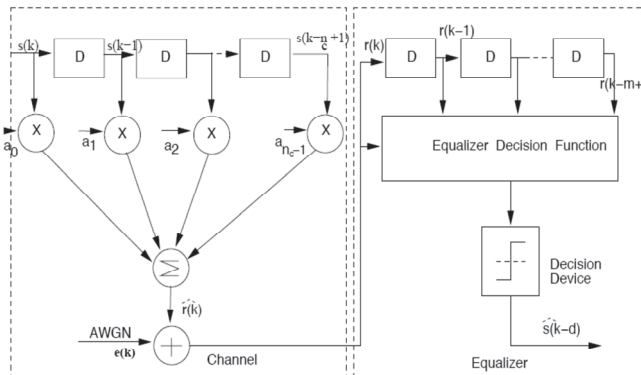


Fig. 2. The Block Diagram of a Typical Digital Communication System.

Typical performance of decision feedback equalizer (DFE), which is a nonlinear system, is shown in Figure 3, below:

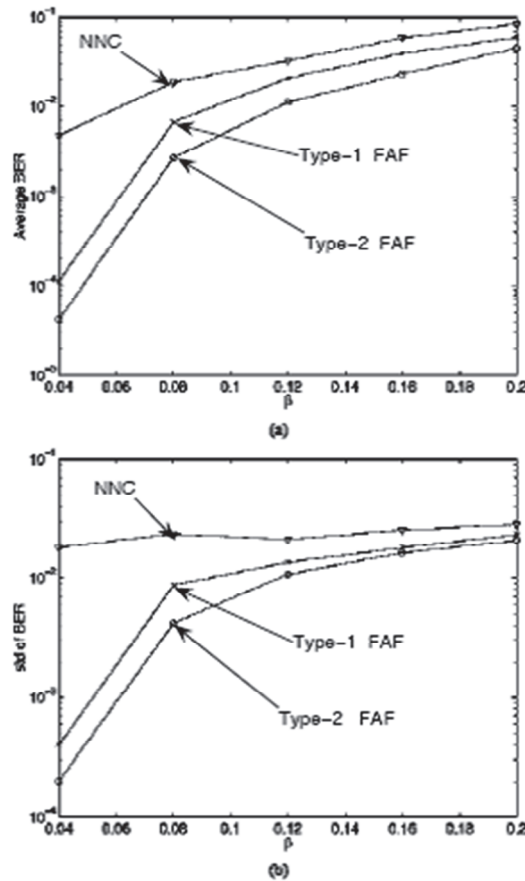


Fig. 3. Performance DFE based on Fuzzy Adaptive Filters (FAF). (a) Mean BER versus std of AWGN. (b) std of BER versus std of AWGN.

We have the following inferences based on the performance curves of type-1 and type-2 FAF (denoted as FAF-I and FAF-II) and neural network classifiers (NNC):

- Both the type-2 FAF TE and DFE perform better than either a type-1 FAF or a Nearest Neighbour Classifier (NNC).
- Since no tuning procedure is used in the design of either type-2 FAF based equalizer, real-time information processing is guaranteed.
- It will be of great advantage to develop a FAF based *Blind Channel Equalizer*.

7. The ANFIS based channel equalizer

In this section, we discuss in detail, the Adaptive Network Based Fuzzy Inference System (ANFIS) and associated equalizer models for mobile cellular and UWB channels. The theory

of ANFIS was introduced by J.S.R. Jang in his seminal paper in (Jang, 1993). It may be noted that the equalization of wireless mobile channels is a non-linear problem, so that a non-linear solution, such as ANFIS, is more appropriate. One has to design the fuzzy *if-then else* rules based on the channel characteristics; namely variances of signal, noise, co-channel (CCI) and adjacent channel interferences (ACI) as well as the transmitted signal (input)-received signal (output) mapping. The equalizer is a non-linear system that effectively undoes the aberrations done to the transmitted signal by the channel due to the noise and co-channel and adjacent channel interferences. Now, the modeling a non-linear system is fairly complex so that conventional methods of system identification techniques cannot be applied to find the inverse system. One possible experimental method to develop a model for indoor wireless channel (viz., the channel impulse response, CIR) is to carry out expensive channel sounding (for example, one could use the RUSK Channel sounder from RF Sub Systems, GmbH, which would cost over a hundred thousand dollars). In this article, we attempt to supplant the expensive channel sounding technique for mobile wireless channel (that too, not restricted to the indoor case) by suitable simulation techniques.

7.1 Introduction to ANFIS

Functionally, there are almost no constraints on the node functions of an adaptive network except piecewise differentiability. Structurally, the only limitation of network configuration is that it should be of feedforward type. Due to these minimal restrictions, the adaptive network's applications are immediate and immense in various areas. For simplicity, we assume the fuzzy inference system under consideration has two inputs x and y and one output z (Jang, 1993). Suppose that the rule base contains two fuzzy if-then rules of Takagi and Sugeno's type:

$$\text{Rule1 : If } x \text{ is } A_1 \text{ and } y \text{ is } B_1, \text{ then } f_1 = p_1x + q_1y + r_1. \quad (2)$$

$$\text{Rule2 : If } x \text{ is } A_2 \text{ and } y \text{ is } B_2, \text{ then } f_2 = p_2x + q_2y + r_2. \quad (3)$$

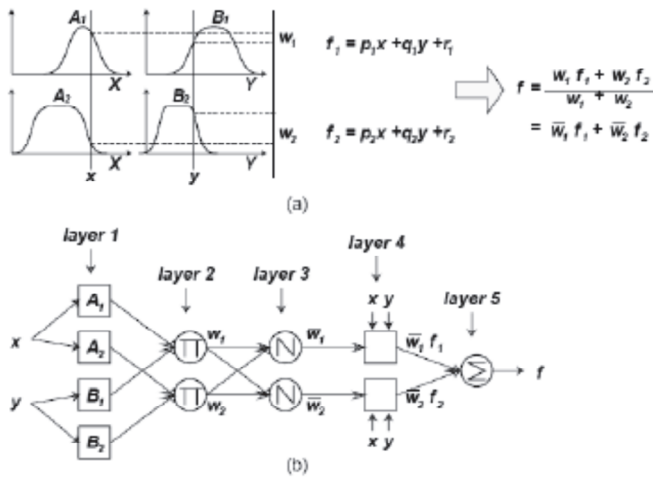


Fig. 4. The Takagi-Sugeno-Kang (TSK) Model of Fuzzy Reasoning. (a) Type-3 Fuzzy Reasoning. (b) Equivalent ANFIS (Type-3 ANFIS).

The type-3 fuzzy reasoning is illustrated in Figure 4(a) and the corresponding equivalent ANFIS architecture (type-3 ANFIS) is shown in Figure 4(b).

7.2 Mobile cellular channel equalization based on ANFIS

One can observe that wireless channel can be modeled as non-linear time-variant (NLTV) when the duration of observation window is fairly long or as non-linear time invariant (NLTI) when the duration of observation window is short. This fact is established by simulation, as it is a hard problem to obtain a rigorous mathematical proof. Conventional channel models available in recent literature were studied to arrive at a suitable paradigm for the wireless channel, consisting of the different variables and parameters. This also enabled us to understand the inadequacies of existing mathematical models for wireless channels. The fuzzy if-then rules are generated by the *inverse system* based on ANFIS (to the channel), which effectively acts as an adaptive equalizer at the receiver side. The ANFIS automatically generate the rule base from a set of input-output data vectors. This is achieved by minimizing the error between actual input signal (at the transmitter of the wireless system) and the estimate of the input (at the receiver). In the simulation, we assume that the external input to the ANFIS equalizer is the output of the channel, which is the sum of the desired channel output plus the weighted sum of the co-channel outputs and the Gaussian noise, which is assumed to be AWGN, with zero mean and standard deviation upto 0.8. In the ensuing sections, we use the following definitions for Signal-to-Noise Ratio (SNR), Signal-to-Interference Ratio (SIR) and Signal-to-Interference Noise Ratio (SINR) (Liang & Mendel 2000):

$$\text{SNR} = 10 \log_{10} \{\sigma_s^2 / \sigma_n^2\} \quad (4)$$

$$\text{SIR} = 10 \log_{10} \{\sigma_s^2 / \sigma_i^2\} \quad (5)$$

$$\text{SINR} = 10 \log_{10} \{\sigma_s^2 / (\sigma_i^2 + \sigma_n^2)\} \quad (6)$$

where σ_s^2 , σ_n^2 , and σ_i^2 are the variances of the signal, AWG noise, and the co-channel and adjacent channel interferences (put together) signal respectively.

Type	Nodes	Linear/Nonlinear Parameters	Fuzzy Rules
ANFIS-17	32	14/14	7
ANFIS-115	64	30/30	15
ANFIS-125	104	50/50	25
ANFIS-25	75	75/20	25
ANFIS-27	131	147/28	49
ANFIS-35	286	500/30	125
ANFIS-37	734	1372/42	343

Table 1. Simulation Parameters for Various ANFIS Based Channel Equalizers

The output of the equalizer is given to a limiter to clip the output levels to limiting values of +1 or -1. The different parameters of the various simulation setups are as tabulated in Table 1. The structure of ANFIS-27 is given in Figure 5. The library function, *anfis*, available in the *Fuzzy Logic Toolbox* of MATLAB® R2010b is used extensively in all simulations.

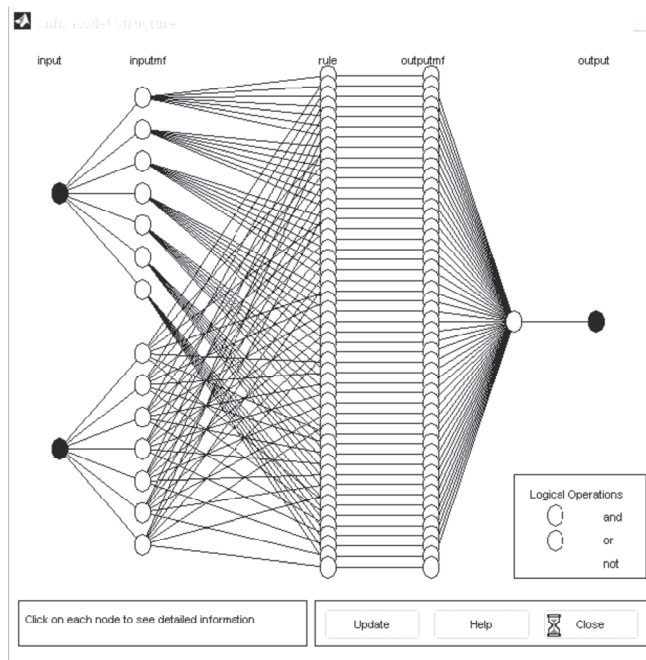


Fig. 5. The Structure of ANFIS–27 Generated Using MATLAB Fuzzy Logic Toolbox.

Note that ANFIS-27 based equalizer has two inputs from multipath components, seven (7) fuzzy rules for each input and one output that feed the receiver subsystem.

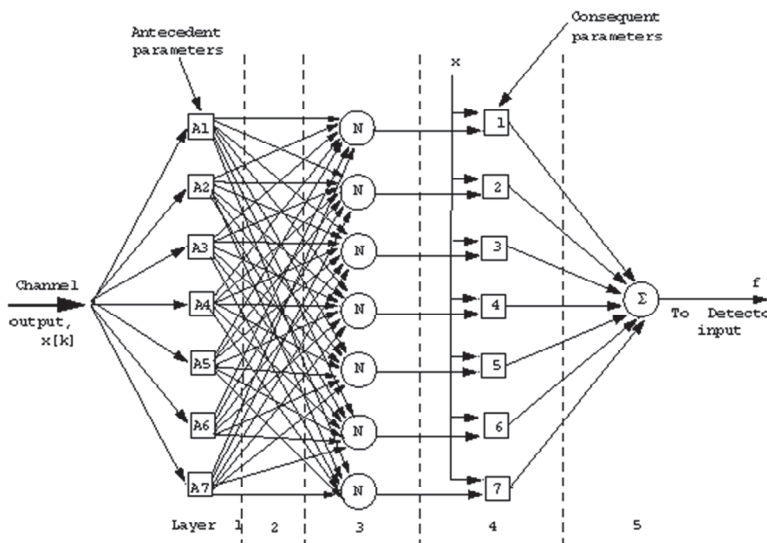


Fig. 6. The Equivalent ANFIS Architecture for Channel Equalizer.

The Figure 6 shows the architecture of the proposed ANFIS based channel equalizer, for 7 fuzzy rules. The wireless channel modeling based on artificial neural networks is capable of depicting the input-output mapping existing in the equalizer system and it does provide us with an exact picture of the variables and parameters defining the system. Moreover, neural network based models do have the learning capability. The fuzzy models, on the other hand, do not possess the learning capability. Therefore, fusing together these two, we can have a model which is capable of both depicting the dynamics of the system in terms of the variables and parameters and is having the self-learning capability. The adaptability of the equalizer under purview is achieved by the learning aspect of neural network. The fuzzy reasoning (especially the TSK model used in ANFIS) maps the input to the output. We follow a first-order ANFIS with the antecedent parameters being the standard deviations of the received signal, CCI and ACI interferences (put together), and the AWGN (σ_s , σ_i , and σ_n , respectively), collectively represented as A_i . The only consequent parameter is the scaling factor of the signal (ρ_i) at the output. The membership functions of A_i , $i = 1, 2, \dots, 7$ are chosen to be *Gaussian*. Some of the rules in the *fuzzy rule base* can be stated as:

$$\text{If } \sigma_s \text{ is } \mathbf{very\ low}, \text{ and } \sigma_i \text{ is } \mathbf{very\ low}, \text{ and } \sigma_n \text{ is } \mathbf{very\ low} \text{ then } y = \rho_1 s. \quad (7)$$

$$\text{If } \sigma_s \text{ is } \mathbf{low}, \text{ and } \sigma_i \text{ is } \mathbf{very\ low}, \text{ and } \sigma_n \text{ is } \mathbf{very\ low} \text{ then } y = \rho_2 s. \quad (8)$$

$$\text{If } \sigma_s \text{ is } \mathbf{medium}, \text{ and } \sigma_i \text{ is } \mathbf{very\ low}, \text{ and } \sigma_n \text{ is } \mathbf{very\ low} \text{ then } y = \rho_3 s. \quad (9)$$

$$\text{If } \sigma_s \text{ is } \mathbf{medium}, \text{ and } \sigma_i \text{ is } \mathbf{low}, \text{ and } \sigma_n \text{ is } \mathbf{low} \text{ then } y = \rho_4 s. \quad (10)$$

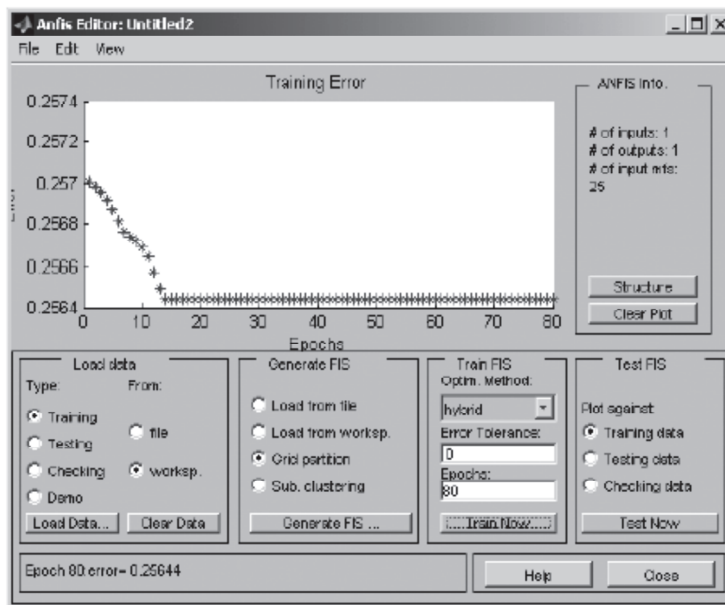


Fig. 7. The Error Plot of ANFIS-125 Training.

The three input variables can assume any one of the 5 possible membership functions from the set, *{very low, low, medium, high, very high}*, leaving us with 125 possible combinations of rules. However, using *fuzzy rule reduction techniques* the total number of rules can be limited to 7 or 25. The steps in the algorithm for simulation of the ANFIS-27 based equalizer are as given below:

1. The standard deviations of CCI and AWGN are logarithmically varied from 0.02 to 0.8. This information is derived from literature.
2. The random binary input data (which represents the input to the channel from the transmitter) is generated and the corrupted data available at the outputs of the two multipaths due to CCI and AWGN is obtained.
3. Set the number of membership functions as 7, membership function type as "Gaussian" and the number of epochs to 80.
4. Simulate the ANFIS (which implements the equalizer) and plot the results.

The error plot of the ANFIS-125 training is illustrated in Figure 7.

We have set the number of epochs as 80 in this case. *As the ANFIS implementation in MATLAB do not lend itself to observe the updation of Antecedent and consequent parameters, while training is under progress, we can consider the training error as a reliable pointer to the step wise updation of the above parameters.* The ANFIS-125 consists of one input, one output, and 25 fuzzy rules for each membership functions.

7.3 The results of simulations of ANFIS based equalizers

The simulated output of the channel, which is the input to the ANFIS based channel equalizer, along with the training data is shown in Figure 8. The output of the channel (received signal), which is a non-linear combination of the signal, the co-channel signals, and the AWG noise, is a random waveform taking values around +1 and -1, as seen from the simulated waveform. The MATLAB code to generate the plot is given below.

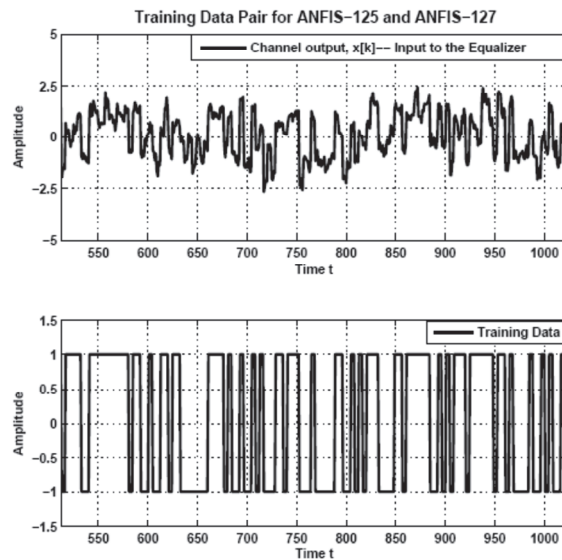


Fig. 8. The Training Data Pair for ANFIS-125 Equalizer Simulation.

The equalized, output after thresholding, will be very much identical to the *training data* as shown in Figure 9.

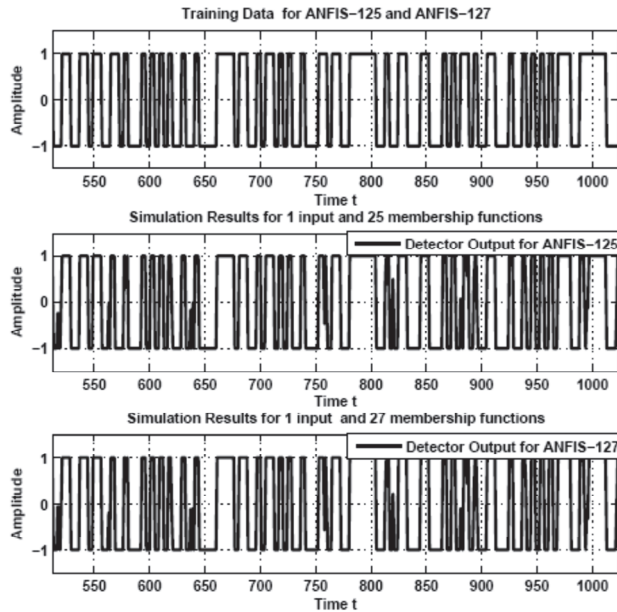


Fig. 9. Simulation Results for ANFIS-125 and ANFIS-127 based Equalizers.

%% MATLAB Code for ANFIS Equalizer Simulation with 1 input and 5
 %% membership functions /1 input and 25 membership functions.
 % Last modified on 13-03-2011.

```
clear all; clf;
tic;
ns=1024;
nb=4;
t=[1:ns*nb];
[x,b] = random_binary(ns,nb);
[x1,b1] = random_binary(ns,nb);
[x2,b2] = random_binary(ns,nb);
[x3,b3] = random_binary(ns,nb);
[x4,b4] = random_binary(ns,nb);
[x5,b5] = random_binary(ns,nb);
[x6,b6] = random_binary(ns,nb);
e=0.2*randn(ns*nb,1);
y1 = x'+0.2*(x1'+x2'+x3'+x4'+x5'+x6')+e;
trnData = [y1 x'];
numMFs = 25;
mfType = 'gaussmf';
epoch_n = 20;
```

```

subplot(211),plot(t(512:1024),y1(512:1024),'k');
axis([512 1024 -5 5]); grid on;
xlabel('Time t');ylabel('Amplitude');
legend('Channel output, x[k]-- Input to the Equalizer');
title('Training Data Pair for ANFIS-125 and ANFIS-127');
subplot(212),plot(t(512:1024),x(512:1024),'k');
axis([512 1024 -1.5 1.5]); grid on;
xlabel('Time t');ylabel('Amplitude');
legend('Training Data');
%%%
%% ANFIS Equalizer Simulation with 1 input and 5 membership
%% functions /1 input and 25 membership functions.
% Last modified on 13-03-2011.
clear all;clf;
tic;
ns=1024;
nb=4;
t=[1:ns*nb];
[x,b] = random_binary(ns,nb);
[x1,b1] = random_binary(ns,nb);
[x2,b2] = random_binary(ns,nb);
[x3,b3] = random_binary(ns,nb);
[x4,b4] = random_binary(ns,nb);
[x5,b5] = random_binary(ns,nb);
[x6,b6] = random_binary(ns,nb);
e=0.2*randn(ns*nb,1);
y1 = x'+0.2*(x1'+x2'+x3'+x4'+x5'+x6')+e;

trnData = [y1 x'];
numMFs = 25;
mfType = 'gaussmf';
epoch_n = 20;
in_fisat = genfis1(trnData,numMFs,mfType);
out_fisat = anfis(trnData,in_fisat,20);
est_x125=evalfis(y1,out_fisat);

est_x125(est_x125<-0.6)=-1.0;
est_x125(est_x125>0.6)=1.0;

trnData = [y1 x'];
numMFs = 27;
mfType = 'gaussmf';
epoch_n = 20;
in_fisat = genfis1(trnData,numMFs,mfType);
out_fisat = anfis(trnData,in_fisat,20);
est_x127=evalfis(y1,out_fisat);

```

```

est_x127(est_x127<-0.6)=-1.0;
est_x127(est_x127>0.6)=1.0;

subplot(311),plot(t(512:1024),x(512:1024),'k');
axis([512 1024 -1.5 1.5]); grid on;
xlabel('Time t');ylabel('Amplitude');
title('Training Data for ANFIS-125 and ANFIS-127');

subplot(312),plot(t(512:1024),est_x125(512:1024),'k');
axis([512 1024 -1.5 1.5]); grid on;
xlabel('Time t');ylabel('Amplitude');
legend('Detector Output for ANFIS-125');
title('Simulation Results for 1 input and 25 membership functions');

subplot(313),plot(t(512:1024),est_x127(512:1024),'k');
axis([512 1024 -1.5 1.5]); grid on;
xlabel('Time t');ylabel('Amplitude');
legend('Detector Output for ANFIS-127');
title('Simulation Results for 1 input and 27 membership functions');
toc;
%%%

```

In one of the simulations, the standard deviation of CCI and AWGN are logarithmically varied between 0.02 and 0.8 and simulation is run on a total of 2048/4096 training data pairs. The results are shown in Figure 10. The MATLAB code to generate the same is appended below.

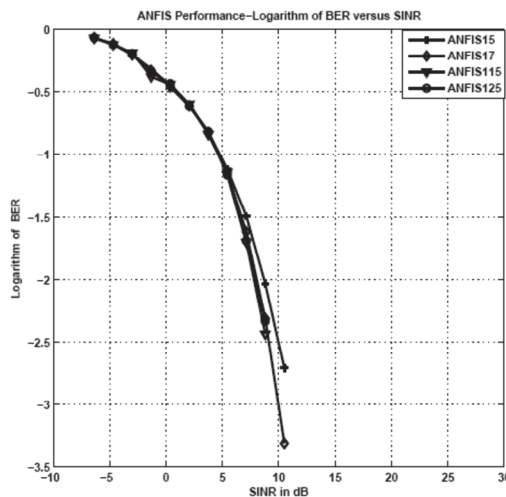


Fig. 10. Performance of ANFIS based Equalizer – Logarithm of BER at output versus SINR in dBs.

```

% %% Modified ANFIS Equalizer Simulation
% %% with more precision. Plots Logarithm of BER versus SINR
% std of CCI varied from 0.02 to 0.8.
% std of AWGN varied from 0.02 to 0.8.
% %% last modified on 13-03-2011.
%anfis15.m
clc;clf;clear;close;
tic;
nb=1024;
ns=4;
it=20;
% t=linspace(0.02,0.8,it);
t=logspace(log10(0.02),log10(0.8),it);
[x,b] = random_binary(nb,ns);
[x1,b1] = random_binary(nb,ns);
[x2,b2] = random_binary(nb,ns);
[x3,b3] = random_binary(nb,ns);
[x4,b4] = random_binary(nb,ns);
[x5,b5] = random_binary(nb,ns);
[x6,b6] = random_binary(nb,ns);
i=1;
cci=x1+x2+x3+x4+x5+x6;
noise=randn(ns*nb,1);
for j=1:it
    e=noise*t(j);
    z=cci*t(j);
    sinr(i)=10*log10(var(x)/(var(z)+var(e)));
    y = x'+z'+e;
    trnData = [y x'];
    numMFs = 5;
    mfType = 'gaussmf';
    epoch_n = 20;
    in_fisimat = genfis1(trnData,numMFs,mfType);
    out_fisimat = anfis(trnData,in_fisimat,20);
    est_x=evalfis(y,out_fisimat);
    est_x(est_x<-0.6)=-1.0;
    est_x(est_x>0.6)=1.0;
    ec=(x~=est_x');
    ber(i)=sum(ec)/(nb*ns);
    i=i+1;
end;
plot(sinr, log10(ber),'-+');hold on;grid on;
xlabel('SINR in dB');
ylabel('Logarithm of BER');
%end of ANFIS15;
i=1;
for j=1:it

```

```

    e=noise*t(j);
    z=cci*t(j);
    sinr(i)=10*log10(var(x)/(var(z)+var(e)));
    y = x'+z'+e;
    trnData = [y x'];
    numMFs = 7;
    mfType = 'gaussmf';
    epoch_n = 20;
    in_fisimat = genfis1(trnData,numMFs,mfType);
    out_fisimat = anfis(trnData,in_fisimat,20);
    est_x=evalfis(y,out_fisimat);
    est_x(est_x<-0.6)=-1.0;
    est_x(est_x>0.6)=1.0;
    ec=(x~=est_x');
    ber(i)=sum(ec)/(nb*ns);
    i=i+1;
end;
plot(sinr, log10(ber),'-d');hold on;grid on;
xlabel('SINR in dB');
ylabel('Logarithm of BER');
%end of ANFIS17;
i=1;
for j=1:it
    e=noise*t(j);
    z=cci*t(j);
    sinr(i)=10*log10(var(x)/(var(z)+var(e)));
    y = x'+z'+e;
    trnData = [y x'];
    numMFs = 15;
    mfType = 'gaussmf';
    epoch_n = 20;
    in_fisimat = genfis1(trnData,numMFs,mfType);
    out_fisimat = anfis(trnData,in_fisimat,20);
    est_x=evalfis(y,out_fisimat);
    est_x(est_x<-0.6)=-1.0;
    est_x(est_x>0.6)=1.0;
    ec=(x~=est_x');
    ber(i)=sum(ec)/(nb*ns);
    i=i+1;
end;
plot(sinr, log10(ber),'-v');hold on;grid on;
xlabel('SINR in dB');
ylabel('Logarithm of BER');
%end of ANFIS115;
i=1;
for j=1:it
    e=noise*t(j);

```

```

    z=cci*t(j);
    sinr(i)=10*log10(var(x)/(var(z)+var(e)));
    y = x'+z'+e;
    trnData = [y x'];
    numMFs = 25;
    mfType = 'gaussmf';
    epoch_n = 20;
    in_fisat = genfis1(trnData,numMFs,mfType);
    out_fisat = anfis(trnData,in_fisat,20);
    est_x=evalfis(y,out_fisat);
    est_x(est_x<-0.6)=-1.0;
    est_x(est_x>0.6)=1.0;
    ec=(x~=est_x');
    ber(i)=sum(ec)/(nb*ns);
    i=i+1;
end;
plot(sinr, log10(ber),'-o');hold on;grid on;
xlabel('SINR in dB');
ylabel('Logarithm of BER');
%end of ANFIS125;
legend('ANFIS15','ANFIS17','ANFIS115','ANFIS125');
title('ANFIS Performance-Logarithm of BER versus SINR');
hold off;
toc;
% % % % % % % % % % % % % % %

```

In another simulation, the $\log(BER)$ at output of the equalizer is calculated for standard deviation of noise varying from 0.02 to 0.8, and the for two versions of ANFIS Equalizers (ANFIS-115 and ANFIS-125) for 2048/4096 training data pairs and standard deviation of AWGN fixed at 0.42, and the results are plotted in Figure 11. The MATLAB code used for the simulation is given below.

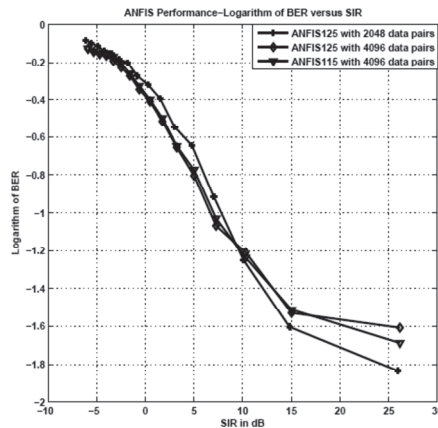


Fig. 11. Performance of ANFIS based Equalizer – Logarithm of BER at output versus SIR in dBs.

```

% %% Modified ANFIS Equalizer Simulation
% %% with more precision. Plots Logarithm of BER versus SIR
% std of CCI varied from 0.02 to 0.8.
% std of AWGN fixed to 0.42.
% %% last modified on 06-02-2010..
%anfis125.m
clc;clf;clear;close;
tic;
nb=512;
ns=4;
it=16;
t=linspace(0.02,0.8,it);
%t=logspace(log10(0.02),log10(0.8),it);
[x,b] = random_binary(nb,ns);
[x1,b1] = random_binary(nb,ns);
[x2,b2] = random_binary(nb,ns);
[x3,b3] = random_binary(nb,ns);
[x4,b4] = random_binary(nb,ns);
[x5,b5] = random_binary(nb,ns);
[x6,b6] = random_binary(nb,ns);
i=1;
cci=x1+x2+x3+x4+x5+x6;
e=randn(ns*nb,1)*0.42;
for j=1:it
    z=cci*t(j);
    sir(i)=10*log10(var(x)/var(z));
    y = x'+z'+e;
    trnData = [y x'];
    numMFs = 25;
    mfType = 'gaussmf';
    epoch_n = 20;
    in_fismat = genfis1(trnData,numMFs,mfType);
    out_fismat = anfis(trnData,in_fismat,20);
    est_x=evalfis(y,out_fismat);
    est_x(est_x<-0.6)=-1.0;
    est_x(est_x>0.6)=1.0;
    ec=(x~=est_x');
    ber(i)=sum(ec)/(nb*ns);
    i=i+1;
end;
plot(sir, log10(ber),'-+');hold on;grid on;
xlabel('SIR in dB');
ylabel('Logarithm of BER');
% end of ANFIS125 with 2048 data pairs..
% %% anfis125.m with 4096 data pairs..
nb=1024;
ns=4;

```

```

it=16;
t=linspace(0.02,0.8,it);
%t=logspace(log10(0.02),log10(0.8),it);
[x,b] = random_binary(nb,ns);
[x1,b1] = random_binary(nb,ns);
[x2,b2] = random_binary(nb,ns);
[x3,b3] = random_binary(nb,ns);
[x4,b4] = random_binary(nb,ns);
[x5,b5] = random_binary(nb,ns);
[x6,b6] = random_binary(nb,ns);
i=1;
cci=x1+x2+x3+x4+x5+x6;
e=randn(ns*nb,1)*0.42;
for j=1:it
    z=cci*t(j);
    sir(i)=10*log10(var(x)/var(z));
    y = x'+z'+e;
    trnData = [y x'];
    numMFs = 25;
    mfType = 'gaussmf';
    epoch_n = 20;
    in_fismat = genfis1(trnData,numMFs,mfType);
    out_fismat = anfis(trnData,in_fismat,20);
    est_x=evalfis(y,out_fismat);
    est_x(est_x<-0.6)=-1.0;
    est_x(est_x>0.6)=1.0;
    ec=(x~=est_x');
    ber(i)=sum(ec)/(ns*nb);
    i=i+1;
end;
plot(sir, log10(ber),'-d');hold on; grid on;
%end of ANFIS125.m
%% anfis115.m with 4096 data pairs..
i=1;
cci=x1+x2+x3+x4+x5+x6;
e=randn(ns*nb,1)*0.42;
for j=1:it
    z=cci*t(j);
    sir(i)=10*log10(var(x)/var(z));
    y = x'+z'+e;
    trnData = [y x'];
    numMFs = 15;%number of membership_rules
    mfType = 'gaussmf';
    epoch_n = 20;
    in_fismat = genfis1(trnData,numMFs,mfType);
    out_fismat = anfis(trnData,in_fismat,20);
    est_x=evalfis(y,out_fismat);

```



```

est_x(est_x<-0.6)=-1.0;
est_x(est_x>0.6)=1.0;
ec=(x~=est_x');
ber(i)=sum(ec)/(nb*ns);
i=i+1;
end;
plot(sir,log10(ber),'-v');hold on;grid on;
%end of ANFIS115.m
legend('ANFIS125 with 2048 data pairs','ANFIS125 with 4096 data pairs','ANFIS115 with
4096 data pairs');
title('ANFIS Performance-Logarithm of BER versus SIR');
hold off;
toc;
% % % % % % % %

```

In this case, we can observe that the $\log(BER)$ reduces as the SIR in dB increases, consistently. Also, when the number of training data pairs is increased, there is a marginal improvement in performance. The performance for the above ANFIS pairs, as regards $\log(BER)$ at output of the equalizer *versus* SNR in dB s for standard deviation of co-channel interference signal fixed at 0.08, is given in Figure 12. The MATLAB code is also given.

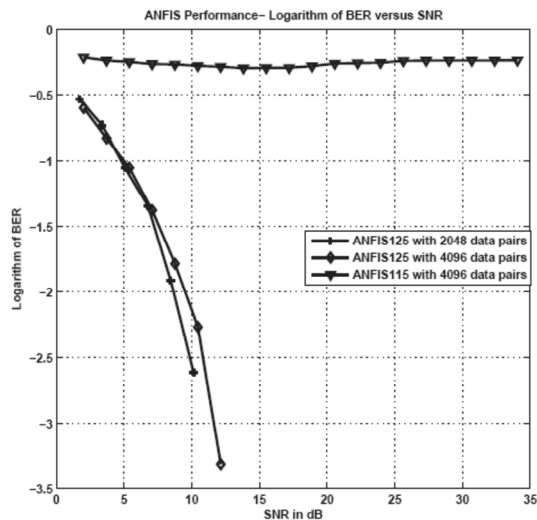


Fig. 12. Performance of ANFIS based Equalizer – Logarithm of BER versus SNR in dB s.

```

% %Modified ANFIS Equalizer Simulation for diff. number of data pairs..
% with more precision. Plots BER versus SNR
% std of CCI fixed at 0.18.
% std of AWGN varied from 0.02 to 0.8.
% % % last modified on 13-03-2011.
%ANFIS125.m
clc;clf;clear;close;

```

```

tic;
nb=512;
ns=4;
it=20;
t=logspace(log10(0.02),log10(0.8),it);% %
[x,b] = random_binary(nb,ns);
[x1,b1] = random_binary(nb,ns);
[x2,b2] = random_binary(nb,ns);
[x3,b3] = random_binary(nb,ns);
[x4,b4] = random_binary(nb,ns);
[x5,b5] = random_binary(nb,ns);
[x6,b6] = random_binary(nb,ns);
i=1;
cci=x1+x2+x3+x4+x5+x6;
noise=randn(ns*nb,1);
for j=1:it
    e=noise*t(j);
    z=cci*0.08; % std of cci is fixed as 0.08
    snr(i)=10*log10(var(x)/var(e));
    y = x'+z'+e;
    trnData = [y x'];
    numMFs = 25;
    mfType = 'gaussmf';
    epoch_n = 20;
    in_fisimat = genfis1(trnData,numMFs,mfType);
    out_fisimat = anfis(trnData,in_fisimat,20);
    est_x=evalfis(y,out_fisimat);
    est_x(est_x<-0.6)=-1.0;
    est_x(est_x>0.6)=1.0;
    ec=(x~est_x');
    ber(i)=sum(ec)/(nb*ns);
    i=i+1;
end;
plot(snr, log10(ber),'-+');hold on;grid on;
xlabel('SNR in dB');
ylabel('Logarithm of BER');
%end of ANFIS125;
nb=1024;
ns=4;

t=logspace(log10(0.02),log10(0.8),it);
[x,b] = random_binary(nb,ns);
[x1,b1] = random_binary(nb,ns);
[x2,b2] = random_binary(nb,ns);
[x3,b3] = random_binary(nb,ns);
[x4,b4] = random_binary(nb,ns);
[x5,b5] = random_binary(nb,ns);

```

```

[x6,b6] = random_binary(nb,ns);
i=1;
cci=x1+x2+x3+x4+x5+x6;
noise=randn(ns*nb,1);
for j=1:it
    e=noise*t(j);
    z=cci*0.08; % std of cci is fixed as 0.18
    snr(i)=10*log10(var(x)/var(e));
    y = x'+z'+e;
    trnData = [y x'];
    numMFs = 25;
    mfType = 'gaussmf';
    epoch_n = 20;
    in_fismat = genfis1(trnData,numMFs,mfType);
    out_fismat = anfis(trnData,in_fismat,20);
    est_x=evalfis(y,out_fismat);
    est_x(est_x<-0.6)=-1.0;
    est_x(est_x>0.6)=1.0;
    ec=(x~=est_x');
    ber(i)=sum(ec)/(nb*ns);
    i=i+1;
end;
plot(snr, log10(ber),'-d');hold on;grid on;
xlabel('SNR in dB');
ylabel('Logarithm of BER');
%end of ANFIS125;
%%anfis115.m

t=logspace(log10(0.02),log10(0.8),it);
[x,b] = random_binary(nb,ns);
[x1,b1] = random_binary(nb,ns);
[x2,b2] = random_binary(nb,ns);
[x3,b3] = random_binary(nb,ns);
[x4,b4] = random_binary(nb,ns);
[x5,b5] = random_binary(nb,ns);
[x6,b6] = random_binary(nb,ns);
i=1;
for j=1:it
    e=noise*t(j);
    z=cci*.08;
    snr(i)=10*log10(var(x)/var(e));
    y1 = x'+z'+e;
    trnData = [y1 x'];
    numMFs = 15;%number of membership_rules
    mfType = 'gaussmf';
    epoch_n = 20;
    in_fismat = genfis1(trnData,numMFs,mfType);

```

```

out_fismat = anfis(trnData,in_fismat,20);
est_x=evalfis(y,out_fismat);
est_x(est_x<-0.6)=-1.0;
est_x(est_x>0.6)=1.0;
ec=(x~=est_x');
ber(i)=sum(ec)/(nb*ns);
i=i+1;
end;
plot(snr,log10(ber),'-v');hold on;grid on;
%end of ANFIS115.m
legend('ANFIS125 with 2048 data pairs','ANFIS125 with 4096 data pairs','ANFIS115 with
4096 data pairs');
title('ANFIS Performance- Logarithm of BER versus SNR');
hold off;
toc;
%%%%%%%%

```

Note that in this case, *ANFIS-115 fails to converge and hence there is no variation in performance even when the SNR is as high as 35dB*. Again, the performance is marginally better, when the number of training data pairs is increased. A plot of the performance of two different ANFIS structures (average BER at the output of the equalizer *versus* SNR and standard deviation of BER at the output of the equalizer *versus* standard deviation of AWGN) based on 100 Monte-Carlo (MC) simulations is given in Figure 12 for ANFIS-115 and ANFIS-125 structures. 4096 training data pairs are used in the simulation. The *Mean(BER) versus SNR* performance improves consistently, as the *SNR* increases. The *std(BER) versus SNR* performance, on the otherhand, deteriorates as the *std(AWGN)* increases, consistently. Performances are marginally better for ANFIS-125 based equalizer. We will now consider the interpretation of the results of simulation of various ANFIS based equalizers.

7.4 Interpretation of results of simulations of ANFIS based equalizers

The following observations are made based on Figures 9, 10, 11, 12 and 13 and Tables 1 as well as results of simulations with less number of data pairs:

- With more number of training data pairs, BER at the output of the equalizer is reduced. This is due to the fact that the ANFIS gets optimally tuned with more training data pairs.
- As shown in Figure 10, performance of all ANFIS Equalizers w.r.t. $\log(BER)$ at the output of the equalizer *versus* SINR, is nearly identical. When the SINR is above $-10dB$, practically the $\log(BER)$ becomes close to zero. However, ANFIS-125 performs slightly better than other structures.
- The performance of ANFIS-125 w.r.t. $\log(BER)$ at the output of the equalizer *versus* SIR is almost identical with 2048 or 4096 data pairs. However for ANFIS-115, performance is slightly poor.
- As it is shown in Figure 12, the performance of ANFIS-125 w.r.t. $\log(BER)$ at the output of the equalizer *versus* SNR is almost identical with 2048 or 4096. data pairs. However for ANFIS-115, performance is very poor even at a SNR of 35dB. This is due to the fact

that equalizer model with ANFIS structure fails to perform when the number of rules is 15. *The AWGN overwhelms the signal, when number of rules for the ANFIS is 15 or less.*

- For MISO or MIMO systems, increasing the number of membership functions is the option for accurate system modeling, since in these cases number of inputs applied to the ANFIS is two or more, and hence it will not be optimal to increase the number of internal inputs in ANFIS.
- An optimal ANFIS structure can be obtained based on the training time and the maximum error that can be tolerated. As indicated in Figure 13, at higher values of standard deviation of AWGN, and that of standard deviation of BER will be less with more number of membership functions. Hence standard deviation of BER at the output of the equalizer can be yet another criterion in selecting a particular ANFIS structure.

8. The equalization of Ultra-Wide Band channels using ANFIS

The Ultra-Wide Band (UWB) is an emerging wireless technology that has recently gained much interest from the communication research and industry (Molisch, 2005a). UWB systems possess unique characteristics and capabilities that make them suitable for short range, high-speed wireless communications (Molisch, 2005b). The UWB systems use signals that are based on repetitive transmissions of short pulses formed by using a single basic pulse shape. The transmitted signals have an extremely low power spectral density and occupy very large bandwidth of several GHz. Thus the UWB systems can operate with negligible interference to the existing radio systems. UWB can provide very high bit rate, low-cost, low-power wireless communication for wide variety of systems: personal computer, TV, VCR, CD, DVD, and MP3 players (Algans et.al., 2002).

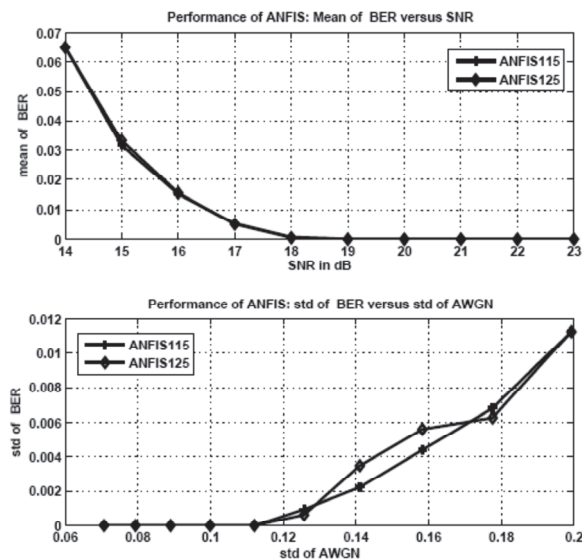


Fig. 13. Performance of ANFIS Equalizers: (a) Mean BER at out of the equalizer *versus* SNR in dBs, and (b) Standard Deviation of BER at output of the equalizer *versus* Standard Deviation of AWGN.

As per FCC recommendations, UWB systems have the following characteristics:

- They have a *relative bandwidth that is larger than 25% of the carrier frequency and/or an absolute band-width more than 500MHz.*
- They occupy a frequency band of 3.1GHz to 10.6GHz.
- FCC have recently allocated 7.5GHz of spectrum for unlicensed commercial UWB communication systems.
- Maximum radiated power is 75nW/MHz (-41.32dBm/MHz) (Molisch, 2005a).

The following are the significant merits of UWB:

1. Accurate position location and ranging, due to the better time resolution.
2. No significant multipath fading due to better time resolution.
3. Multiple access due to wide transmission bandwidths.
4. Possibility of extremely high data rates.
5. Covert communications due to low transmission power operation.
6. Possible easier material penetration due to the presence of components at different frequencies.

8.1 The channel covariance matrix formulation for UWB channels

The Shadow-fading fluctuations of the average received power are known to be lognormally distributed. Recently, for macrocell scenarios, the fluctuations in delay and angle spread are shown to behave similarly (Algans et.al., 2002). The reason is that these quantities are sums of powers of individual sub-paths times the square of their corresponding delay times or angles. Since the powers are log-normally distributed and *sums of lognormal variables are approximately log-normal*, this implies that angle-spreads and delay-spreads have log-normal distributions (Beaulieu et.al., 1995). This motivation of how angle spread and delay spread are log-normally distributed also suggests that they will be correlated with shadow fading and each other. Let us assume that $X_{1n}, X_{2n}, X_{3n}, \dots$ are zero-mean, unit-variance Gaussian random variables, representing the signals received at base station n . Then we define:

$$\rho_{DA} = E [X_{1n}, X_{2n}] \quad (11)$$

$$\rho_{DF} = E [X_{2n}, X_{3n}] \quad (12)$$

$$\rho_{AF} = E [X_{3n}, X_{1n}] \quad (13)$$

$$\zeta = E [X_{3n}, X_{3m}] \quad (14)$$

In particular, $\sigma_{SF,n}$ (variance of shadow fading component w.r.t. to base station, n) is negatively correlated with $\sigma_{DS,n}$ (variance of *delay spread*) and $\sigma_{AS,n}$ (variance of *angle spread*), while the latter two have positive correlations with each other. It should be noted that this relationship does not hold for the angle-spread at the mobile since the different paths with distinct angles do not necessarily lead to such pronounced differences in the delays. These correlations can be expressed in terms of a covariance matrix A , whose A_{ij} component represents the correlations between X_{in} and X_{jn} , with $i, j = 1, 2, 3$. Note that the matrix A is symmetrical. Measurements of cross-correlations of these parameters between different base stations are more difficult. In particular, only correlations between shadow-fading components have been adopted. These correlations are assumed to be the same between any two different base-stations and are denoted by ζ . For simplicity and due to lack of further data, the cross-correlation matrix between the X_{in} triplet ($i = 1, 2, 3$) of different base-stations are assumed to be given by the following matrix B .

$$A = \begin{bmatrix} 1 & \rho_{DA} & \rho_{DF} \\ \rho_{DA} & 1 & \rho_{AF} \\ \rho_{DF} & \rho_{AF} & 1 \end{bmatrix}, \text{ and } B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \zeta \end{bmatrix} \quad (15)$$

8.2 Simulation of ANFIS equalizer for UWB channels

We will now consider the simulation setup for UWB channels based on the Channel Covariance Matrices (CCM). The following extended channel covariance matrix was used in the simulations:

$$A = \begin{bmatrix} 1 & 0.8 & -0.7 & 0.6 \\ 0.8 & 1 & -0.6 & 0.5 \\ -0.7 & -0.6 & 1 & 0.5 \\ 0.6 & 0.5 & 0.5 & 1 \end{bmatrix}, \text{ and } A' = \begin{bmatrix} 1 + \alpha & 0.8 + \beta & -0.7 + \gamma & 0.6 + \delta \\ 0.8 + \beta & 1 + \alpha & -0.6 + \varepsilon & 0.5 + \eta \\ -0.7 + \gamma & -0.6 + \varepsilon & 1 + \alpha & 0.5 + \zeta \\ 0.6 + \delta & 0.5 + \eta & 0.5 + \zeta & 1 + \alpha \end{bmatrix} \quad (16)$$

A' indicate the modified CCM corrupted by CCI and AWGN. We use an ANFIS with the following parameters in the equalizer:

- One-input One-output ANFIS.
- 20 Rules.
- Gaussian membership functions.
- Maximum spread in CCM parameters is 0.5 ([0.0:0.1:0.5]).

The structure of the ANFIS is given in Figure 14. The simulation results are given in Figure 15. The MATLAB code for the simulation is appended below.

% Program to model the wideband channel

% Using the covariance matrix.

% Last modified on 13-03-2011.

clc; clf; clear all;

tic;

covm=[1 .8 -.7 .6; .8 1 -.6 .5;

-.7 -.6 1 .5; .6 .5 .5 1];

for mc=1:4% for 4 MC simulations.

mxerr=[];

for spr=.1:.01:.5 %for .. values of spread.

x=covm(:);

y(1)=x(1)+randn+spr;%first element with spread..

y(2)=x(2)+randn+spr;%second element with spread..

y(3)=x(3)+randn+spr;%third element with spread..

y(4)=x(4)+randn+spr;%fourth element with spread..

y(5)=x(2); y(6)=x(1); y(11)=x(1); y(16)=x(1);

y(7)=x(7)+randn+spr;

y(8)=x(8)+randn+spr;

y(9)=x(3);

y(10)=x(10)+randn+spr;

y(12)=x(12)+randn+spr;

y(13)=x(4); y(14)=x(12); y(15)=x(12);

y=y(1:16);

trnData = [y' x];

numMFs = 20;% No of Membership functions..

```

mfType = 'gaussmf';
epoch_n = 20;
in_fismat = genfis1(trnData,numMFs,mfType);
out_fismat = anfis(trnData,in_fismat,20);
est_x=evalfis(y,out_fismat);
err=(x-est_x);
mxerr=[mxerr,max(err)];
end;
covm
covm_eqln=reshape(est_x, 4,4)
error_abs=reshape(err,4,4)
plot([.1:.01:.5],mxerr,'LineWidth',mc);hold on; grid on;
end;
xlabel('Spread of Covariance matrix elements');
ylabel('Maximum Error in Estimation ');
legend('MC=1','MC=2','MC=3','MC=4');
title('Simulation Results: Covariance Spread versus Error');
toc;

```

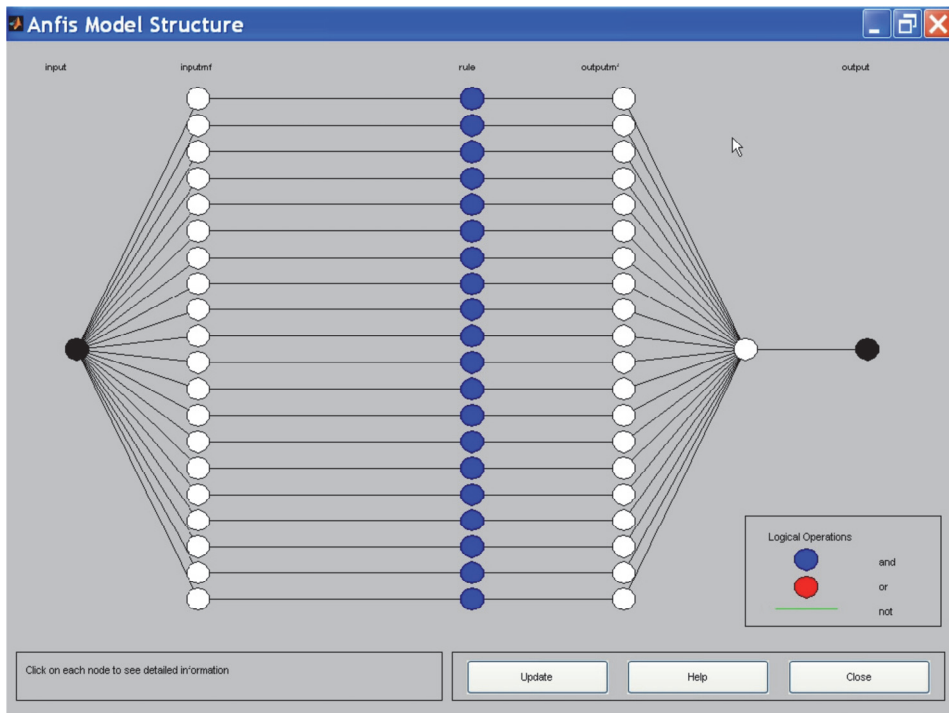


Fig. 14. ANFIS Model Structure Used for UWB Channel Equalization: Number of Inputs=1, Number of Outputs=1, Number of Rules=20, and Type of Membership Function- Gaussian.

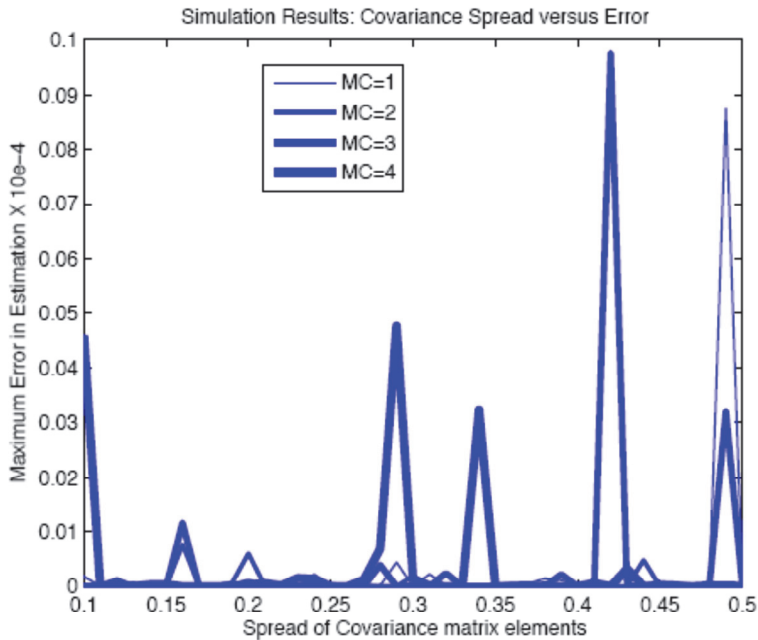


Fig. 15. Results of Simulation – ANFIS based Equalizer for UWB Channels.

It shows that the ANFIS model is capable of estimating the CCM parameters with almost negligible error.

9. Concluding remarks on ANFIS equalizers

In this chapter, we considered an alternative solution to the non-linear channel equalization problem. Several ANFIS equalizer structures are considered, with varying number of inputs and membership functions. It was found that the *BER versus SINR performances* of all of them were almost the same. However, at low values of SNR, ANFIS structure with more number of nodes (or more number of rules) performed slightly better. But as the number of nodes in the ANFIS structure was increased, convergence time was also increased. The number of nodes in the ANFIS structure is a function of the number of inputs, membership functions and outputs. The time for convergence increases, as the number of inputs or membership functions increases.

It is also shown that equalizers based on ANFIS can be suitably adapted for UWB channels as well. A channel co-variance matrix (CCM) formulation was used to model the UWB channel. It was shown that the estimate of the CCM was better when the *spread* in parameters was *small*.

10. Acknowledgements

The author is most thankful to Prof M Harisankar, Adjunct Professor, Department of IT, Indian Institute of Management, Kozhikode for his valuable guidance, mentoring and technical support while authoring the chapter.

11. References

- B.Widrow et.al. (1975). The Complex LMS Algorithm, *Proceedings of the IEEE*, Vol.63, pp.719–720, April.
- John G.Proakis and Masoud Salehi, (2002). *Communication Systems Engineering*, 2nd edition, ISBN 978-0130617934, Pearson Education.
- Y. Okumura, E. Ohmori, T. Kawano, and K. Fukua, (1968). Field Strength and its Variability in UHF and VHF Land-mobile Radio Service, *Rev. Elec. Commun. Lab.*, Vol. 16, No. 9.
- M. Hata, (1980). Empirical Formula for Propagation Loss in and Mobile Radio Services, *IEEE Transactions on Vehicular Technology*, Vol. 29, pp. 317–325, August.
- M.S. Smith, J.E.J. Dalley, (2000). A New Methodology for Deriving Pathloss Models from Cellular Drive Test Data, *Proceedings of AP2000 Conference*, Davos, Switzerland, April.
- Cheng Jian Lin and Wen Hao Ho, (2004). Blind Equalization Using Pseudo-Gaussian-Based Compensatory Neuro-Fuzzy Filters, *International Journal of Applied Science and Engineering*, Vol.2, No.1, pp.72–89, January.
- Li-Xin Wang and Jerry M. Mendel, (1993). An RLS Fuzzy Adaptive Filter, with Application to Nonlinear Channel Equalization, *IEEE Transactions on Fuzzy Systems*, Vol.1, pp.895–900, August.
- Qilian Liang and Jerry M. Mendel, (2000). Overcoming Time-Varying Co-Channel Interference Using Type-2 Fuzzy Adaptive Filters, *IEEE Transactions on Circuits and Systems-II:Analog and Digital Signal Processing*, Vol.47, No.12, pp.1419–1428, December.
- Jyh-Shing Roger Jang, (1993). ANFIS: Adaptive-Network-Based Fuzzy Inference System, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol.23, No.3, pp.665–685, May/June.
- Andreas F. Molisch, (2005a). Ultra-wideband Propagation Channels-Theory, Measurement, and Modeling, *IEEE Transactions on Vehicular Technology*, Vol. 54, No. 5, pp. 1528–1545, September.
- Andreas F. Molisch, (2005b). *Wireless Communications*, ISBN 81-265-1056-0, John Wiley.
- A. Algans, K. I. Pedersen, and P. E. Mogensen, (2002). Experimental Analysis of the Joint Statistical Properties of Azimuth Spread, Delay Spread, and Shadow Fading, *IEEE Journal on Selected Areas in Communication*, Vol. 20, No. 3, pp. 523–531, April.
- N. C. Beaulieu, A. A. Abu-Dayya, and P. J. McLane, (1995). Estimating the Distribution of a Sum of Independent Lognormal Random Variables, *IEEE Transactions on Communications*, Vol. 43, No. 12, pp. 2869–2873, December.

Adaptive Channel Estimation in Space-Time Coded MIMO Systems

Murilo B. Loiola¹ and Renato R. Lopes² and João M. T. Romano²

¹*Universidade Federal do ABC (UFABC)*

²*University of Campinas (UNICAMP)*

Brazil

1. Introduction

Space-time codes are an effective and practical way to exploit the benefits of spatial diversity in multiple-input, multiple-output (MIMO) systems. As the name suggests, the coding of space-time codes is performed in both the spatial and the temporal domains, introducing correlation between signals transmitted by different antennas at different time instants. At the receiver, this space-time correlation is exploited to mitigate the detrimental effects of fading and to improve the quality of received signal. The use of space-time codes allows the system to profit from transmit diversity without any power increment and with no channel knowledge at the transmitter.

Among the existing space-time coding schemes, orthogonal space-time block codes (OSTBCs) (Alamouti, 1998; Duman & Ghayeb, 2007; Larsson & Stoica, 2003; Tarokh et al., 1999) are of particular interest because they achieve full diversity at low receiver complexity. More specifically, the maximum likelihood (ML) receiver for OSTBCs consists of a linear receiver followed by a symbol-by-symbol decoder.

In order to correctly decode the received signals, the ML receiver for OSTBCs must have perfect channel knowledge. Unfortunately, this channel information is not normally available to the receivers; therefore channel estimation techniques are essential for the system to work properly. When the channel is static, methods such as those presented in (Vucetic & Yuan, 2003) and (Larsson et al., 2003) can be successfully used. However, the channel may be time-varying due to the mobility of the transmitter and/or receiver, to changes in the environment, or to carrier frequency mismatch between transmitter and receiver. In these cases, the estimation algorithm must be able to track the channel variations. One of the most widely known approaches to channel tracking is Kalman filtering (Enescu et al., 2007; Kailath et al., 2000; Komninakis et al., 2002; Piechocki et al., 2003; Simon, 2006). An important characteristic of the Kalman filter (KF) is its inherent ability to deal with nonstationary environments. In (Komninakis et al., 2002) for instance, a Kalman filter that uses the outputs of a minimum mean squared error (MMSE) decision-feedback equalizer (DFE) is developed to track Rice MIMO frequency-selective channels. Channel estimation using Kalman filters for MIMO-OFDM systems is studied in (Enescu et al., 2007; Piechocki et al., 2003). The use of Kalman filters to estimate channels in space-time block coded MIMO systems is also developed in the literature. In (Liu et al., 2002), a KF is used to estimate fast flat fading MIMO channels in Alamouti-based schemes. Therefore, it is limited to the case of two transmit

antennas. An extension of (Liu et al., 2002) for any type of OSTBCs is presented in (Balakumar et al., 2007). It is also shown in (Balakumar et al., 2007) that the KF can be significantly simplified due to the orthogonality of OSTBC codewords.

The algorithms developed in (Liu et al., 2002) and (Balakumar et al., 2007) assume that the channel coefficients are uncorrelated and independent from each other. However, spatial correlation between transmit and/or receive antennas usually exists in practical scenarios. This can occur, for instance, if the separation of adjacent antennas is not sufficient to allow the signals undergo different channel fades. Besides affecting the performance of OSTBCs, spatial correlations reduce the capacity gains of MIMO channels. Furthermore, the complexity reduction in (Balakumar et al., 2007) cannot be achieved if the channels were correlated. A generalization of (Balakumar et al., 2007) for correlated channels is presented in (Loiola et al., 2009) with a complexity similar to that in (Balakumar et al., 2007).

As with most Kalman channel estimators (KCEs), the KCE in (Loiola et al., 2009), is a time-varying filter whose coefficients need to be computed anew for each time instant. This computation increases the complexity of the filter, especially because it involves a matrix inversion. However, in many scenarios with constant modulus signal constellations the filter coefficients converge quickly. Taking advantage of this fact, the authors in (Loiola et al., 2009) derive a steady-state Kalman channel estimator (SS-KCE) (Simon, 2006), where the asymptotic value of the filter coefficients are computed at the initialization and used in a time-invariant filter. In spite of the significant complexity reduction, it is shown in (Loiola et al., 2009) that the SS-KCE suffers negligible performance degradation compared to the regular KCE, especially when channel variations are fast. However, the SS-KCE of (Loiola et al., 2009) depends on the solution of a Riccati equation. This chapter extends (Loiola et al., 2009) by providing an explicit expression for the SS-KCE and by proving that, under mild conditions, the SS-KCE is stable. It also proves that, at worst, the SS-KCE is marginally stable, but it is never unstable.

In this chapter, we also present a KCE with improved robustness to channel model mismatch. In fact, all the KCEs mentioned so far rely on an autoregressive model of the channel dynamics (Enescu et al., 2007; Komninakis et al., 2002; Li & Wong, 2007). This model is an approximation of Bello's wide-sense stationary uncorrelated scattering (WSSUS) model (Jakes, 1974), which in turn is an approximation of what actually governs the channel variations. In other words, the KCEs rely on a channel dynamics that may be far from true. To mitigate this problem, in this chapter we develop a fading-memory Kalman channel estimator (FM-KCE) (Simon, 2006) for estimating MIMO channels with OSTBC. This filter artificially increases the process noise in the state equations, so that the filter must rely more on the measurement than on the prediction step of the KCE. As a result, the filter is more robust to model mismatch.

An outline of the chapter is as follows. We begin with a basic channel model and some definitions. We then proceed to developed the KCEs mentioned so far, including the SS-KCE and the FM-KCE. We also prove that the SS-KCE is never unstable. Simulations results are presented in the sequel. Finally, the last section presents a summary of the chapter.

2. System model

We consider a MIMO system with N_T transmit antennas sending data blocks of length T to N_R receive antennas. The channel is assumed to be flat and constant during the transmission of each data block and can change between consecutive blocks. The relationship between transmitted and received signals for a data block k can be expressed as (Duman & Ghayeb,

2007; Larsson & Stoica, 2003)

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X}_k + \mathbf{N}_k, \quad (1)$$

where $\mathbf{Y}_k$ is an $N_R \times T$ matrix with the received signals, $\mathbf{X}_k$ is an $N_T \times T$ matrix containing the transmitted signals at block k , the $N_R \times T$ matrix $\mathbf{N}_k$ contains samples of independent, zero mean, circularly symmetric, white Gaussian noise with covariance σ_n^2 , and the frequency-flat channel is represented by the $N_R \times N_T$ matrix $\mathbf{H}_k$.

In space-time block coding, the matrix $\mathbf{X}_k$ represents a mapping that transforms a block of M complex symbols, $\mathbf{x}_k = [x_{k,1} \ x_{k,2} \ \cdots \ x_{k,M}]^T$, to an $N_T \times T$ complex matrix. The space-time codeword $\mathbf{X}_k$ is then used to transmit these M symbols in T time slots, achieving a rate of M/T . The matrix $\mathbf{X}_k$ is an OSTBC if (Duman & Ghayeb, 2007; Larsson & Stoica, 2003; Tarokh et al., 1999; Vucetic & Yuan, 2003): 1) all elements of $\mathbf{X}_k$ are linear functions of symbols of $\mathbf{x}_k$ and their complex conjugates and 2) for an arbitrary $\mathbf{x}_k$, the matrix $\mathbf{X}_k$ satisfies $\mathbf{X}_k \mathbf{X}_k^H = \|\mathbf{x}_k\|^2 \mathbf{I}_{N_T}$, where $\mathbf{I}_{N_T}$ is the identity matrix of order N_T , $\|\cdot\|$ represents the Euclidean norm and $(\cdot)^H$ denotes the conjugate transpose of a matrix. From the definition of OSTBC the space-time codeword $\mathbf{X}_k$ can be written as (Duman & Ghayeb, 2007; Larsson & Stoica, 2003)

$$\mathbf{X}_k = \sum_{m=1}^M \left(x_{k,m} \mathbf{A}_m + x_{k,m}^* \mathbf{B}_m \right), \quad (2)$$

where $x_{k,m}$, $m = 1, \dots, M$ is the m^{th} symbol of the data vector $\mathbf{x}_k$ and the $N_T \times T$ matrices $\mathbf{A}_m$ and $\mathbf{B}_m$ describe the code.

One of the main characteristics of OSTBCs is the simplicity of the decoder. More specifically, the ML space-time decoder for OSTBCs consists of a linear receiver followed by a symbol-by-symbol decoder (Duman & Ghayeb, 2007; Larsson & Stoica, 2003; Tarokh et al., 1999; Vucetic & Yuan, 2003). The linear receiver generates sufficient statistics through a linear combination of the received signals. Then, these sufficient statistics are passed to a symbol-by-symbol decision device. It is important to highlight that the computation of the sufficient statistics depends on the channel knowledge at the receiver.

As mentioned before, the channel is assumed to be fixed during the transmission of a space-time codeword $\mathbf{X}_k$, but can vary between consecutive codewords. According to the widely used wide-sense stationary uncorrelated scattering (WSSUS) model (Jakes, 1974), the channel coefficients are modeled as independent, zero-mean, complex Gaussian random variables with time autocorrelation function

$$\mathbb{E} \left[h_{k,i,j} h_{t,i,j}^* \right] \approx \mathcal{J}_0(2\pi f_D T_s |k - t|), \quad (3)$$

where $h_{k,i,j}$, $i = 1, \dots, N_R$, $j = 1, \dots, N_T$ is the (i, j) element of matrix $\mathbf{H}_k$, $\mathcal{J}_0$ is the zero-order Bessel function of the first kind, $f_D T_s$ is the normalized Doppler rate (assumed the same for all transmit-receive antenna pairs) and T_s is the symbol period.

Although exact modeling of channel dynamics by finite length autoregressive (AR) processes is impossible because the time autocorrelation function (3) is nonrational and its spectrum is bandlimited, we can approximate the time evolution of channel coefficients by low-order AR processes. This is possible because the first few correlation terms of (3), for small lags, capture most of the channel dynamics (Komninakis et al., 2002; Li & Wong, 2007). Therefore, following (Komninakis et al., 2002; Li & Wong, 2007), we herein approximate the MIMO channel variations by a first order AR process. Thus, defining $\text{vec}(\cdot)$ as the operator that

stacks the columns of a matrix on top of each other, the time evolution of the channel is given by

$$\mathbf{h}_k^{\text{ind}} = \beta \mathbf{h}_{k-1}^{\text{ind}} + \mathbf{w}_k, \quad (4)$$

where

$$\mathbf{h}_k^{\text{ind}} = \text{vec}(\mathbf{H}_k), \quad (5)$$

$\beta = \mathcal{J}_0(2\pi f_D T_s)$, $\mathbf{w}_k$ is a vector of length $N_R N_T$ containing independent samples of circularly symmetric, zero-mean, Gaussian excitation noise with covariance matrix $\mathbf{Q} = \sigma_w^2 \mathbf{I}_{N_R N_T}$, and $\sigma_w^2 = (1 - \beta^2)$. The superscript ind indicates that the channel coefficients are independent and spatially uncorrelated. In general, however, the channel coefficients present a spatial correlation that depends on the propagation environment as well as the polarization of the antenna elements and the spacing between them. Among the spatially correlated channel models, one of the most used splits the fading correlation into two independent components called receive correlation and transmit correlation, respectively. This amounts to modeling the channel as (Larsson & Stoica, 2003)

$$\mathbf{H}_k = \mathbf{R}_R^{1/2} \mathbf{H}_k^{\text{ind}} \left(\mathbf{R}_T^{1/2} \right)^T, \quad (6)$$

where $\mathbf{R}_R$ models the correlation between receive antennas, $\mathbf{R}_T$ models the correlation between transmit antennas, $(\cdot)^{1/2}$ stands for the Hermitian square root of a matrix (Gantmacher, 1959; Golub & Van Loan, 1996) and $\mathbf{H}_k^{\text{ind}}$ is a MIMO channel with independent, uncorrelated and unit variance Gaussian elements.

As the model (6) is calculated from the channel with uncorrelated coefficients, it would be natural to try to formulate the time evolution of spatially correlated channel in a way analog to (4). Therefore, in the next subsection we derive an AR model for time-varying, spatially correlated MIMO channels.

2.1 Time-varying spatially correlated state-space channel model

If we apply the $\text{vec}(\cdot)$ operator to (6), we obtain

$$\text{vec}(\mathbf{H}_k) = \mathbf{h}_k = \left(\mathbf{R}_T^{1/2} \otimes \mathbf{R}_R^{1/2} \right) \text{vec}(\mathbf{H}_k^{\text{ind}}) = \mathbf{G} \mathbf{h}_k^{\text{ind}}, \quad (7)$$

where $\mathbf{G} = \left(\mathbf{R}_T^{1/2} \otimes \mathbf{R}_R^{1/2} \right)$, $\otimes$ represents the Kronecker product and we use the fact that, for any matrices $\mathbf{D} = \mathbf{ABC}$ of compatible sizes, $\text{vec}(\mathbf{D}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B})$ (Golub & Van Loan, 1996).

Using the properties of Kronecker products (Golub & Van Loan, 1996; Horn & Johnson, 1991), we can define the matrix $\mathbf{R}_h$ as

$$\begin{aligned} \mathbf{R}_h &= \mathbf{G} \mathbf{G}^H = \left(\mathbf{R}_T^{1/2} \otimes \mathbf{R}_R^{1/2} \right) \left(\mathbf{R}_T^{1/2} \otimes \mathbf{R}_R^{1/2} \right)^H = \left(\mathbf{R}_T^{1/2} \otimes \mathbf{R}_R^{1/2} \right) \left(\mathbf{R}_T^{H/2} \otimes \mathbf{R}_R^{H/2} \right) \\ &= \left(\mathbf{R}_T^{1/2} \mathbf{R}_T^{H/2} \otimes \mathbf{R}_R^{1/2} \mathbf{R}_R^{H/2} \right) = (\mathbf{R}_T \otimes \mathbf{R}_R). \end{aligned} \quad (8)$$

Pre-multiplying (4) by $\mathbf{G}$ results in

$$\mathbf{G} \mathbf{h}_k^{\text{ind}} = \beta \mathbf{G} \mathbf{h}_{k-1}^{\text{ind}} + \mathbf{G} \mathbf{w}_k. \quad (9)$$

Comparing (9) to (7), we can describe the dynamics of spatially correlated MIMO channels as

$$\mathbf{h}_k = \beta \mathbf{h}_{k-1} + \mathbf{G} \mathbf{w}_k. \quad (10)$$

This model is similar to (4), but in (10) the excitation noise is correlated.

3. Adaptive channel estimation using Kalman filters

In order to formulate the problem of channel estimation as one of state estimation, we need two equations named process and measurement equations, respectively (Haykin, 2002; Kailath et al., 2000; Simon, 2006). The process equation describes the dynamic behavior of the state variables to be estimated, while the measurement equation presents the relationship between the state variables and the observed system output. As we focus on channel tracking, we can use (10) as the process equation and $\mathbf{h}_k$ as the state vector. The system output, in our case, is the channel output $\mathbf{Y}_k$ in (1). Thus, the measurement equation can be formed by stacking the columns of $\mathbf{Y}_k$, $\mathbf{H}_k$ and $\mathbf{N}_k$ in (1) on top of each other, resulting in

$$\mathbf{y}_k = \mathcal{X}_k \mathbf{h}_k + \mathbf{n}_k, \quad (11)$$

where $\mathcal{X}_k = \mathbf{X}_k^T \otimes \mathbf{I}_{N_R}$ and $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$ is the covariance matrix of the measurement noise $\mathbf{n}_k$. As mentioned in section 2, one of the main properties of any OSTBC is the orthogonality of the codewords. This characteristic is still valid for the transformed codeword $\mathcal{X}_k$, as shown in the following lemma:

Lemma 1 (Orthogonality of $\mathcal{X}_k$). *The matrix $\mathcal{X}_k$ satisfies*

$$\mathcal{X}_k^H \mathcal{X}_k = \|\mathbf{x}_k\|^2 \mathbf{I}_{N_R N_T}. \quad (12)$$

Proof. Given that $\mathbf{X}_k$ is an OSTBC and using the properties of Kronecker products (Golub & Van Loan, 1996; Horn & Johnson, 1991), we can write

$$\begin{aligned} \mathcal{X}_k^H \mathcal{X}_k &= (\mathbf{X}_k^T \otimes \mathbf{I}_{N_R})^H (\mathbf{X}_k^T \otimes \mathbf{I}_{N_R}) = (\mathbf{X}_k^* \otimes \mathbf{I}_{N_R}) (\mathbf{X}_k^T \otimes \mathbf{I}_{N_R}) = \mathbf{X}_k^* \mathbf{X}_k^T \otimes \mathbf{I}_{N_R} \\ &= (\mathbf{X}_k \mathbf{X}_k^H)^T \otimes \mathbf{I}_{N_R} = \|\mathbf{x}_k\|^2 \mathbf{I}_{N_T} \otimes \mathbf{I}_{N_R} = \|\mathbf{x}_k\|^2 \mathbf{I}_{N_R N_T} \end{aligned} \quad (13)$$

where $(\cdot)^*$ is the conjugate of a matrix. $\square$

The state-space formulation of the problem of estimation of flat, time-varying and spatially correlated MIMO channels is then given by (10) and (11). As both (10) and (11) are linear functions of the state vector $\mathbf{h}_k$ and the noises $\mathbf{w}_k$ and $\mathbf{n}_k$ are independent, white and Gaussian, the Kalman filter provides the optimal recursive estimates, in the MMSE sense, for the channel coefficients (Kailath et al., 2000; Simon, 2006). Hence, the classical KF for estimating the MIMO channel is given by

Prediction

$$\hat{\mathbf{h}}_{k|k-1} = \beta \hat{\mathbf{h}}_{k-1|k-1} \quad (14a)$$

$$\mathbf{P}_{k|k-1} = \beta^2 \mathbf{P}_{k-1|k-1} + \mathbf{G} \mathbf{Q} \mathbf{G}^H \quad (14b)$$

Filtering

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathcal{X}_k^H (\mathcal{X}_k \mathbf{P}_{k|k-1} \mathcal{X}_k^H + \mathbf{R}_n)^{-1} \quad (15a)$$

$$\boldsymbol{\alpha}_k = \mathbf{y}_k - \mathcal{X}_k \hat{\mathbf{h}}_{k|k-1} \quad (15b)$$

$$\hat{\mathbf{h}}_{k|k} = \hat{\mathbf{h}}_{k|k-1} + \mathbf{K}_k \boldsymbol{\alpha}_k \quad (15c)$$

$$\mathbf{P}_{k|k} = [\mathbf{I} - \mathbf{K}_k \mathcal{X}_k] \mathbf{P}_{k|k-1} \quad (15d)$$

In these equations, $\mathbf{P}_{k|k-1}$ is the prediction error covariance matrix and $\hat{\mathbf{h}}_{k|k-1}$ is the vector with estimated channel coefficients, at a data block k , computed from the observation of previous blocks.

As done in (Balakumar et al., 2007), the complexity of (14a)–(15d) can be reduced. To that end, as shown in (Loiola et al., 2009), we exploit the orthogonality of OSTBCs as well as the model characteristics. We begin by substituting (8) into (14b)

$$\mathbf{P}_{k|k-1} = \beta^2 \mathbf{P}_{k-1|k-1} + \mathbf{G} \mathbf{Q} \mathbf{G}^H = \beta^2 \mathbf{P}_{k-1|k-1} + \sigma_w^2 \mathbf{G} \mathbf{G}^H = \beta^2 \mathbf{P}_{k-1|k-1} + \sigma_w^2 \mathbf{R}_h, \quad (16)$$

where we use the fact that $\mathbf{Q} = \sigma_w^2 \mathbf{I}_{N_R N_T}$.

Remembering that $\mathbf{R}_n = \sigma_n^2 \mathbf{I}_{N_R N_T}$ and using the matrix inversion lemma and lemma 1, it is possible to rewrite (15a) as

$$\begin{aligned} \mathbf{K}_k &= \mathbf{P}_{k|k-1} \mathbf{x}_k^H \left[\mathbf{R}_n^{-1} - \mathbf{R}_n^{-1} \mathbf{x}_k \left(\mathbf{x}_k^H \mathbf{R}_n^{-1} \mathbf{x}_k + \mathbf{P}_{k|k-1}^{-1} \right)^{-1} \mathbf{x}_k^H \mathbf{R}_n^{-1} \right] \\ &= \mathbf{P}_{k|k-1} \mathbf{x}_k^H \left[\frac{1}{\sigma_n^2} \mathbf{I}_{N_R N_T} - \frac{1}{\sigma_n^4} \mathbf{x}_k \left(\frac{1}{\sigma_n^2} \mathbf{x}_k^H \mathbf{x}_k + \mathbf{P}_{k|k-1}^{-1} \right)^{-1} \mathbf{x}_k^H \right] \\ &= \mathbf{P}_{k|k-1} \left[\frac{1}{\sigma_n^2} \mathbf{x}_k^H - \frac{1}{\sigma_n^4} \mathbf{x}_k^H \mathbf{x}_k \left(\frac{\|\mathbf{x}_k\|^2}{\sigma_n^2} \mathbf{I}_{N_R N_T} + \mathbf{P}_{k|k-1}^{-1} \right)^{-1} \mathbf{x}_k^H \right] \\ &= \frac{1}{\sigma_n^2} \mathbf{P}_{k|k-1} \left[\mathbf{x}_k^H - \frac{\|\mathbf{x}_k\|^2}{\sigma_n^2} \left(\frac{\|\mathbf{x}_k\|^2}{\sigma_n^2} \mathbf{I}_{N_R N_T} + \mathbf{P}_{k|k-1}^{-1} \right)^{-1} \mathbf{x}_k^H \right] \\ &= \frac{1}{\sigma_n^2} \mathbf{P}_{k|k-1} \left[\mathbf{I}_{N_R N_T} - \frac{\|\mathbf{x}_k\|^2}{\sigma_n^2} \left(\frac{\|\mathbf{x}_k\|^2}{\sigma_n^2} \mathbf{I}_{N_R N_T} + \mathbf{P}_{k|k-1}^{-1} \right)^{-1} \right] \mathbf{x}_k^H. \end{aligned} \quad (17)$$

Employing the matrix inversion lemma once more, it is possible to write the inverse matrix of the last expression of (17) as

$$\begin{aligned} &\left(\frac{\|\mathbf{x}_k\|^2}{\sigma_n^2} \mathbf{I}_{N_R N_T} + \mathbf{I}_{N_R N_T} \mathbf{P}_{k|k-1}^{-1} \mathbf{I}_{N_R N_T} \right)^{-1} = \\ &\frac{\sigma_n^2}{\|\mathbf{x}_k\|^2} \mathbf{I}_{N_R N_T} - \frac{\sigma_n^2}{\|\mathbf{x}_k\|^2} \left(\frac{\sigma_n^2}{\|\mathbf{x}_k\|^2} \mathbf{I}_{N_R N_T} + \mathbf{P}_{k|k-1} \right)^{-1} \left(\frac{\sigma_n^2}{\|\mathbf{x}_k\|^2} \mathbf{I}_{N_R N_T} \right) = \\ &\frac{\sigma_n^2}{\|\mathbf{x}_k\|^2} \mathbf{I}_{N_R N_T} - \left(\frac{\sigma_n^2}{\|\mathbf{x}_k\|^2} \right)^2 \left(\frac{\sigma_n^2}{\|\mathbf{x}_k\|^2} \mathbf{I}_{N_R N_T} + \mathbf{P}_{k|k-1} \right)^{-1}. \end{aligned} \quad (18)$$

Substituting now (18) into (17), the Kalman gain adaptation equation is expressed as

$$\begin{aligned} \mathbf{K}_k &= \frac{1}{\sigma_n^2} \mathbf{P}_{k|k-1} \left[\frac{\sigma_n^2}{\|\mathbf{x}_k\|^2} \left(\frac{\sigma_n^2}{\|\mathbf{x}_k\|^2} \mathbf{I}_{N_R N_T} + \mathbf{P}_{k|k-1} \right)^{-1} \right] \mathbf{x}_k^H \\ &= \frac{1}{\|\mathbf{x}_k\|^2} \mathbf{P}_{k|k-1} \left(\frac{\sigma_n^2}{\|\mathbf{x}_k\|^2} \mathbf{I}_{N_R N_T} + \mathbf{P}_{k|k-1} \right)^{-1} \mathbf{x}_k^H = \frac{1}{\|\mathbf{x}_k\|^2} \mathbf{A}_k \mathbf{x}_k^H, \end{aligned} \quad (19)$$

where we define

$$\mathbf{A}_k = \mathbf{P}_{k|k-1} \left(\frac{\sigma_n^2}{\|\mathbf{x}_k\|^2} \mathbf{I}_{N_R N_T} + \mathbf{P}_{k|k-1} \right)^{-1}. \quad (20)$$

Using lemma 1, (14a), (15b), and (19) into (15c), we have

$$\begin{aligned}
 \hat{\mathbf{h}}_{k|k} &= \hat{\mathbf{h}}_{k|k-1} + \frac{1}{\|\mathbf{x}_k\|^2} \mathbf{A}_k \mathcal{X}_k^H (\mathbf{y}_k - \mathcal{X}_k \hat{\mathbf{h}}_{k|k-1}) \\
 &= \hat{\mathbf{h}}_{k|k-1} - \frac{1}{\|\mathbf{x}_k\|^2} \mathbf{A}_k \mathcal{X}_k^H \mathcal{X}_k \hat{\mathbf{h}}_{k|k-1} + \frac{1}{\|\mathbf{x}_k\|^2} \mathbf{A}_k \mathcal{X}_k^H \mathbf{y}_k \\
 &= \hat{\mathbf{h}}_{k|k-1} - \frac{1}{\|\mathbf{x}_k\|^2} \mathbf{A}_k \left(\|\mathbf{x}_k\|^2 \mathbf{I}_{N_R N_T} \right) \hat{\mathbf{h}}_{k|k-1} + \frac{1}{\|\mathbf{x}_k\|^2} \mathbf{A}_k \mathcal{X}_k^H \mathbf{y}_k \\
 &= (\mathbf{I}_{N_R N_T} - \mathbf{A}_k) \hat{\mathbf{h}}_{k|k-1} + \frac{1}{\|\mathbf{x}_k\|^2} \mathbf{A}_k \mathcal{X}_k^H \mathbf{y}_k = (\mathbf{I}_{N_R N_T} - \mathbf{A}_k) (\beta \hat{\mathbf{h}}_{k-1|k-1}) + \frac{1}{\|\mathbf{x}_k\|^2} \mathbf{A}_k \mathcal{X}_k^H \mathbf{y}_k \\
 &= \beta \mathbf{B}_k \hat{\mathbf{h}}_{k-1|k-1} + \frac{1}{\|\mathbf{x}_k\|^2} \mathbf{A}_k \mathcal{X}_k^H \mathbf{y}_k,
 \end{aligned} \tag{21}$$

where $\mathbf{B}_k$ is defined as

$$\mathbf{B}_k = \mathbf{I}_{N_R N_T} - \mathbf{A}_k. \tag{22}$$

Finally, using (19) into (15d) we obtain

$$\mathbf{P}_{k|k} = \left(\mathbf{I}_{N_R N_T} - \frac{1}{\|\mathbf{x}_k\|^2} \mathbf{A}_k \mathcal{X}_k^H \mathcal{X}_k \right) \mathbf{P}_{k|k-1} = (\mathbf{I}_{N_R N_T} - \mathbf{A}_k) \mathbf{P}_{k|k-1} = \mathbf{B}_k \mathbf{P}_{k|k-1}. \tag{23}$$

Putting together (16), (20), (21), (22) and (23), the reduced complexity Kalman channel estimator (KCE) for correlated MIMO-OSTBC systems is given by (Loiola et al., 2009)

$$\mathbf{P}_{k|k-1} = \beta^2 \mathbf{P}_{k-1|k-1} + \sigma_w^2 \mathbf{R}_h \tag{24a}$$

$$\mathbf{A}_k = \mathbf{P}_{k|k-1} \left(\frac{\sigma_n^2}{\|\mathbf{x}_k\|^2} \mathbf{I}_{N_R N_T} + \mathbf{P}_{k|k-1} \right)^{-1} \tag{24b}$$

$$\mathbf{B}_k = \mathbf{I}_{N_R N_T} - \mathbf{A}_k \tag{24c}$$

$$\hat{\mathbf{h}}_{k|k} = \beta \mathbf{B}_k \hat{\mathbf{h}}_{k-1|k-1} + \frac{1}{\|\mathbf{x}_k\|^2} \mathbf{A}_k \mathcal{X}_k^H \mathbf{y}_k \tag{24d}$$

$$\mathbf{P}_{k|k} = \mathbf{B}_k \mathbf{P}_{k|k-1} \tag{24e}$$

It is important to note that one of the key assumptions to the complexity reduction in (Balakumar et al., 2007) is the uncorrelated nature of the channel coefficients. In this case, and supposing that the initial value $\mathbf{P}_{0|0}$ is also a diagonal matrix, it is shown in (Balakumar et al., 2007) that $\mathbf{P}_{k|k-1}$ is always diagonal, which simplifies all subsequent calculations. However, for a general spatial correlation matrix $\mathbf{R}_h$, it is not possible to simplify the computation of the matrix inversion in (24b). For this reason, the approach taken in (Loiola et al., 2009) to reduce the complexity of KCE (24a)–(24e) is the development of a steady-state Kalman channel estimator, which is presented in section 4. It will be shown in section 4 that the steady-state Kalman channel estimator has a complexity order less than or equal to that of the algorithm in (Balakumar et al., 2007) and works also for non-diagonal spatial correlation matrices.

It is also worth observing that the channel estimates produced by the Kalman filter (24a)–(24e) correspond to weighted sums of instantaneous ML channel estimates. To see this, first

consider the instantaneous ML channel estimates, i.e., the estimates computed by using only the k^{th} data block, which is given by (Kaiser et al., 2005)

$$\hat{\mathbf{h}}_k^{(\text{ML})} = \left(\mathcal{X}_k^H \mathcal{X}_k \right)^{-1} \mathcal{X}_k^H \mathbf{y}_k. \quad (25)$$

For OSTBCs, thanks to lemma 1, (25) reduces to

$$\hat{\mathbf{h}}_k^{(\text{ML})} = \left(\|\mathbf{x}_k\|^2 \mathbf{I}_{N_R N_T} \right)^{-1} \mathcal{X}_k^H \mathbf{y}_k = \frac{1}{\|\mathbf{x}_k\|^2} \mathcal{X}_k^H \mathbf{y}_k. \quad (26)$$

Thus, using (26) the channel estimate (24d) can be rewritten as

$$\hat{\mathbf{h}}_{k|k} = \beta \mathbf{B}_k \hat{\mathbf{h}}_{k-1|k-1} + \mathbf{A}_k \hat{\mathbf{h}}_k^{(\text{ML})}. \quad (27)$$

Consequently, the KF proposed in (Loiola et al., 2009) updates the channel estimates through weighted sums of instantaneous maximum likelihood channel estimates. It is important to note that the weights are time-varying and optimally calculated, in the MMSE sense, for each data block.

Considering communication systems where pilot sequences are periodically inserted between information symbols, the algorithm in (24a)–(24e) can operate in both training and decision-directed (DD) modes. First, when pilot symbols are available, the matrix $\mathcal{X}_k$ in (24d) is constructed from them. Once the transmission of pilot symbols is finished, the algorithm enters in decision-directed mode and the matrix $\mathcal{X}_k$ is then formed by the decisions provided by the ML space-time decoder. Note that these decisions are based on the channel estimates generated by the algorithm in the previous iteration.

4. Steady-state Kalman channel estimator

The measurement equation (11) represents a time-varying system, since the matrix $\mathcal{X}_k$ changes at each transmitted data block. However, in the Kalman channel estimator (24a)–(24e), only (24d) has an explicit dependence on $\mathcal{X}_k$. Because of the orthogonality of OSTBC codewords, all other expressions in this recursive estimator depend only on the energy of the uncoded data block, i.e. $\|\mathbf{x}_k\|^2$. Now, for constant modulus signal constellations such as M-PSK, $\|\mathbf{x}_k\|^2$ is a constant. In this case, (24a)–(24c) and (24e) are just functions of the initial estimate of $\mathbf{P}_{k|k}$, the normalized Doppler rate, the spatial correlation matrix, a constant equal to the energy of the constellation symbols and the variance of the measurement noise.

These parameters can be estimated ahead of time using, for example, the methods proposed in (Jamoos et al., 2007) and in the references therein. Thus, we assume that the parameters in (24a)–(24c) and (24e) are known. Furthermore, we can analyze the state-space model (10) and (11) to check if the matrices $\mathbf{P}_{k|k}$, $\mathbf{A}_k$ and $\mathbf{B}_k$ converge to steady-state values. If this is the case, and if these values can be found, the time-varying matrices could be replaced by constant matrices, originating a low complexity sub-optimal estimator known as the steady-state Kalman channel estimator (SS-KCE) (Loiola et al., 2009). As pointed out in (Simon, 2006), the steady-state filter often performs nearly as well as the optimal time-varying filter.

To determine the SS-KCE, we begin by substituting (24e) into (24a), which yields

$$\mathbf{P}_{k|k-1} = \beta^2 \mathbf{B}_{k-1} \mathbf{P}_{k-1|k-2} + \sigma_w^2 \mathbf{R}_h. \quad (28)$$

Now substitute (24c) into (28) to obtain

$$\mathbf{P}_{k|k-1} = \beta^2 (\mathbf{I}_{N_R N_T} - \mathbf{A}_{k-1}) \mathbf{P}_{k-1|k-2} + \sigma_w^2 \mathbf{R}_h. \quad (29)$$

Taking into account (24b), we can rewrite (29) as

$$\mathbf{P}_{k|k-1} = \beta^2 \mathbf{P}_{k-1|k-2} - \beta^2 \mathbf{P}_{k-1|k-2} \left(\frac{\sigma_n^2}{n_s} \mathbf{I}_{N_R N_T} + \mathbf{P}_{k-1|k-2} \right)^{-1} \mathbf{P}_{k-1|k-2} + \sigma_w^2 \mathbf{R}_h, \quad (30)$$

where $n_s = \|\mathbf{x}\|^2$ corresponds to the energy of each uncoded data block $\mathbf{x}_k$, assumed to be a constant.

If $\mathbf{P}_{k|k-1}$ converges to a steady-state value, then $\mathbf{P}_{k|k-1} = \mathbf{P}_{k-1|k-2}$ for large k . Denoting this steady-state value as $\mathbf{P}_\infty$, we rewrite (30) as

$$\mathbf{P}_\infty = \beta^2 \mathbf{P}_\infty - \beta^2 \mathbf{P}_\infty \left(\mathbf{P}_\infty + \frac{\sigma_n^2}{n_s} \mathbf{I}_{N_R N_T} \right)^{-1} \mathbf{P}_\infty + \sigma_w^2 \mathbf{R}_h. \quad (31)$$

Equation (31) is a discrete algebraic Riccati equation (DARE) (Kailath et al., 2000; Simon, 2006). If it can be solved, we can use $\mathbf{P}_\infty$ in (24b) and (24c) to calculate the steady-state values of matrices $\mathbf{A}$ and $\mathbf{B}$, denoted $\mathbf{A}_\infty$ and $\mathbf{B}_\infty$, respectively. Hence, the steady-state Kalman channel estimator proposed in (Loiola et al., 2009) is given simply by

$$\hat{\mathbf{h}}_{k|k} = \beta \mathbf{B}_\infty \hat{\mathbf{h}}_{k-1|k-1} + \frac{1}{n_s} \mathbf{A}_\infty \mathcal{X}_k^H \mathbf{y}_k. \quad (32)$$

As in (27), the steady-state KF generates channel estimates by averaging instantaneous ML channel estimates. However, as opposed to (27), the weights in (32) are not time-varying. The problem now is to determine the solution of (31). As the DARE is highly nonlinear, its solutions $\mathbf{P}_\infty$ may or may not exist, they may or may not be unique or indeed they may or may not generate a stable steady-state filter. In the next subsection, we present the solution to (31), and discuss the stability of the resulting filter (32).

4.1 Existence of DARE solutions

To show one possible solution of the DARE in (31), let $\mathbf{R}_h = \mathbf{Q}_U^H \mathbf{\Lambda} \mathbf{Q}_U$ be the eigendecomposition of $\mathbf{R}_h$. Since $\mathbf{Q}_U$ is unitary, it is easy to verify that $\mathbf{P}_\infty = \mathbf{Q}_U^H \mathbf{\Sigma} \mathbf{Q}_U$ is a solution of the DARE, as long as the diagonal matrix $\mathbf{\Sigma}$ satisfies

$$\mathbf{\Sigma} = \beta^2 \mathbf{\Sigma} - \beta^2 \mathbf{\Sigma} \left(\mathbf{\Sigma} + \frac{\sigma_n^2}{n_s} \mathbf{I}_{N_R N_T} \right)^{-1} \mathbf{\Sigma} + \sigma_w^2 \mathbf{\Lambda}. \quad (33)$$

Now let σ_i and λ_i be the i -th diagonal element of $\mathbf{\Sigma}$ and $\mathbf{\Lambda}$, respectively. Then, since all the matrices in (33) are diagonal, σ_i must satisfy

$$\sigma_i^2 + b\sigma_i + c = 0, \quad (34)$$

where $b = \sigma_n^2(1 - \beta^2)/n_s - \sigma_w^2 \lambda_i$ and $c = -\sigma_n^2 \sigma_w^2 \lambda_i / n_s$.

Equation (34) has two possible solutions. We now show that only one of these solutions is valid, in the sense that the resulting $\mathbf{P}_\infty$ is a valid autocorrelation matrix. To that end, we need to show that the eigenvalues of $\mathbf{P}_\infty$ are real and non-negative. We begin by noting that $\mathbf{R}_h$ is a correlation matrix, so $\lambda_i \geq 0$. As the remaining terms of c also are positive, we conclude

that $c \leq 0$. Thus, the discriminant of (34), given by $b^2 - 4c$, is non-negative. We identify two possibilities. First, the discriminant is zero if and only if $b = c = 0$. This happens if and only if there is no mobility, in which case $\beta = 1$ and $\sigma_w^2 = 0$. In this case, $\sigma_i = 0$, so $\mathbf{P}_\infty$ does not have full rank. On the other hand, if there is mobility, the discriminant of (34) is strictly positive. In this case, the quadratic equation in (34) has two distinct real solutions. Furthermore, since $c \leq 0$, we have that $b^2 - 4c \geq b^2$, so the solution given by $(-b + \sqrt{b^2 - 4c})/2$ is non-negative, which concludes the proof.

We also need to prove that the SS-KCE in (32) is stable. To that end, note that stability holds as long as the eigenvalues of $\mathbf{I} - \mathbf{A}_\infty$ have magnitude less than one. Now, using the fact that $\mathbf{P}_\infty = \mathbf{Q}_U^H \Sigma \mathbf{Q}_U$, it is easy to verify that the eigenvalues of $\mathbf{I} - \mathbf{A}_\infty$, ρ_i , are given by

$$\rho_i = \frac{\sigma_n^2/n_s}{\sigma_n^2/n_s + \sigma_i^2}. \quad (35)$$

Note that $\sigma_i \geq 0$, so that $0 < \rho_i \leq 1$. Also, note that $\rho_i = 1$ if and only if $\sigma_i = 0$, which happens if and only if $\lambda_i = 0$, i.e., when the spatial correlation matrix $\mathbf{R}_h$ does not have full rank. In this case, the SS-KCE is marginally stable. In all other cases, the filter is stable.

Finally, we note that the SS-KCE does not work very well in low mobility. In fact, we will show that, as $\beta \rightarrow 1$, the SS-KCE in (32) tends to $\hat{\mathbf{h}}_{k|k} = \hat{\mathbf{h}}_{k-1|k-1}$. In other words, as $\beta \rightarrow 1$, the SS-KCE does not update the channel estimate, simply keeping the initial guess for all iterations while ignoring the channel output. This makes intuitive sense. Indeed, as $\beta \rightarrow 1$, the state equation (10) tends to $\mathbf{h}_k = \mathbf{h}_{k-1}$, i.e., the channel becomes static. In this case, as we have more and more observations, the variance of the estimation error in the Kalman filter tends to zero. Thus, in steady-state, the filter stops updating the channel estimates. To prove this result in our case, we note that, as $\beta \rightarrow 1$, $\sigma_w^2 \rightarrow 0$, so the solution of (34) tends to $\sigma_i = 0$. Using again the fact that $\mathbf{P}_\infty = \mathbf{Q}_U^H \Sigma \mathbf{Q}_U$, we see that the eigenvalues of $\mathbf{A}_\infty$ are given by $\sigma_i / (\sigma_n^2/n_s + \sigma_i^2)$. Thus, as $\beta \rightarrow 1$, these eigenvalues tend to zero, so that $\mathbf{A}_\infty \rightarrow \mathbf{0}$, and the result follows.

5. Fading-memory Kalman channel estimator

As mentioned in Section 2, the first order AR model used in (10) is only an approximate description of the time evolution of channel coefficients. This modeling error can degrade the performance of Kalman-based channel estimators. One possible solution to mitigate this performance degradation in the KCE is to give more emphasis to the most recent received data, thus increasing the importance of the observations and decreasing the importance of the process equation (Anderson & Moore, 1979; Simon, 2006). To understand how this can be done, we consider the state-space model (10) and (11). For this model, it is possible to show (Anderson & Moore, 1979; Simon, 2006) that the sequence of estimates produced by the KCE minimizes $E[J_N]$, where the cost function J_N is given by

$$J_N = \sum_{k=1}^N \left[\left(\mathbf{y}_k - \mathcal{X}_k \hat{\mathbf{h}}_{k|k-1} \right)^H \mathbf{R}_n^{-1} \left(\mathbf{y}_k - \mathcal{X}_k \hat{\mathbf{h}}_{k|k-1} \right) + \mathbf{w}_k^H \left(\sigma_w^2 \mathbf{R}_h \right)^{-1} \mathbf{w}_k \right]. \quad (36)$$

The importance of the most recent observations can be increased if they receive a higher weight than past data. This can be accomplished with an exponential weight, controlled by a scalar $\alpha \geq 1$. In this case, the cost function can be rewritten as (Anderson & Moore, 1979;

Simon, 2006)

$$\tilde{J}_N = \sum_{k=1}^N \left[\left(\mathbf{y}_k - \mathcal{X}_k \hat{\mathbf{h}}_{k|k-1} \right)^H \alpha^{2k} \mathbf{R}_n^{-1} \left(\mathbf{y}_k - \mathcal{X}_k \hat{\mathbf{h}}_{k|k-1} \right) + \mathbf{w}_k^H \alpha^{2k+2} \left(\sigma_w^2 \mathbf{R}_h \right)^{-1} \mathbf{w}_k \right]. \quad (37)$$

Following (Anderson & Moore, 1979; Simon, 2006), it is possible to show that the minimization of $E[\tilde{J}_N]$ for OSTBC systems leads to the fading-memory Kalman channel estimator (FM-KCE), given by

$$\mathbf{P}_{k|k-1} = (\alpha\beta)^2 \mathbf{P}_{k-1|k-1} + \sigma_w^2 \mathbf{R}_h \quad (38a)$$

$$\mathbf{A}_k = \mathbf{P}_{k|k-1} \left(\frac{\sigma_n^2}{\|\mathbf{x}_k\|^2} \mathbf{I}_{N_R N_T} + \mathbf{P}_{k|k-1} \right)^{-1} \quad (38b)$$

$$\hat{\mathbf{h}}_{k|k} = \beta (\mathbf{I}_{N_R N_T} - \mathbf{A}_k) \hat{\mathbf{h}}_{k-1|k-1} + \mathbf{A}_k \frac{\mathcal{X}_k^H \mathbf{y}_k}{\|\mathbf{s}_k\|^2} \quad (38c)$$

$$\mathbf{P}_{k|k} = (\mathbf{I}_{N_R N_T} - \mathbf{A}_k) \mathbf{P}_{k|k-1} \quad (38d)$$

The only difference between the KCE and the FM-KCE is the existence of the scalar α^2 in the update equation of prediction error covariance matrix of the FM-KCE in (38a). This increases the variance of the prediction error, to which the filter responds by giving less importance to the system equation. The same could also be accomplished by using a system equation with a noise term of increased variance. It is worth noting that when $\alpha = 1$, the FM-KCE reduces to the KCE. On the other hand, when $\alpha \rightarrow \infty$, the channel estimates provided by the FM-KCE are solely based on the received signals and the system model is not taken into account. As an aside, we note that the FM-KCE can be interpreted as a result of adding a fictitious process noise (Anderson & Moore, 1979; Simon, 2006), which in consequence reduces the confidence of the KCE in the system model and increases the importance of observed data. To see that this fictitious process noise addition is mathematically equivalent to the FM-KCE, we rewrite (38a) as

$$\begin{aligned} \mathbf{P}_{k|k-1} &= (\alpha\beta)^2 \mathbf{P}_{k-1|k-1} + \sigma_w^2 \mathbf{R}_h = (\alpha^2 - 1 + 1) \beta^2 \mathbf{P}_{k-1|k-1} + \sigma_w^2 \mathbf{R}_h \\ &= \beta^2 \mathbf{P}_{k-1|k-1} + \sigma_w^2 \mathbf{R}_h + (\alpha^2 - 1) \beta^2 \mathbf{P}_{k-1|k-1} = \beta^2 \mathbf{P}_{k-1|k-1} + \tilde{\mathbf{Q}}, \end{aligned} \quad (39)$$

where

$$\tilde{\mathbf{Q}} = \sigma_w^2 \mathbf{R}_h + (\alpha^2 - 1) \beta^2 \mathbf{P}_{k-1|k-1} \quad (40)$$

and $(\alpha^2 - 1) \beta^2 \mathbf{P}_{k-1|k-1}$ corresponds to the covariance matrix of the fictitious process noise. Due to the similarity between the KCE (24a)–(24d) and the FM-KCE (38a)–(38d), one could think that the FM-KCE should also have a steady-state version. Following the same steps described in Section 4 to the derivation of (31), it is not hard to show that the Riccati equation for the FM-KCE is given by

$$\mathbf{P}_\infty = (\alpha\beta)^2 \mathbf{P}_\infty - (\alpha\beta)^2 \mathbf{P}_\infty \left(\mathbf{P}_\infty + \frac{\sigma_n^2}{n_s} \mathbf{I}_{N_R N_T} \right)^{-1} \mathbf{P}_\infty + \sigma_w^2 \mathbf{R}_h. \quad (41)$$

Its solution is also of the form $\mathbf{P}_\infty = \mathbf{Q}_U^H \mathbf{\Sigma} \mathbf{Q}_U$. The elements of the diagonal matrix $\mathbf{\Sigma}$ are given by $\sigma_i = -b + \sqrt{b^2 - 4c}$, where $b = \sigma_n^2(1 - \alpha^2\beta^2)/n_s - \sigma_w^2\lambda_i$ and $c = -\sigma_n^2\sigma_w^2\lambda_i/n_s$. Since $c \geq 0$,

we conclude that $\sigma_i \geq 0$, so the solution leads to a valid autocorrelation matrix, as before. Also, as before, we see that the steady-state filter is stable as long as $\sigma_i > 0$. Now, $\sigma_i = 0$ if and only if $c = 0$, which happens if $\lambda_i = 0$, i.e., $\mathbf{R}_h$ does not have full rank, or if $\sigma_w^2 = 0$, i.e., if there is no mobility. In either of these cases, the steady-state filter is marginally stable. Otherwise, the filter is stable.

Finally, we note that the DARE (41) could also be derived from the process equation

$$\mathbf{h}_k = \alpha \beta \mathbf{h}_{k-1} + \mathbf{G} \mathbf{w}_k. \quad (42)$$

Comparing (10) to (42), we see that the state transition matrix in (42) is modified by the scalar $\alpha \geq 1$, while the variance of the process noise remains the same. As shown in (Simon, 2006), this could be interpreted as an artificial increase in the process noise variance and hence equivalent to that done in (40).

6. Simulation results

In this section, we present some simulation results to illustrate the performance of the presented channel estimation algorithms. In all simulations the correlated channels are generated by (7), where the elements of $\mathbf{h}_k^{\text{ind}}$ are Rayleigh distributed with time autocorrelation function given by (3). It is worth emphasizing that the estimators presented in this chapter approximate the channel dynamics by the first order AR model (10). The receiver operates in decision-directed mode, i.e. after a certain number of space-time training codewords, the channel estimators employ the decisions provided by the ML space-time decoder. Unless stated otherwise, we insert 25 OSTBC training codewords between every 225 OSTBC data codewords.

Supposing that the spatial correlation coefficient between any two adjacent receive (transmit) antennas is given by p_r (p_t), it is possible to express each (i, j) element of the spatial correlation matrices R_R and R_T as $p_r^{|i-j|}$, $i, j = 1, \dots, N_R$ and $p_t^{|i-j|}$, $i, j = 1, \dots, N_T$, respectively. We assume that the receiver has perfect knowledge of the variances of process and measurement noises, the spatial correlation matrix and the normalized Doppler rate $f_D T_s$. The simulation results presented in the sequel correspond to averages of 10 channel realization, in each of which we simulate the transmission of 1×10^6 orthogonal space-time codewords. For comparison purposes, we also simulate a channel estimator implemented by the well known RLS adaptive filter (Haykin, 2002), with a forgetting factor of 0.98. This value was determined by trial and error to yield the best performance of the RLS.

To verify if there is any performance degradation of the SS-KCE (32) compared to the KCE (24a)–(24e), we simulate the transmission of 8-PSK symbols from $N_T = 2$ transmit antennas to $N_R = 2$ receive antennas using the Alamouti space-time block code (Alamouti, 1998). We also assume $p_t = 0.4$, $p_r = 0$ and different normalized Doppler rates. Fig. 1 shows the estimation mean squared error (MSE) for KCE and SS-KCE as a function of $f_D T_s$. We observe that the smaller the value of $f_D T_s$ (i.e. the smaller the relative velocity between transmitter and receiver), the greater the gap between KCE and SS-KCE. In the limit when $f_D T_s = 0$, the channel is time-invariant, the solution of (31) is null and the SS-KCE does not update the channel estimates. On the other hand, for channels varying at typical rates, both algorithms have equivalent performances. This can be seen in Fig. 2, which presents the symbol error rates at the output of ML space-time decoders fed with channel state information (CSI) provided by KCE and SS-KCE, as well as at the output of an ML decoder with perfect channel knowledge. Clearly, SS-KCE has the same performance of the KCE for the two values of $f_D T_s$ considered while demanding just a fraction of the complexity.

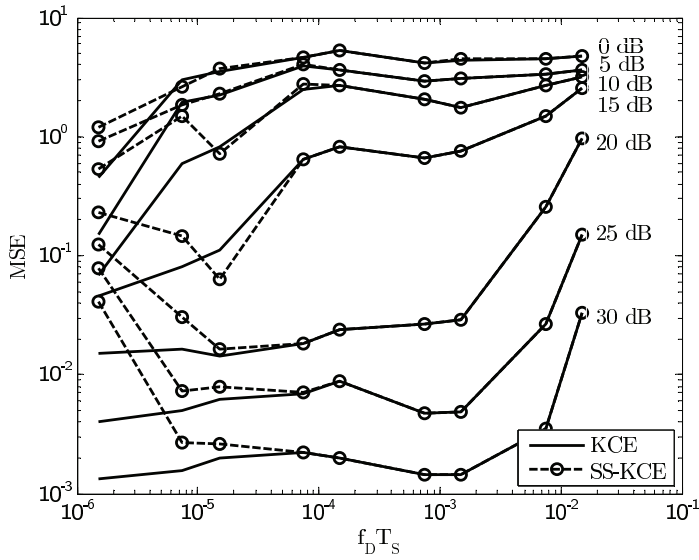


Fig. 1. Estimation mean squared error for KCE and SS-KCE.

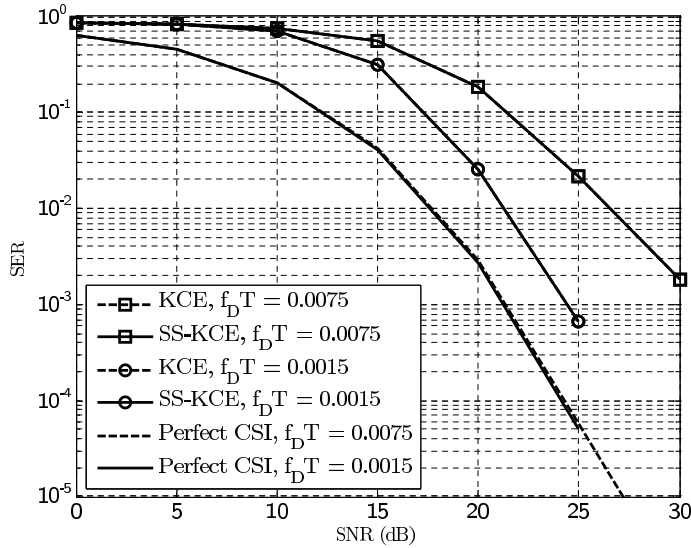


Fig. 2. Symbol error rates of ML decoders fed with channel estimates provided by KCE and SS-KCE.

We can explain the performance equivalence of KCE and SS-KCE by the fast convergence of the matrix $\mathbf{P}_{k|k-1}$ to its steady-state value. This means that the SS-KCE uses the optimal

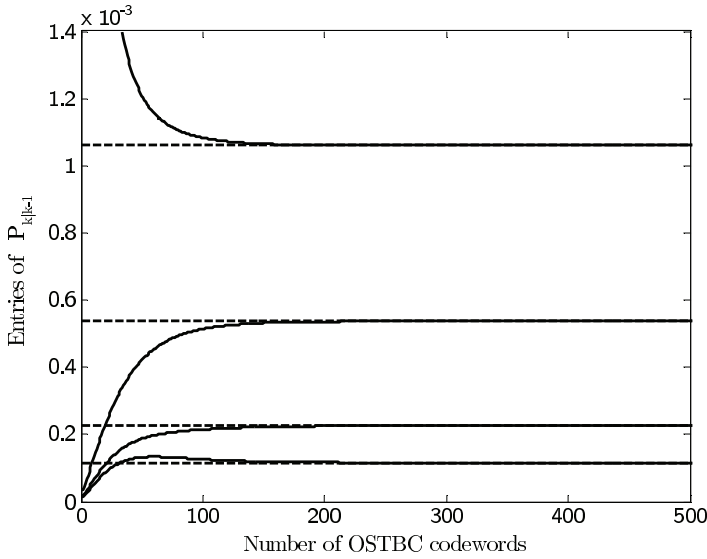


Fig. 3. Evolution of the entries of $\mathbf{P}_{k|k-1}$.

values of $\mathbf{A}_k$ and $\mathbf{B}_k$ after just a few blocks. Consequently, after these few blocks, the estimates provided by the SS-KCE are the same as those generated by the optimal KCE. To exemplify the fast convergence of $\mathbf{P}_{k|k-1}$, Fig. 3 shows the evolution of the values of the elements of $\mathbf{P}_{k|k-1}$ for an 8-PSK, Alamouti coded system with $N_R = N_T = 2$, $f_D T_s = 0.0015$, $p_r = 0.4$, $p_t = 0.8$, SNR = 15 dB and with the initial condition $\mathbf{P}_{0|0} = \mathbf{I}_{N_R N_T}$. It is clear from this figure that the elements of the matrix $\mathbf{P}_{k|k-1}$ reach their steady-state values before the transmission of 200 blocks. As the simulated system inserts 25 training blocks between 225 data blocks, we see that $\mathbf{P}_{k|k-1}$ converges even before the second training period. Due to the similar performances of KCE and SS-KCE, we hereinafter present just SS-KCE results.

It is important to observe that the gap in the symbol error rate curves of Fig. 2, between the decoders with perfect CSI and with estimated CSI, is due in great part to the use of the first order AR approximation to the channel dynamics. To show this, in Fig. 4 we present the symbol error rates at the output of decoders with perfect CSI and with SS-KCE estimates for the same scenario used in Fig 2, except that in Fig. 4 the channel is also generated by a first order AR process. As we can see, for $f_D T_s = 0.0015$, the receiver composed by SS-KCE and the space-time decoder has the same performance as the ML decoder with perfect CSI. For $f_D T_s = 0.0075$ and an SER of 10^{-3} , the receiver using SS-KCE is about 5 dB from the decoder with perfect CSI. This value is half of that shown in Fig. 2.

To analyze the impact of spatial channel correlation in the performance of the channel estimation algorithms, the next scenario simulates the transmission of QPSK symbols to 2 receive antennas using Alamouti's code for a normalized Doppler rate of 0.0045. The receiver correlation coefficient p_r is set to zero while the transmitter correlation coefficient p_t assumes values of 0.2 and 0.8. Fig. 5 presents the channel estimation MSE for SS-KCE and RLS algorithms for both p_t considered. From this figure, we note that the performances of the estimation algorithms are hardly affected by transmitter spatial correlation and that the

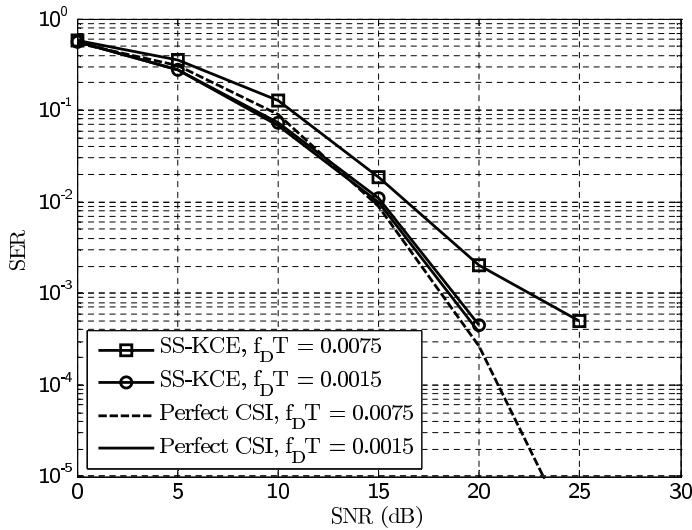


Fig. 4. Symbol error rates of ML space-time decoders for a first order AR channel.

curves for RLS are indistinguishable. It is also clear that the SS-KCE performs much better than the classical RLS algorithm. The symbol error rates at the output of ML decoders using the channel estimates provided by SS-KCE and RLS filters are shown in Fig. 6. Since the simulated RLS adaptive filter is not able to track the channel variations, the decoder can not correctly decode the space-time codewords, leading to a poor receiver performance. On the other hand, the receiver fed with SS-KCE estimates is 3 dB from the decoder with perfect CSI for both values of p_t at an SER of 10^{-4} .

In the previous simulations, the channel estimators tracked simultaneously the 4 possible channels between 2 transmit and 2 receive antennas. If the number of antennas increases, the number of channels to be tracked simultaneously also increases. To illustrate the capacity of the KF-based algorithms to track a larger number of channels, we simulate a system sending QPSK symbols from $N_T = 4$ transmit to $N_R = 4$ receive antennas. We employ the 1/2-rate OSTBC of (Tarokh et al., 1999) and assume $p_t = 0.8$ and $p_r = 0.4$. The MSE for the RLS and the SS-KCE is shown in Fig. 7. We observe that the estimates produced by the RLS algorithm are affected by the rate of channel variation. Moreover, the RLS MSE flattens out for SNR's greater than 10 dB. On the other hand, for this scenario, the SS-KCE has the same performance for both values of $f_D T_s$ considered and the MSE presents a linear decrease with the SNR. The similar performances of SS-KCE for $f_D T_s = 0.0015$ and $f_D T_s = 0.0045$ are also reflected in the symbol error rates at the output of the ML decoders, as shown in Fig. 8. For an SER of 10^{-3} , the decoders using the channels estimates provided by the SS-KCE are about 1 dB from the curves of the ML decoders with perfect CSI. For an SER of 10^{-3} and $f_D T_s = 0.0015$ the decoder fed with RLS channels estimates is approximately 4 dB from the optimal decoder, while for $f_D T_s = 0.0045$ the RLS-based decoder presents an SER no smaller than 10^{-1} in the simulated SNR range.

To cope with the modeling error introduced by the use of the first-order AR channel model, we show the FM-KCE in Section 5. Hence, to illustrate the performance improvement of FM-KCE

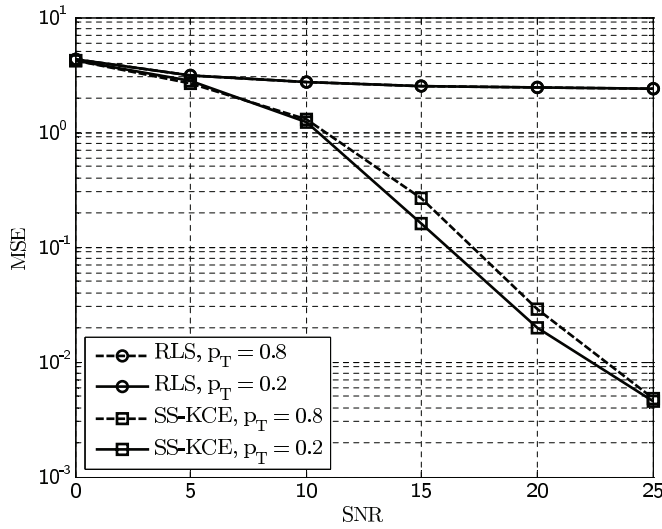


Fig. 5. Estimation mean square error for different transmitter correlation coefficient.

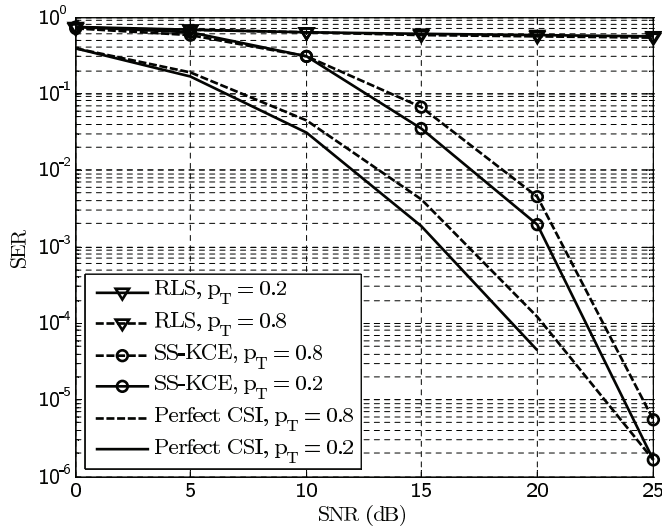


Fig. 6. Symbol error rate for different transmitter correlation coefficient.

in comparison to the SS-KCE, we simulate a MIMO system with 2 transmit antennas sending Alamouti-coded QPSK symbols to 2 receive antennas. The normalized Doppler rate is set to 0.0015, the receiver correlation coefficient p_r is set to zero while the transmitter correlation coefficient assumes the value $p_t = 0.4$. We vary the number of training codewords from 4

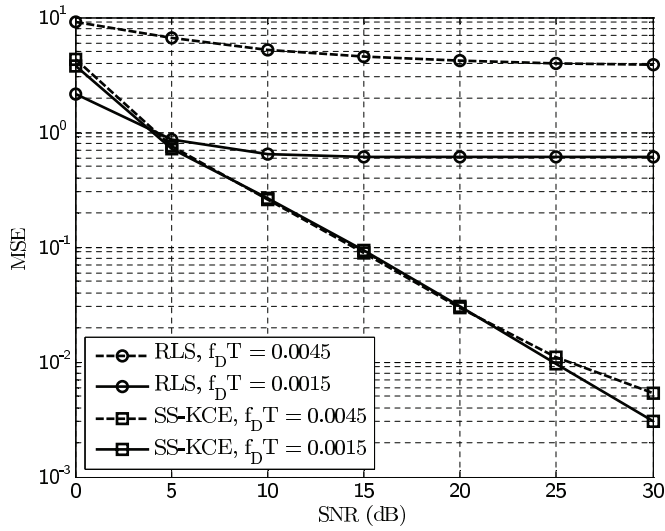


Fig. 7. Estimation mean square error for different values of $f_D T$.

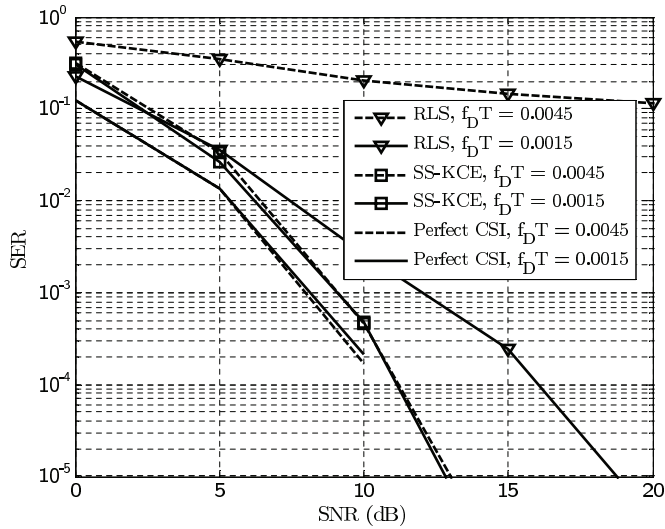


Fig. 8. Symbol error rate for different values of $f_D T$.

to 32 while maintaining the total number of blocks (training + data) fixed to 160 codewords. Also, we assume the weight of the FM-KCE $\alpha = 1.1$.

In Fig. 9 we present the estimation MSE for SS-KCE and for the steady-state version of FM-KCE, computed from the solution of the Riccati equation (41), with 4, 8, 12, 16, 20, 24, 28 and 32 training codewords. The arrows in this figure indicate the number of training

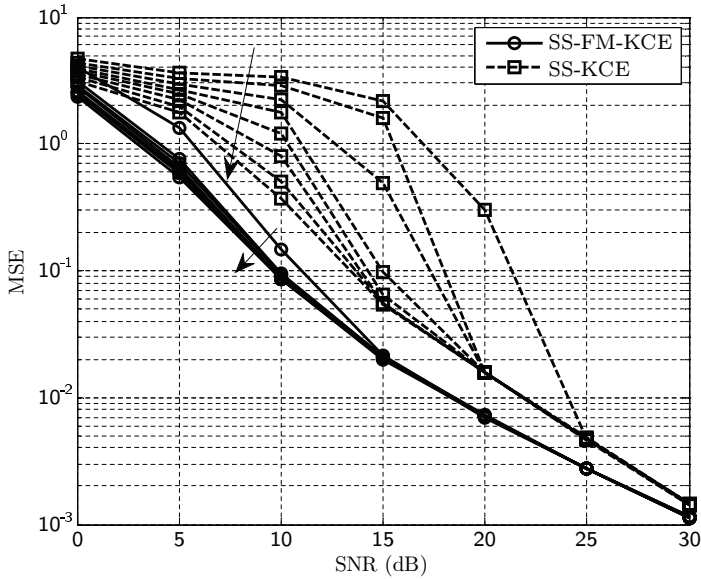


Fig. 9. Estimation mean square error for SS-KCE and FM-KCE.

codewords in ascending order. From Fig. 9, it is evident the superiority of FM-KCE over SS-KCE. Differently from SS-KCE, whose performance improves with the increase in the number of training codewords, the FM-KCE presents similar performances for the whole range of training codewords considered. For instance, for an MSE of 10^{-2} the FM-KCE performs 5 dB better than the SS-KCE with 4 training codewords and about 3.5 dB better than the SS-KCE with 32 training codewords.

The superior performance of the FM-KCE can also be observed in Fig. 10, which shows the SER at the output of ML decoders fed with CSI provided by SS-KCE and FM-KCE, as well as with perfect channel knowledge, for different training sequence lengths. For an SER of 10^{-3} , the receiver with the FM-KCE is about 0.8 dB from the decoder with perfect CSI, while the receiver using channel estimates provided by the SS-KCE presents performance losses of 3 and 5.5 dB from the decoder with perfect CSI for 32 and 4 training codewords, respectively. For an SER of 10^{-4} , the receiver with the FM-KCE performs 2 and 3.5 dB better than the receiver with SS-KCE for 32 and 4 training codewords, respectively, and presents a loss of 0.5 dB from the ML space-time decoder with perfect CSI. Thus, from Figs. 9 and 10, we see that the FM-KCE allows the use of a small number of training codewords without compromising the performance of the receiver.

7. Summary

In this chapter, we presented channel estimation algorithms intended for systems employing orthogonal space-time block codes. Before developing the channel estimators, we construct a state-space model to describe the dynamic behavior of spatially correlated MIMO channels. Using this channel model, we formulate the problem of channel estimation as one of state

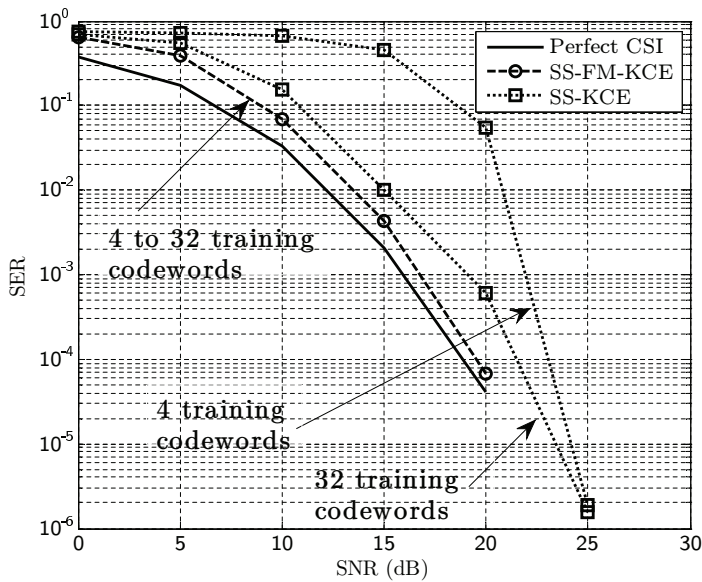


Fig. 10. Symbol error rate for SS-KCE and FM-KCE.

estimation. Thus, by applying the well-known Kalman filter to that state-space model, and using the orthogonality of OSTBCs, we arrive at a low-complexity optimal Kalman channel estimator. We also show that the channel estimates provided by the KCE in fact correspond to weighted sums of instantaneous maximum likelihood channel estimates.

For constant modulus signal constellations, a reduced complexity estimator is given by the steady-state Kalman filter. This filter also generates channel estimates by averaging instantaneous ML channel estimates. The existence and stability of the steady-state Kalman channel estimator is intimately related to the existence of solutions to the discrete algebraic Riccati equation derived from the KCE.

Simulation results indicate that the SS-KCE performs nearly as well as the optimal KCE, while demanding just a fraction of the calculations. They also show that the fading memory estimator outperforms the traditional Kalman filter by as much as 5 dB for a symbol error rate of 10^{-3} .

8. Acknowledgments

We acknowledge the financial support received from CAPES.

9. References

- Alamouti, S. M. (1998). A Simple Transmit Diversity Technique for Wireless Communications, *IEEE Journal on Selected Areas in Communications* 16(10): 1451–1458.
- Anderson, B. D. O. & Moore, J. B. (1979). *Optimal Filtering*, Prentice-Hall.

- Balakumar, B., Shahbazpanahi, S. & Kirubarajan, T. (2007). Joint MIMO Channel Tracking and Symbol Decoding Using Kalman Filtering, *IEEE Transactions on Signal Processing* 55(12): 5873–5879.
- Duman, T. M. & Ghayeb, A. (2007). *Coding for MIMO Communication Systems*, John Wiley and Sons.
- Enescu, M., Roman, T. & Koivunen, V. (2007). State-Space Approach to Spatially Correlated MIMO OFDM Channel Estimation, *Signal Processing* 87(9): 2272–2279.
- Gantmacher, F. R. (1959). *The Theory of Matrices*, Vol. 1, AMS Chelsea Publishing.
- Golub, G. H. & Van Loan, C. F. (1996). *Matrix Computations*, 3 edn, John Hopkins University Press.
- Haykin, S. (2002). *Adaptive Filter Theory*, 4 edn, Prentice-Hall.
- Horn, R. A. & Johnson, C. R. (1991). *Topics in Matrix Analysis*, Cambridge University Press.
- Jakes, W. C. (1974). *Microwave Mobile Communications*, John Wiley and Sons.
- Jamoos, A., Grivel, E., Bobillet, W. & Guidorzi, R. (2007). Errors-In-Variables-Based Approach for the Identification of AR Time-Varying Fading Channels, *IEEE Signal Processing Letters* 14(11): 793–796.
- Kailath, T., Sayed, A. H. & Hassibi, B. (2000). *Linear Estimation*, Prentice Hall.
- Kaiser, T., Bourdoux, A., Boche, H., Fonollosa, J. R., Andersen, J. B. & Utschick, W. (eds) (2005). *Smart Antennas – State of the Art*, Hindawi Publishing Corporation.
- Komninakis, C., Fragouli, C., Sayed, A. H. & Wesel, R. D. (2002). Multi-Input Multi-Output Fading Channel Tracking and Equalization Using Kalman Estimation, *IEEE Transactions on Signal Processing* 50(5): 1065–1076.
- Larsson, E. & Stoica, P. (2003). *Space-Time Block Coding for Wireless Communications*, Cambridge University Press.
- Larsson, E., Stoica, P. & Li, J. (2003). Orthogonal Space-Time Block Codes: Maximum Likelihood Detection for Unknown Channels and Unstructured Interferences, *IEEE Transactions on Signal Processing* 51(2): 362–372.
- Li, X. & Wong, T. F. (2007). Turbo Equalization with Nonlinear Kalman Filtering for Time-Varying Frequency-Selective Fading Channels, *IEEE Transactions on Wireless Communications* 6(2): 691–700.
- Liu, Z., Ma, X. & Giannakis, G. B. (2002). Space-Time Coding and Kalman Filtering for Time-Selective Fading Channels, *IEEE Transactions on Communications* 50(2): 183–186.
- Loiola, M. B., Lopes, R. R. & Romano, J. M. T. (2009). Kalman Filter-Based Channel Tracking in MIMO-OSTBC Systems, *Proceedings of IEEE Global Communications Conference, 2009 – GLOBECOM 2009*, IEEE, Honolulu, HI.
- Piechocki, R. J., Nix, A. R., McGeehan, J. P. & Armour, S. M. D. (2003). Joint Blind and Semi-Blind Detection and Channel Estimation for Space-Time Trellis Coded Modulation Over Fast Faded Channels, *IEE Proceedings on Communications* 150(6): 419–426.
- Simon, D. (2006). *Optimal State Estimation - Kalman, H_∞ , and Nonlinear Approaches*, John Wiley and Sons.
- Tarokh, V., Jafarkhani, H. & Calderbank, A. R. (1999). Space-Time Block Codes from Orthogonal Designs, *IEEE Transactions on Information Theory* 45(5): 1456–1467.
- Vucetic, B. & Yuan, J. (2003). *Space-Time Coding*, John Wiley and Sons.

Adaptive Filtering for Indoor Localization using ZIGBEE RSSI and LQI Measurement

Sharly Joana Halder¹, Joon-Goo Park² and Wooju Kim¹

¹*Yonsei University, Seoul*

²*Kyungpook National University, Daegu
Republic of Korea*

1. Introduction

The term “filter” is often used to describe a device in the form of a piece of physical hardware or computer software that is applied to a set of noisy data in order to extract information about a prescribed quantity of interest [25], [26]. Filter has been designed to take noisy data as input to reduce the effects of noise as much as possible.

A Wireless Sensor Network (WSN) is a network that consists of numerous small devices that are in fact tiny computers. These so-called nodes are composed of a power supply, a processor, different kinds of memory and a radio transceiver for communication. WSNs are generally used to observe or sense the environment in a non-intrusive way. In order to perform this task, nodes are often extended with sensors, like infrared, ultrasonic or temperature sensors, hence the names sensor nodes and sensor networks. The domain of WSNs is still very young. During the last few years, new developments in the area of communication, computing and sensing have enabled and stimulated the miniaturization and optimization of computer hardware. These evolutions have led to the emergence of WSNs. Despite the increasing capabilities of hardware in general, sensor nodes are still very restricted devices. They have a limited amount of processing power, memory capacity and most importantly energy. This makes WSNs a challenging research topic.

Despite current restrictions, several applications for WSNs have already been designed. WSNs are currently found in very different domains [3]. The large literature can be classified by relying on several criteria. One of these is the physical means used for localization, e.g., through the RF attenuation in the Electro-Magnetic (EM) waves [4], [11], [13] (Received Signal Strength Indicator - RSSI - based techniques) or the time required to cover the distance between transmitter and receiver (Ultra Wide Band); if using ultrasonic pulses, one could also use the time of arrival or time-difference of arrival of the waves [10]. This can even be extended to Audible-frequency sounds [9]. Another classification is based on the ranging feature, where distinguish between Range-free and Range-based localization techniques [11]. Moreover, it can be classified according to the Single-hop [11] and Multi-hop [14] localization scheme. Finally, it can differentiate between centralized [14] and distributed [9] localization systems.

A common consensus among localization researchers is that indoor localization requires room-level accuracy. Indoor localization uses many different sensors such as infrared, RFID, Ultrasound, Ultra-Wide Band, Bluetooth, and WLAN. Different sensors provide different

range of accuracy from centimeters to room-level. It seems like the accuracy is smaller than a room. But in practice, when we directly transform (x,y) , it coordinates to room-level information and causes mistakes. The reason being is, wireless signal is easy to suffer disturbance that makes localization unusual. This causes jumping in a split second or over a short span. Such situation may effect the location estimation from one room to another.

Ubiquitous indoor environments often contain substantial amounts of metal and other such reflective materials that affect the propagation of radio frequency signals in non-trivial ways, causing severe multipath effects, dead-spots, noise and interference. The main focus of this scheme is to represents a cheap and enhanced ranging technique to measure the radio strength by using two useful radio hardware link quality metrics named Link Quality Indicator (LQI) and Received Signal Strength Indicator (RSSI). In this scheme the mobile device itself calculate the position. Moreover, the device calculates its own position based on its own measurements. The proposed protocol tries to improve the existing algorithms [4], [27] using RSSI and LQI values. The indoor localization systems presented in this report are based on the RSSI as a strength indicator and LQI as a quality indicator of received packets. It can also be used to estimate a distance from a node to reference points. This system uses the LQI and RSSI in a different way and therefore it could lead to better and more predictable results than the other existing system. Several experiments were conducted to investigate the performance of the proposed scheme. At first, this system performs with respect to the signal analysis to understand the characteristic of the LQI and RSSI values on three types of environments to decide how the environment effects on RSSI and LQI strength. The effect of distance on received signal strength can be measured by RSSI and LQI provided by the radio. Secondly, this scheme performs with respect to the signal analysis is to filter the original signals in order to remove the noise. Besides, the noise could be estimated by using adaptive filtering algorithms. Sudden peaks and gaps in the signal strength are removed and the whole signal is smoothed, which eases the analysis process. We used two different types of new filtering to smooth the real RSSI, 'LQI' filtering and 'BOTH' filtering, and compared the results. And we found that 'BOTH' filter smooth more the raw RSSI value than existing 'Fusion' filtering [27].

In our research we used an adaptive filter as it performs well to track an object under such changing conditions in the RF signal environment. In this chapter, the proposed protocol will try to improve the existing algorithms using RSSI and LQI values. The localization systems presented in this report are based on RSSI as a strength indicator and LQI as a quality indicator of a received packets, it can also be used to estimate a distance from a node to reference points.

The remainder of this Chapter consists of six sub chapters. Chapter 2 describes some properties of ZigBee, RSSI and LQI. Chapter 3 reveals the previous works based on indoor location and WSN. Chapter 4 provides the proposed model of "Adaptive Filtering for Indoor Localization using ZIGBEE RSSI and LQI Measurement" and its probability of returning the correct location. Chapter 5 describes the analytical results obtained from the model of location system. And Chapter 6 concludes the chapter with conclusions.

2. ZigBee, RSSI and LQI

2.1 ZigBee

There are several standards that address mid to high data rates for voice, PC LANs, video, etc. and until recently there has not been a wireless network standard that meets the unique

needs of devices such as sensors and control devices. Sensors and control devices which are mostly used in industries and homes distinguish them with low data rates and in needs of very low energy consumption. A standards-based wireless technology needed having the performance characteristics that closely meet the requirements for reliability, security, low power and low cost.

Table 1 presented the IEEE 802.15 Task Group 4 is chartered to investigate a low data rate solution with multi-month to multi-year battery life and very low complexity. It is intended to operate in an unlicensed, international frequency band.

Since low total system cost is a main issue in industrial and home wireless applications, a highly integrated single-chip approach is the preferred solution of semiconductor manufacturers developing IEEE 802.15.4 compliant transceivers. The IEEE standard at the PHY is the significant factor in determining the RF architecture and topology of ZigBee enabled transceivers. For these optimized short-range wireless solutions, the other key element above the Physical and MAC Layer is the Network/Security Layers for sensor and control integration. The ZigBee group was organized to define and set the typical solutions for these layers for star, mesh, and cluster tree topologies.

Feature(s)	IEEE 802.11b	Bluetooth	ZigBee
Power Profile	Hours	Days	Years
Complexity	Very Complex	Complex	Simple
Nodes/Master	32	7	64000
Latency	Enumeration upto 3 sec	Enumeration upto 10 sec	Enumeration 30 ms
Range	100 m	10m	70m~300m
Extendability	Roaming possible	No	Yes
Data Rate	11Mbps	1Mbps	250Kbps
Security	Authentication Service Set ID (SSID)	64 bit, 128 bit	128 bit AES and Application Layer user defined

Table 1. Comparison of key features of complementary wireless technologies [4]

ZigBee Applications:

ZigBee is the wireless technology that:

- Enables broad-based deployment of wireless networks with low cost, low power solutions [5].
- Provides the ability to run for years on inexpensive primary batteries for a typical monitoring application [5].
- Addresses the unique needs of remote monitoring & control, and sensory network applications [5].

Figure 1 shows the ZigBee application areas. However, ZigBee technology is well suited to a wide range of building automation, industrial, medical and residential control & monitoring applications. Essentially, applications that require interoperability and/or the RF performance characteristics of the IEEE 802.15.4 standard would benefit from a ZigBee solution.

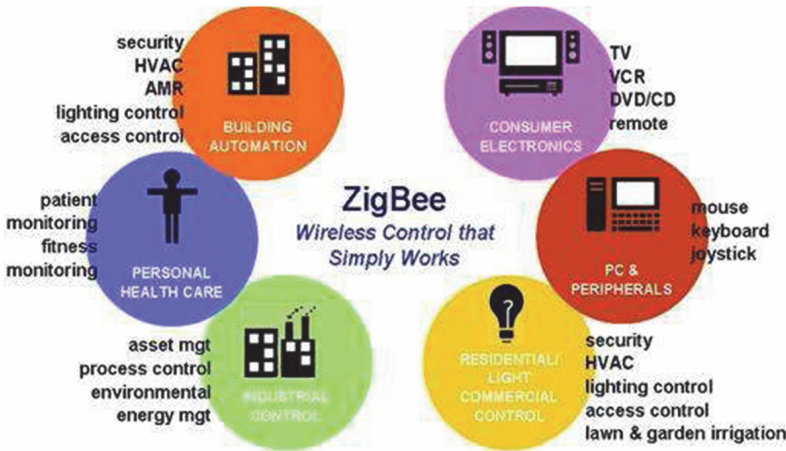


Fig. 1. ZigBee applications [5].

2.2 Received signal strength indicator (RSSI)

Majority of the existing methods leverage the existence of IEEE 802.11 base stations with powerful radio transmit powers of approximately 100mW per base station. Such radios are in a different class from the low power IEEE 802.15.4 compliant radios that typically transmit at low power levels ranging from 52mW to 29mW. The wide availability of larger number of IEEE 802.15.4 radios has revived the interest for signal strength based localization in sensor network. Despite of rapidly increasing popularity of IEEE 802.15.4 radios and signal strength localization, there is a lack of detailed characterization of the fundamental factors contributing to large signal strength variation. The analysis of RSSI values is needed to understand the underlying features of location dependent RSSI patterns and location fingerprints. An understanding of the properties of the RSSI values for location can assist in improving the design of positioning algorithms and in deployment of indoor positioning systems. The characteristics of RSS, received signal strength will decrease with increased distance as the equation below shows:

$$\text{RSSI} = - (10n\log_{10} d + A) \quad (1)$$

Where,

n = signal propagation constant, also named propagation exponent.

d = distance from sender.

A = received signal strength at a distance of one meter.

Lots of localization algorithms require a distance to estimate the position of unknown devices. One possibility to acquire a distance is measuring the received signal strength of the incoming radio signal.

The idea behind RSS is that the configured transmission power at the transmitting device (PTX) directly affects the receiving power at the receiving device (PRX). According to Friis' free space transmission equation, the detected signal strength decreases quadratically with the distance to the sender (Figure 2.a).

$$\text{PRX} = \text{PTX} * \text{GTX} * \text{GRX} (\lambda / 4\pi d)^2 \quad (2)$$

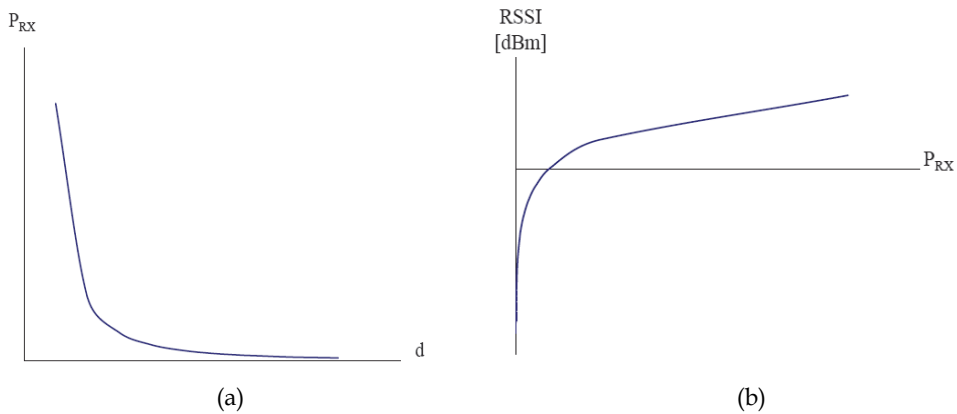


Fig. 2. (a) Received power P_{RX} versus distance to the transmitter. (b) RSSI as quality identifier of the received signal power P_{RX} .

Where,

P_{TX} = Transmission power of sender

P_{RX} = Remaining power of wave at receiver

G_{TX} = Gain of transmitter

G_{RX} = Gain of receiver

λ = Wave length

d = Distance between sender and receiver

In embedded devices, the received signal strength is converted to a received signal strength indicator (RSSI) which is defined as ratio of the received power to the reference power (P_{Ref}). Typically, the reference power represents an absolute value of $P_{Ref}=1mW$.

$$RSSI = 10 * \log PRX/PRF \quad [RSSI] = dBm \quad (3)$$

An increasing received power results a rising RSSI. Figure 2.b illustrates the relation between RSSI and the received signal power. Plotting RSSI versus distance d results in a graph, which is in principle axis symmetric to the abscissa. Thus, distance d is indirect proportional to RSSI. In practical scenarios, the ideal distribution of PRX is not applicable, because the propagation of the radio signal is interfered with a lot of influencing effects.

2.3 Link quality indicator (LQI)

For communications IEEE 802.15.4 radios provide applications with information about the incoming signal [17]. The effect of distance on received signal strength (RSS) can be measured by the packet success rate, RSSI and LQI provided by the radio. LQI is a metric introduced in IEEE 802.15.4 that measures the error in the incoming modulation of successfully received packets (packets that pass the CRC criterion). The LQI metric characterizes the strength and quality of a received packet. It is introduced in the 802.15.4 standard [1] and is provided by CC2430 [17]. LQI measures each successfully received packet and the resulting integer ranges from 0x00 to 0xff (0-255), indicating the lowest and highest quality signals detectable by the receiver (between -100dBm and 0dBm). The correlation value of LQI range from 50 to 110 where 50 indicates the minimum value and

110 represents the maximum. The 50 is typically the lowest quality frames detectable by CC2430. Software must convert the correlation value to the range 0-255, e.g. by calculating:

$$LQI = (CORR - a) \cdot b \quad (4)$$

Where,

CORR= correlation value, a and b are found empirically

The CORR (correlation value) is the raw LQI value which can be obtained from the last byte of the message. The raw value can get from CC2430 (CORR) is between 40 and 110.

Limited to the range 0-255, where a and b are found empirically based on PER measurements as a function of the correlation value. A combination of RSSI and correlation values may also be used to generate the LQI value. LQI values are uniformly distributed between these two limits. Different from RSSI, LQI measures the qualities of links while RSSI measures the strengths of links. LQI is a measure of the error in the signal, not the strength of the signal. A “weak” signal may still be a very crisp signal with no errors and thus a potentially good routing neighbor. If there is no interference from other 2.4 GHz devices, then LQI will generally be good over distance. Note that, scaling the link quality to a LQI, compliant with IEEE 802.15.4, must be done by software. This can be done on the basis of the RSSI value, the correlation value or a combination of those two. Signal strength and link quality values are not necessarily linked. But if the LQI is low, it is more likely that the RSSI will be low as well. Nevertheless, they also depend on the emitting power. A research group had the following results:

	Low RF	High RF
LQI	105	108
RSSI	- 75dBm	- 25dBm

Even though they do not describe how far from each other the sender and the receiver are located, it illustrates perfectly that both low and high power emissions guarantee a good link quality. The low RF emissions could be more sensitive to external disturbances. LQI exhibits a very good correlation with packet loss, and is therefore a good link quality indicator. However, one of the contributions of the present work is to show that RSSI is a reasonable metric if it is processed correctly, and if interference can be distinguished from noise. Given that LQI is a superior metric, it should not be forgotten that it is only made available by 802.15.4-compliant devices. It therefore makes sense to make the most out of RSSI.

3. Related works

This chapter introduces the area of ubiquitous computing and the underlying sensing technologies such as ultrasonic, infrared, Global Positioning Systems, and radio frequency identification. At first, an brief overview of each of the systems is given and then the similarities and differences to the approach are discussed.

3.1 Terminology and principles

There are numbers of existing location systems which utilize a variety of sensing technologies and system architectures. These systems have varying characteristics, such as accuracy, scalability, range, power consumption and cost. This section describes some of the

sensor terminology and principles used with reference to location systems. There are a number of different sensor technologies have been used in location systems.

3.1.1 Light

Light is a widely-used medium in location systems, varying from the use of simple infrared LEDs and sensors to tag-less vision based tracking. Infrared has been popularly used for containment-based location systems [20]. Infrared location systems can suffer in strong sunlight and under fluorescent tube lighting as both of these are sources of infrared light. Video cameras can be used both to recognize objects within the environment, allowing the device to calculate its position, or as an infrastructure to track mobile objects which may or may not be augmented. The processing power required to track objects, especially if they are untagged, using image-based methods can be large compared to other methods.

3.1.2 Radio-based localization

Localization in sensor networks can be achieved using knowledge about the radio signal behavior and the reception characteristics between two different sensor nodes. The quality of a radio signal, i.e. its strength at reception time, is expressed by RSSI: the higher the RSSI-value, the better the signal reception. The main advantage of using radio-based localization techniques is that no additional hardware for the sensor nodes is required. The main disadvantage of the technique is that the measured signal strengths are generally unstable and variable over time, which leads to localization errors. In this section, two common localization techniques using radio signal strength information are presented. Afterwards, the proximity idea is discussed, a technique that takes into account the range of radio communication rather than its quality. Finally, a technique for analyzing the RSSI behavior over time is presented. The technique cannot be used for localization itself, but it can provide useful mobility information about the node to be located. Following are three types of radio-based localization systems:

3.1.2.1 Converting signal strength to distance

In theory, there exists an exponential relation between the strength of a signal sent out by a radio and the distance the signal has traveled. In reality, this correlation has proven to be less perfect, but it still exists. Reference nodes broadcast a message to inform their position at regular intervals. Unknown nodes receive the broadcast message from reference nodes and measure the strength of the received signal [4], [27], [28]. Localization errors for this method range from two to three meters at average, with indoor errors being larger than outdoor ones. The main reason for the large number of errors is that the effective radio-signal propagation properties differ from the perfect theoretical relation that is assumed in the algorithm. Reflections, fading and multipath effects largely influence the effective signal propagation. The distance estimates, which are based on the theoretical relation, are thus inaccurate and lead to high errors in the calculated locations.

3.1.2.2 Fingerprinting signal strength

The second method that uses RSSI for localization is called Fingerprinting. This technique is based on the specific behavior of radio signals in a given environment, including reflections, fading and so on, rather than on the theoretical strength-distance relation. The fingerprinting technique [11], [12], [15] is an anchor-based technique that consists of two separate phases. During the first phase, called the Offline Phase, a fingerprint database of

the environment is constructed. During the next phase, called the Online Phase, real-time localization is performed. The greatest disadvantage of the fingerprints method is that an offline phase is required for the system to work. The offline phase is very time consuming. Moreover, the fingerprinting database that is created during the offline phase is location dependent. If one wants to use the same system in another environment or if radical changes to the current environment are made, the offline phase has to be repeated.

3.1.2.3 Proximity-based localization

Proximity-based localization systems are an anchor-based solution to the localization problem. These systems derive their location data from connectivity information of the network [4], [7], [8], [11], [12], [23]. Knowledge about whether two devices, i.e. an unknown node and an anchor, in the network are within communication range is transformed into an assumption about their mutual distance and location.

3.1.3 Ultrasound

The propagation speed of ultrasound waves in air is slow compared to that of RF. Sound waves are generally reflected by objects in the environment, which also makes position by containment possible. Utilizing the differential time-of-flight between RF and ultrasound pulses allows position to be estimated to within a few centimeters of the ground truth. The attenuation of sound in air limits range to several meters. Sound waves generally take about 20ms to die out; this therefore limits the update rate ultrasound location systems can obtain. The prevalent frequency used in ultrasound ranging is ultrasound. A lot of ultrasound location systems have been developed using narrow band 40 kHz transducers [10], [21]. Following are three types of ultrasound based localization systems:

3.1.3.1 The bat ultrasonic location system

Bat system provides fine-grain 3D location and orientation information which its predecessor, the Active Badge System, did not. Position is calculated using trilateration. The Bat emitter will transmit a short ultrasound pulse and receivers placed at 1.5m apart at known locations on the ceiling will pick up the signals [16].

3.1.3.2 Cricket

Cricket [10] is an indoor location system developed at MIT and utilizes RF and ultrasound using static transmitters and mobile receivers. The first iteration of the system is containment-based allowing for areas of arbitrary size to be created via careful placement of transmitters in the environment. A later iteration, called Cricket Compass [22], set out to allow orientation as well as position to be determined.

3.1.3.3 Dolphin

The Dolphin system [21], developed at the University of Tokyo, utilizes both RF and ultrasound to create a peer-to-peer system, providing co-ordinate based positioning. The aim is to develop a system which is easy to configure and provides a high degree of accuracy in three dimensions.

3.1.4 Adhoc positioning system (APS) using AoA

Niculescu and Nath [6] aim to create an algorithm and simulate a system where nodes have highly directional detection capabilities and there exist a small number of seeded nodes. Different algorithms were simulated in this chapter to gain some insight into how systems

with different properties would behave. The data suggests that higher node densities increase the probability of node connectivity sufficiently to calculate location and orientation. Smaller angles lower the error. However this reduces the percentage of nodes for which locations are estimated.

3.1.5 Global positioning system (GPS)

The GPS system consists of twenty seven satellites that orbit the Earth [14], [24]. GPS use the distance and angle measurements to the reference points are used to compute the position of the object by triangulation. GPS uses the time of flight of RF signals to estimate the distance between GPS satellites and receiver [24]. In indoor environments, GPS satellites signals get attenuated and reflected by various metallic objects [24]. Indoor GPS performance has fundamental limitations that result in much larger position estimation errors compared to outdoors.

RSSI can provide us with the cheapest localization system possible, while the form factor of the sensor nodes is not increased. The technique is applicable to indoor environment and the errors achieved with a RSSI-based system seem to be promising compared to the more expensive systems. In this chapter, we decided to design and implement a RSSI-based system to solve the localization problem listed above. The main reason is that it can be developed with small modifications to the existing systems.

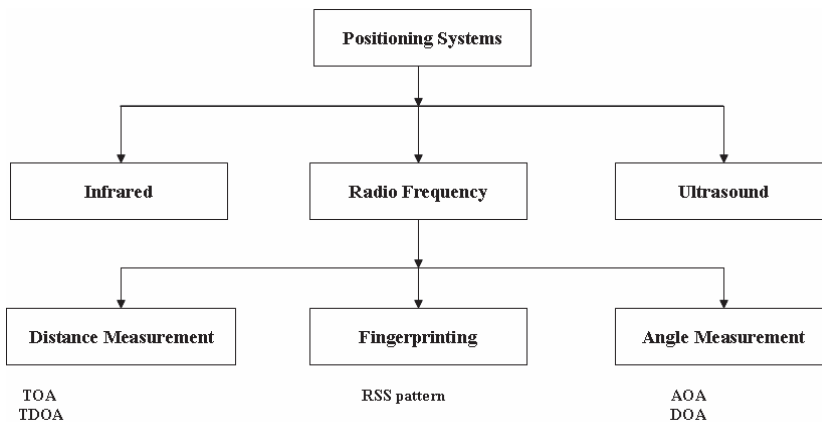


Fig. 3. Taxonomy of positioning system.

3.1.6 Comparison

The following table provides a comparison of the surveyed sensing systems from the point of accuracy and precision, scale, cost and limitations:

From Table 2, we concluded that RSSI can provide us with the cheapest localization system possible, while the form factor of the sensor nodes is not increased. The technique is applicable to indoor environment and the errors achieved with a RSSI-based system seem to be promising compared to the more expensive systems. In this chapter, we decided to design and implement a RSSI-based system to solve the localization problem listed here. The main reason is that it can be developed with small modifications to the existing systems. However, the aspects of accuracy and coverage area are still to be investigated in Chapter 4 and 5.

	GPS	Infrared	Ultrasound	RSSI
Applicable indoor	Not recommended	Yes	Yes	Yes
Need for extra hardware	Yes	Yes	Yes	No
Cost of extra hardware	High	Low	High	N.A
Size of extra hardware	Average	Average	Large	N.A
Average expected error	± 10 meters	± 5 meters	± 10 meters	1~3 meters

Table 2. Comparison of different location sensing technologies [13]

4. Proposed scheme

This chapter will focus on how this effective protocol has been implemented, and implementation issues are considered. In our experiments, to measure the radio strength, two useful radio hardware link quality metrics were used: (i) LQI and (ii) RSSI. Specifically, RSSI is the estimate of the signal power and is calculated over 8 symbol periods, while LQI can be viewed as chip error rate and is calculated over 8 symbols following the start frame delimiter (SFD). The specific point in a system where position estimates are calculated is an important design parameter. In this scheme the mobile device itself calculate the position. The device calculates its own position based on its own measurements.

4.1 Selected location system architecture

This scheme decides to use a private and scalable system. It features an active base station that transmits both RSSI and LQI signals. The mobile devices receive the signals, but they do not transmit anything themselves. The base station transmits the RSSI and LQI signals at the same moment in time. A mobile device measures signal, and is able to calculate the distance to the transmitter. By this scheme the location privacy of the user, who carries the mobile device, can be easily guaranteed because the mobile device does not send out any signals that might disclose its presence or its location. A further advantage of this architecture is scalability to many mobile devices. Because the mobile devices do not transmit any signals, there can be an unlimited number of mobile devices in principle. Due to its privacy and scalability features, this architecture might be particularly suitable for large-scale professional location systems or systems in public spaces. Each mobile device calculates its own position, based on the received signals. This scheme has divided into two subsystems.

As we know, for environmental changes the log model also change, so the proposed system uses a scaling factor for adjusting the log model with the measured data. This system includes a scaling factor s with the basic RSSI log model equation.

$$\text{RSSI} = -10 \log_{10} (sd + 1) + A \quad (5)$$

Where,

s = scaling factor

For our experiment, we use a filtering process for smoothing the RSSI values. We proposed a LQI filtering and BOTH filtering of RSSI and LQI values, for smoothing the measured

RSSI. From our experiment, we determine the filtering factor a for filtering and we used the following equation for smoothing the measured RSSI.

$$\text{smooth_RSSI}_{t(\text{BOTH})} = a * \text{RSSI}_t + (1-a) * \text{RSSI}_{t-1} \quad (6)$$

5. Experimental result

Adaptive filter contains a set of adjustable parameters. In design problem the requirement is to find the optimum set of filter parameters from knowledge of relevant signal characteristics according to some criterion.

This mathematical system combines the general principles of a proximity-based localization system with the analysis of the radio signal strength behavior over distance. This system uses the link quality indicator and radio signal strength indicator in a different way and therefore it could lead to better and more predictable results than the other existing system. Several experiments had conducted to investigate the performance of the proposed scheme.

The first step of this system performs with respect to the signal analysis to understand the characteristic of LQI and RSSI values on three types of environments. The effect of distance on received signal strength can be measured by RSSI and LQI provided by the radio. Equation 1 describes the basic model formula for RSSI where RSS decrease with increased distance. As for environmental change the log model also changed, this scheme decided to use a scaling factor s in the basic log model equation to adjust the log model with measured RSSI values. So, to find the accurate log model for specific environment we use a scaling factor s in equation 5. The experiment is conducted on three types of following environment to decide how the environment effects on RSSI and LQI strength. The first experiment is conducted in close space indoor environment.

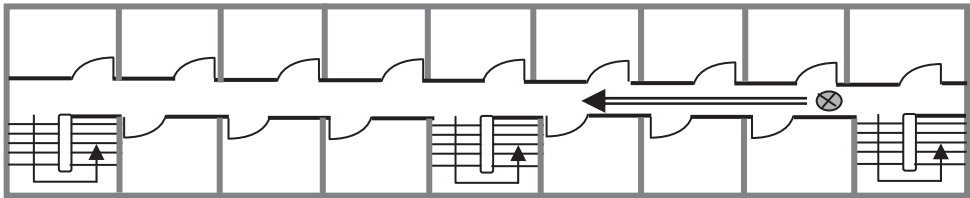


Fig. 4. Close space indoor environment (path 1).

The second experiment is deployed in half open space indoor environment, where few meters of the corridor was open.

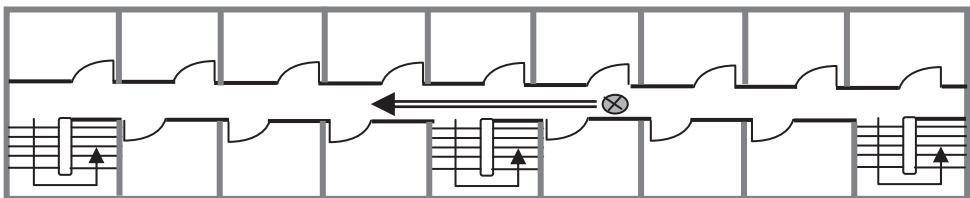


Fig. 5. Half open space indoor environment (path 2).

The third experiment is conducted on the open space indoor environment to decide the variation of the RF and LQI from other two experiments.

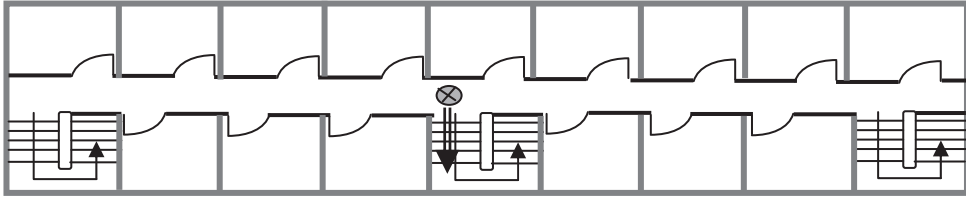


Fig. 6. Open space indoor environment (path 3).

To determine the accurate distance, the following distance equation has been used:

$$\text{Distance} = \frac{\frac{RSSI - A}{10 - 10n} - 1}{s} \quad (7)$$

And for measuring the signal attenuation factor the following equation has used:

$$N = \frac{RSSI - A}{-10 \log_{10}(sd + 1)} \quad (8)$$

The following figures represent the measured RSSI and LQI values. These figures are also representing how the log model curve adjusted with the measured RSSI values in close space indoor environments.

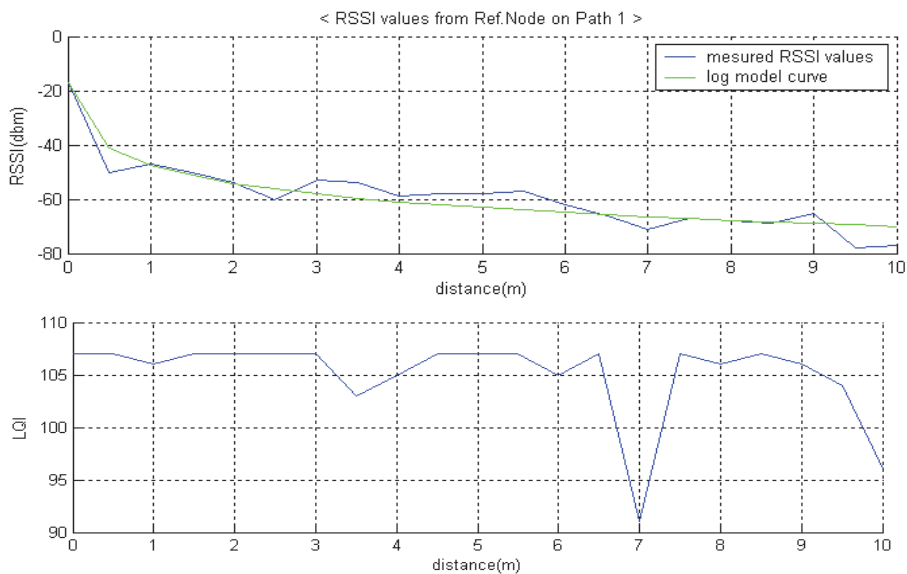


Fig. 7. Characteristics of RSSI and LQI in close space indoor environment.

The proposed system was implemented with TinyOS 2.x and the ZigBee device (Hybus, Hmote 2430), which were designed to operate in environments where they are approximately co-planar, therefore constrained in 3 of their 6 degrees of freedom (pitch, roll and z-axis). Hardware had already been developed as a research platform. This section will describe these devices. The performance of the systems has been analyzed by implementation and simulation using Matlab.

LQI and BOTH Filter

The proposed scheme performs with respect to the signal analysis to filter the original signals in order to remove the noise. Sudden peaks and gaps in the signal strength are removed and the whole signal is smoothed, which eases the analysis process.

Various filters can be used to smooth the RSSI value. Two common filters are, simple averaging and feedback filters. Averaging is the most basic filter type, but it requires more data packets to be sent. Feedback filters uses only a small part of the most recent RSSI value for each calculation. This requires less data, but increases the latency when calculating a new position.

Averaging filter: The average RSSI value is simply calculated by requiring a few packets from each reference node, each time the RSSI value are measured and calculated according to the equation below:

$$\overline{RSSI} = \frac{1}{n} \sum_{i=0}^{i=n} RSSI_i \quad (9)$$

Feedback filter: If a filter approximation is used, this can be done as shown below. In this equation the variable a is typically 0.75 or above. This approach ensures that a large difference in RSSI values will be smoothed. Therefore, it is not advisable if the assets that should be tracked can move long distance between each calculation.

$$RSSI_n = a * RSSI_n + (1-a) * RSSI_{n-1} \quad (10)$$

This means that the averaged RSSI value corresponding to the signal strength at distance depends on both the previous averaged value and the most recently measured value. As the value of a , which should be between zero and one, determines the degree of filtering if a is chosen to be close to one, the new measurement barely plays a role in the calculation of the new average. If on the other hand the value of a is nearly zero, virtually no filtering is performed. An optimal filter, i.e., value for a , specific for this project will be determined in this section. In this section we are going to propose a new filtering process which is a fusion of RSSI and LQI where we decided to use LQI as a reference aid when the RSSI or LQI is below RSSI and LQI threshold, respectively.

Figures 8 and 9 represent the average distance error according to RSSI and LQI values for all three testbeds. Figure 8 shows that in 8 meters the LQI value is 100. From our experiment we found that when LQI=100, it gives 20% reliability. Whereas, Figure 9 shows that in 8 meters RSSI value is -77, from our experiment we found that it gives 10% reliability. So we can come to a decision that over long distance LQI reliability is better than RSSI. Depend on this decision; we decide to use LQI as an assistance filtering factor for RSSI, which we are going to discuss next.

From Fig. 8, it is found that LQI gives best performance when the value was 108 in about 2 meters distance, which indicate that it gives 80% reliability. The results show that when the

value is 100, it gives lowest performance in about 8 meters. As our measurement testbed was 10 meters, we decided to determine the value below 10 meters. So, we determined the LQI reliability from 100 to 108, where the reliability varies from 20% to 80%, which means LQI filtering factor a varies between 0.8 to 0.2 for LQI filtering. And we also determine that if a value is below 101 then it should be negligible. The following equation has used for LQI filtering:

$$\text{smooth_RSSI}_{t(LQI)} = a \cdot \text{RSSI}_t + (1-a) \cdot \text{RSSI}_{t-1} \quad (11)$$

$$a = 0.8 - 0.6 \cdot \frac{108 - LQI}{8}$$

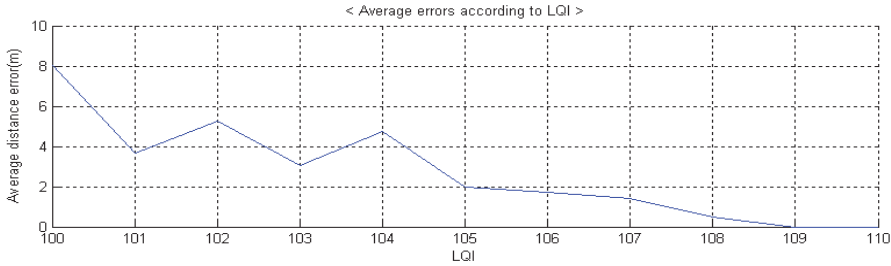


Fig. 8. Average errors according to LQI in all three paths.

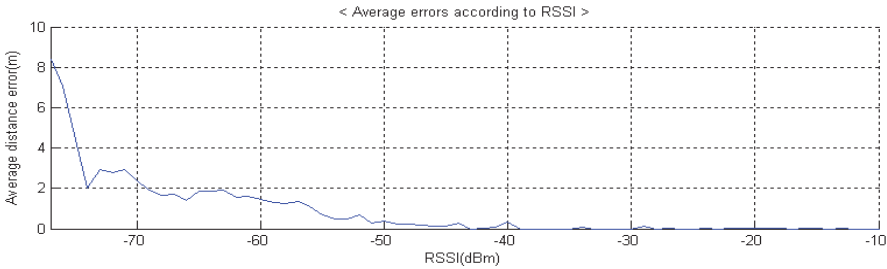


Fig. 9. Average errors according to RSSI in all three paths.

From Fig. 9, we found that RSSI gave best performance when its' value was -15 in about 0 meter distances, by which we determine that it gives 100% reliability. It has also seen that when its' value was -75 it gave lowest performance in about 5 meters, which means 50% reliability. So we decided the reliability for RSSI between -15 to -75, where the reliability varies from 50% to 100%. So this system decided to use the RSSI filtering factor a around 1 to 0.5 according to RSSI value. We also determine that if RSSI value is below -75 then the RSSI reliability will be 10%, which could be neglected. However, the following equation was used for RSSI filtering:

$$\text{smooth_RSSI}_{t(RSSI)} = a \cdot \text{RSSI}_t + (1-a) \cdot \text{RSSI}_{t-1} \quad (12)$$

$$a = 1 - 0.5 \cdot \frac{-15 - \text{RSSI}}{60}$$

With this experiment, we want to determine an optimal α value for filtering the radio signal strengths. We want to have a filter that is able to remove the noise i.e., the sudden peaks, gaps and shaded signal, but that should preserve the typical signal behavior of both stillness and movement.

For smoothing the real RSSI values, we use a fusion of RSSI and LQI filtering. For this BOTH filtering we use the LQI filtering in case of sudden peaks and shaded signals. In this BOTH filter we decided to measure the difference between present RSSI value and previous RSSI value. If the difference of present and previous RSSI value smaller than RSSI threshold value or if the LQI value is smaller than LQI threshold value then the signal will be filtered by RSSI filter otherwise it will use LQI filter. To determine the LQI threshold value, we use the defined LQI threshold value 105. For determining RSSI threshold value, we use ADD (average distance difference) value as a RSSI threshold average distance difference (ADD) [27].

At first we used LQI filtering technique to smooth the absolute RSSI values and after that we used BOTH filtering process.

LQI Filter

For smoothing the real RSSI values we use LQI filtering, so the real RSSI has smoothed very well.

if $LQI < 100$

$\alpha = 0$

else

$$\text{smooth_RSSI}_{t(LQI)} = \alpha \cdot \text{RSSI}_t + (1-\alpha) \cdot \text{RSSI}_{t-1}$$

After applying the filter, the program analyses the behavior of the filtered signal strengths over distance. Following figures provide the results:

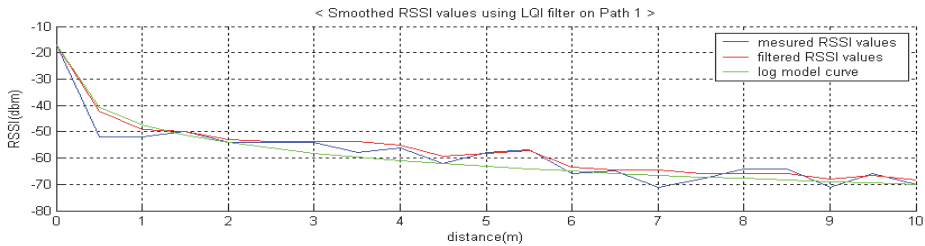


Fig. 10. Smoothed RSSI values using LQI filter for path 1.

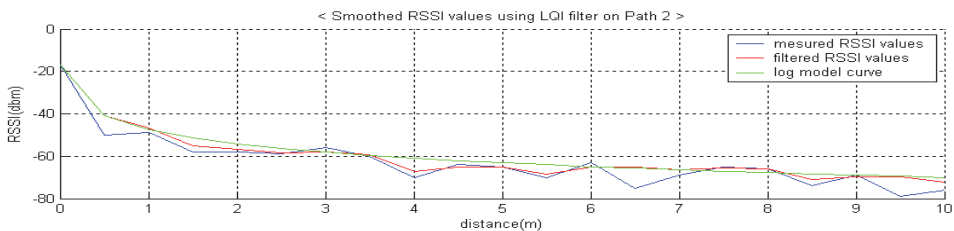


Fig. 11. Smoothed RSSI values using LQI filter for path 2.

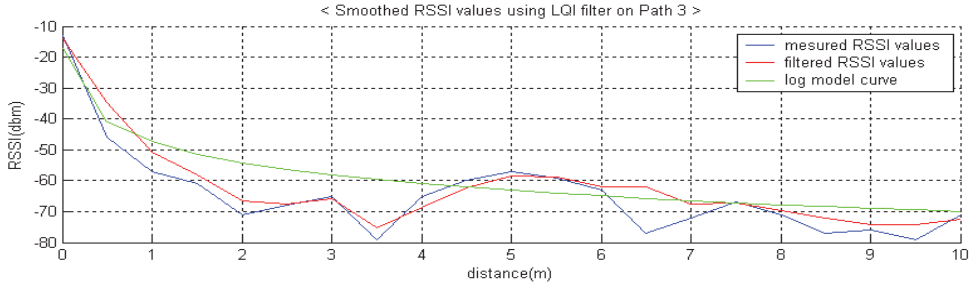


Fig. 12. Smoothed RSSI values using LQI filter for path 3.

BOTH Filter

The following equation has used to filter both RSSI and LQI values:

$$\text{RSSI} \rightarrow [\text{LQI filter}] \rightarrow [\text{RSSI filter}] \rightarrow \text{smooth_RSSI}_{\text{t(Both)}}$$

$$\text{smooth_RSSI}_{\text{(BOTH)}} = a \cdot \text{RSSI}_t + (1-a) \cdot \text{RSSI}_{t-1}$$

if $(\text{abs}(\text{RSSI}_t - \text{RSSI}_{t-1}) < \text{RSSI}_{\text{Threshold}}) \mid (\text{LQI}_t < \text{LQI}_{\text{Threshold}})$

LQI Filter

$\text{smooth_RSSI}_1 = \text{LQI Filter}$

RSSI Filter

$\text{smooth_RSSI}_2 = \text{RSSI Filter}$

$\text{smooth_RSSI}_{\text{t(BOTH)}} = \text{smooth_RSSI}_2$

After applying the filter, the program analyses the behavior of the filtered signal strengths over distance. Based on analysis, this system decides LQI values perform better than RSSI values over distance in a fully indoor environment. But on the other two environments, RSSI values perform better than LQI values. The best performance is happened when the RSSI values smoothed by BOTH filtering. Following figures provide the results:

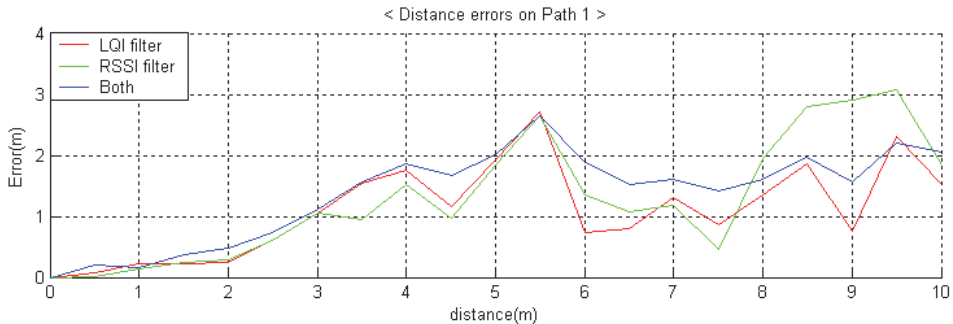


Fig. 13. Smoothed RSSI values using LQI, RSSI and BOTH filter for path 1.

From the above analyses, we can determine that BOTH filters perform better than the existing RSSI filter [4], Fusion filter [27] and our proposed LQI filter.

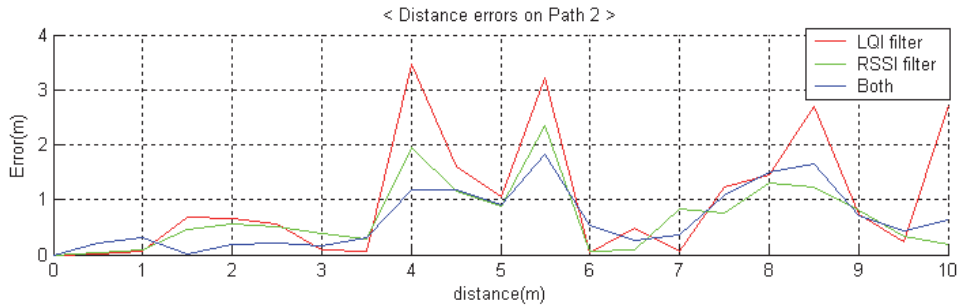


Fig. 14. Smoothed RSSI values using LQI, RSSI and BOTH filter for path 2.

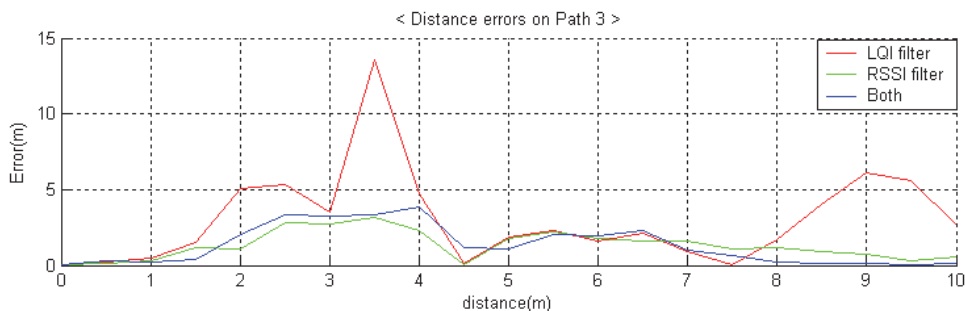


Fig. 15. Smoothed RSSI values using LQI, RSSI and BOTH filter for path 3.

Experimental performance

From the above findings, we can say that the proposed BOTH filtering algorithm could reduce more average error and maximum error distance than other existing algorithm [4], [12], [27], [28]. The average reduction of distance error deduced by using the proposed BOTH filtering is 56% and 68% average error and maximum error, respectively than the existing RSSI filtering [4] and Fusion filtering[27]. So we can say that our proposed method can perform well over other existing algorithm.

Table 3 shows the experimental results of our proposed filtering algorithm. From adaptive adjustment we found that this process could reduce a high number of errors from without filter measured signal.

From above discussion, we can draw the conclusion that our new enhancement technique gives a significantly improved performance over other existing techniques.

6. Conclusion

The technology is developing faster and faster, the expanding of the Internet makes people connected. There have been many forms of connections in cyberspace i.e., wired connection, wireless network, structured network, Ad-hoc network, etc. Life will be rather different without any form of such communication. Certainly the security will be a great concern in

PATH	FILTER NAME	AVG. ERROR	MAX ERROR	RSSI FILTER		LQI FILTER		FUSION FILTER		BOTH FILTER	
1	Non filter	1.2750	4.0577	Avg. RDC. of Avg- Error	Avg. RDC. of Avg- Error	Avg. RDC. of Avg- Error	Avg. RDC. of Avg- Error	Avg. RDC. of Avg- Error	Avg. RDC. of Avg- Error	Avg. RDC. of Avg- Error	Avg. RDC. of Avg- Error
	RSSI Filter	1.0911	3.1117								
	LQI Filter	1.0962	2.7127								
	Fusion Filter	1.2807	3.0742								
	BOTH filter	1.0959	2.6453								
2	Non filter	3.0142	15.183	33%	50%	41%	49%	49%	62%	56%	68%
	RSSI Filter	1.0943	5.4847								
	LQI Filter	1.0040	3.4574								
	Fusion Filter	0.8408	3.3281								
	BOTH filter	0.6495	1.8214								
3	Non filter	5.3016	21.183	33%	50%	41%	49%	49%	62%	56%	68%
	RSSI Filter	2.3972	7.2861								
	LQI Filter	3.0213	13.609								
	Fusion Filter	1.3088	3.1259								
	BOTH filter	1.3174	3.8351								

Table 3. Error reduction comparison of RSSI filter, Fusion filter, proposed LQI filter and BOTH filter

such beneficial technology. The security measures to provide Confidentiality and Integrity have been taken into account in the design of such technology. This chapter investigates the use of RF location systems for indoor domestic applications. Based on the assumption, low cost and minimal infrastructure are important for consumers, the concept of RF location system for Integrated In-door Location Using RSSI and LQI provided by ZigBee module is introduced.

This chapter addresses the problem of tracking an object. This chapter discuss about how to overcome the problems in the existing methods calculating the distance in indoor environment. This chapter has presented a new Mathematical Method for reducing the error in the location identification due to interference within the infrastructure based sensor

network. The proposed Mathematical Method calculates the distance using LQI and RSSI predicted based on the previously measured values. The calculated distance corrects the error induced by interference. The experimental results show that the proposed Mathematical Method can reduce the average error around 25%, and it is always better than the other existing interference avoidance algorithms. This technique has been found to work well in instances modeled on real world usage and thereby minimizing the effect of the error and hope that the findings in this chapter will be helpful for design and implementation of object tracking system in indoor environment.

7. Acknowledgements

This work is financially supported by Korea Minister of Ministry of Land, Transport and Maritime Affairs (MLTM) as U-City Master's and Doctoral Course Grant Program. And special thanks to Yen Sethia for her kind cooperation.

8. References

- [1] IEEE Standard for Information Technology. (October 2003). Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs), Local and Metropolitan Area Networks, Part 15.4
- [2] Kamran, J. (January 2005). ZigBee Suitability for Wireless Sensor Networks in Logistic Telemetry Applications. Master's Thesis in Electrical Engineering, School of Information Science, Computer and Electrical Engineering, Halmstad University, Sweden
- [3] Liu, H.; Darabi, H.; Banarjee, P. & Liu, J. (2007). Survey of Wireless Indoor Positioning Techniques and Systems. *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, Vol.37, No.6, (2007), pp. 1067-1080
- [4] Tae Young, C. (December 2007). A Study on In-door Positioning Method Using RSSI Value in IEEE 802.15.4 WPAN. Master's Thesis in School of Electronical Engineering & Computer Science, Kyungpook National University, Korea
- [5] <http://www.ZigBee.org/en/about/faq.asp>
- [6] Dragos, N. & Badri, N. (April 2001). Ad-hoc Positioning System, Technical Report DCS-TR-435, Rutgers University, also in *Symposium on Ad-Hoc Wireless Networks*, pp. 2926-2931, San Antonio, Texas, USA, November 2001
- [7] Lorincz, K. & Welsh, M. (2005). Motetrack: A Robust, Decentralized Approach to RF-based Location Tracking, *Proceedings of the International Workshop on Location- and Context-Awareness (LoCA '05)*, Munich, Germany, May 12-13, 2005
- [8] Vehbi Cagri, G. (August 2007). Real-Time and Reliable Communication In wireless Sensor and Actor Networks. PhD Thesis in School of Electrical and Computer Engineering, Georgia Institute of Technology, USA
- [9] Zhang, J.; Yan, T.; Stankovic, J. & Son, S. (2005). Thunder: A Practical Acoustic Localization Scheme for Outdoor Wireless Sensor Networks. Technical Report CS-2005-13, Department of Computer Science, University of Virginia, USA
- [10] Priyantha, N.; Chakraborty, A. & Balakrishnan, H. (2000). The Cricket Location-Support System, *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking*, pp. 32-43, Boston, MA, USA, August 6-11, 2000

- [11] Alippi, C. & Vanini, G. (2005). A RF Map-based Localization Algorithm for Indoor Environments, *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 652-655, Kobe, Japan, May 23-26, 2005
- [12] Bahl, P. & Padmanabhan, V. RADAR: An In-building RF-based User Location and Tracking System. INFOCOM, Vol.2, pp. 775-784, Tel Aviv, Israel
- [13] Kumar, S. (February 2006). Sensor System for Positioning and Identification in Ubiquitous Computing. Final Thesis
- [14] Bulusu, N.; Heidemann, J. & Estrin, D. (2000). GPS-less Low Cost Outdoor Localization for Very Small Devices, *Personal Communications Magazine*, Vol.7, No.5, pp. 28-34, October 2000
- [15] Kaemarungsi, K. (2005). Design of Indoor Positioning System Based on Location Fingerprint Technique. Master's thesis, University of Pittsburgh, USA
- [16] <http://www.uk.research.att.com/bat/>
- [17] CC2430 datasheet. Available from <http://www.chipcon.com/>
- [18] Want, R.; Hopper, A.; Falcao, V. & Gibbons, J. The Active Badge Location System. Technical Report 92.1, Olivetti Research Limited (ORL), ORL, 24a Trumpington Street, Cambridge CB2 1QA, UK
- [19] Krohn, A.; Beigl, M.; Hazas, M.; Gellersen, H. & Schmidt, A. (2005). Using Fine-grained Infrared Positioning to Support the Surface Based Activities of Mobile Users, *Fifth International Workshop on Smart Appliances and Wearable Computing (IWSAWC)*, Columbus, Ohio, USA, June 10, 2005
- [20] <http://www.ubisense.net/>
- [21] Fukuju, Y.; Minami, M.; Morikawa, H. & Aoyama, T. (2003). Dolphin: An Autonomous Indoor Positioning System in Ubiquitous Computing Environment, *IEEE Workshop on Software Technologies for Future Embedded Systems (WSTFES2003)*, pp. 53-56, Hakodate, Hokkaido, Japan, May 2003
- [22] Priyantha, N.; Miu, A.; Balakrishnan, H. & Teller, S. (2001). The Cricket Compass for Context-aware Mobile Applications, *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, pp. 1-14, Rome, Italy, July 16-21, 2001
- [23] Bahl, P.; Padmanabhan, V. & Balacgandran, A. (2000). Enhancements to the RADAR User Location and Tracking System. Microsoft Research Technical Report, February 2000.
- [24] Getting, I. The Global Positioning System. *IEEE Spectrum*, Vol.30, No.12, (December 1993), pp. 36- 47,
- [25] Simon, H. (1984). *Introduction to Adaptive Filters*, ISBN 0029494605, Collier Macmillan Publishers, London
- [26] Sayed, A. (2003). *Fundamentals of Adaptive Filtering*, ISBN 0471461261, IEEE Press Wiley-Interscience, New York
- [27] Halder, S. J.; Choi, T.; Park, J.; Kang, S.; Park, S. & Park, J. (2008). Enhanced Ranging Using Adaptive Filter of ZIGBEE RSSI and LQI Measurement, *Proceedings of The 10th International Conference on Information Integration and Web-based Applications & Services (iiWAS2008)*, pp. 367-373, Linz, Austria, November 24-26, 2008
- [28] Halder, S. J.; Choi, T.; Park, J.; Kang, S.; Yun, S. & Park, J. (2008). On-line Ranging for Mobile Objects Using ZIGBEE RSSI Measurement. *Proceedings of The 3rd International Conference on Pervasive Computing and Applications (ICPCA2008)*, pp. 662-666, Alexandria, Egypt, October 06-08, 2008

Part 4

Other Applications

Adaptive Filters for Processing Water Level Data

Natasa Reljin¹, Dragoljub Pokrajac¹ and Michael Reiter²

¹*Delaware State University,*

²*Bethune-Cookman University*
USA

1. Introduction

Salt marshes are composed of various habitats contributing to high levels of habitat diversity and increased productivity (Kennish, 2002; Zharikov et al., 2005), making them among the most productive ecosystems on the Earth. The salt marsh consists of a halophytic vegetation community growing near saline waters (Mitsch & Gosselink, 2000) characterized by grasses, herbs, and low shrubs (Adam, 2002). Salt marshes exist between the upper limit of the high tide and the lower limit of the mean high water tide (Adam, 2002). They represent an important factor in the support of surrounding food chains, and due to the high level of productivity their economic and aesthetic value is increasing (Delaware Department of Natural Resources and Environmental Control, 2002; Zharikov et al. 2005). The survival and reproduction of many species of commercial fish and shellfish is dependent upon salt marshes (Zharikov & Skilleter, 2004). In addition, salt marshes provide critical habitat and food supply to crustaceans (Zharikov et al., 2005) and shorebirds (Potter et al., 1991). They are often considered as a primary indicator of the ecosystem health (Zhang et al., 1997). Because of their ability to transfer and store nutrients, salt marshes are an important factor in the maintenance and improvement of water quality (Delaware Department of Natural Resources and Environmental Control, 2002; Zhang et al., 1997). In addition, they provide significant economic value as a cost-effective means of flood and erosion control (Delaware Department of Natural Resources and Environmental Control, 2002; Morris et al., 2004). This economic value makes coastal systems the site of elevated human activity (Kennish, 2002).

Determining the effects of sea level rise on tidal marsh systems is currently a very popular research area (Temmerman et al., 2004). While average sea level has increased 10-25 cm in the past century (Kennish, 2002), the Atlantic coast has experienced a sea level rise of 30 cm (Hull & Titus, 1986). Local relative sea level has risen an average rate of 0.12 cm yr⁻¹ in the past 2000 years, but at Breakwater Harbor in Lewes, DE sea level is rising at the average rate of 0.33 cm yr⁻¹, nearly three times that rate (Kraft et al., 1992). According to the National Academy of Sciences and the Environmental Protection Agency, sea level rise within the next century could increase 60 cm to 150 cm (Hull & Titus, 1986).

The changes in sea level rise are particularly affecting tidal marshes, since they are located between the sea and the terrestrial edge (Adam, 2002; Temmerman et al., 2004). The prediction is that sea level rise will have the most negative effect on marshes in the areas where the landward migration of the marsh is restricted by dams and levees (Rooth & Stevenson, 2000).

If sea level rises the almost certain prediction of 0.5 m by 2100 and marsh migration is prevented, then more than 10,360 km² of wetlands will be lost (Kraft et al., 1992). If the sea level rises 1 m then 16,682 km² of coastal marsh will be lost, which is approximately 65% of all extant coastal marshes and swamps in the United States (Kraft et al., 1992).

Due to an imminent potential threat which can jeopardize the Mid-Atlantic salt marshes, it is very important to examine the effect of sea level rise on these marshes. The marshes of the St. Jones River near Dover, DE, can be considered to be typical Mid-Atlantic marshes. These marshes are located in developing watersheds characterized by dams, ponds, agricultural lands, and increasing urbanization, providing an ideal location for studying the impacts of sea level rise on salt marsh extent and location. In order to determine the effect of sea level rise on the salt marshes of the St. Jones River, the change in salt marsh composition was quantified. Unfortunately, as for most marsh locations along the Atlantic seaboard, the data on sea level rise for this area was not available for comparison with marsh condition. However, a wide data set for this area is available through a water quality monitoring program, and if it could be properly processed and analyzed it could result in sea level rise data for the location of the interest.

In this chapter, we describe the application of signal processing on the water level data from the St. Jones River watershed. The emphasis is on adaptive filtering in order to remove the influence of upstream water level on the downstream levels.

2. Data

The St. Jones River, in central Delaware, is 22.3 km long (Pokrajac et al., 2007a). It has an average mean high water depth (MHW) of 4 m in the main stem, and an average width of 15 feet. The site's watershed area is 19,778 ha, and the tidal reaches are influenced by fresh water runoff from the urbanized area upstream. An aerial photo of the St. Jones River is shown in Fig. 1.



Fig. 1. Aerial photo of St. Jones River.

The data used in this research were obtained from the Delaware National Estuarine Research Reserve (DNERR), which collected the data as part of the System Wide Monitoring Program (SWMP) under an award from the Estuarine Reserves Division, Office of Ocean and Coastal Resource Management, National Ocean Service, and the National Oceanic and Atmospheric Administration (Pokrajac et al. 2007a, 2007b). Through SWMP, researchers collect long term water quality data from coastal locations along Delaware Bay and elsewhere in order to track trends in water quality.

The original dataset contained 57,127 measurements, taken approximately every thirty minutes using YSI 6600 Data Probes (Fig. 2) (Pokrajac et al., 2007a, 2007b). The measurements were taken from January 31, 2002 through October 31, 2005. In order to determine if sea level rise is influencing the St. Jones River, the water level data were collected from two SWMP locations: Division Street and Scotton Landing (Pokrajac et al., 2007b). Probes were left in the field for two weeks at a time, collecting measurements of water level, temperature ($^{\circ}\text{C}$), specific conductivity (mS cm^{-1}), salinity (ppt), depth (m), turbidity (NTU), pH (pH units), dissolved oxygen percent saturation (%), and dissolved oxygen concentration (mg L^{-1}). We used only the water level (depth) data for this study, which were collected using a non-vented sensor with a range from 0 to 9.1 m, an accuracy of ± 0.18 m, and a resolution of 0.001 m. Due to the fact that the probes are not vented, changes in atmospheric pressure appear as changes in depth, which results in an error of approximately 1.03 cm for every millibar change in atmospheric pressure (Mensing, 2005). However, the exceptionally large dataset (57,127 data points) overwhelms this data error.



© Fondriest Environmental, Inc.

Fig. 2. YSI 6600 Data Probe.

The downstream location, Scotton Landing, is located at coordinates latitude 39 degrees 05' 05.9160" N, longitude 75 degrees 27' 38.1049" W (Fig. 3). It has been monitored by SWMP since July 1995. The average MHW depth is 3.2 m, and the river is 12 m wide (Mensing, 2005). This location possesses a clayey silt sediment with no bottom vegetation, and has a salinity range from 1 to 30 ppt. The tidal range is from 1.26 m (spring mean) to 1.13 m (neap mean). The data collected at the Scotton Landing site are referred as *downstream data* (see Fig. 4).

The water level data from the Scotton Landing site alone were not sufficient. In addition to tidal forces, this site is influenced by upstream freshwater runoff, so changes in depth could not be isolated to sea level change. However, the data from a non-tidal upstream sampling site could be used for removing the upstream influence at Scotton Landing. Therefore, the data from an upstream location, Division Street, was included in the analysis. Its coordinates are latitude 39 degrees 09' 49.4" N, longitude 75 degrees 31' 8.7" W (see Fig. 3.). The Division Street sampling site is located in the mid portion of the St. Jones River, upstream from the Scotton Landing site. At this location, the river's average depth is 3 m and width is 9 m. The site possesses a clayey silt sediment with no bottom vegetation, and has a salinity in the range from 0 to 28 ppt. The tidal range at this location varies from 0.855 m (spring mean) to 0.671 m (neap mean). The data were monitored from January 2002 (Mensing, 2005). The data collected at the Division Street site are referred to as *upstream data* (see Fig. 4).

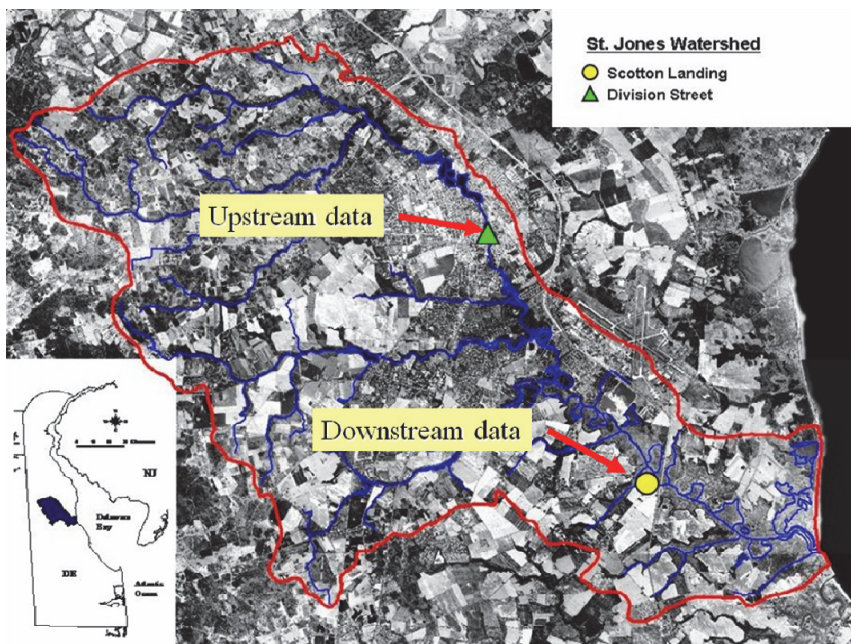


Fig. 3. Sampling locations for St. Jones data: ▲ Division Street (upstream data); ● Scotton Landing (downstream data).

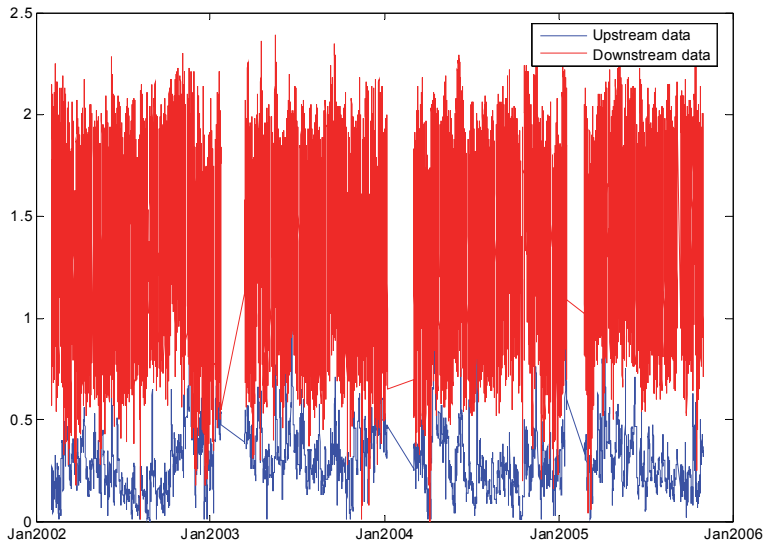


Fig. 4. Original dataset (upstream and downstream data).

3. Data pre-processing

The data were sampled every $T_s = 30$ minutes, and the dataset consisted of “chunks” of continuous measurements. Some of the measurements were missing due to maintenance or malfunction of the equipment, probe replacement, etc. The length of the intervals with missing measurements varied between 1 h (1 missing measurement) and 1517.5 h (3036 missing measurements), but the majority of the intervals were shorter than 10 h.

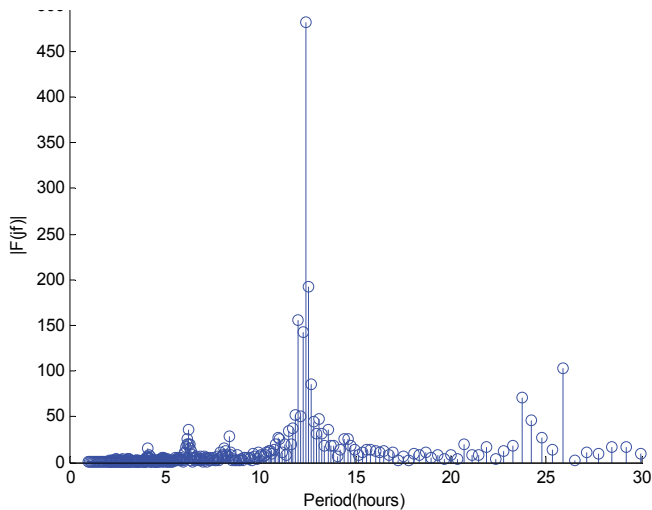


Fig. 5. Spectrum of collected data before filtering (chunk 99, downstream data).

The discrete Fourier spectra (Proakis & Manolakis, 2006) of all the chunks contained three prominent peaks, which is shown in Fig. 5 using chunk 99 from the downstream data. The first peak corresponds to lunar semi-diurnal tides with a period of approximately 12.4 h, and the diurnal tides with a period of approximately 24.8 h. In addition, there is a peak that corresponds to solar tides, which have a period of approximately 12 h. These periodicities are also shown in Fig. 6.

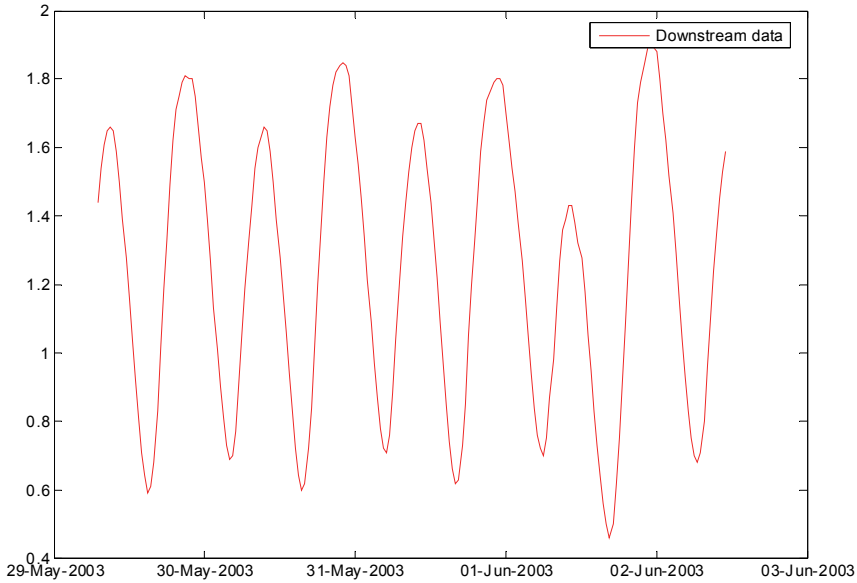


Fig. 6. The periodicities of the downstream data.

The dataset had several problems that had to be rectified before further processing. One data sample (Sep 28, 2004, 09:00:00) had an incorrect time, which was located sometime between Sep 27, 2004, 23:30:00 and Sep 28, 2004, 00:30:00, and was corrected. Four data samples (Jul 24, 2003, 07:30:00; Jun 10, 2005, 09:00:00; Aug 11, 2005, 15:00:00; Aug 11, 2005, 15:30:00) had missing values. In addition, the number of intervals with no measurements (total of 99 “gaps” in experiment) represented a problem for signal processing (for example, for filtering). Fig. 7 shows the number of chunks as a function of the duration of the missing measurements. Due to the properties of the used data and the shortest period of 12 h, we decided to interpolate intervals shorter than 12 h. Also, we interpolated all the above mentioned samples with missing data values. The treatment of the missing values is shown in Fig. 8.

In order to interpolate data for each interval of missing measurements, first we approximated the existing data within 20 samples from the interval. We used a least squares approximation followed the combination of the 4th order polynomial and trigonometric functions:

$$x(t) = \sum_{j=0}^4 a_j t^j + \sum_{j=1}^3 A_j \sin\left(\frac{2\pi t}{T_j} + \theta_j\right) \quad (1)$$

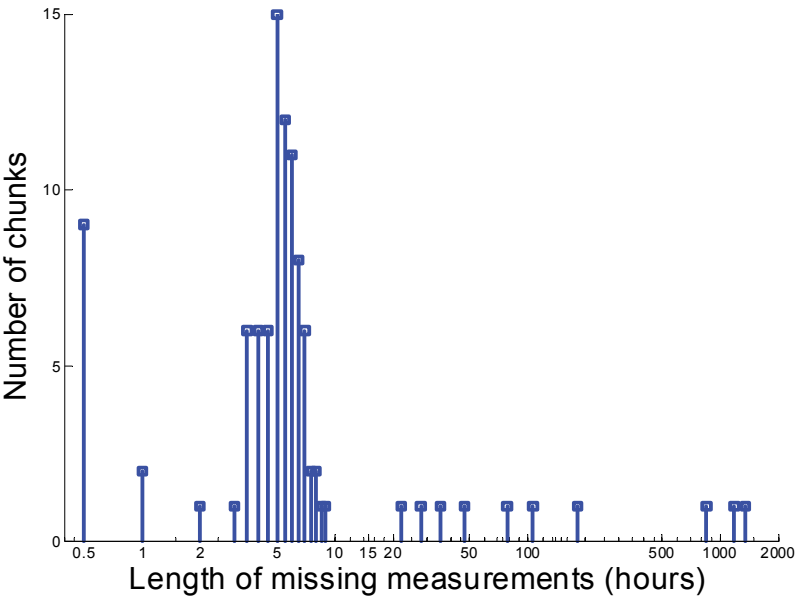


Fig. 7. The number of chunks as function of the duration of missing measurements.

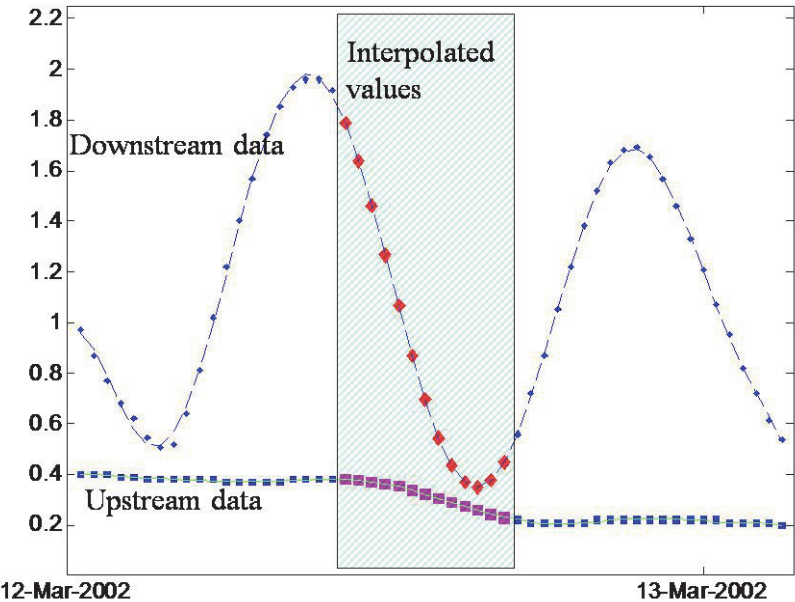


Fig. 8. The treatment of the missing values.

where $T_1 = 12.4$ h, $T_2 = 24.8$ h and $T_3 = 12$ h. Then, we interpolated missing values using the computed approximation functions. The interpolation was performed on 866 samples, which represented less than 2% of the original number of samples. One example of the interpolated intervals is depicted in Fig. 9.

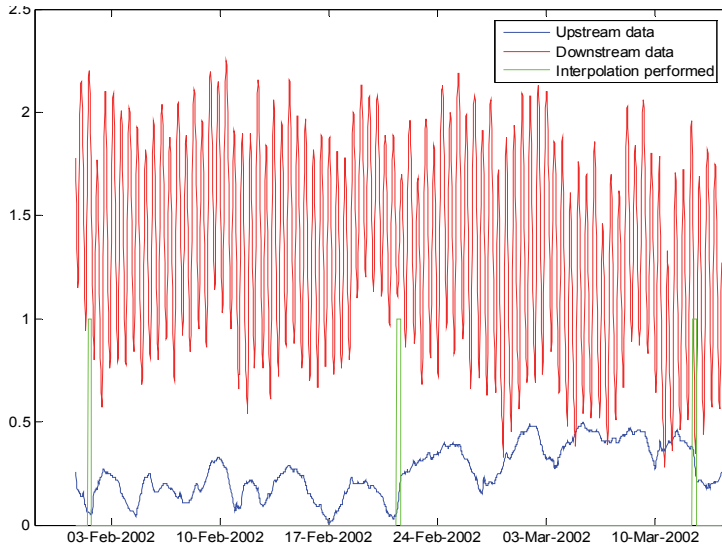


Fig. 9. An example of interpolated intervals.

The interpolation resulted in the merging of the majority of chunks, thus giving us only 11 chunks. The sizes of the new chunks were as follows: 4105, 5422, 4, 4, 7154, 14357, 10750, 5, 4491, 9423, and 2278. Three of those chunks (3, 4 and 8) have very small size, which made them suitable for discarding. Therefore, the interpolation process left us with only 8 chunks.

4. Filtering of the tidal components

We performed discrete filtering of both upstream and downstream data using the Filter Design and Analysis (FDA) Tool in Matlab Signal Processing Toolbox, v.6.2 in order to remove the tidal periodic components from the data. The first idea was to create and use the infinite impulse response (IIR) filter (Proakis & Manolakis, 2006), because it can potentially meet the design specifications with lower order than the corresponding finite impulse response (FIR) filter, which would also result in shorter time to buffer the data. However, several attempts (using the Yule-Walker method, notch or elliptic filters) didn't achieve the expected results – the order was too high and the attenuation was less than specified (Pokrajac et al., 2007a). Hence, we designed the FIR filter. Since the spectrum of the data had peaks in two bands (see Fig. 5), two stopband filters were designed. Both of them had a passband ripple of 0.05, and the sampling frequency $f_s = (1/30) \text{ min}^{-1} = 0.556 \text{ mHz}$ (Pokrajac et al., 2007a). In order to have a stopband attenuation of at least 20 dB in the 11 – 11.4 μHz band, which corresponds to a 24.8 h period, the first created filter was of order 168. The attenuation of 40 dB in the 22.401 – 23.148 μHz band (which corresponds to periods of 12 and 12.4 h) was achieved with the second filter of order $N_{\text{filter}} = 354$. Here, more attenuation was needed due to the very high corresponding peak in the spectrum. In Figs. 10 and 11,

magnitude responses of the first and the second filters are shown. The result of applying both filters on chunk 99 and downstream data is illustrated in Fig. 12. At the beginning of each chunk, we had to discard $N_{filter}-1$ data samples in order to perform filtering. This led to discarding less than 5% of the data. The standard deviation of the downstream data after the filtering was $std(y_{FIR}(t)) = 0.200$. Also, we tried the alternative approach by applying a moving average (MA) filter of length $Q = 25$, which corresponds to a period of 12.4 h. Standard deviation of the downstream data after the MA filter was $std(y_{MA}(t)) = 0.223$. The result of filtering the downstream data is shown in Fig. 13.

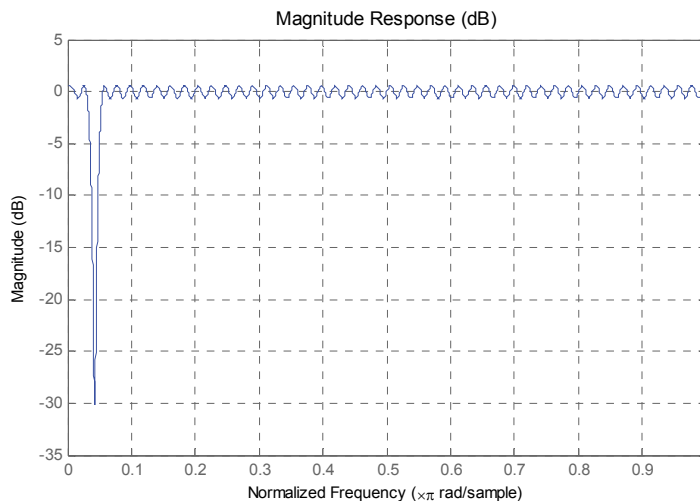


Fig. 10. Magnitude response of the first filter.

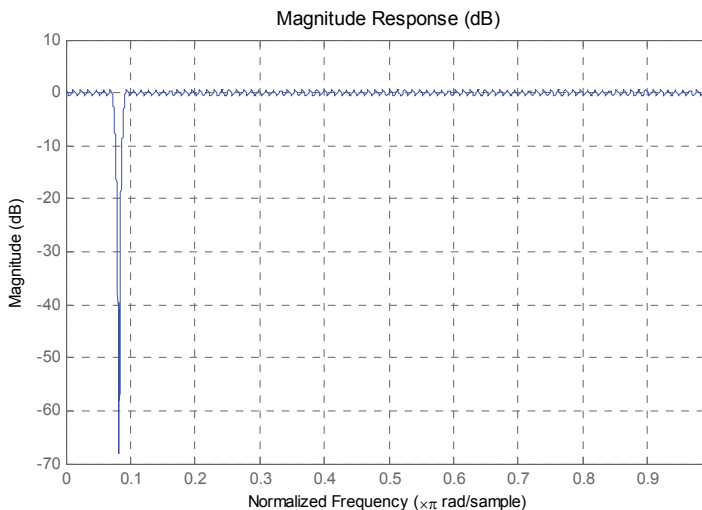


Fig. 11. Magnitude response of the second filter.

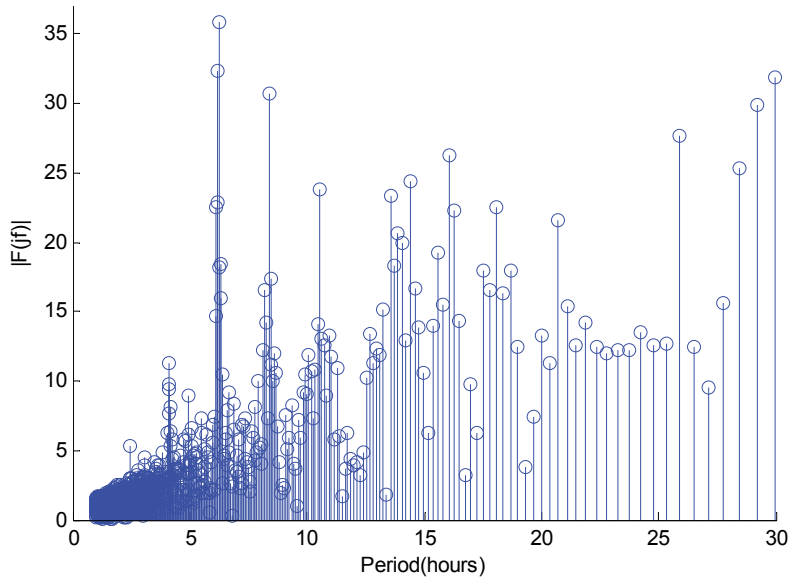


Fig. 12. Spectrum after filtering (chunk 99 and downstream data).

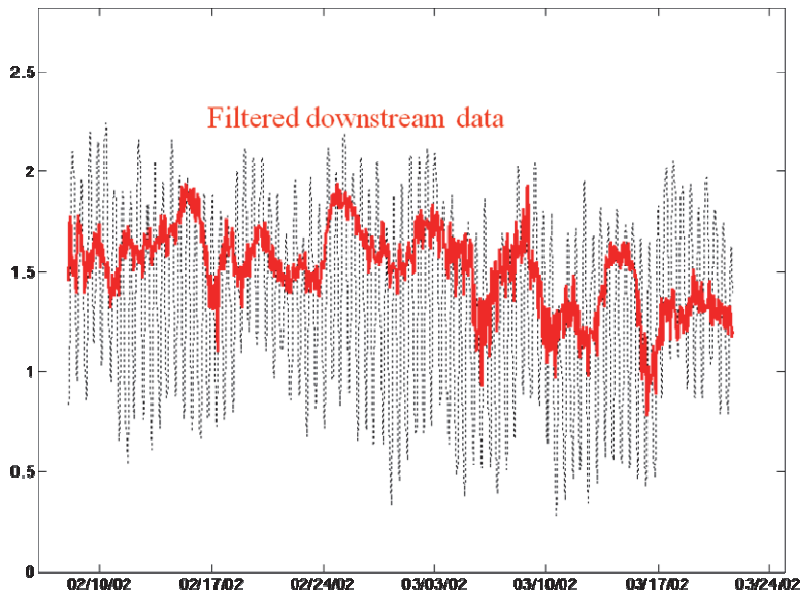


Fig. 13. Filtered downstream data.

5. Application of the adaptive filters

The downstream data y_t can be considered as a non-stationary function of the delayed upstream data x_t (see Fig. 14) (Pokrajac et al., 2007a, 2007b). It can be described as the discrete model $y_t = f_t(x_t, x_{t-T_s}, \dots, x_{t-(L-1)T_s}) + r_t$, where L is the maximal delay of the model and r_t is the residual corresponding to the portion of the downstream data which cannot be explained by the upstream data.

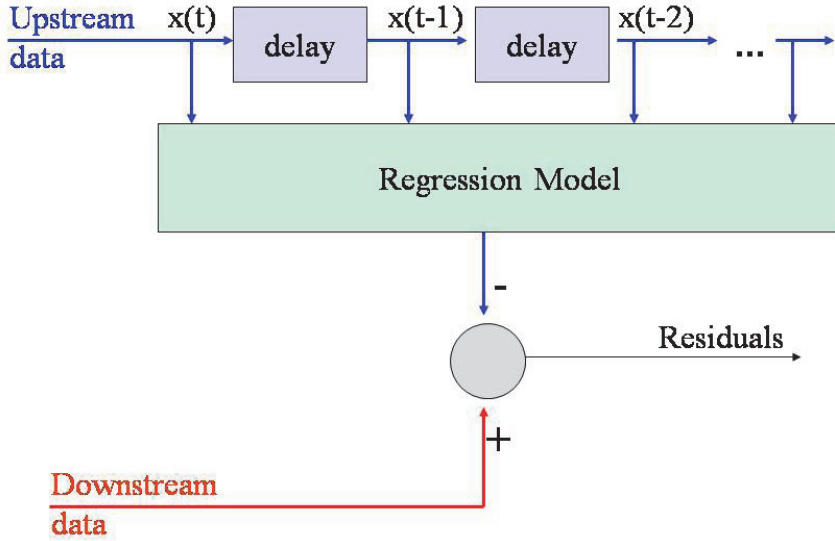


Fig. 14. Removal of the upstream data influence.

If a function f_t is linear, the adaptive linear model can be represented as follows:

$$y_t = \mathbf{w}_t^T \mathbf{x}_t + r_t \quad (2)$$

where $\mathbf{w}_t = [w_{0,t} \dots w_{L-1,t}]^T$ are coefficients and $\mathbf{x}_t = [x_t \dots x_{t-(L-1)T_s}]^T$ is the upstream data vector. A linear regression model could be obtained if the coefficients $\mathbf{w}$ are held constant (Devore, 2007):

$$y_t = \mathbf{w}^T \mathbf{x}_t + r_t \quad (3)$$

The coefficient of determination, R^2 , is usually used to measure the accuracy of the model, (Devore, 2007). It is defined as a function of averaged squared residuals and the standard deviation of the response:

$$R^2 = 1 - \frac{\overline{\hat{r}_t^2}}{\text{std}(y_t)^2} \quad (4)$$

where the residuals are estimated with:

$$\hat{r}_t = y_t - \mathbf{w}_t^T \mathbf{x}_t \quad (5)$$

The updating of the coefficients $\mathbf{w}_t$ in Eq. (2) is performed using the Widrow-Hoff least mean squares (LMS) algorithm (Widrow & Stearns, 1985):

$$\mathbf{w}_{t+1} = \mathbf{w}_t + 2\mu\hat{r}_t\mathbf{x}_t \quad (6)$$

where μ represents the adjustable learning rate, and $\hat{r}_t$ is estimated using Eq. (5). In addition to the Widrow-Hoff LMS algorithm, we applied time notching by adjusting the coefficients only when all the time instants, $t, \dots, t-(L-1)T_s$, belonged to the same chunk of interpolated data (Pokrajac et al., 2007a, 2007b), see Fig. 15.

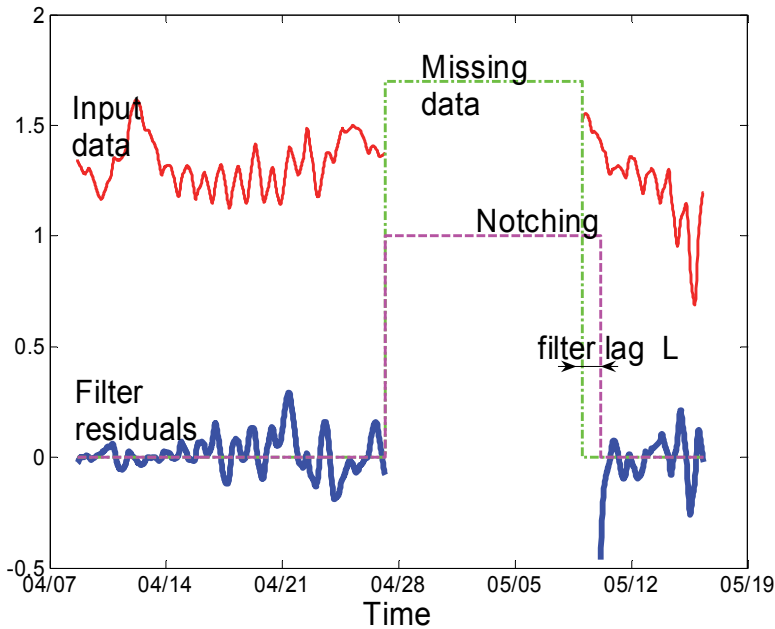


Fig. 15. Time notching in adaptive filtering.

Using the linear regression given with Eq. (3) on the data $y_{MA}(t)$, which is processed by the MA filter, we were able to explain only 6% of the variance, i.e. $R^2 = 0.06$ for $L = 55$. In Table 1 are shown the results of obtained $std(\hat{r}(t))$, for different values of the learning rate and the filter delay, when the adaptive filter given with Eqs. (2), (5) and (6) is used. Useful models were obtained when $std(\hat{r}(t)) < std(y_{MA}(t)) = 0.200$, and are shown in the shaded boxes in Table 1. As can be seen, the best results were obtained for $L = 55$, $\mu = 0.015$, which yielded to $R^2 = 0.37$. Small values of the learning rate, combined with small filter length, lead to unsatisfactory results. On the other hand, the learning becomes unstable if the filter length and learning rates are getting large.

μ/L	30	35	40	45	50	55
0.01	0.226	0.213	0.201	0.190	0.187	0.180
0.015	0.204	0.190	0.174	0.164	0.160	0.157
0.02	0.183	0.170	0.159	1.846	4.8e5	1.2e13

Table 1. *Std* of residuals for different values of learning rate μ and the filter length L , for MA filter. Useful models are in shaded boxes.

When the designed FIR filter was used on the same data $y_{MA}(t)$, we received the results shown in Table 2. As can be seen, the results given in Table 1 are better. However, if $L = 45$ and $\mu = 0.015$ (which yields to $R^2 = 0.24$), the best performance of the designed FIR filter is achieved.

μ/L	20	30	35	40	45	50	55
0.01	0.25	0.23	0.21	0.21	0.19	0.19	0.17
0.015	0.23	0.20	0.20	0.18	0.17	24.9	5e4
0.02	0.21	0.19	0.47	2e6	5e16	3e30	1e48

Table 2. *Std* of residuals for different values of learning rate μ and the filter length L , for designed FIR filter. Useful models are in shaded boxes.

Fig. 16 shows the residuals in time of the three useful adaptive filters, applied on data $y_{MA}(t)$ and processed by using the MA filter. The particular combination of the filter parameters was different, but the residuals showed similar behavior. Table 3 provides the time intervals when the relative residuals are larger than four standard deviations for the adaptive filter with $L = 55$, $\mu = 1.5 \times 10^{-2}$ (Pokrajac et al., 2007b). At the beginning of the learning process, the filter

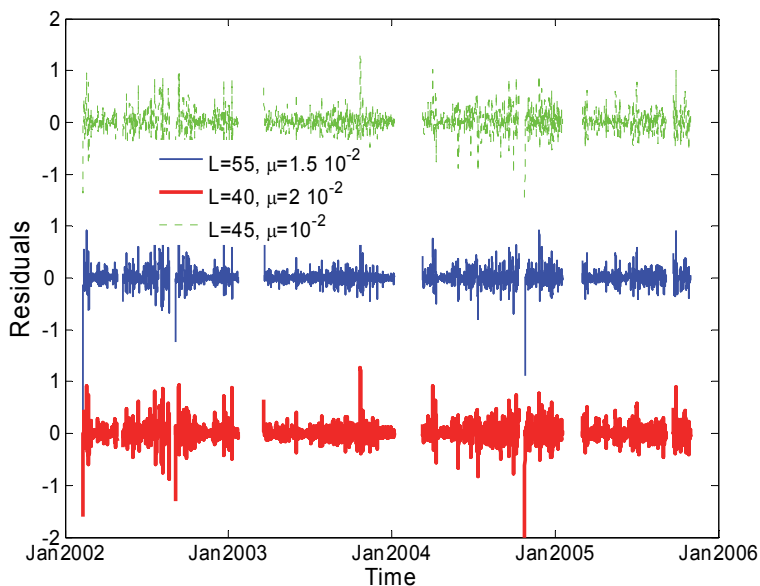


Fig. 16. Residuals of the adaptive filters applied on $y_{MA}(t)$ data using three different combinations of the learning rate and the filter length.

coefficients were not adapted fully, which caused the identified peaks. These peaks corresponded to observations from February 2002. In addition, there are two other peaks, corresponding to Sep 4, 2002, and Oct 25, 2004, which can be explained by a transient behavior after the notching interval.

Year	Begin		End		Note	Year	Begin		End		Note
	Date	Time	Date	Time			Date	Time	Date	Time	
2002	02/08	23:00	02/10	00:00	Learning	2004	04/01	15:30	04/02	18:30	
	02/17	06:30	02/18	01:30	Learning		04/09	17:00	04/09	18:00	
	07/20	00:30	07/20	09:00			05/29	12:30	05/29	19:00	
	08/07	01:00	08/07	03:30			07/12	20:30	07/13	02:00	
	08/07	16:30	08/07	17:30			09/19	00:00	09/19	06:00	
	08/21	14:30	08/21	21:30			09/29	20:30	09/30	06:30	
	09/04	16:00	09/04	19:30	Transient		10/05	15:30	10/05	20:30	
	09/11	15:00	09/12	11:00			10/25	03:00	10/25	06:00	Transient
	12/21	09:30	12/21	10:30			11/26	09:00	11/26	15:00	
2003	01/09	02:00	01/09	04:00		2005	09/27	20:00	09/28	04:30	
	10/22	12:00	10/24	09:00							

Table 3. Identified intervals of large residuals (adaptive filtering on $y_{MA}(t)$ data, $L = 55$, $\mu = 1.5 \text{ e-}2$)

6. Conclusion

We have described the application of the adaptive filtering for analyzing river hydrographic data. When determining the portion of the downstream data that is not influenced by the upstream data, the numerical results show that adaptive filtering is superior to linear regression.

7. Acknowledgement

This work was partially supported by the US Department of Commerce (award #NA06OAR4810164), NOAA (#NA06OAR4810164), NIH (#2P20RR016472-04), DoD/DoA (#45395-MA-ISP, #54412-CI-ISP, W81XWH-09-1-0062), and NSF (#0320991, CREST grant #HRD-0630388, #HRD-0310163).

8. References

- Adam, P. (2002). Saltmarshes in a time of change. *Environmental Conservation*, Vol. 29, No. 1, pp. 39-61
- Delaware Department of Natural Resources and Environmental Control (DNREC) (2002). *Technical Background Report Silver Lake Watershed*, pp. 81, Dover, Delaware, USA

- Devore, J. (2007). *Probability and Statistics for Engineering and the Sciences* (7th Edition), Duxbury Pr., ISBN 0-495-55744-7
- Fondriest Environmental Inc, <http://www.fondriest.com>
- Hull, C. & Titus, J. (1986). Greenhouse Effect, Sea Level Rise, and Salinity in the Delaware Estuary, *EPA report 230-05-86-010*
- Kennish, M. (2002). Environmental Threats and Environmental Future of Estuaries. *Environmental Conservation*, Vol. 29, No. 1, pp. 78-107
- Kraft, J., Hi-Il, T. & Khalequzzaman, M. (1992). Geologic and Human Factors in the Decline of the Tidal Salt Marsh Lithosome: the Delaware Estuary and Atlantic Coastal Zone. *Sedimentary Geology*, Vol. 80, pp. 232-246
- Mensing, M. (2005). SWMP Metadata, In: *National Estuarine Research Reserve Centralized Data Management Office*, http://cdmo.baruch.sc.edu/QueryPages/data_summary_progress.cfm
- Mitsch, W. & Gosselink, J. (2000). *Wetlands*, John Wiley and Sons Inc., ISBN 0-471-29232-x, New York
- Morris, R., Reach, I., Duffy, M., Collings, T. & Leafe, R. (2004). On the Loss of Saltmarshes in South-East England and the Relationship with *Nereis Diversicolor*. *Journal of Applied Ecology*, Vol. 41, pp. 787-791
- Pokrajac, D., Reljin, N., Reiter, M., Stotts, S. & Scarborough, R. (2007). Signal Processing of St. Jones River, Delaware Water Level Data, *Proceedings of ETRAN 2007*, Herceg Novi, Montenegro, Jun 2007
- Pokrajac, D., Reljin, N., Reiter, M., Stotts, S., Scarborough, R. & Nikolic, J. (2007). Adaptive Filters for Water Level Data Processing, *Proceedings of TELSIKS 2007*, Nis, Serbia, September 2007
- Potter, I.C., Manning, R. & Lonergan, N. (1991). Size, Movements, Distribution and Gonadal Stage of the Western King Prawn in a Temperate Estuary and Local Marine Waters. *Journal of Zoology*, Vol. 223, pp. 419-445
- Proakis, J. & Manolakis, D. (2006). *Digital Signal Processing* (4th Edition), Prentice Hall, ISBN 0-131-87374-1
- Rooth, J. & Stevenson, J. (2000). Sediment Deposition Patterns in *Phragmites Australis* Communities: Implications for Coastal Areas Threatened by Rising Sea-level. *Wetlands Ecology and Management*, Vol. 8, pp. 173-183
- Temmerman, S., Govers, G., Wartel, S. & Meire, P. (2004). Modeling Estuarine Variations in Tidal Marsh Sedimentation: Response to Changing Sea Level and Suspended Sediment Concentrations. *Marine Geology*, Vol. 212, pp. 1-19
- Widrow, B. & Stearns, S. (1985). *Adaptive Signal Processing*, Prentice Hall, ISBN 0-130-04029-0
- Zhang, M., Ustin, S., Rejmankova, E. & Sanderson, E. (1997). Monitoring Pacific Coast Salt Marshes Using Remote Sensing. *Ecological Applications*, Vol. 7, No. 3, pp. 1029-1053
- Zharikov, Y., Skilleter, G., Lonergan, N., Tarant, T. & Cameron, B. (2005). Mapping and Characterizing Subtropical Estuarine Landscapes using Aerial Photography and GIS for Potential Application in Wildlife Conservation and Management. *Biological Conservation*, Vol. 125, pp. 87-100

- Zharikov, Y. & Skilleter, G. (2004). Potential Interactions between Humans and non breeding Shorebirds on a Subtropical Intertidal Flat. *Australian Ecology*, Vol. 29, pp. 647-660

Nonlinear Adaptive Filtering to Forecast Air Pollution

Luca Mesin, Fiammetta Orione and Eros Pasero
Department of Electronics, Politecnico di Torino
Italy

1. Introduction

Air pollution is an important problem concerning the quality of living and health conditions of the population in urban areas (Pasero & Mesin, 2010). Indeed, the issue of air quality is now a major concern for many governments worldwide.

Since the early 1970s, the EU has been working to improve air quality by controlling emissions of harmful substances into the atmosphere, by improving fuel quality and by integrating environmental protection requirements into the transport and energy sectors. As the result of EU legislation, much progress has been made in tackling air pollutants. However air quality continues to cause problems. As an example, photochemical smog, particularly active in sunny days, regularly exceeds a safe limit over main European metropolitan areas. EU legislation established an hourly average of $180\mu\text{g}/\text{m}^3$ as the threshold of safe limit for ozone (Directive 2002/3/EC) beyond which authorities have to inform population.

According to the indications of the Environment Commissioner Janez Potočnik (Potočnik, 2010), air pollution is responsible for 370.000 premature deaths in EU each year. Airborne particles (e.g. Particulate Matter with diameter lower or equal to $10\mu\text{m}$, PM_{10}) are mainly present in pollutant emissions from industry, traffic and domestic heating. They can cause asthma, cardiovascular problems, lung cancer and premature death. The EU Directive 2008/50/EC requires Member States to ensure that certain limit values for PM_{10} are met. These limits, which were to be met by 2005, impose both an annual average concentration value ($40\mu\text{g}/\text{m}^3$), and a daily concentration value ($50\mu\text{g}/\text{m}^3$) which must not be exceeded more than 35 times per calendar year (European Environmental Bureau, 2005). The World Health Organization pointed out that USA traffic fatalities are over 40.000 per year, while air pollution claims 70.000 lives annually (Air Quality Guidelines, WHO, 2006). Experimental studies carried on Oslo urban site pointed out the temporal connection between cardiopulmonary inflammatory state of sick citizens and high concentration of PM in the area, especially in winter and early spring time (Schwarze P.E. et al, 2010). The impact of air pollution is on human health and on ecosystems also, as marine, freshwater, grassland, heat or forest ecosystem. The air pollution is directly linked to acidification of forests and water ecosystems, and eutrophication of soils and waters, leading to limited supply of oxygen in rivers and lakes. The benefits of air pollution abatement in the heat and grassland ecosystem, typical European farm environment, have been estimated as two hundred millions of euro each year saved (DeSmet et al, 2007). Indeed, air pollution not only affects

health care systems, as we cited above, but even local economy in terms of costs of medication, absences from work, and child and old people care expenses (Brown et al., 2002).

European environmental regulations demand the implementation of automatic procedures to prevent principal air pollutants concentration from exceeding specific alarm-thresholds in urban or suburban areas. Technology assists decision and action processes in controlling air pollution problem, implementing monitoring and forecasting techniques or automatic operating procedures in order to prevent the risk for the principal air pollutants to be above alarm thresholds. Increased efforts are made by governmental authorities for benchmarking approaches in data exchange and multi-model capabilities for air quality forecast and (near) real-time information systems, allowing information exchange between meteorological services, environmental agencies, and international initiatives.

2. Air pollution components

There are two main sources contributing to the formation of atmospheric particulate. One source can be attributed to natural phenomena such as soil erosion, volcanic eruption, presence of sea salt above oceans and other particles above water expanses. The other source is anthropogenic. Human activities impacting on the atmosphere, either directly (as primary emissions) or indirectly (as secondary emissions), bring unexpected change into the environmental processes of the ecosystem. The alteration of these natural processes is not easily absorbed by the ecosystem, bringing serious damages on it and adverse consequences on human health.

Pollutants are the elements that change the energy balance of the ecosystem. Depending on whether pollutants are introduced in the atmosphere, they can be considered primary or secondary. Primary pollutants are directly introduced in the air. They usually react with solar radiation and with other elements already in the atmosphere, changing their normal molecular structure. Secondary pollutants are not directly emitted as such, but are produced when primary pollutants react in the atmosphere (see the official site of EPA, United States Environmental Protection Agency, www.epa.gov).

From the indications of EPA, primary pollutant class includes carbon monoxides (CO), sulphur dioxides (SO₂), nitrogen oxides (NO_x), and hydrocarbons (Volatile Organic Compounds, VOCs). CO is a colorless and odorless gas, toxic for human health. CO is mainly discharged in the atmosphere by petrol fumes, as well as from steelworks and refineries, whose energy processes don't achieve complete carbon combustion. This pollutant is particularly dangerous because of its higher affinity for hemoglobin than oxygen. This could determine hypoxia in CO poisoning. High levels of CO generally occur in areas with heavy traffic congestion. SO₂ is a colorless, sour gas, which affects human respiratory system. The largest sources of SO₂ emissions are from fossil fuel combustion at power plants (73%) and other industrial facilities (20%). NO_x form quickly from volcanic and thunderstorm activity and from emission by vehicles and power plants. They are also used in fertilizers-manufacturing processes, because they can improve yield, stimulating the action of pre-existing nitrates in the ground. VOCs are organic compounds. They are divided into methane and not methane category. VOCs are not toxic, but they cause long-term chronic health effect, as reduction in visual or audible senses. Anthropogenic sources may be paints and coating, chlorofluorocarbons and fossil fuels combustion. These air pollutants are usually trapped near the ground beneath a layer of warm air, especially

under atmospheric stable conditions. Atmospheric stability occurs when the air near the surface is unable to rise, too heavy compared to the layer right above. It usually happens under high pressure atmospheric conditions. The high pressure above traps the pollutants in stagnant air surface layer (Sokhi, 2008).

Particulate matter (PM) is a complex mixture of extremely small particles and liquid droplets. Particle pollution is made up of a number of components, including acids (such as nitrates and sulfates), organic chemicals, metals, and soil or dust particles. PM can be primary or secondary (Perkins, 1974). Such powders are constituted from various pollutants, as lead, nickel, copper, cadmium and asbestos. They are easily inhaled and, depending on their dimension, they can reach and intoxicate various levels of breathing apparatus, down to alveolus, where the oxygen enhancement of hemoglobin occurs.

Ozone (O_3) is naturally present in the stratosphere, portion of the atmosphere around 12000 up to 45000 m asl, where it acts as a filter to block hazardous ultraviolet radiations. O_3 density is highest at around 25000 m asl in the ozonosphere, the low stratosphere. Thus the troposphere is not affected from the whole solar radiation load that impacts on the ozonosphere. The mechanism for high altitudes ozone formation was suggested by Sidney Chapman in Thirties (Chapman, 1932). Though the concentration of tropospheric ozone is controlled by stratosphere-troposphere exchange, in situ photochemical ozone formation plays an increasing role, especially in urban and industrial areas. At ground-level O_3 is a secondary pollutant. It is linked with adverse effects on the respiratory system (bronchitis, allergic asthma, irritations up to pulmonary edemas) and irritates mucous (Bard, 2010). At ground level ozone is created as a byproduct of the oxidation of carbon monoxide (CO) and hydrocarbons. These chemicals are called ozone precursors, and are often emitted simultaneously in the atmosphere via vehicle exhaust, industrial emissions, and other man-made sources. Photochemical ozone is catalyzed in the presence of sunlight by nitrogen oxides ($NO_x = NO + NO_2$) (Lelieveld, 2000) (Perkins, 1974). The process is called photolytic cycle for the oxides of nitrogen and causes photochemical ozone synthesis. It is summarized in the following system:



First equation of the system describes the absorption of solar radiation from NO_2 , absorption that modifies the visible wavelength of the compound. Thus the gas appears reddish-brown color over metropolitan areas, for example, in sunny, weather-steady days. Atmospheric stable conditions arise higher tropospheric O_3 density, because the high pressure traps the pollutants in stagnant air surface layer (Sokhi, 2008), inhibiting the environmental dispersion by means of wind advection and air turbulence (Geller, 2001). Second equation shows the extremely reactive O, which can react with hydrocarbon in order to have it involved in reactions that fuel nitrogen synthesis. M is any non reactive species that can absorb chemicals released in the reaction, in order to stabilize O_3 . The M body will probably be N_2 or O_2 , since these are the most available elements in the air. Third equation takes the process back to the beginning with a renewed NO_2 .

In a NO_x rich environment, the net reactions describing the oxidation of CO and VOC into respectively carbon dioxide (CO_2) and water vapor (plus carbonyl compounds), can be described as follows:



where CARB indicates Carbonyl Compounds.

In the chain of reactions resulting in (2) and (3), peroxy radicals (HO_2) are produced, which in turn oxidize nitric oxide (NO) to give nitrogen dioxide (NO_2). NO_2 is then photolyzed to give highly reactive atomic oxygen, and through reaction with oxygen a molecule of ozone (3).

3. Air pollution prediction models

Some mathematical models have been developed to simulate the time evolution of the concentration of air pollutants (Hass et al, 1995), determined by the atmospheric advection (the air transport due to wind) and by occurring chemical reactions. These models intend to simulate always more thoroughly the physical, chemical and dynamical processes which control the emission, transport and deposition of atmospheric trace species, in order to forecast their concentrations. The interaction between meteorological variables and pollutants concentration in the air is described. Reciprocal influence depends especially from wind advection and from wet and dry deposition. However sophisticated they are, these models are difficult to develop because of the wide variety of phenomena that have to be properly described for producing meaningful results, and the need of skilled personnel for setting up the experiments. These models have to take into account also the presence of urban areas, not only environmental variables, when providing local concentration values to be compared with safety thresholds.

A typical air quality physical based model is the European Air Pollution Dispersion (EURAD, http://www.eurad.uni-koeln.de/index_e.html), developed by the Rhenish Institute of Environmental Research of Germany. The EURAD system is able to simulate the evolution of pollutant tracers' in the atmosphere, from local to continental scale, by employing a nesting technique. It involves many atmospheric and chemical dedicated submodels. Physical laws that rule deterministic system models are described through systems of partial differential equations. The model can provide concentrations of pollutants of interest in the troposphere (portion of atmosphere from ground surface up to 10 km) over Europe and their transfer from air to soil or water ecosystems. Chaotic systems, as environmental one, are usually described as deterministic, so if their initial state was known exactly, their future state can be predicted. However, the precision with which the initial state can be studied strongly affects the capacity of modeling its future state (see Section 6.1).

A quicker and effective way to produce reliable outcomes of future concentrations of air pollutants can be achieved by employing black-box models for time-series forecasting (Sjöberg et al, 1994). These models are able to learn from environmental time-series, as meteorological and air pollution-concentrations datasets, the deterministic rules governing the process. No information regarding the complex physical and chemical dynamics involved in the process is used to develop such models. These methods allow real time and low cost local numerical predictions based on the analysis of extensive datasets. Black-box models perform a mathematical mapping between independent and dependent variables in a data set, by employing parameters which are not related to the physics of the complex system under study. These models can "learn" the underlying relationship between the

variables from past recordings, and they can produce forecasts of future output values accordingly. A model with specific topology and a finite number of free parameters is first chosen so that a large set of input-output relationships can be represented by proper choices of the parameters. As an example, Artificial Neural Networks (ANN) satisfy the universal approximation property, which means that they can approximate a nonlinear map as precisely as needed by increasing the number of parameters. Once chosen the model, an optimization algorithm is used to select the parameters in order to fit as close as possible to the data.

The major requirement for employing black box models for urban air quality forecasting is the availability of sufficiently long and accurate time-series for all the input and output variables involved in the study. This entails that when meteorological data are employed as input for predicting future concentrations of pollutants, the related observations should be recorded as close as possible to the air quality monitoring station, and with a similar sampling frequency.

Compared to complex physically based models, black box methods are faster, simpler and make use of a minor number of assumptions. Essentially, black box methods are based on the statistical assumption that conditions similar to those to be predicted can be found in a database of past measurements used to train the model (the training set). This requires the training set to be large enough that all specific events, including rare ones, are represented. Deterministic and black box models are not completely disjointed, even if their implementation is very different. The very fact that a certain predicted output can match a specific system condition is a deterministic hypothesis assumed by both techniques.

4. Determinism and stationarity

Air pollution data are affected by many time varying processes, like daily and seasonal atmospheric effects and further emissions. This poses the problem of stationarity. A time-series is formally defined as stationary if joint probabilities of finding it in certain conditions are independent of time (Papoulis, 1984). This means that a set of time-series should be recorded in the same conditions to define joint probabilities. Nevertheless, usually ergodicity of the series is assumed, and ensemble averages are substituted by time averages. Even under such an assumption, when we are working with finite series, inferring joint probabilities is a challenging task. It is in fact needed that all possible phenomena pertaining to the dynamics of the system can be properly represented in the time-series at disposal.

Another approach to discuss stationarity concerns the properties of the system from which the series is selected. Suppose that the environment can be described as a set of complicated, unknown deterministic rules. We could describe the dynamics of such a complex system with a vector field acting on an unknown number of state variables:

$$\frac{d\vec{x}}{dt} = \vec{F}(\vec{x}, t) \quad (4)$$

where $\vec{x}$ is the vector of state variables and $\vec{F}$ is a set of functions defining the vector field. A time-series, as the concentration of a pollutant sampled over time, can be considered as a measure extracted from the system. Such a measure can be modeled by an unknown function of the state variables:

$$y(t) = g(\tilde{x}(t)) \quad (5)$$

The system is said to be autonomous if the vector field does not change in time, which means:

$$\tilde{F} = \tilde{F}(\tilde{x}) \quad (6)$$

If the system is autonomous, we can expect that a sort of stationarity characterizes also the time-series that can be selected by making a measure. Nevertheless, we cannot expect that the statistical properties of an arbitrary measure are neither constant in time nor that they have only periodic trends. Indeed, the complex system determining the evolution of the pollutant of interest is nonlinear and can show chaotic behaviour (Kantz & Schreiber, 1997). Moreover, when working with finite series, we have again problems in understanding if the series can be assumed as extracted from a stationary, autonomous deterministic system or not. Indeed, we can say that the concept of stationarity for a finite series depends on the considered time period (Kantz & Schreiber, 1997) and on its relation with the time constants of the considered system. If the investigated time period is short enough that the system did not manifest any trend or evolution within it, the series appears to be stationary. On the other hand, if the system is affected by a trend which has a significant effect on the time period considered, the time-series can not be considered stationary even if it was indeed extracted from an autonomous system. In such a case, extending the period of investigation, the system could manifest a deterministic variation (as in the case of seasonal effects) demonstrating a stable stationary dynamics (called ciclo-stationary, in case of periodic variations). Thus, we could say that, in order to consider stationary a time-series, it should be measured over a time range which is much longer than the longest characteristic time scale which is relevant for the evolution of the system.

If the system has a deterministic flow which is constant in time (i.e., it is autonomous), the time-series extracted from it have dynamics determined by deterministic rules. In such a case we could, at least in principle, use past events to learn the dynamics of the considered time-series. Assuming that the space of possible events is densely covered by the trajectory $y(t)$ used to train the model, we could build a good predictor (if measurement noise is negligible) of future events by exploring what happened in the past in situations similar to the investigated one.

Sometimes we are interested in studying a time-series extracted from a system which is not stationary at all. This is the case in which we are trying to predict air pollution dynamics close to a place in which a new polluting plant is installed, for example. Another instance is the case in which a system for air pollution prediction is trained in a place and then it is moved in a nearby region. A further example could be the variation of the data introduced by a calibration of the sensor in which an offset is eliminated by adding a constant. In all these cases the dynamics of the environment from which the air pollution data are extracted or the measuring apparatus undergo an abrupt change. Thus, the prediction system is required to *adapt* to the new data, as limited information can be found on the training examples contained in the database (Widrow & Winter, 1988).

5. Spatial and temporal adaptation

From the discussion above, we conclude that adaptation can be useful at least for two reasons.

First, when we need a data-driven model to perform reliable predictions on time-series measured from an environment which is stationary. Data-driven models adapt to the specific data set or to the specific environment from which the air pollution data are recorded. The prediction model can be trained on past measurements and free parameters can be fixed, minimising a cost function which penalizes the prediction mistakes on the training data. The same methodology can be used to fit the same model to different data extracted from different environments. Nevertheless, the optimal choice of the parameters of the model will be different when considering data collected from different places. We could say that the forecast tool built in this way is adaptive in the sense that it is able to adapt to and it is determined by the specific data to be processed. We'll use the term "spatial adaptation" to indicate this property of the filter.

Second, when we need an adaptive filtering approach to perform tried and true predictions on data recorded from a non stationary environment. In such a case we need a filter with parameters that are not fixed, but can vary when the environment undergoes a change. The parameters of the model will be changed dynamically reducing a cost function, which is not built using only training data but also using the new acquired data (as soon as they are available). In this way the filter can learn new dynamics of the system, which were not represented in the training data. We can define this second feature of the adaptive filter as "temporal adaptation", as the filter is not fixed in time, but is defined by parameters which are dynamically changed to adapt to temporal variations of the data.

6. Nonlinear local adaptive prediction methods

6.1 Introduction

Many works in the literature (Foxall et al. 2001, Marra et al 2003) indicate that the air pollution time-series are nonlinear. Let us assume that the air pollution series is a measure taken from a system that can be described with a vector flow, as shown by equations (4) and (5). An autonomous, nonlinear system with more than two state variables can show chaotic behavior (Strogatz, 1994). This is surely the case for air pollution applications. This implies that similar past events could evolve into very different conditions. Thus, we expect that the dynamics of the system from which the time-series was extracted is sensitive to small differences amongst initial conditions. Such small differences can be amplified by the nonlinear dynamics toward very different final conditions.

The Lyapunov exponent is a measure of the rate of exponential divergence of trajectories starting from neighboring points (Kantz & Schreiber, 1997). The prediction horizon is usually very short and related to the inverse of the Lyapunov exponent (Haykin, 1999). Indeed, initial conditions are known with some uncertainty due to additive noise, always present on experimental, real data. The inaccuracy of measuring the initial conditions is amplified exponentially for a chaotic system, determining a larger and larger imprecision on predicted values.

Nonlinear dynamics could be poorly approximated using linear methods, so that nonlinear models must be built. Both meteorological and air pollution data are useful to adapt the model and learn how to predict pollutants concentration. The following steps are in general used to develop an adaptive local prediction tool (Pasero & Mesin, 2010):

- a. measurement of information on the investigated environment through specific sensors;
- b. pre-processing of raw time-series data (to reduce noise content or to extract optimal features);

- c. selection of a model representing the dynamics of the investigated process;
- d. choice of optimal parameters of the model in order to minimize a cost function measuring the error in forecasting the data of interest; the mean square error is usually chosen as cost function;
- e. validation of the prediction, which possibly guides the selection of an alternative model.

The predicted concentration is the result of a nonlinear processing of the possibly linearly filtered input data.

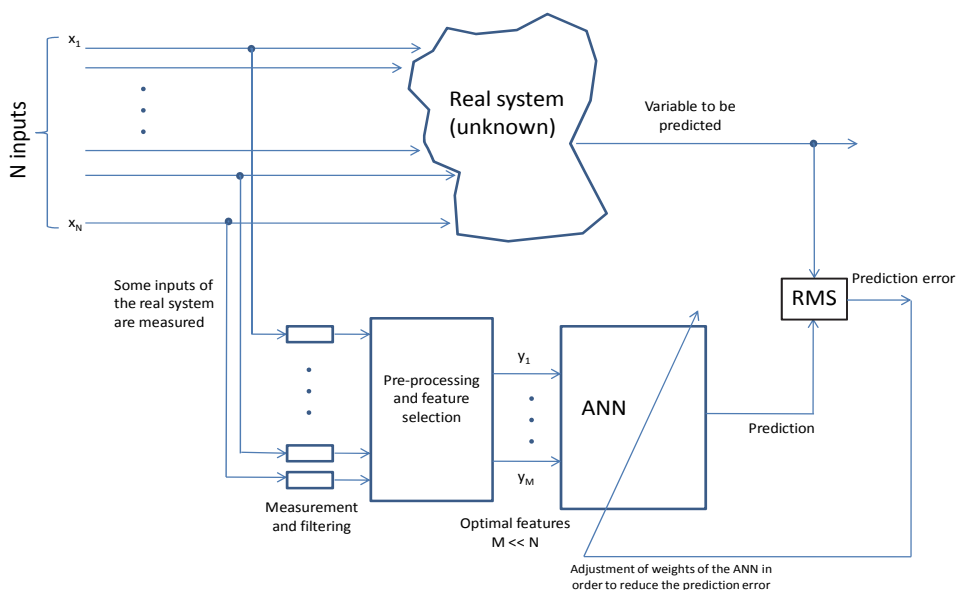


Fig. 1. Sketchy representation of the training of the nonlinear filter. The details of the real system are neglected. Some of the state variables of the real system are measured and linearly filtered in order to reduce noise or to remove undesired trends. The measured variables are then processed (e.g. by PCA or ICA) and essential features are selected. The ANN processes such features and produces a prediction which is compared with the actual value of the variable. The prediction error guides the update of the weights (i.e. free parameters of the ANN) in order to optimize the prediction.

In this chapter we focus on ANNs, which have been often used as a prognostic tool for air pollution (Perez et al., 2000; Božnar et al., 2004; Cecchetti et al., 2004; Slini et al., 2006; Karatzas et al., 2008). We discuss in the following an application of ANN to the regression and system identification analysis in order to predict the dynamics of local ozone concentration. Ozone is an important pollutant due to its extensive occurrence in EU territory (not only metropolitan) and many efforts have been played up to provide appreciable data sets.

A nonlinear filter based on an ANN is designed updating adaptively the weights in order to minimize the prediction error (Figure 1). A short introduction of methods for features

selection is discussed below, together with an elementary description of the properties of ANNs and methods for training them and for system identification.

6.2 Features selection

The selection of optimal features that are going to be used as the input of the forecasting tool is of paramount importance. Selecting good features is useful to facilitate data visualization and understanding, to reduce the measurement and storage requirements, to bring down the noise content, to defy the curse of dimensionality, to improve prediction or classification performance.

Simple methods to select the best input variables are based on adding one feature at a time and exploring the approximation error or starting with a complex network with many input variables and pruning it removing the variables with lower influence. For example, a pruning algorithm is described in (Corani, 2005) to develop an air pollution predictor model. A model considered large enough to capture the desired input-output relationship was initially tested. Then a measure of the contribution of each variable to the network efficiency is computed each iteration. A pruning algorithm removes the less influential feature from the network architecture. The procedure continues iteratively until just a unique input data remains in the network. The optimal brain surgeon (OBS) algorithm was used to determine the variable to be eliminated, based on estimate of the increase of error on the training set resulting from the removal of each input variable from the network architecture. The network providing the minimum error on a validation set is finally chosen.

Input features may also be pre-processed before being applied to the network. Linear filters may be used in order to remove noise or a trend from the input data. In order to remove redundant information, but still preserving the main energy or information contained in the data, the input variables can be linearly combined using Principal Component Analysis (PCA) or Independent Component Analysis (ICA). The components obtained by such linear combinations can be ranked in order of the energy or of the amount of information they preserve of the input variables. The components preserving the maximum energy or information are usually selected, whereas the others are neglected, obtaining data compression. Moreover, in case in which the neglected components have high noise content, such a data compression allows also improving the signal to noise ratio.

Specifically, PCA determines the amount of redundancy in the measured data by cross-correlation and estimates a linear transformation W , which reduces this redundancy to a minimum. The first principal component is the direction of maximum variance in the data. The other components can be obtained iteratively searching for the directions of maximum variance in the space of data orthogonal to the subspace spanned by already determined principal directions. As an alternative, principal components and their energy content can be obtained by solving the eigenvalue problem for the cross-correlation matrix:

$$\hat{\mathbf{R}}_{xx} = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mm} \end{bmatrix} \quad (7)$$

where r_{ij} is the correlation between the i^{th} and the j^{th} data (Pasero & Mesin, 2010; Mesin et al., 2011).

ICA extracts from input data some features which are statistically independent, preserving the total information content and minimizing the mutual information. ICA performs a linear

transformation between the data and the features to be determined. Central limit theorem guarantees that a linear combination of variables has a distribution that is “closer” to a Gaussian than that of any individual variable. Assuming that the features to be estimated are independent and non-Gaussian (but possibly one of them), the independent components can be determined by applying to the data the linear transformation that maps them into features with distribution which is as far as possible from Gaussian. Thus a measure of non-Gaussianity is used as an objective function to be maximized by a given numerical optimization technique with respect to possible linear transformations of the input data. Different methods have been developed considering different measures of Gaussianity. The most popular methods are based on measuring kurtosis, negentropy or mutual information (Hyvarinen, 1999; Mesin et al., 2011).

Another interesting algorithm was proposed in (Koller and Sahami, 1996). The mutual information of the features is minimized (in line with ICA approach), using a backward elimination procedure where at each state the feature which can be best approximated by the others is eliminated iteratively (see Pasero & Mesin, 2010 for an air pollution application of this method). Thus in this case the mutual information of the input data is explored, but there is no transformation of them (as done instead by ICA).

A further method based on mutual information is that of looking for the optimal input set for modelling a certain system selecting the variables providing maximal information on the output. Thus, in this case the information that the input data have on the output is explored, and features are again selected without being transformed or linearly combined. However, selecting the input variables in term of their mutual information with the output raises a major redundancy issue. To overcome this problem, an algorithm was developed in (Sharma, 2000) to account for the interdependencies between candidate variables exploiting the concept of Partial Mutual Information (PMI). It represents the information between a considered variable and the output that is not contained in the already selected features. The variables with maximal PMI with the output are iteratively chosen (Mesin et al, 2010).

Many of the methods indicated above for feature selections are based on statistical processing of the data, requiring the estimation of probability density functions from samples. Different methods have been proposed to estimate the probability density function (characterizing a population), based on observed data (which is a random sample extracted from the population). Parametric methods are based on a model of density function which is fit to the data by selecting optimal values of its parameters. Other (not parametric) methods are based on a rescaled histogram. Kernel density estimation or Parzen method (Parzen, 1962; Costa et al., 2003) was proposed as a sort of a smooth histogram.

A short introduction to feature selection and probability density estimation is discussed in (Pasero & Mesin, 2010).

6.3 ANN

Our approach exploits ANNs to map the unknown input-output relation in order to provide an optimal prediction in the least mean squared (LMS) sense (Haykin, 1999). ANNs are biologically inspired models consisting of a network of interconnections between neurons, which are the basic computational units. A single neuron processes multiple inputs and produces an output which is the result of the application of an activation function (usually nonlinear) to a linear combination of the inputs:

$$y_i = \varphi_i \left(\sum_{j=1}^N w_{ij} x_j + b_i \right) \quad (8)$$

where $\{x_j\}$ is the set of inputs, w_{ij} is the synaptic weight connecting the j^{th} input to the i^{th} neuron, b_i is a bias, $\varphi_i(\cdot)$ is the activation function, and y_i is the output of the i^{th} neuron considered. Fig. 2A shows a neuron. The synaptic weights w_{ij} and the bias b_i are parameters that can be changed in order to get the input-output relation of interest.

The simplest network having the universal approximation property is the feedforward ANN with a single hidden layer, shown in Fig. 2B.

The training set is a collection of pairs $\{\bar{x}_k, d_k\}$, where $\bar{x}_k$ is an input vector and d_k is the corresponding desired output. The parameters of the network (synaptic weights and bias) are chosen optimally in order to minimize a cost function which measures the error in mapping the training input vectors to the desired outputs. Usually, the mean square error is considered as cost function:

$$E(\bar{w}, \bar{b}) = \sum_{i=1}^N (d_i - y(\bar{x}_i; \bar{w}, \bar{b}))^2 \quad (9)$$

Different optimization algorithms were investigated to train ANNs. The main problems concern the velocity of training required by the application and the need of avoiding the entrapment in a local minimum. Different cost functions have also been proposed to speed up the convergence of the optimization, to introduce a-priori information on the nonlinear map to be learned or to lower the computational and memory load. For example, in the sequential mode, the cost function is computed for each sample of the training set sequentially for each step of iteration of the optimization algorithm. This choice is usually preferred for on-line adaptive training. In such a case, the network learns the required task at the same time in which it is used by adjusting the weights in order to reduce the actual mistake and converges to the target after a certain number of iterations. On the other hand, when working in batch mode, the total cost defined on the basis of the whole training set is minimized.

An ANN is usually trained by updating its free parameters in the direction of the gradient of the cost function. The most popular algorithm is backpropagation, a gradient descent algorithm for which the weights are updated computing the gradient of the errors for the output nodes and then propagating backwards to the inner nodes. The Levenberg-Marquardt algorithm (Marquardt, 1963) was also used in this study. It is an iterative algorithm to estimate the synaptic weights and the bias in order to reduce the mean square error selecting an update direction which is between the ones of the Gauss-Newton and the steepest descent methods. The optimal update of the parameters $\bar{\delta}_{opt}$ is obtained solving the following equation:

$$(J^T J + \lambda I) \bar{\delta}_{opt} = J^T (\bar{d} - y(\bar{x}, \bar{W})) \quad \text{where} \quad J_i = \frac{\partial y(\bar{x}_i, \bar{W})}{\partial \bar{W}}, \quad \bar{W} = (\bar{w}, \bar{b}) \quad (10)$$

where λ is a regularization term called damping factor. If reduction of the square error E is rapid, a smaller damping can be used, bringing the algorithm closer to the Gauss-Newton

method, whereas if an iteration gives insufficient reduction in the residual, λ can be increased, giving a step closer to the gradient descent direction. A few more details can be found in (Pasero & Mesin, 2010).

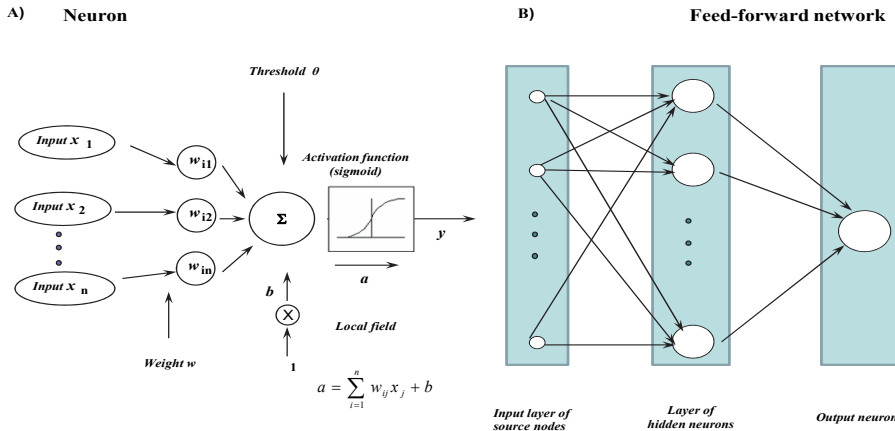


Fig. 2. A) Sketchy representation of an artificial neuron. B) Example of feedforward neural network, with a single hidden layer and a single output neuron. It is the simplest ANN topology satisfying the universal approximation property.

Due to the universal approximation property, the error in the training set can be reduced as much as needed by increasing the number of neurons. Nevertheless, it is not needed to follow also the noise, which is always present in the data and is usually unknown (even no information about its variance is assumed in the following). Thus, reducing the approximation error beyond a certain limit can be dangerous, as the ANN learns not only the determinism hidden within the data, but also the specific realization of the additive random noise contained in the training set, which is surely different from the realization of the noise in other data. We say that the ANN is overfitting the data when a number of parameters larger than those strictly needed to decode the determinism of the process are used and the adaptation is pushed so far that the noise is also mapped by the network weights. In such a condition, the ANN produces very low approximation error on the training set, but shows low accuracy when working on new realizations of the process. In such a case, we say that the ANN has poor generalization capability, as cannot generalize to new data what it learns on the training set. A similar problem is encountered when too much information is provided to the network by introducing a large number of input features. Proper selection of non redundant input variables is needed in order not to decrease generalization performance (see Section 6.2).

Different methods have been proposed to choose the correct topology of the ANN that provides a low error in the training data, but still preserving good generalization performances. In this work, we simply tested more networks with different topology (i.e., a different number of neurons in the hidden layer) on a validation set (i.e., a collection of pairs of inputs and corresponding desired responses which were not included in the training set). The network with minimum generalization error was chosen for further use.

6.4 System identification

For prediction purposes, time is introduced in the structure of the neural network. For immediately further prediction, the desired output y_n at time step n is a correct prediction of the value attained by the time-series at time $n+1$:

$$y_n = \hat{x}_{n+1} = \varphi(\bar{w} \cdot \bar{x} + b) \quad (11)$$

where the vector of regressors $\bar{x}$ includes information available up to the time step n . Different networks can be classified on the basis of the regressors which are used. Possible regressors are the followings: past inputs, past measured outputs, past predicted outputs and past simulated outputs, obtained using past inputs only and the current model (Sjöberg et al., 1994). When only past inputs are used as regressors for a neural network model, a nonlinear generalization of a finite impulse response (FIR) filter is obtained (nonlinear FIR, NFIR). A number of delayed values of the time-series up to time step n is used together with additional data from other measures in the nonlinear autoregressive with exogenous inputs model (NARX). Regressors may also be filtered (e.g., using a FIR filter). More generally, interesting features extracted from the data using one of the methods described in Section 2 may be used. Moreover, if some of the inputs of the feedforward network consist of delayed outputs of the network itself or of internal nodes, the network is said to be recurrent. For example, if previous outputs of the network (i.e., predicted values of the time-series) are used in addition to past values of input data, the network is said to be a nonlinear output error model (NOE). Other recursive topologies have also been proposed, e.g. a connection between the hidden layer and the input (e.g. the simple recurrent networks introduced by Elman, connecting the state of the network defined by the hidden neurons to the input layer; Haykin, 1999). When the past inputs, the past outputs and the past predicted outputs are selected as regressors, the model is recursive and is said to be nonlinear autoregressive moving average with exogenous inputs (NARMAX). Another recursive model is obtained when all possible regressors are included (past inputs, past measured outputs, past predicted outputs and past simulated outputs): the model is called nonlinear Box Jenkins (NBJ).

7. Example of application

7.1 Description of the investigated environment and of the air quality monitoring station

To coordinate and improve air quality monitoring, the London Air Quality Network (LAQN) was established in 1993, which is managed by the King's College Environmental Research Group of London. Recent studies commissioned by the local government Environmental Research Group (ERG) estimated that more than 4300 deaths are caused by air pollution in the city every year, costing around £2bn a year. Air pollution persistence or dispersion is strictly connected to local weather conditions. What are typical weather conditions over London area? Precipitation and wind are typical air pollution dispersion factor. Nevertheless rainy periods don't guarantee optimal air quality, because rain only carries down air pollutants, that still remain in the cycle of the ecosystem. Stable, hot weather is typical air pollution persistence factor. From MetOffice reports we deduce rainfall is not confined in a special season. London seasons affect the intensity of rain, not the incidence. Snow is not very common in London area. It is most

likely when Arctic and Siberian winds occur from north, north-east. In the summer there are usually a few days of particularly hot weather in London. They are often followed by a thunderstorm.

In this study, we used the air quality data from the LAQN Harlington station situated in the Hillingdon borough. London Hillingdon-Harlington (LHH, 51,488 lat, -0, 416 lon) is an urban background air quality station located in Heathrow Airport zone. The station is north-east the main Heathrow runway, around 21 km west far from London City. The borough of Hillingdon is on the outskirts of the densely populated London area and its air quality is affected by the airport and road traffic, urban heating and suburb manufacturing. There are some expanses of water, small lakes, and green zones around 10 km west from LHH. The area is plain. CO, NO, NO₂ and NO_x, O₃, PM₁₀ and PM_{2.5} are the pollutants species monitored. Meteorological data was obtained by a nearby LAQN monitoring station located in the Heathrow Airport (LHA).

LHA-LHH zone should experience ozone, nitrogen oxides and carbon monoxide pollution. As we mentioned above, nitrogen oxides are in fact synthesized from urban heating, manufacturing processes and motor vehicle combustion, especially when revs are kept up, over fast-flowing roads and motorways. There are a motorway (A4) at about 2 km north from Heathrow runway and another perpendicular fast-flowing road (M4). Nitrogen oxides, especially in the form of nitrate ions, are used in fertilizers-manufacturing processes, to improve yield by stimulating the action of pre-existing nitrates in the ground. As we mentioned above, the study area is on the borderline of a green, cultivated zone west from London metropolitan area. Carbon monoxide, a primary pollutant, is directly emitted especially from exhaust fumes and from steelworks and refineries, whose energy processes don't achieve complete carbon combustion.

7.2 Neural network design and training

The study period ranged from January 2004 to December 2009, though it was reduced to only those days where all the variables employed in the analysis were available. All data considered, 725 days, were at disposal for the study and 16 predictors were selected: daily maximum and average concentration of O₃, up to three days before (6 predictors); daily maximum and average concentration for CO, NO, NO₂ and NO_x of the previous day (8 predictors); daily maximum and daily average of solar radiation of the previous day (2 predictors). Predictors have been selected according to literature (Corani, 2005; Lelieveld & Dentener, 2000), completeness of the recorded time-series, and a preliminary trial and error procedure. Efficient air pollution forecasting requires the identification of predictors from the available time-series in the database and the selection of essential features which allow obtaining optimal prediction. It is worth noticing that, by proceeding by trials and errors, the choice of including O₃ concentration up to three days before was optimal. This time range is in line with that selected in (Kocak, 2000), where a daily O₃ concentration time-series was investigated with nonlinear analysis techniques and the selected embedding dimension was 3.

Data were divided into training, validation and test set.

The training set is used to estimate the model parameters. The first 448 days and those with maximum and minimum of each selected variable were included in the training set. Different ANN topologies were considered, with number of neurons in the hidden layer

varying in the range 3 to 20. The networks were trained with the Levenberg-Marquardt algorithm in batch mode. Different numbers of iterations (between 10 and 200) were used for the training.

The validation set was used to compute the generalization error and to choose the ANN with best generalization performances. The validation data set was made of the 277 remaining days, except for 44 days. The latter represents the longest uninterrupted sequence and it has been therefore used as test dataset (see Section 7.3).

The network with best generalization performances (i.e., minimum error in the validation set) was found to have 4 hidden neurons, and it was trained for 30 iterations. Once the optimal ANN has been selected, it is employed on the test data set. The test set is used to run the chosen ANN on previously unseen data, in order to get an objective measure of its generalization performances.

Another neural network was developed from the first one, changing dynamically the weights using the new data acquired during the test. The initial weights of the adapted ANN are those of the former ANN, selected after the validation step. The adaptive procedure is performed using backpropagation batch training. For the prediction of the (n+1) observation in the data set, all the previous n-data patterns in test data set are used to update the initial weights. Also this neural network was employed on the test data set, as shown in the following section.

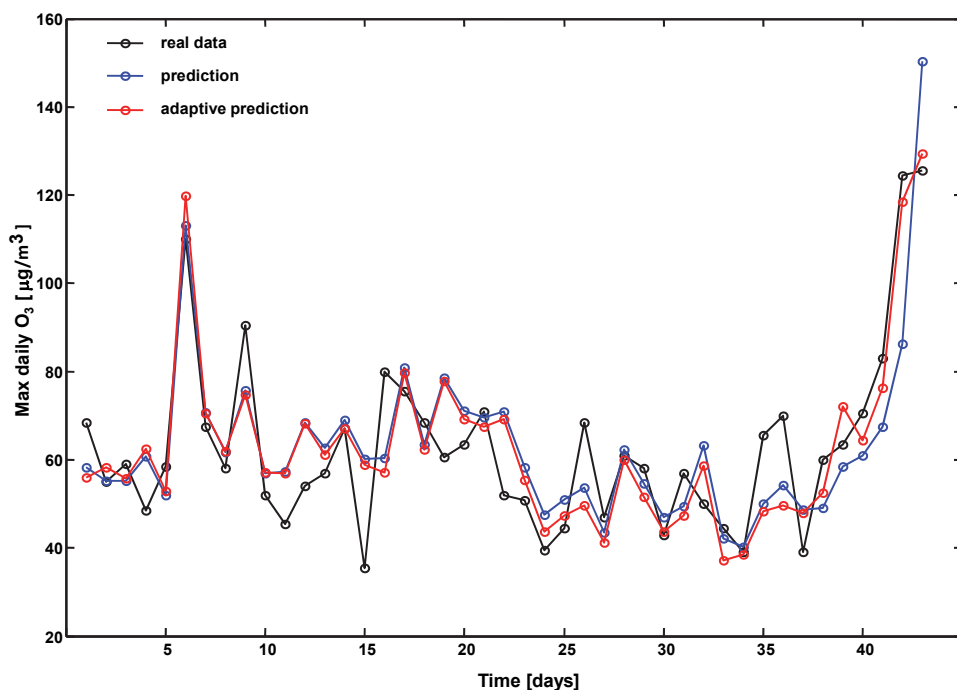


Fig. 3. Application of two ANNs to the test set.

7.3 Results

Two different ANNs are considered, as discussed in Section 7.2. The first one has weights which are fixed. This means that the network was adapted to perform well on the training set and then was applied to the test set. This requires the assumption that the system is stationary, so that no more can be learned from the new acquired data. Such an ANN is spatially adapted to the data (referring to Section 5). The second network has the same topology as the first one, but the weights are dynamically changed considering the new data which are acquired. The adaptation is obtained using backpropagation batch training, considering the data of the test set preceding the one to be predicted. Thus, temporal adaptation is used (refer to Section 5).

The results of the first ANN on the test data set are shown in Figure 3 and in Table 1 in terms of linear correlation coefficient (R^2), root mean square error (RMSE) and ratio between the RMSE and the data set standard deviation (STD). It emerges that the performances on the training and validation data set are generally good; the RMSE is below half the standard deviation of the output variable and R^2 around 0.90. A drop in the performances is noticeable on the test data set, meaning that some of the dynamics are not entirely modeled by the ANN. Performing a temporal adaptation by changing the ANN weights, a slight improvement in prediction performances is noticed as shown in Table 1. The adapted network is obtained using common backpropagation as described before. The optimal number of iterations and the adaptive step were respectively found to be 14 and 0.0019, low enough to prevent instabilities due to overtraining.

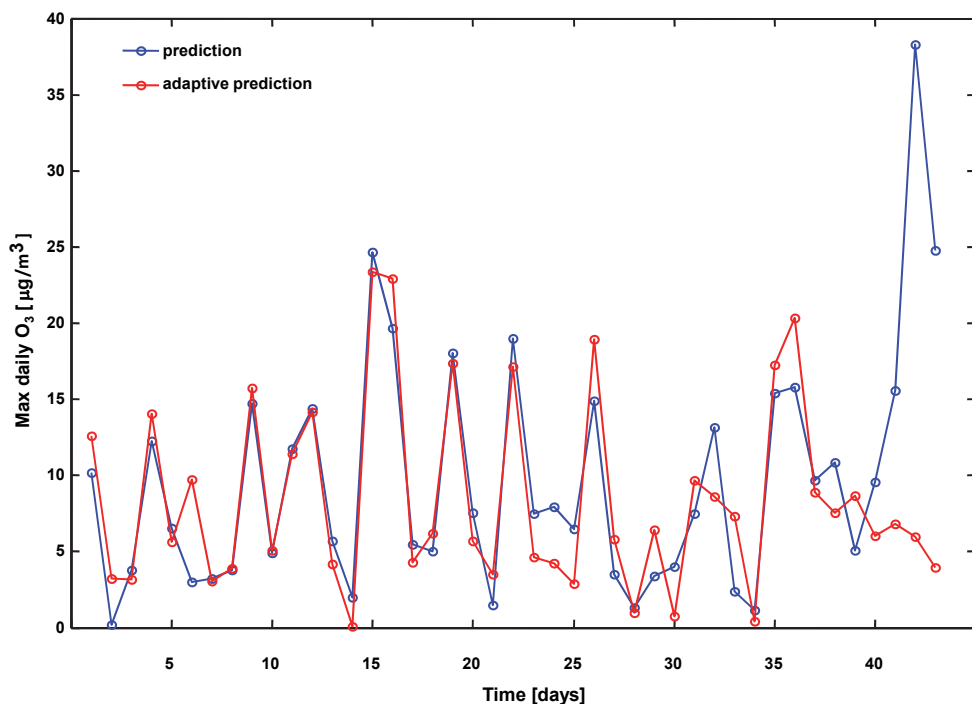


Fig. 4. Absolute error of two ANNs when applied to the test set.

DATASET	RMSE [$\mu\text{g}/\text{m}^3$]	RMSE/STD	R ²
TRAINING SET	11.19	0.45	0.89
VALIDATION SET	11.41	0.41	0.91
TEST SET (FIXED WEIGHTS)	12.35	0.62	0.79
TEST SET (TEMPORAL ADAPTATION)	10.42	0.52	0.86

Table 1. Results of application of two ANNs to the data

From the comparison of predictions in Figure 3 and most notably from the plot of the absolute errors in Figure 4, it can be seen that the adaptive network performs sensibly better towards the end of the data set, i.e. when more data is available for the adaptive training. The accuracy of the ANN model can also be compared to the performances of the persistence method, shown in Table 2. The persistence method assumes that the predicted variable at time $n+1$ is equal to its value at time n . Although very simple, this method is often employed as a benchmark for forecasting tools in the field of environmental and meteorological sciences. For example, many different nonlinear predictor models were compared to linear ones and to the persistence method in forecasting air pollution concentration in (Ibarra-Berastegi et al, 2009). Surprisingly, in many cases persistence of level was not outperformed by any other more sophisticated method. Concerning this study, however, it can be seen comparing the results in Tables 1 and 2 that the considered ANNs outperforms the persistence method in each data set considered, with improvements in terms of RMSE ranging from around 40% to 50% .

DATASET	RMSE [$\mu\text{g}/\text{m}^3$]	R ²
TRAINING SET	19.82	0.66
VALIDATION SET	20.66	0.70
TEST SET	19.85	0.43

Table 2. Results of application of the persistence method to the data

7.4 Discussion

Two predictive tools for tropospheric ozone in urban areas have been developed. The performances of the models are found to be satisfactory both in terms of absolute and relative goodness-of-fit measures, as well as in comparison with the persistence method. This entails that the choice of the exogenous predictors (CO, nitrogen oxides, and solar radiation) was appropriate for the task, though it would be interesting to assess the change in performances that can be obtained by including other reactants (VOC) involved in the formation of tropospheric ozone.

In terms of model efficiency, it has been shown that further adaptive training on the test data set may result in increased accuracy. This could indicate that the dynamics of the environment is not stationary or, more probably, that the training set was not long enough for the ANN model to learn the dynamics of the environment. However, a thorough analysis of the benefits of adaptive training can be carried out on longer uninterrupted time-

series. For instance, such a study could give insights on the optimal number of previous data patterns to be used for the adaptive steps.

Adaptive training could also be employed to improve pollutant prediction on nearby sampling stations. Since the development of air quality forecasting tools with ANNs is a data-driven process, the quantity as well as the quality of the information at disposal is of primary importance. This may severely hinder the development of accurate local models for recently installed sampling stations, or for those nodes of the monitoring network where the amount of missing/non validated data is considerable. To overcome these problems, one could first develop an ANN model for another node of the network, close enough to the one of interest and with a sufficient number of reliable data for training and validation. Once the major dynamics of the process are mapped into the ANN architecture using the former dataset, the model can be fine tuned with adaptive training to match the conditions of the chosen node, such as different reactants concentrations or local meteorological conditions.

8. Final remarks and conclusion

Many applications are not feasible to be processed with static filters with a fixed transfer function. For example, noise cancellation, when the frequency of the interference to be removed is slightly varying (e.g., power line interference in biomedical recordings), cannot be performed efficiently using a notch filter. For such problems, the filter transfer function can not be defined a-priori, but the signal itself should be used to build the filter. Thus, the filter is determined by the data: it is data-driven.

Adaptive filters are constituted by a transfer function with parameters that can be changed according to an optimization algorithm minimizing a cost function defined in terms of the data to be processed. They found many applications in signal processing and control problems like biomedical signal processing (Mesin et al., 2008), inverse modeling, equalization, echo cancellation (Widrow et al, 1993), and signal prediction (Karatzas et al, 2008; Corani, 2005).

In this chapter, a prediction application is proposed. Specifically, we performed 24-hour maximal daily ozone-concentrations forecast over London Heathrow airport (LHA) zone. Both meteorological variables and air pollutants concentration time-series were used to develop a nonlinear adaptive filter based on an artificial neural network (ANN). Different ANNs were used to model a range of nonlinear transfer functions and classical learning algorithms (backpropagation and Levenberg-Marquardt methods) were used to adapt the filter to the data in order to minimize the prediction error in the LMS sense. The optimal ANN was chosen with a cross-validation approach. In this way, the filter was adapted to the data. We indicated this process with the term "spatial adaptation". Indeed, the specific choice of network topology and weights was fit to the data detected in a specific location. If prediction is required for a nearby region, the same adaptive methodology may be applied to develop a new filter based on data recorded from the new considered region. Thus, a specific filter is adapted to the data of the specific place in which it should be used. Hence, in a sense, the filter is specific to the spatial position in which it is used. For this case, the concept of "spatial adaptation" was introduced in order to stress the difference with respect to what can be called "temporal adaptation". Indeed, once the filter is adapted to the data, two different approaches can be used to forecast new events: the transfer function of the filter could be fixed (which means that the weights of the ANN are fixed) and the prediction tool can be considered as a static filter; on the other hand, the filter could be dynamically updated considering the new

data. In the latter case, the filter has an input-output relation which is not constant in time, but it is temporally adapted exploiting the information contained in the new detected data. Both approaches have found applications in the literature. For example, in (Rusanovskyy et al. 2007), video compression coding was performed both within single frames using a “spatial adaptation” algorithm and over different frames using a “temporal adaptation” method. Both spatial and temporal adaptations were also implemented here for the representative application on ozone concentration forecast. The “spatial adaptation” of the ANN (on the basis of the training set) was sufficient to obtain prediction performances that overcome those of the persistence method when the filter was applied to the new data contained in the test set. This indicates that the training was sufficient for the filter to decode some of the determinism that relates the future ozone concentration to the already recorded meteorological and air pollution data. Moreover, applying to new data the same deterministic rules learned from the database used for training, the predictions are reliable. Nevertheless, when the filter was updated based on the new data (within the “temporal adaptation” framework), the performances were still greater. This indicates that new information was contained in the test data. The same outcome is expected in all cases in which the investigated system is not stationary or when it is stationary, but the training dataset did not span all possible dynamics.

The specific application presented in this work showed the importance of having consistent datasets in order to implement reliable tools for air quality monitoring and control. These datasets have to be filled with information from weather measurement stations (equipped with solar radiation, temperature, pressure, wind, precipitation sensors) and air quality measurement stations (equipped with a spectrometer to determine particle matters size and sensors to monitor concentration of pollutants like O_3 , NO_x , SO_2 , CO). It is important that different environmental and air pollution variables are measured over the same site, as all such variables are related by physical, deterministic laws imposing their diffusion, reaction, transport, production or removal. Indeed, local trend of air pollutants can cause air quality differences in a range of 10-20 km.

As all statistical approaches, also our filter would benefit of increasing the amount of training and test data, unavoidable condition to give the work more and more significance. Long time-series could be investigated in order to assess possible non stationarities, which temporally adapted filters could decode and counteract in the prediction process. Different sampling stations could also be investigated in order to assess the spatial heterogeneities of air pollution distribution. Moreover, the work could be extended to other consistent air pollutant datasets, in order to provide a more complete air quality analysis of the chosen site.

In conclusion, local air pollution investigation and prediction is a fertile field in which adaptive filters can play a crucial role. Indeed, data-driven approaches could provide deeper insights on pollution dynamics and precise local forecasts which could help preventing critical conditions and taking more efficient countermeasures to safeguard citizens health.

9. Acknowledgments

We are deeply indebted to Riccardo Taormina for his work in processing data and for his interesting comments and suggestions. This work was sponsored by the national project AWIS (Airport Winter Information System), funded by Piedmont Authority, Italy.

10. References

- Bard, D.; Laurent, O.; Havard, S.; Deguen, S.; Pedrono, G.; Filleul, L.; Segala, C.; Lefranc, A.; Schillinger, C.; Rivière, E. (2010). Ambient air pollution, social inequalities and asthma exacerbation in Greater Strasbourg (France) metropolitan area: the PAISA study, Artificial Neural Networks to Forecast Air Pollution, Chapter 15 of "Air Pollution", editor V. Villaniy, SCIYO Publisher, ISBN 978-953-307-143-5.
- Božnar, M.Z.; Mlakar, P.J.; Grašič, B. (2004). Neural Networks Based Ozone Forecasting. Proceeding of 9th Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, June 1-4, 2004, Garmisch-Partenkirchen, Germany.
- Brown, L.R.; Fischlowitz-Roberts, B.; Larsen, J.; (2002). The Earth Policy Reader, Earth Policy Institute, ISBN 0-393-32406-0.
- Cecchetti, M.; Corani, G.; Guariso, G. (2004). Artificial Neural Networks Prediction of PM10 in the Milan Area, Proc. of IEMSs 2004, University of Osnabrück, Germany, June 14-17.
- Chapman, S. (1932). Discussion of memoirs. On a theory of upper-atmospheric ozone, Quarterly Journal of the Royal Meteorological Society, vol. 58, issue 243, pp. 11-13.
- Corani, G. (2005). Air quality prediction in Milan: neural networks, pruned neural networks and lazy learning, Ecological Modelling, Vol. 185, pp. 513-529.
- Costa, M.; Moniaci, W.; Pasero, E. (2003) INFO: an artificial neural system to forecast ice formation on the road, Proceedings of IEEE International Symposium on Computational Intelligence for Measurement Systems and Applications, pp. 216-221.
- De Smet, L.; Devoldere, K.; Vermoote, S. (2007). Valuation of air pollution ecosystem damage, acid rain, ozone, nitrogen and biodiversity - final report. Available online: http://ec.europa.eu/environment/air/pollutants/valuation/pdf/synthesis_report_final.pdf.
- Environmental Research Group, King's College London. (2010).z Air Quality project [Online]. Available: <http://www.londonair.org.uk/london/asp/default.asp>.
- European Environmental Agency EEA, (2008) Annual European Community LRTAP Convention emission inventory report 1990-2006, Technical Report 7/2008, ISSN 1725-2237.
- European Environmental Bureau EEB, (2005) Particle reduction plans in Europe, EEB Publication number 2005/014, Editor Responsible Hontelez J., December.
- European Communities, (2002) Directive 2002/3/EC of the European Parliament and of the Council of 12 February 2002 relating to ozone ambient air, Official Journal of European Community, OJ series L, pp. L67/14-L67/30. Available: <http://eur-lex.europa.eu/JOIndex.do>.
- Foxall, R.; Krcmar, I.; Cawley, G.; Dorling, S.; Mandic, D.P. (2001). Nonlinear modelling of air pollution time-series, *icassp*, Vol. 6, pp. 3505-3508, IEEE International Conference on Acoustics, Speech, and Signal Processing.
- Geller, R. J.; Dorevitch, S.; Gummin, D. (2001). Air and water pollution, Toxicology Secrets, 1st edition, L. Long et al Ed., Elsevier Health Science, pp.237-244.
- Hass, H.; Jakobs, H.J. & Memmesheimer, M. (1995). Analysis of a regional model (EURAD) near surface gas concentration predictions using observations from networks, Meteorol. Atmos. Phys. Vol. 57, pp. 173-200.

- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*, Prentice Hall.
- Hyvarinen, A. (1999). Survey on Independent Component Analysis, *Neural Computing Surveys*, Vol. 2, pp. 94-128.
- Ibarra-Berastegi, G.; Saenz, J.; Ezcurra, A.; Elias, A.; Barona, A. (2009). Using Neural Networks for Short-Term Prediction of Air Pollution Levels. *International Conference on Advances in Computational Tools for Engineering Applications (ACTEA '09)*, July 15-17, Zouk Mosbeh, Lebanon.
- Kantz, H. & Schreiber, T. (1997). *Nonlinear Time-series Analysis*, Cambridge University Press.
- Karatzas, K.D.; Papadourakis, G.; Kyriakidis, I. (2008). Understanding and forecasting atmospheric quality parameters with the aid of ANNs. *Proceedings of the IJCNN*, Hong Kong, China, pp. 2580-2587, June 1-6.
- Koller, D. & Sahami, M. (1996). Toward optimal feature selection, *Proceedings of 13th International Conference on Machine Learning (ICML)*, pp. 284-292, July 1996, Bari, Italy.
- Kocak, K.; Saylan, L.; Sen, O. (2000). Nonlinear time series prediction of O₃ concentration in Istanbul. *Atmospheric Environment*, Vol. 34, pp. 1267-1271.
- Lelieveld, J.; Dentener, F.J. (2000). What controls tropospheric ozone?, *Journal of Geophysical Research*, Vol. 105, n. d3, pp. 3531-3551.
- London Air Quality Network, Environmental Research Group of King's College, London. Web page : <http://www.londonair.org.uk/london/asp/default.asp>.
- Marra, S.; Morabito, F.C. & Versaci M. (2003). Neural Networks and Cao's Method: a novel approach for air pollutants time-series forecasting, *IEEE-INNS International Joint Conference on Neural Networks*, July 20-24, Portland, Oregon.
- Marquardt, D (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics* 11: 431-441. doi:10.1137/0111030.
- Mesin, L.; Kandoor, A.K.R.; Merletti, R. (2008). Separation of propagating and non propagating components in surface EMG. *Biomedical Signal Processing and Control*, Vol. 3(2), pp. 126-137.
- Mesin, L.; Orione, F.; Taormina, R.; Pasero, E. (2010). A feature selection method for air quality forecasting, *Proceedings of the 20th International Conference on Artificial Neural Networks (ICANN)*, Thessaloniki, Greece, September 15-18.
- Mesin, L.; Holobar, A.; Merletti, R. (2011). Blind Source Separation: Application to Biomedical Signals, Chapter 15 of "Advanced Methods of Biomedical Signal Processing", editors S. Cerutti and C. Marchesi, Wiley-IEEE Press, ISBN: 978-0-470-42214-4.
- Met Office UK climate reports, <http://www.metoffice.gov.uk/climate/uk/averages/ukmapavge.html>.
- Papoulis, A. (1984). *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York.
- Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics* 33: 1065-1076. doi:10.1214/aoms/1177704472.
- Pasero, E.; Mesin L. (2010). Artificial Neural Networks to Forecast Air Pollution, Chapter 10 of "Air Pollution", editor V. Villaniy, SCIYO Publisher, ISBN 978-953-307-143-5.

- Perez, P; Trier, A.; Reyes, J. (2000). Prediction of PM_{2.5} concentrations several hours in advance using neural networks in Santiago, Chile. *Atmospheric Environment*, Vol. 34, pp. 1189-1196.
- Perkins H.C. (1974) *Air Pollution*, International Student Edition, Mc Graw Hill.
- Potočník, J. (2010).
<http://europa.eu/rapid/pressReleasesAction.do?reference=IP/10/1265&format=HTML>.
- Rusanovskyy, D.; Gabbouj, M.; Ugur, K. (2007). Spatial and Temporal Adaptation of Interpolation Filter For Low Complexity Encoding/Decoding. *IEEE 9th Workshop on Multimedia Signal Processing*, pp.163-166.
- Science Encyclopedia. <http://science.jrank.org/pages/6028/Secondary-Pollutants.html>.
- Schwarze, P.E.; Totlandsdal, A.L.; Herseth, J.L.; Holme, J.A.; Låg, M; Refsnes, M.; Øvrevik, J.; Sandberg, W.J.; Bølling, A.K.; (2010). Importance of sources and components of particulate air pollution for cardio-pulmonary inflammatory responses, Chapter 3 of "Air Pollution", editor V. Villaniy, SCIYO Publisher, ISBN 978-953-307-143-5.
- Sharma, A. (2000). Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: 1 - A strategy for system predictor identification. *Journal of Hydrology*, Vol. 239, pp. 232-239.
- Sjöberg, J.; Hjalmeron, H. & L. Ljung (1994). *Neural Networks in System Identification*. Preprints 10th IFAC symposium on SYSID, Copenhagen, Denmark. Vol.2, pp. 49-71.
- Sokhi R. S. (2007), *World Atlas of Atmospheric Pollution*, Ed. Anthem Press, London.
- Slini, T.; Kaprara, A.; Karatzas, K.; Moussiopoulos, N. (2006), PM₁₀ forecasting for Thessaloniki, Greece, *Environmental Modelling & Software*, Vol. 21(4), pp. 559-565.
- Strogatz, S.H. (1994). *Nonlinear Dynamics and Chaos*, Addison-Wesley.
- Widrow, B.; Winter, R.G. (1988), *Neural Nets for Adaptive Filtering and Adaptive Pattern Recognition*. *IEEE Computer Magazine*, Vol. 21(3), pp. 25-39.
- Widrow, B.; Lehr, M.A.; Beaufays, F.; Wan, E.; Bilello, M. (1993). *Adaptive signal processing*. *Proceedings of the World Conference on Neural Networks*, IV-548, Portland.
- World Health Organization (2006). *Air quality guidelines. Global update 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide*, ISBN 92 890 2192 6.

A Modified Least Mean Square Method Applied to Frequency Relaying

Daniel Barbosa¹, Renato Machado Monaro², Ricardo A. S. Fernandes²,
Denis V. Coury² and Mário Oleskovicz²

¹Salvador University (UNIFACS)

²Engineering School of São Carlos / University of São Paulo (USP)
Brazil

1. Introduction

In an Electrical Power System (EPS), a fast and accurate detection of faulty or abnormal situations by the protection system are essential for a faster return to the normal operation condition. With this objective in mind, protective relays constantly monitor the voltage and current signals, including their frequency.

The frequency is an important parameter to be monitored in an EPS due to suffer significant alterations during a fault or undesired situations. In practice, the equipment are designed to work continuously between 98% and 102% of nominal frequency (IEEE Std C37.106, 2004). However, variations on these limits are constantly observed as a consequence of the dynamic unbalance between generation and load. The larger variations may indicate fault situations as well as a system overload. Considering the latter, the frequency relay can help in the load shedding decision and, consequently, in the power system stability. In this way, a prerequisite for stable operation has become more difficult to maintain considering the large expansion of electrical systems (Adanir, 2007; Concordia et al., 1995).

The importance of correct frequency estimation for EPS is then observed, especially if the established limits for its normal operation are not reached. This can cause serious problems for the equipment connected to the power utility, such as capacitor banks, generators and transmission lines, affecting the power balance. Therefore, frequency relays are widely used in the system to detect power oscillations outside the acceptable operation levels of the EPS.

Due to the technological advances and considerable increase in the use of electronic devices of the last decades, the frequency variation analyses in EPS were intensified, since the modern components are more sensitive to this kind of phenomenon.

Taking this into account, the study of new techniques for better and faster power system frequency estimation has become extremely important for a power system operation. Thus, some researchers have proposed different techniques to solve the frequency estimation problem. Algorithms based on the phasor estimation, using the LMS method, the Fast Fourier Transform (FFT), intelligent techniques, the Kalman Filter, the Genetic Algorithms, the Weighted Least Square (WLS) technique, the three-phase Phase-Locked Loop (3PLL) and the Adaptive Notch Filter (Dash et al., 1999; 1997; El-Naggar & Youssed, 2000; Girgis & Ham, 1982; Karimi-Ghartemani et al., 2009; Kusljevic et al., 2010; Mojiri et al., 2010; Phadke et al., 1983; Rawat & Parthasarathy, 2009; Sachdev & Giray, 1985). The adaptive filter based on the

LMS proposed by Pradhan et al. (2005) should be outlined. The LMS was first introduced by Widrow and Hoff (Farhang-Boroujeny, 1999) for digital signal processing and has been widely used because of its simplified structure, efficiency and computing robustness.

In this chapter, will be presented the LMS in its complex form (Widrow et al., 1975) with an adaptive step-size (Aboulnasr & Mayyas, 1997) and some practical aspects of the algorithm implementation, which provides an increased convergence speed. It must be emphasized that the complex signal analyzed is formed by the power system three-phase voltages applying the $\alpha\beta$ -Transform (Barbosa et al., 2010).

The algorithm performance was tested by computer simulations using ATP (Alternative Transients Program) software (EEUG, 1987). Some EPS equipment was modelled, including: a synchronous generator with speed governor and voltage control, transmission lines with parameters dependent on the frequency and power transformers. Extreme operational situations were tested in order to observe the behaviour of the proposed technique and validate the obtained results. It must be highlighted that the results of frequency estimation were compared to results of a commercial frequency relay (function 81), and it shows that the adaptive filter theory applied to the digital protection is fast.

2. The algorithm based on LMS

The algorithm based on the LMS method, presented in Fig. 1, is a combination of the adaptive process with digital filtering. In this Figure, $\bar{u}(n-1) = [u(n-1) \ u(n-2) \ \dots \ u(n-M)]$ is the vector with M past values of the input $u(n)$; $\bar{w}(n) = [w_1(n) \ w_2(n) \ \dots \ w_N(n)]^T$ is the vector with the filter coefficients; $y(n)$ is the desired signal (the output filter) and $e(n)$ is the error associated to the filter approximation.

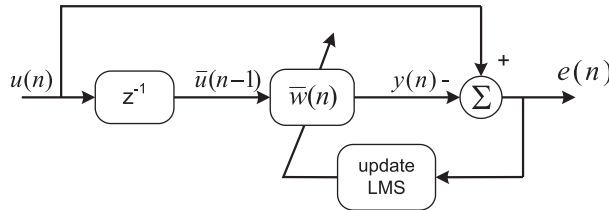


Fig. 1. Adaptive filter based on LMS.

The input signal of the filter can be estimated by minimizing the squared error by the coefficient adaptations ($\bar{w}(n)$), which are recursively adjusted to obtain optimal values. At each iteration, the coefficients can be calculated by:

$$\bar{w}(n+1) = \bar{w}(n) + \mu (-\nabla(n)), \quad (1)$$

where μ is the convergence parameter and ∇ is the gradient of error performance surface and is responsible for determining the adjustment of coefficients.

The LMS algorithm is very sensitive to μ . This can be mainly observed by the speed of the estimation and processing time. The lesser μ value, the longer the time to reach the aimed error and vice-versa. However, it is important to respect the convergence interval given by (Haykin, 2001):

$$0 < \mu < \frac{1}{NS_{max}}, \quad (2)$$

where N is the filter size and S_{max} is the maximum power spectral density value of the input signal.

3. The adaptive algorithm and the frequency estimation

The study of digital filters is a consolidated research area. Regarding digital protection, digital filters provide the frequency component extraction used in digital relay algorithms. The information contained in the input data from a three-phase system can be processed simultaneously, making it possible to obtain more precise results if compared to conventional methods.

It must be highlighted that the proposed algorithm, called Frequency Estimation Algorithm by the Least Mean Square (FEALMS), uses three-phase signals as inputs. It is considered that the three-phase voltages from EPS can be represented by:

$$\begin{aligned} V_a(n) &= A_{max}\cos(\omega n\Delta t + \phi) + \xi_{n_a} \\ V_b(n) &= A_{max}\cos(\omega n\Delta t + \phi - \frac{2\pi}{3}) + \xi_{n_b}, \\ V_c(n) &= A_{max}\cos(\omega n\Delta t + \phi + \frac{2\pi}{3}) + \xi_{n_c} \end{aligned} \quad (3)$$

where, A_{max} is the peak value; ω is the signal angular frequency¹; n is the sample number of the discrete signal; Δt is the time between two consecutive samples; ϕ is the signal phase and ξ_n is the error between two consecutive samples. Fig. 2 illustrates the proposed relay algorithm.

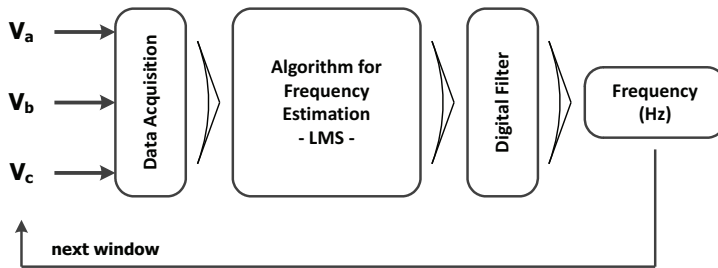


Fig. 2. Basic relay algorithm.

3.1 The data acquisition

All the stages in data acquisition are performed with the aim of having a more realistic analysis of the obtained results. The input voltage signals simulated are characterized by a high sampling rate in order to represent the analog signals more realistically.

A flowchart that represents the procedure of data acquisition can be visualized by means of Fig. 3. A second order Butterworth low pass filter with a cut-off frequency of 200Hz was utilized. A sample rate of 1,920Hz and an analog-to-digital converter (ADC) of 16 bits were also used.

The low pass filter was used to avoid spectral spreading and to make sure that the digital representation after ADC conversion is a good representation of the original signal. It is worth commenting that a low cut-off frequency of 200Hz was used in order to stabilise the method and ensure that the LMS algorithm will converge. Due to this situation, most of the harmonic components were eliminated, increasing the precision of the proposed method.

¹ $\omega = 2\pi f$, where f is the fundamental frequency of the electrical system

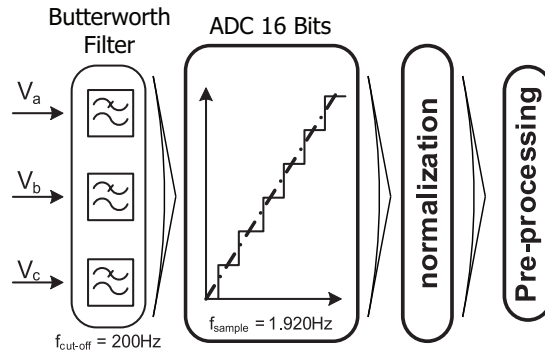


Fig. 3. Data acquisition flowchart.

The data acquisition was performed in a moving window with one sample step. All the filter processing should be performed on one data window, respecting the available time for processing, which is the time between two consecutive samples. Fig. 4 illustrates this process.

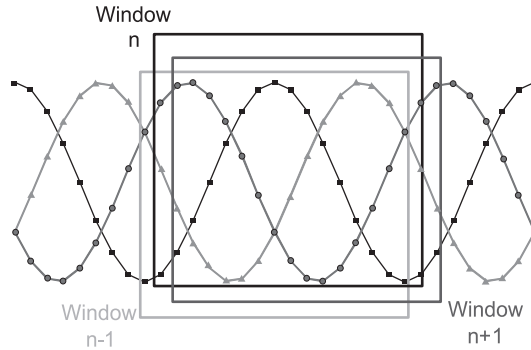


Fig. 4. Moving window process.

3.2 The normalization process

Normalization standardises the data obtained from the electrical system, regardless of the voltage level analysed. Consequently, if either a sag or a swell occurs in any phase, the algorithm will maintain its estimation without loss of precision or speed. Fig. 5 illustrates the normalisation process implemented.

3.3 The pre-processing process

After normalisation, a pre-processing stage was performed, obtaining the signal in its complex form for the digital filter. This was obtained by applying the $\alpha\beta$ -Transform on the three-phase voltages, as represented in the following equation (Akke, 1997):

$$\begin{bmatrix} V_\alpha(n) \\ V_\beta(n) \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} [V(n)], \quad (4)$$

where $V(n) = [V_a(n) \ V_b(n) \ V_c(n)]^T$.

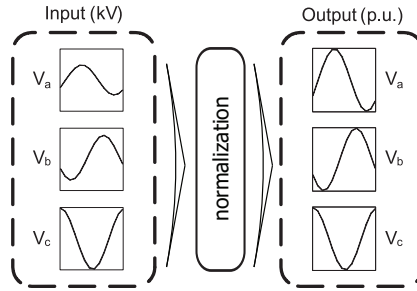


Fig. 5. Data normalization flowchart.

After the pre-processing stage, in order to obtain the α and β components by (4), the complex voltage is defined as:

$$u(n) = V_{\alpha}(n) + jV_{\beta}(n). \quad (5)$$

3.4 The coefficient generator

Adapting the filter coefficients is simple and inherent to the algorithm. This adjustment is performed sample by sample in order to make sure that the squared average error is minimised. However, to improve the algorithm performance and minimise the processing time, the estimation filter coefficients ($\hat{w}(n)$) are initialised with the estimation of the previous window (Barbosa et al., 2010). The first window is initialised with the fundamental frequency of the electrical system. The aim of this procedure is to increase the speed of the estimation process. The coefficient generator flowchart is shown in Fig. 6.

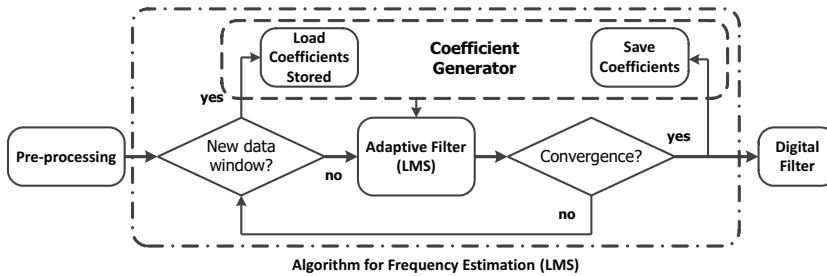


Fig. 6. Coefficient generator flowchart.

3.5 The adaptive filter

In the adaptive filter, the coefficients are updated recursively to minimise the squared error. The error is calculated as the difference between desired and estimated values, given by:

$$e(n) = u(n) - y(n), \quad (6)$$

where $y(n)$ represents the estimated value. The complex voltage ($u(n)$) can be described by:

$$\begin{aligned} u(n) &= U_{max}e^{j(\omega n\Delta T + \phi)} + \zeta(n) \\ &= y(n) + \zeta(n), \end{aligned} \quad (7)$$

where U_{max} is the amplitude of the complex signal, ζ is the noise component, ΔT is the sampling interval, ϕ is the phase of signal, n is the sample number and ω is the angular frequency of the analyzed signal. The estimated complex voltage ($y(n)$) can be represented by equation (8).

$$y(n) = y(n-1)e^{j\omega\Delta T}. \quad (8)$$

The equations (8) and (7) are the base of used model for proposed frequency estimation. Although the output filter can be represented by previous model, it is a linear combination between the input vector, lagged by one sample, and the vector with filter coefficients as illustrated below.

$$y(n) = \bar{w}^H(n)\bar{u}(n-1), \quad (9)$$

where H is the Hilbert transform and $\bar{w}$ is the vector with filter coefficients. This vector denotes the difference between two consecutive samples, ie, the phase difference between the samples being analyzed, according to the equation (10).

$$\bar{w}(n) = e^{j\bar{\omega}(n-1)\Delta T}, \quad (10)$$

where, $\bar{\omega}$ is the estimated angular frequency.

It must be emphasised that the LMS task is to find the filter coefficients that minimize the error. Following this procedure, the filter coefficients are updated until the error is sufficiently small. The complex weight vector at each sampling instant is given by (Widrow et al., 1975):

$$\bar{w}(n+1) = \bar{w}(n) + \mu(n)e(n)^*\bar{u}(n-1), \quad (11)$$

where the (*) symbol denotes the complex conjugate and μ is the convergence factor controlling the stability and rate of convergence of the algorithm. Fig. 7 shows the evolution of the adaptive filter coefficients of eighth order during the iterations.

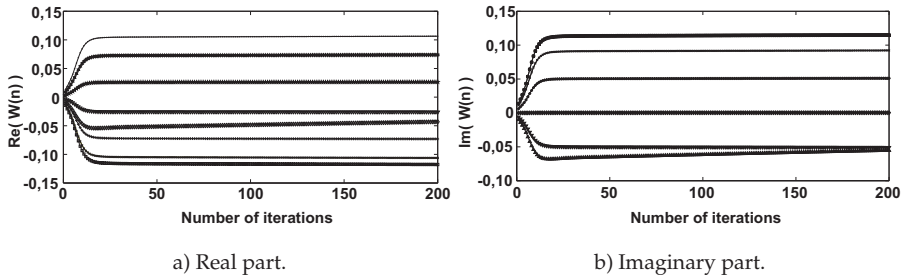


Fig. 7. The adaptive filter coefficient update.

The step size $\mu(n)$ is modified for better convergence in the presence of noise and its equation is given by (Aboulnasr & Mayyas, 1997):

$$\mu(n+1) = \lambda\mu(n) + \gamma p(n)p(n)^*, \quad (12)$$

where $p(n)$ represents the autocorrelation error and it is calculated as:

$$p(n) = \rho p(n-1) + (1-\rho)e(n)e(n-1). \quad (13)$$

In the equation, ρ is the exponential weighting parameter. The ρ ($0 < \rho < 1$), λ ($0 < \lambda < 1$) and γ ($0 < \gamma$) are constants that control the convergence time and they are determined by statistical studies (Kwong & Johnston, 1992).

3.6 The stability of the proposed algorithm

The stability is a critical factor in proposed algorithm implementation, especially if the convergence factor (μ) is out of range associated. Due to these problems, a continuous monitoring of data window samples is performed, providing a self-tuning range of convergence. The behavior of μ is controlled by equation (14) (Wies et al., 2004):

$$0 < \mu < \frac{1}{\frac{N}{M} \sum_{n=0}^{M-1} u(n)u(n)^*}, \quad (14)$$

where N and M are the filter and window sizes, respectively. Fig. 8 shows the proposed algorithm flowchart with the stability control.

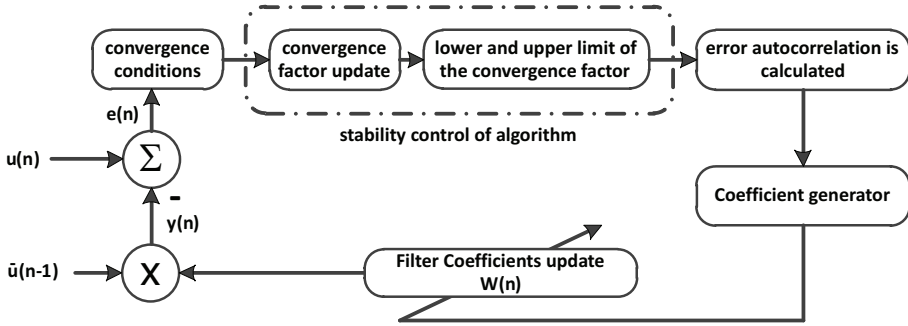


Fig. 8. Proposed algorithm flowchart.

3.7 The frequency estimation

The frequency estimation was performed according to Begovic et al. (1993). To find the phase difference, the complex variable Γ was defined as:

$$\Gamma = y(n)y(n-1)^*. \quad (15)$$

The relationship between Γ and the system frequency is obtained by equation expansion above. This expanding is shown by:

$$\begin{aligned} \Gamma &= U_{max}^2 \left[e^{j(\omega n \Delta T + \phi)} e^{-j(\omega(n-1) \Delta T + \phi)} \right] \\ &= U_{max}^2 \left[e^{j(\omega n \Delta T + \phi - \omega n \Delta T + \omega \Delta T - \phi)} \right] \\ &= U_{max}^2 e^{j2\pi f_{est} \Delta T} \\ &= U_{max}^2 \{ \cos(2\pi f_{est} \Delta T) + j \sin(2\pi f_{est} \Delta T) \}, \end{aligned} \quad (16)$$

where $U_{max} = 1$, once the input signal is normalized, and f_{est} is the estimated frequency. The frequency of the estimated signal ($y(n)$) was calculated in function of the phase difference between two consecutive samples, and the latter was provided by the equation below:

$$f_{est} = \frac{f_s}{2\pi} \arctan \left(\frac{\Im(\Gamma)}{\Re(\Gamma)} \right), \quad (17)$$

where f_s is the sampling frequency and $\Re()$ and $\Im()$ are the real and imaginary parts, respectively.

3.8 The convergence process

The stop rule adopted was the maximum number of iterations (1,000) or error smaller than 10^{-5} . This error can be estimated by:

$$e_{relat} = abs(y(n) - u(n)), \quad (18)$$

where e_{relat} is the relative error between samples, $abs()$ is the absolute value, $y(n)$ is the estimated value and $u(n)$ is the desired value or input sample.

3.9 The post-processing process of the output signal

The output signal (estimated frequency) is additionally filtered by a second order Butterworth low pass digital filter with a cut-off frequency of 5Hz. This procedure reduces the oscillation present in the proposed method output, avoiding errors due to abrupt variations of the frequency. It is important to observe that the delay of the low pass filter does not influence the algorithm performance negatively as can be seen in the results.

4. The power system simulation

Fig. 9 shows the representation of the simulated electrical system, taking into account load switching and permanent faults in order to evaluate the frequency estimation technique proposed in this work.

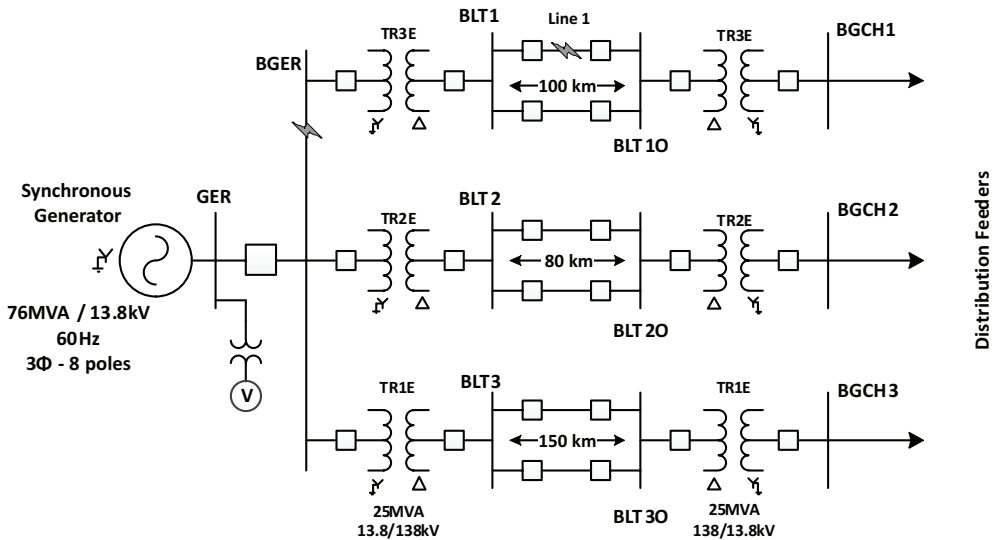


Fig. 9. The power system representation using ATP software.

The electrical system consists of a 13.8 kV and 76 MVA (60Hz) synchronous generator, 13.8:138 kV /138:13.8 kV and 25 MVA three phase power transformers, transmission lines between 80 and 150 km in length and loads between 5 and 25 MVA with a 0.92 inductive power factor. Power transformers have a delta connection in the high voltage winding and a star connection in the low voltage winding. The power transformers were modeled using ATP software (saturable transformer component) considering their saturation curves as illustrated in Fig. 10. Tables 1 and 2 show the parameters used in order to simulate the power system components using ATP software.

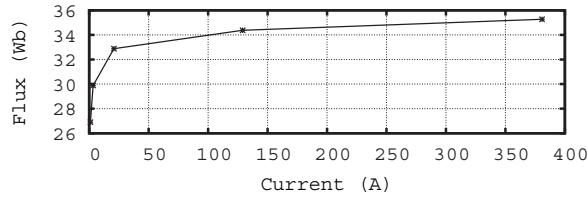


Fig. 10. Saturation curve of power transformers.

Description	Value (unit)	Description	Value (unit)
S	76 (MVA)	N_p	8
V_L	13.8 (kV _{rms})	f	60 (Hz)
I_{FD}	250 (A)	R_a	0.004 (p.u.)
X_l	0.175 (p.u.)	X_o	0.132 (p.u.)
X_d	1.150 (p.u.)	X_q	0.685 (p.u.)
X'_d	0.310 (p.u.)	X''_d	0.210 (p.u.)
X''_q	0.182 (p.u.)	τ'_{do}	5.850 (sec)
τ''_{do}	0.036 (sec)	τ''_{qo}	0.073 (sec)

Table 1. Synchronous generator data used in the simulation.

In Table 1, S is the total three-phase volt-ampere rating of the machine, N_p is the number of poles which characterise the machine rotor, V_L is the rated line-to-line voltage of the machine, f is the electrical frequency of generator, I_{FD} is the field current, R_a is the armature resistance, X_l is the armature leakage reactance, X_o is the zero-sequence reactance, X_d is the direct-axis synchronous reactance, X_q is the quadrature-axis synchronous reactance, X'_d is the direct-axis transient reactance, X''_d is the direct-axis subtransient reactance, X''_q is the quadrature-axis subtransient reactance, τ'_{do} is the direct-axis open-circuit transient time constant, τ''_{do} is the direct-axis open-circuit subtransient time constant and τ''_{qo} is the quadrature-axis open-circuit transient time constant.

Element	$R_+(\Omega)$	$L_+(mH)$
Primary impedance of transformer	1.7462	151.37
Secondary impedance of transformer	0.0175	1.514

Table 2. Power Transformer data used in the simulation.

It is important to emphasise that the transmission line model used was JMARTI from ATP. This was because it is possible to have a variation of the line parameters in function of the frequency and consequently obtain a better representation of the system's behavior when facing disturbances resulting from unbalance between generation and load.

It must also be emphasised that the synchronous generator was simulated with an automatic speed control for hydraulic systems (Boldea, 2006) and automatic voltage regulation (AVR) (Boldea, 2006; Lee, 1992; Mukherjee & Ghoshal, 2007), considering various electrical and mechanical parameters from the generator. Equation 19 shows the transfer function of the speed regulator used:

$$\frac{\eta(s)}{\Delta F(s)} = -\frac{1}{R} \cdot \frac{1 + sT_r}{(1 + sT_g)(1 + s\frac{r}{R}T_r)} \quad (19)$$

where $\eta(s)$ is the servomotor position, $\Delta F(s)$ is the frequency deviation, R is the steady-state speed droop, r is the transient speed droop, T_g is the main gate servomotor time constant and T_r is the reset time. Table 3 presents the parameters concerning the speed regulator.

Description	Value (unit)
Main gate servomotor time constant (T_g)	0.600 (sec)
Reset time (T_r)	0.838 (sec)
Transient speed droop (r)	0.279
Steady-state speed droop (R)	0.100
Moment of inertia (M)	1.344 (sec)
Water starting constant (T_W)	0.150 (sec)

Table 3. Parameters concerning the speed regulator.

Fig. 11 shows the block diagram of the excitation control system which was used. The basic function of the excitation control system automatically adjusts the magnitude of the DC field current of the synchronous generator to maintain the terminal voltage constant as the output varies according to the capacity of the generator (Kundur, 1994).

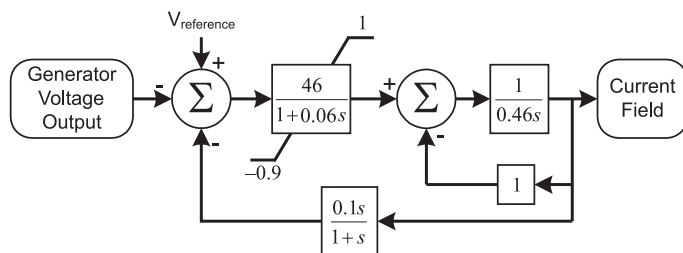


Fig. 11. Block diagram of the excitation control system.

The field voltage control can improve the transient stability of the power system after a major disturbance. However, the extent of the field voltage output is limited by the exciter's ceiling voltage, which is restricted by generator rotor insulation (Kundur, 1994; Leung et al., 2005).

5. Test cases

This section presents results of the proposed scheme. Although a great deal of data was used to test the proposed technique, only four cases of abnormal operation were carefully chosen to

illustrate the technique performance concerning the electrical system presented in Fig. 9. Each condition imposes a particular dynamic behavior in the power balance and, consequently, in the variation of the power system frequency. Measurements from a commercial relay (function 81) were obtained by using the simulated voltage signals from ATP in order to compare the results. Moreover, the actual frequency of the EPS was measured directly from the angular speed of the synchronous generator. It should be emphasized that the sample rates of 1,920Hz and 1,000Hz were used in the FEALMS software and a commercial relay (function 81), respectively.

Due to a great influence from the adjustment of the filter parameters in the results, these parameters were selected according to Kwong & Johnston (1992), and they are: $\mu_{max} = 0.18$, $\mu_{min} = 0.001$, $p_{inicial} = 0$, $\lambda = 0.97$, $\gamma = 0.01$ and $\rho = 0.99$. Based on Fig. 9, the simulated situations were:

- a sudden connection of load blocks;
- a permanent fault involving phase A and ground (AG) on the BGER busbar at 2s;
- a sudden disconnection of TR1E and TR3E transformers at 1s;
- a permanent fault at 50% of line 1;
- the generator overexcitation;
- the TR3E transformer energization with full load.

5.1 A sudden connection of load blocks

Fig. 12(a) shows the estimation of the synchronous generator frequency using the FEALMS, the ATP software reference curve and the commercial frequency relay responses considering the connection of load blocks in the BGCH3 busbar. In the figure, a slight delay in the frequency estimation by the relay can be observed, when compared to the correct result given by the ATP software curve. In this situation, a very good precision concerning the FEALMS can be observed, even in critical points of the system's behavior. The error concerning the application of the proposed technique is also presented as illustrated in Fig. 12(b).

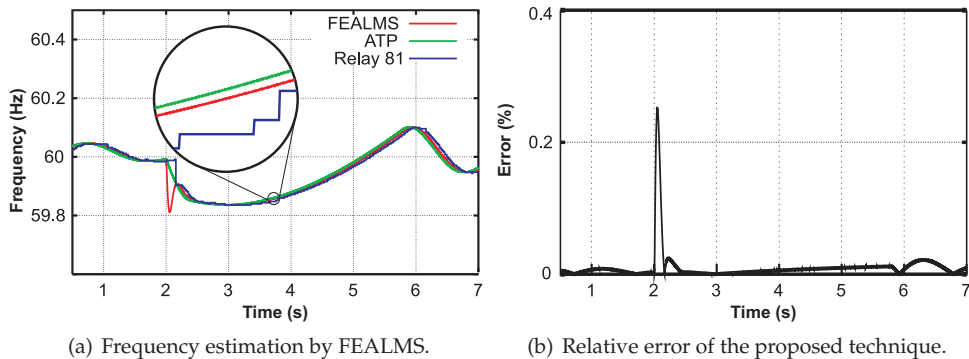


Fig. 12. Connection of load blocks on the BGCH3 busbar at 2s.

5.2 A permanent fault involving phase A and ground (AG) on the BGER busbar at 2s

Fig. 13(a) shows the estimation given by the proposed technique, the ATP reference curve, as well as the commercial frequency relay performances for an AG fault on the BGER busbar at 2s.

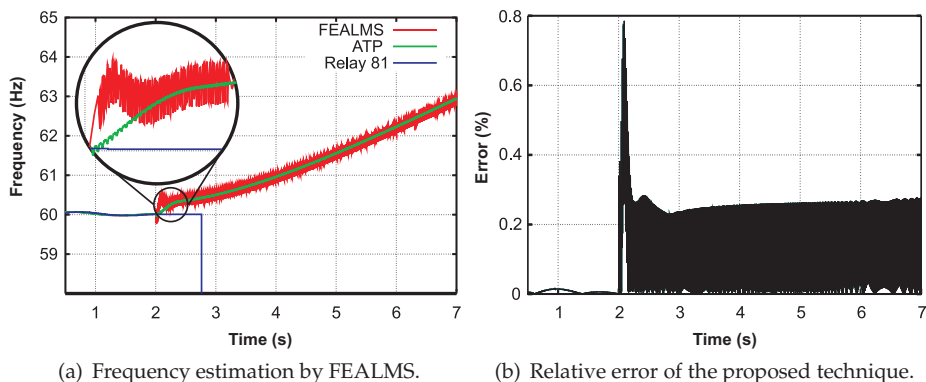


Fig. 13. AG fault on BGER busbar at 2s.

It can be seen that the commercial frequency relay tested loses the reference voltage, hindering the frequency estimation. It should be emphasized that even in this unfavorable condition, the FEALMS estimated the power system frequency satisfactorily.

5.3 A sudden disconnection of TR1E and TR3E transformers at 1s

Fig. 14(a) also illustrates the FEALMS responses, the ATP reference curve, as well as the commercial frequency relay for a disconnection of TR1E and TR3E transformers. The analysis of this condition is fundamental to test the algorithm's robustness facing practical field situations, taking into account the high level of distortion presented in the input signals.

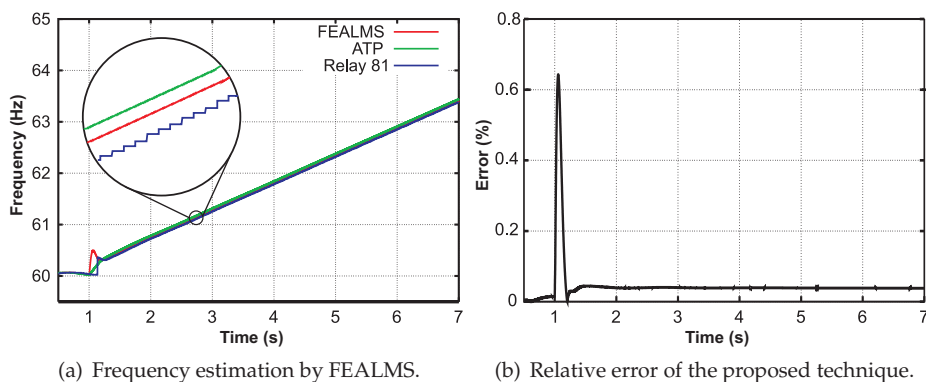


Fig. 14. A sudden disconnection of TR1E and TR3E transformers at 1s.

It can be clearly observed that the commercial relay (function 81), as well as the FEALMS, presented similar satisfactory responses. Once more, a delay can be observed for the commercial relay (function 81), possibly due to the filtering process.

5.4 A permanent fault at 50% of line 1

In Fig. 15, a frequency variation and relative error of frequency estimation by the proposed technique are observed considering an AG fault at 50% of the transmission length of line 1. In this situation, the correspondent three-phase circuit breaker opened after 80ms from the fault inception.

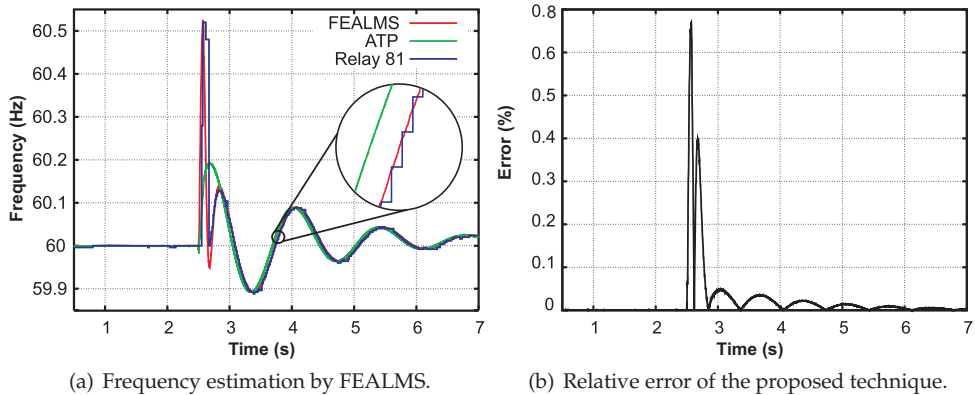


Fig. 15. AG fault at 50% of the line 1 length.

It should also be observed that the recovery of the machine synchronization was achieved in this situation.

5.5 The generator overexcitation

Fig. 16(a) illustrates the FEALMS and commercial relay (function 81) responses, as well as the ATP reference curve for the generator overexcitation. This situation illustrates the ability of the algorithm to estimate the frequency in the presence of harmonics in the input signals.

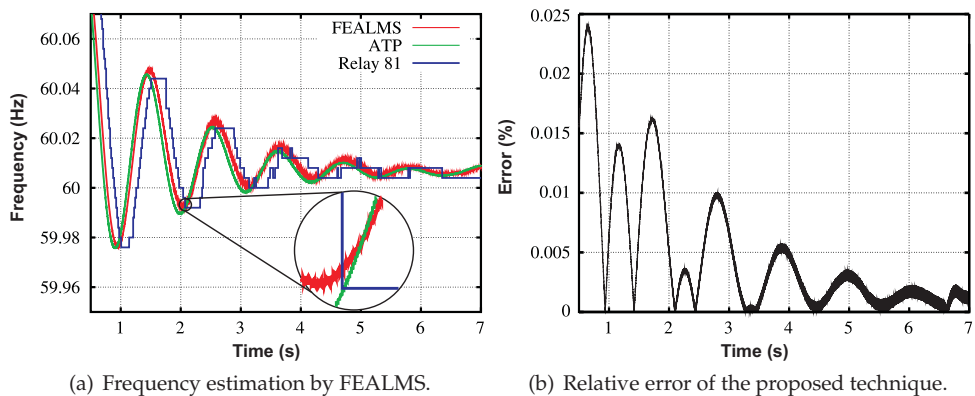


Fig. 16. Frequency estimation for generator overexcitation.

5.6 The TR3E transformer energization with full load

Fig. 17(a) illustrates the frequency estimation of the proposed technique together with the commercial frequency relay performance, as well as the ATP reference curve for a case of frequency drop concerning TR3E transformer energization with full load.

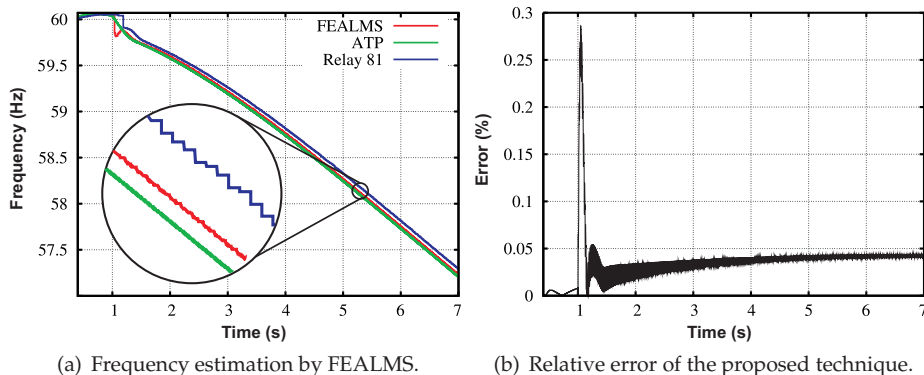


Fig. 17. TR3E transformer energization with full load.

It can be seen from Fig. 17(b) that most of the estimation errors for FEALMS are below 0.05%.

6. Conclusions

This chapter presented an alternative method for the frequency estimation in electrical power systems using the modified LMS algorithm. The implementation of digital filtering and the $\alpha\beta$ -Transform in the proposed technique made the simultaneous use of the three-phase power system voltages for the estimation purpose possible.

In the complex LMS algorithm considered, the adjustment of the step size was used as being adaptive based on the estimations of the error.

The simulations used for testing the proposed algorithm were obtained using ATP software. The adaptive filter theory applied to the digital protection was fast and reliable. Some points should be observed:

1. The FEALMS algorithm can be applied to various situations and voltage levels, as it is not influenced by the magnitude of the input waveforms;
2. Three-phase voltages were analysed contrary to a single phase (used in commercial relays), making the proposed algorithm more robust;
3. The technique is easy to implement and does not need adjustments and knowledge of further functions, as is the case of the commercial relay (function 81) which was used;
4. Applying FEALMS algorithm, the average error in all cases studied was 0.08%.

It is also important to highlight the feasibility and computing efficiency of this method, make it suitable for commercial applications.

7. References

Aboulnasr, T. & Mayyas, K. (1997). A robust variable step-size LMS-type algorithm: Analysis and simulations, *IEEE Transactions on Signal Processing* 45(3): 631–639.

- Adanir, T. (2007). Extremely short term frequency estimation (estfe) algorithm for underfrequency protection, *International Journal of Electrical Power & Energy Systems* 29(4): 329–337.
- Akke, M. (1997). Frequency estimation by demodulation of two complex signals, *IEEE Transactions on Power Delivery* 12(1): 157–163.
- Barbosa, D., Monaro, R. M., Coury, D. V. & Oleskovicz, M. (2010). Digital frequency relaying based on the modified least mean square method, *International Journal of Electrical Power & Energy Systems* 32(3): 236 – 242.
URL: <http://www.sciencedirect.com/science/article/B6V2T-4X076Y8-2/2/fe189bb77b1dd0aedab3d7bf7e306fee>
- Begovic, M. M., Djuric, P. M., Dunlap, S. & Phadke, A. G. (1993). Frequency tracking in power networks in the presence of harmonics, *IEEE Transactions on Power Delivery* 8: 480–486.
- Boldea, I. (2006). *Synchronous Generators*, CRC Press, Boca Raton. ISBN: 084935725X.
- Concordia, C., Fink, L. H. & Poullikkas, G. (1995). Load shedding on an isolated system, *IEEE Transactions on Power Systems* 10(3): 1467–1472.
- Dash, P. K., Padhan, A. K. & Panda, G. (1999). Frequency estimation of distorted power system signals using extended complex Kalman filter, *IEEE Transactions on Power Delivery* 14(3): 761–766.
- Dash, P. K., Swain, D. P., Routray, A. & Liew, A. C. (1997). An adaptive neural network approach for the estimation of power system frequency, *Electric Power Systems Research* 41: 203–210.
- EEUG (1987). *Alternative Transients Program Rule Book*, LEC.
- El-Naggar, K. M. & Youssed, H. K. M. (2000). A genetic based algorithm for frequency-relaying applications, *Electric Power Systems Research* 55: 173–178.
- Farhang-Boroujeny, B. (1999). *Adaptive Filters: Theory and Applications*, John Wiley & Sons, Inc.
- Girgis, A. A. & Ham, F. M. (1982). A new FFT-based digital frequency relay for load shedding, *IEEE Transactions on Power Apparatus and Systems* PAS-101(2): 433–439.
- Haykin, S. (2001). *Adaptive Filter Theory*, 4 edn, Prentice Hall, New Jersey.
- IEEE Std C37.106 (2004). IEEE guide for abnormal frequency protection for power generating plants.
- Karimi-Ghartemani, M., Karimi, H. & Bakhshai, A. R. (2009). A filtering technique for three-phase power systems, *IEEE Transactions on Instrumentation and Measurement* 58(2): 389–396.
- Kundur, P. (1994). *Power system stability and control*, The EPRI power system engineering series, McGraw-Hill, New York. ISBN: 007035958X.
- Kusljevic, M. D., Tomic, J. J. & Jovanovic, L. D. (2010). Frequency estimation of three-phase power system using weighted-least-square algorithm and adaptive fir filtering, *IEEE Transactions on Instrumentation and Measurement* 59(2): 322–329.
- Kwong, R. H. & Johnston, E. W. (1992). A variable step size LMS algorithm, *IEEE Transactions on Signal Processing* 40: 1633–1642.
- Lee, D. C. (ed.) (1992). *IEEE Recommended Practice for Excitation System Models for Power System Stability Studies (IEEE Std 421.5-1992)*, Energy Development and Power Generating Committee of the Power Engineering Society.
- Leung, J. S. K., Hill, D. J. & Ni, Y. (2005). Global power system control using generator excitation, PSS, FACTS devices and capacitor switching, *Electric Power Systems Research* 27(5–6): 448–464.

- Mojiri, M., Yazdani, D. & Bakhshai, A. (2010). Robust adaptive frequency estimation of three-phase power systems, *IEEE Transactions on Instrumentation and Measurement* 59(7): 1793–1802.
- Mukherjee, V. & Ghoshal, S. (2007). Intelligent particle swarm optimized fuzzy PID controller for AVR system, *Electric Power Systems Research* 77(12): 1689–1698.
- Phadke, A. G., Thorp, J. S. & Adamiak, M. G. (1983). A new measurement technique for tracking voltage phasors, local system frequency, and rate of change of frequency, *IEEE Transactions on Power Apparatus and Systems* PAS-102(5): 1025–1038.
- Pradhan, A. K., Routray, A. & Basak, A. (2005). Power system frequency estimation using least mean square technique, *IEEE Transactions on Power Delivery* 20(3): 1812–1816.
- Rawat, T. K. & Parthasarathy, H. (2009). A continuous-time least mean-square adaptive filter for power system frequency estimation, *International Journal of Electrical Power & Energy Systems* 31: 111–115.
- Sachdev, M. S. & Giray, M. M. (1985). A least error squares technique for determining power system frequency, *IEEE Transactions on Power Apparatus and Systems* PAS-104(2): 437–444.
- Widrow, B., McCool, J. & Ball, M. (1975). The complex lms algorithm, *Proceedings of the IEEE* 63: 719–720.
- Wies, R. W., Pierre, J. W. & Trudnowski, D. J. (2004). Use of least mean squares (lms) adaptive filtering technique for estimating low-frequency electromechanical modes in power systems, *Power Engineering Society General Meeting*, Vol. 2, pp. 1863–1870.

Anti-Multipath Filter with Multiple Correlators in GNSS Receivers

Chung-Liang Chang

*Department of Biomechatronics Engineering
National Pingtung University of Science and Technology
Pingtung County
Taiwan*

1. Introduction

The positioning technique of global navigation satellite system (GNSS) has become mature and also been applied to a variety of navigation vehicles, whether it be the application to ground vehicle or aircraft. Nevertheless, the precision of GNSS is susceptible to intentional or unintentional factors such as interference or jammer, etc. The influences range from minor effect like the positioning precision of satellite signal to significant impact like the misleading information to users or malfunction of receivers. The ionosphere or troposphere in environment or the noise in receiver itself are the source of positioning error when satellite passes through ionosphere or troposphere, the change of media results in the delay of wave transmission rate and yields error. The adoption of dual frequency receiver can decrease error but it presents no significant improvement in terms of the error generated by multipath.

The effect of multipath is because the satellite signal is reflected or diffracted by obstacle prior to its reception by antenna. Most of the time, it results in the decrease of signal propagation power and delay of time. In 1973, Hagerman employed conventional code tracking to analyze the effect of multipath on the coarse/acquire (C/A) code in carrier L1 using one chip of early-late spacing. He also estimated that under different delay, phase and signal magnitude, the effect may result in 70–80 m tracking error [Hagerman, 1973]. With the growing application of global positioning system (GPS), many researches investigating multipath effect have been proposed to effectively reduce its impact and have provided various estimation algorithms for implementation in hardware.

The most effective solution for multipath effect is the location of antenna. Assume the antenna is placed above the highest source reflection, the reflected signal will not be received. In antenna design, it can reduce the gain of received signal coming through lower elevation. Generally, the receiver will setup up the minimum elevation capable of receiving satellite signal. The design of choke ring antenna is used to mitigate multipath. The choke ring antenna circles the antenna with vertical concentric rings, whose function is to reduce the gain of received reflected signal. However, such a function is strongly related to the location of antenna.

In addition, the change of structure in internal correlator design of receiver is also a solution for multipath. The conventional GPS receiver typically adopts one chip early-late

spacing of correlator. The use of narrow correlator to reduce chip spacing can effectively mitigate multipath and noise, which cuts down the error of 70–80 m to 8–10 m (van Dierendonck et al., 1992). Note that the use of narrow correlator technique in coherent discriminator may lead to the lock failure in code delay locked loop without the cooperation of phase locked loop (PLL).

The strobe correlator and edge correlator are both solutions for multipath mitigation (Garin et al., 1996). The strobe correlator is implemented using two different narrow correlator discriminators. The strobe correlator and edge correlator developed by Ashtech only provide code correlation for C/A code. The enhanced strobe correlator (Garin and Rousseau, 1997) offers carrier phase correction and code correction for C/A code. With the additional carrier phase correction in terms of multipath its real-time dynamic processing outperforms previous methods. Note that the narrow correlator and strobe correlator do not encompass carrier phase correction. Thus, their sensitivity approaches that of conventional correlator.

Another discriminator design is early 1/ early 2 (E1/E2) correlator (Mattos, 1996; van Dierendonck and Braasch, 1997). The method utilizes part of correlation coefficients not subject to multipath effect for multipath mitigation. That is, it employs two correlators with the spacing and location at the front end of correlation function. However, this method is a choice between noise mitigation and multipath mitigation.

The multipath estimation method initial estimates multipath signal and then subtracts it from received signal so that the signal approaches direct signal. Literature review that resembles this algorithm are MEDLL, MET. (van Nee, 1992; van Nee et al., 1994), which utilize maximum likelihood estimation technique and recursive least square method to estimate the magnitude, delay, phase and erase it from received signal. Though the above estimation methods can not completely eliminate multipath signal, they present significant improvement in terms of multipath delay within certain range.

Nevertheless, these techniques have difficulties in mitigating short-delay multipath signals (less than 0.1 PN code chip or approximately 30 m). Scholars have proposed methods on short-delay multipath mitigation (Sleewaegen et al., 2001; Stone and Chansarkar, 2004). However, these techniques still have drawbacks. The method proposed by Sleewaegen requires a scaling factor, depending on multipath environment, to link the signal amplitude with the range error. The method proposed by Stone and Chansarkar is to estimate the pseudorange error on the basis of a statistical model, which requires large numbers of collected data. Consequently, the performances of these two methods are significantly influenced by multipath environment.

The author has proposed an adaptive filter in 2008 (Chang and Juang, 2008), which adopts five tap-delay to effectively mitigate short-delay multipath. Though this method is efficient in short-delay multipath mitigation, it does not guarantee that the receiver will not receive multipath signal at different time delay under variable environment. Moreover, the correlator technique of conventional receiver is not quite capable of accurately describing the data distribution of correlated signal, which results in longer period of time to estimate multipath parameter. Thus, this paper utilizes multi-correlator technique in combination with proposed method to mitigate the mystical multipath signal. Simulation results show that the multi-correlator technique can clearly present the output distribution of correlator, make adaptive filter rapidly estimate multipath parameter and cope with multipath signal at different time delay.

2. Methodology

2.1 Multipath overview

Multipath effect is caused by the reflection of satellite signal by obstacles when the receiver receives the reflected signal, it leads to positioning error and the lock failure of signal for receiver, which renders positioning function void. In GPS, the desired signal consists of only the direct path signal. All other signals distort the desired signal and result in ranging measurement errors. To understand the effect of multipath in measurement process, let's consider the heart of the GPS code tracking loop. The pseudorange measurement originates from a locally generated pseudorandom noise (PRN) code which is kept phase-locked to the received code. The discriminator is formed based on the difference between early correlator output and late correlator output. The output of the discriminator is fed back to the local code generator to keep synchronism between the local code and incoming code. This generates the so-called delay-locked loop (DLL). When multipath is present, the incoming code, correlation function and discriminator functions are distorted. Analytically, the direct and multipath components may be conducted separately. Note that for the direct-path case, the discriminator function passes through zero when the code-tracking error (local-code delay) is zero. This is the ideal case. However, when multipath is present, the distorted function has a zero-crossing at non-zero code tracking error. Fig. 1 demonstrates the tracking errors of the early-late discriminator output due to multipath in the DLL. The tracking errors result from distortion of the correlation function with the received IF signal. In the direct-path case, the ideal case is when the discriminator function passes through zero while the code tracking error is zero. However, with the presence of multipath, the distorted function has a zero-crossing at a non-zero code tracking error. With the direct signal, when the relative multipath phase is 0 radians, the multipath component is 'in-phase'. With π radians, the multipath component is 'out-of phase'.

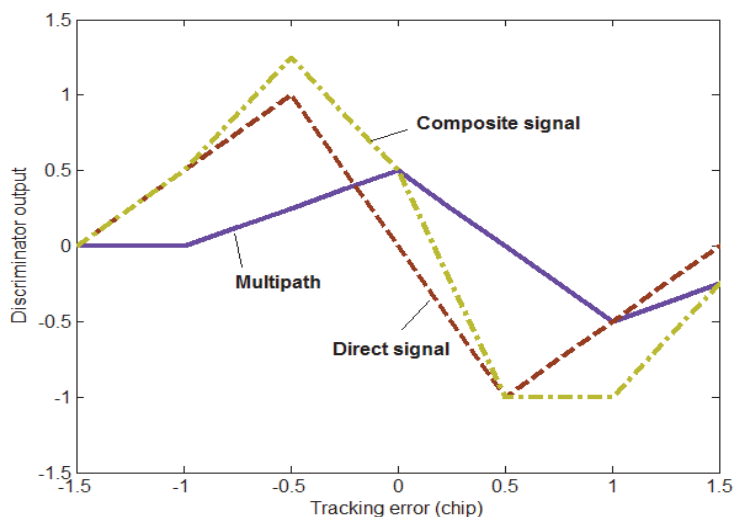


Fig. 1. Composite distorted of early-late discriminator.

Thus, pseudorange multipath analysis encompasses simulation of direct and indirect path signals and determination of zero-crossing of distorted discriminator function. There are three multipath parameters to consider: strength, delay and phase. The absolute value of each parameter is irrelevant. The upper and lower bounds of the multipath error can be determined, for a given multipath-to-direct ratio, by fixing the relative multipath phase at 0 and π radians, respectively, and varying the relative multipath delay. At each delay point, the distorted discriminator curve is determined and the resulting zero-crossing point and pseudorange error are calculated. The result of an example is presented in Fig. 2, which illustrates result of the theoretical multipath error envelope versus the multipath delay. The code autocorrelation sidelobes have been ignored. This simulation is offered in the case of 24 MHz bandwidth receiver filter, 1-chip, 0.5-chip, and 0.2-chip early-late (E-L) spacing and unaltered multipath amplitude. A conventional GPS receiver adopts a delay-lock loop with a 1-chip spacing between early and late correlators. The smaller E-L spacing is regarded as narrow-correlator architecture. Narrow-correlator receivers typically utilize spacings in the range of 0.05 to 0.2 PRN chips.

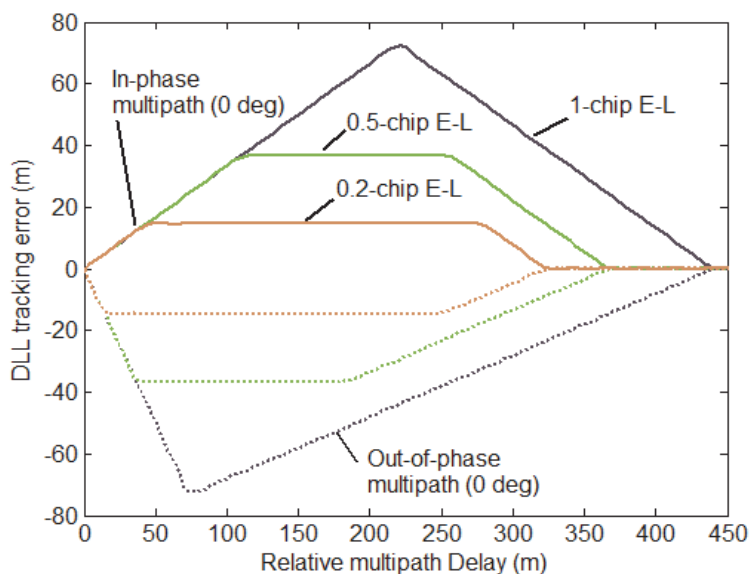


Fig. 2. Multipath error envelope for a conventional, 1-chip early-to-late (E-L), 0.5-chip E-L, and 0.2-chip E-L DLL receiver; Multipath component is half the strength of direct signal.

2.2 Signal model

A GPS receiver may receive a number of reflected signals and direct signal from the satellite. The error source of GPS consist of ionosphere delay, troposphere delay, receiver noise and multipath effect. Except for multipath, the other errors can be significantly decreased through advanced prediction and differential correction method. It is hard to depict the statistical model of the received signal in the presence of multipath. However,

many hypotheses can still be proposed. One hypothesis describes that the multipath signals are delayed with respect to direct GPS signal. Thus, let's consider only these reflected signals with a delay of less than one chip. This is because signals with a code delay larger than one chip are uncorrelated with the direct signals. Otherwise, the multipath signal is assumed to have the lower power than the direct one. The composite baseband signal, ignoring the navigation data bit, is given by

$$z[n] = \sum_{k=0}^M \alpha_k p(nT_s - \tau_k) \exp(-j(\omega nT_s + \theta_k)) + v[n] \quad (1)$$

where α_k , θ_k and τ_k denote amplitude, carrier phase, and code delay of k -th delayed signal. M represents the number of multipath component. $p(\bullet)$ indicates spread-spectrum code. ω denotes the IF angular frequency. The notation $z[n] = z(nT_s)$ is employed to denote a digital sequence sampled at the frequency $f_s = 1/T_s$ where T_s indicates period of sampling and n is the discrete time index. The 0-th delayed signal corresponds to the direct signal. $v[n]$ is modeled as white Gaussian noise distribution. The positioning error caused by the reception of multipath and direct signal is not only associated with the hardware design of receiver but also the detection algorithm. The literature review has provided several solutions for multipath effect. The following chapter will describe the proposed algorithm to counteract multipath.

2.3 Multiple correlator concept

The design of multi-correlator is seldom implemented due to the consideration of processing speed of hardware and cost. Owing to the promotion of hardware speed, decrease of cost and emergence of software wireless, the application of multi-correlator technique to receiver has become more prevalent. In fact, the strobe correlator described above is one of multi-correlator technique, which utilizes the linear combination of two correlators as discriminator output and adjusts chip spacing to track signal. Multi-correlator technique can depict the signal distribution after correlation process. In other words, this technique can present the process of correlation output in detail. Fig. 3 demonstrates the correlation output using 1 and 32 correlators, respectively. This figure illustrates that the multipath component can not be apparent if it adopts one set of correlator (early, prompt, and late). On the contrary, the 32 sets of correlator can better present the distribution of correlation output. Assume there are five correlators and the correlation of received signal is known. The linear combination of the five correlators can constitute received signal, which is expressed as:

$$v_j = \sum_{i=1}^Q r_i u_i \quad (2)$$

v denotes each measurement value of correlator, Q indicates the number of correlator, r is corresponding correlation value and u is the scaled value of correlation center itself. Take the five correlators as example. Assume five correlators are located at -0.5, -0.25, 0, 0.25, 0.5, respectively. The combination of five correlators can be employed to accomplish the measurement value of each correlator. Equation (1) is rewritten as follows:

$$\begin{aligned}
 \mathbf{v} &= \mathbf{R}\mathbf{\Lambda} \\
 \Rightarrow \begin{bmatrix} 0.5 \\ 0.75 \\ 1 \\ 0.75 \\ 0.5 \end{bmatrix} &= \begin{bmatrix} 1 & 0.75 & 0.5 & 0.25 & 0 \\ 0.75 & 1 & 0.75 & 0.5 & 0.25 \\ 0.5 & 0.75 & 1 & 0.75 & 0.5 \\ 0.25 & 0.5 & 0.75 & 1 & 0.75 \\ 0 & 0.25 & 0.5 & 0.75 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad (3)
 \end{aligned}$$

The makeup of $\mathbf{v}$ is constituted by third correlation (location as 0) because the $\mathbf{\Lambda}$ of the rest correlators is 0. The makeup of $\mathbf{R}$ is based on the location of selected correlator. With the $\mathbf{v}$ and $\mathbf{R}$ matrix known apriori, the magnitude of the signal in terms of the distribution set up by correlator can be known based on $\mathbf{\Lambda} = \mathbf{R}^{-1}\mathbf{v}$. The more the correlators, the clearer the distribution of the signal.

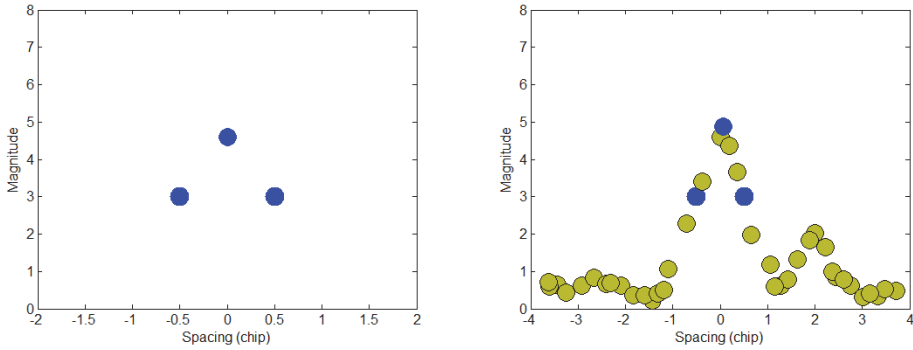


Fig. 3. Comparison of single correlator and multi-correlator.

It is known that the multi-correlator can depict the makeup of signal. Thus, we will see if multi-correlator can estimate direct signal with the direct signal plus multipath signal. Assume the multipath delay as 0.25 chip, signal magnitude as 0.5 and five correlators are shown as Fig. 4. Based on $\mathbf{\Lambda} = \mathbf{R}^{-1}\mathbf{v}$, the distribution of signal is known. Apparently, a correlation value exists between third and fourth correlator and the $\mathbf{\Lambda}$ of fourth correlator is lower. Using the negative correlation value from fourth correlator, we can eliminate multipath. Fig. 5 illustrates the multipath mitigation when the location of time delay is at the location of set multi-correlator.

$$\begin{aligned}
 \mathbf{\Lambda} &= \mathbf{R}^{-1}\mathbf{v} \\
 &= \begin{bmatrix} 1 & 0.75 & 0.5 & 0.25 & 0 \\ 0.75 & 1 & 0.75 & 0.5 & 0.25 \\ 0.5 & 0.75 & 1 & 0.75 & 0.5 \\ 0.25 & 0.5 & 0.75 & 1 & 0.75 \\ 0 & 0.25 & 0.5 & 0.75 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.625 \\ 1 \\ 1.375 \\ 1.25 \\ 0.875 \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0.5 \\ 0 \end{bmatrix} \quad (4)
 \end{aligned}$$

When the multipath delay is not at the set correlator, the calculated value after the above deduction approximates direct signal with little gap. Fig. 6 demonstrates the scenario when the location of multipath time delay is not at the location of set correlator.

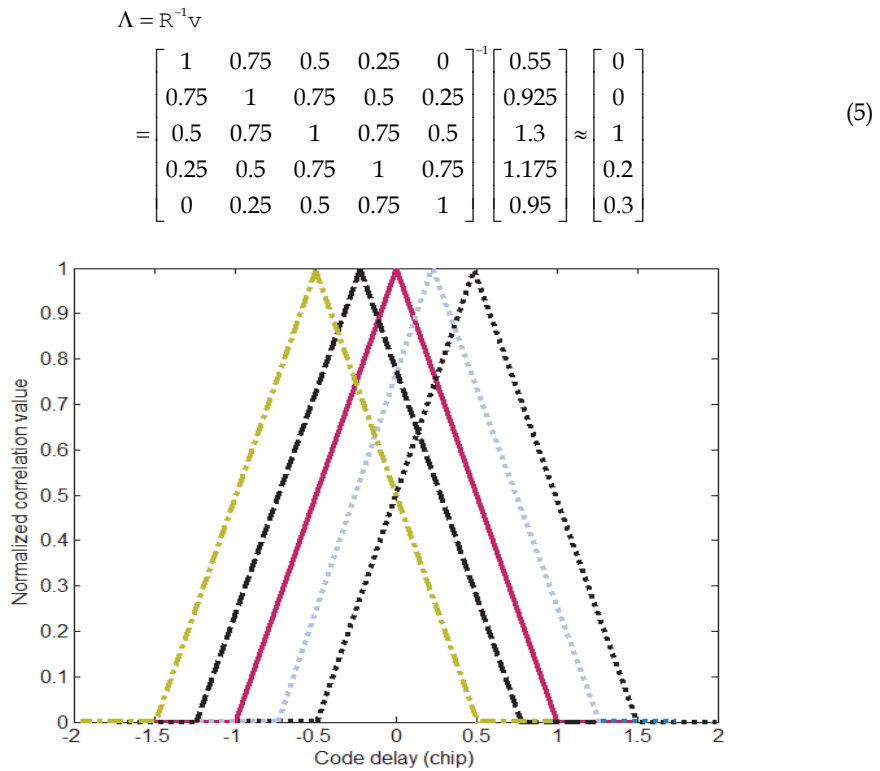


Fig. 4. Distribution of five correlators.

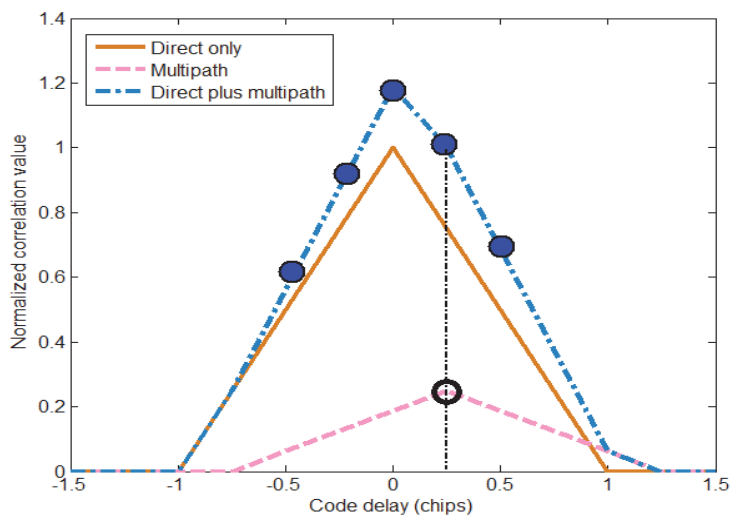


Fig. 5. Multipath delay is at the set correlator.

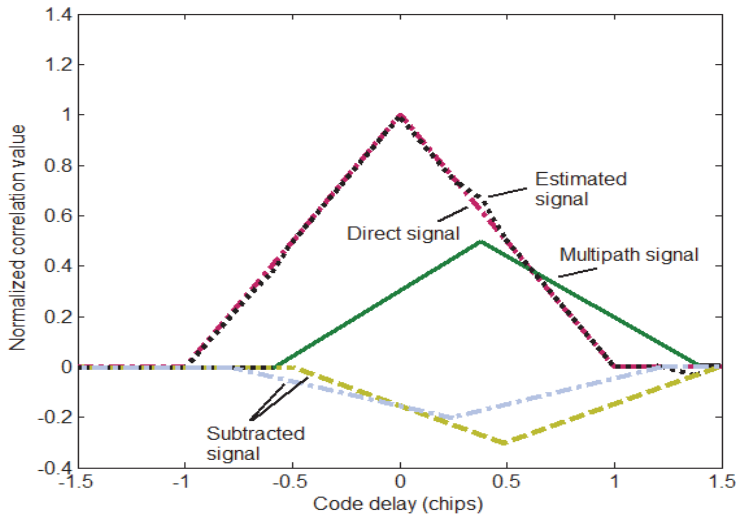


Fig. 6. Multipath delay is not at the set correlator.

2.4 Anti-multipath filter with multiple correlator

The previous chapter has clearly presented the advantage of multi-correlator method and its operation process. This chapter will elaborate how to constitute an anti-multipath filter based on multi-correlator. Fig. 5 shows the block diagram of the multipath mitigation system. The received signal is processed in a RF filter, then downconverted and sampled to a digital IF signal.

The tracking module consists of multiple correlator, code/carrier generator, discriminator and filter. The purpose of this module is to acquire accurate code phase and the carrier phase from PLL and DLL. The multipath estimator is used to estimate the correlation parameter of multipath, on the basis of the adaptive filter by employing duplicated signal and digital IF signal. Fig. 7 demonstrates that the estimated signal parameters are sent to the

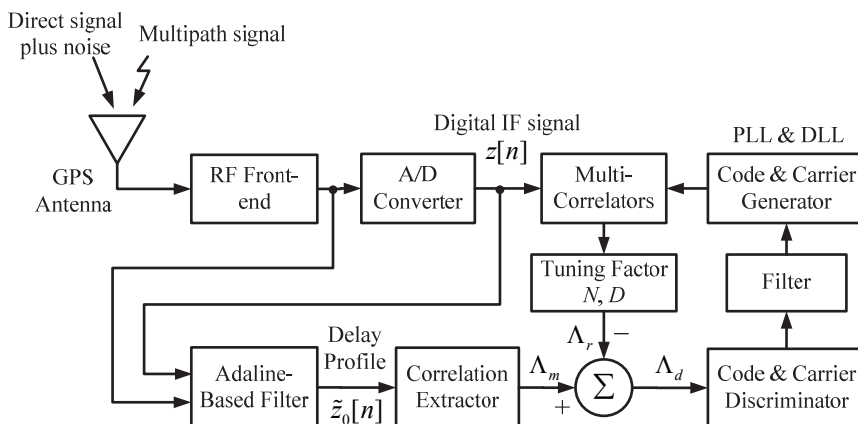


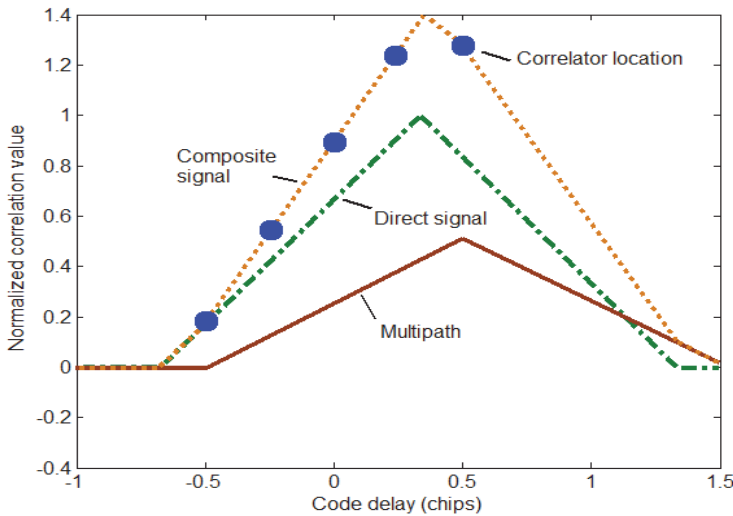
Fig. 7. Multipath Mitigation System Block Diagram.

correlation decomposer and the correlation value of multipath signal is determined in the multipath cancellation area.

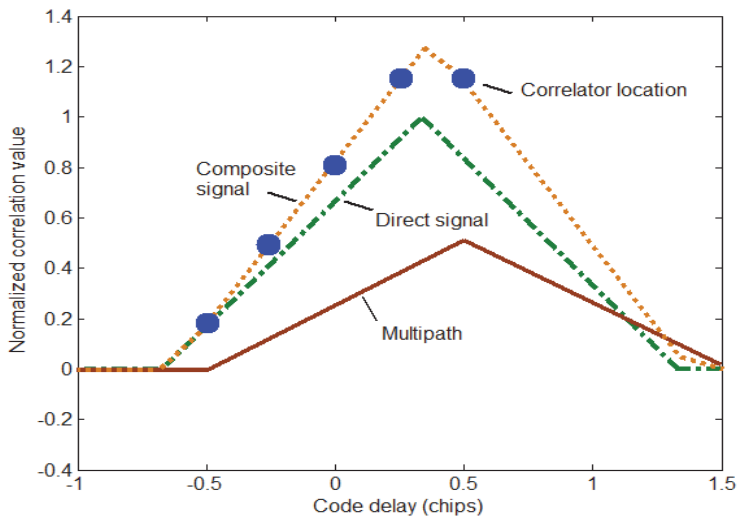
The estimated delayed signal is recreated at the Adaline-based filter and is subtracted from the correlation value of the received signal. The process of multi-correlators, multipath estimator, correlation value decomposer, and multipath cancellation will be elaborated in the following subsections.

2.4.1 Multi-correlators techniques

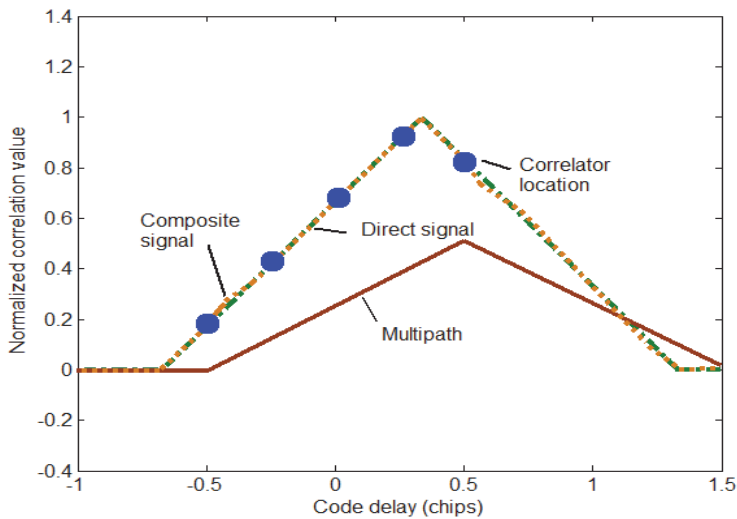
The concept of multi-correlator and the process of this method have been detailed in previous chapter. What we consider for the time being is that initial point of code delay of received signal and the local replica is not identical and multipath does not take place at the set correlator. Thus, parallel shift method is utilized to change the element of R matrix, such as shift the correlator location. Based on the simulation, assume the code shift of received signal as 0.3 chip and multipath delay as 0.5 chip. Using the above method, we add two variable as shift times N and shift range D , respectively. The purpose is to acquire the received direct signal and counteract multipath. The circle in red in the following figure are the code shift of direct signal, the shift times and range of correlator. The following will present the process. Fig. 8 (a) denotes the correlator output without shift operation. The color green is direct signal, the dark brown is multipath and the brown denotes composite signal. Afterward, the correlator is shifted 0.1 chip ($N=1$ and $D=0.1$), and Fig. 8 (b) is derived. However, this figure reveals that the performance does not meet our expectation. Fig 8(c) illustrates that after shift 0.3 chip ($N=1$ and $D=0.3$), the brown signal and green direct signal almost overlaps. The program is to simulate the location of set correlator in order to acquire v . Through the variation of R , multipath is mitigated. The result presents that acquired signal of correlator output using parallel shift method (shift 0.3 chip) is a more efficient strategy in multipath mitigation as opposed to shift 0.1 chip without shift.



(a)



(b)



(c)

Fig. 8. Multi-correlators technique simulation results. (a) $N=0$, $D=0.1$ (b) $N=1$, $D=0.1$ (c) $N=3$, $D=0.1$.

2.4.2 Adaline-based filter

The function of a multipath estimator is to estimate the multipath delay using Adaline-based filter, shown in Fig. 9. It adopts the tap-delay line with an Adaline network (Widrow and Hoff, 1960) to constitute this structure without a non-linear element. An adaptive

algorithm such as the LMS algorithm or the Back-Propagation (BP) learning algorithm is often employed to adjust the weights of the Adaline so that it responds accurately to as many patterns as possible in a training set. It is the simplest and most intelligent self-learning system which adapts itself to achieve an optimal solution (Rumelhart, D. E. et al, 1986). In this paper, the BP with an adaptive learning rate algorithm serves as a substitute for the LMS algorithm so as to prevent inherent limitations in LMS and to improve filter convergence rate (Schalkoff, R. J., 1997). The multipath estimator offers the multipath delay profile. Suppose the estimated digital IF signal is given by:

$$\tilde{z}[n] = \sum_{k=0}^{\tilde{M}} \tilde{\alpha}_k p(nT_s - \tilde{\tau}_k) \exp(-j(\omega nT_s + \tilde{\theta}_k)) + v[n] \quad (6)$$

Where the parameter with the symbol “ $\sim$ ” denotes the estimated parameter. A reference signal is a replica of code and carrier deriving from the output of DLL and PLL, which is shown as:

$$c_k[n] = p(nT_s - kd_r - \varepsilon_r) \exp(-j(\omega nT_s - \varepsilon_\theta)) \quad (k = 0, \dots, K) \quad (7)$$

where ε_r and ε_θ denote the measured group delay and carrier phase consisting of multipath error. d_r indicates sample period of delay of the multipath signals and Kd_r denotes the maximum delay of multipath signals. It is difficult to determine the parameters directly without any assumption about multipath signals. Thus, (6) is adopted in estimation process and modified by using reference signal and replacing $\tilde{M}$ with $\tilde{K}$ where the output signal of the filter is expressed as:

$$\tilde{z}[n] = \sum_{k=0}^{\tilde{K}} w_k c_k[n] + w_b c_b + \tilde{v}[n] \quad (8)$$

where $w_k = \tilde{\alpha}_k \exp(j\tilde{\theta}_k)$ represents the adjustable weight. The filter weight is employed to minimize the cost function, called squared error energy function and defined by using Equation (1) and (6):

$$L[n] = E\{(z[n] - \tilde{z}[n])(z[n] - \tilde{z}[n])^H\} \quad (9)$$

The filter minimizing the cost function is chosen by its tap weights to be the optimal solution to the normal equation (Haykin, S., 1986).

$$w_k^{\text{opt}} = C_k G_k^{-1}, \quad (10)$$

where C_k denotes the autocorrelation, $E\{x_i[n]x_k^H[n]\}$, of two reference signals ($c_i[n]$ and $c_k[n]$). g is the cross-correlation, $E\{z[n]c_i^H[n]\}$, of the digital IF signal $z[n]$ and reference signal $c_i[n]$. Where $E\{\cdot\}$ indicates an expectation operator. The filter solves (10) recursively using the BP with adaptive learning rate algorithm. This learning rule performs a gradient descent on the energy function to derive a minimum:

$$\begin{aligned} w_k[n] &= w_k[n-1] + \mu\delta[n-1]c_k[n-1], \\ w_b[n] &= w_b[n-1] + \mu\delta[n-1]c_b \end{aligned} \quad (11)$$

where $\delta[n]$ denotes the output layer error term. $\tilde{\alpha}_k$, $\tilde{\theta}_k$, and $\tilde{\tau}_k$ are estimated as the absolute value of weight $|w_k|$, the phase angle of weight $\arg(w_k)$ and the value of delay element kd_τ , respectively. The bias weight w_b connected to a constant input $c_b = +1$, effectively controls the input signal level of the filter. Note that the digital IF signal given in (1) is adopted as the desired signal and the output of DLL and PLL serves as the filter input signal. The reference signal is determined by (7) which generates the output of each delay element. Therefore, the estimated delay parameters from the filter weights and the delay element can be derived on condition that the learning algorithm has converged. The learning rate coefficient μ determines stability and convergence rate and a BP trained reference signal is adopted to obtain the minimum of (9) (Widrow 1986; Jacobs 1988). Suppose the learning rate is too large, the search path will oscillate about the desired path and converge more slowly than a direct descent. Nevertheless, the descent will progress in small steps if the learning rate is too small. It will greatly significantly increases the total time to convergence. Consequently, an adaptive coefficient where the value of μ is a function of the error derivation is adopted as the solution (Schalkoff, R. J., 1997).

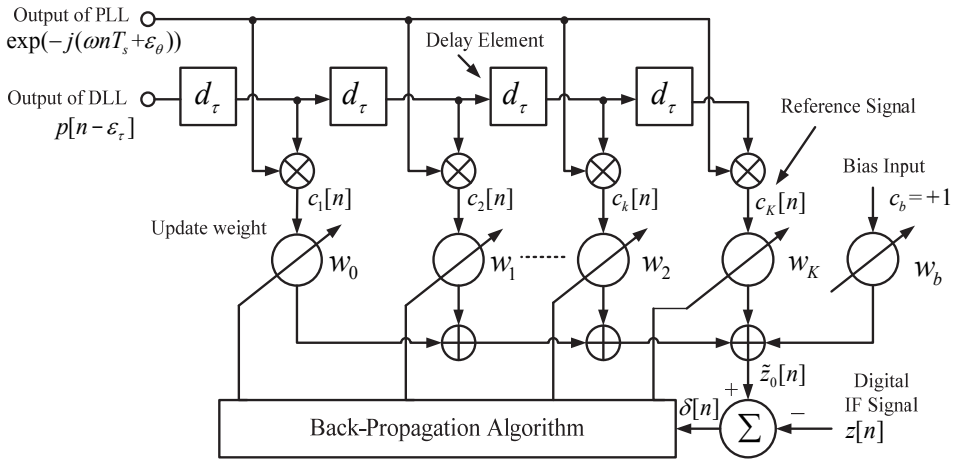


Fig. 9. Structure of the Adaline-based filter used in the multipath estimator.

2.4.3 Correlation extractor

After the use of adaptive filter, the estimated parameters can be obtained and the correlation decomposer divides the estimated parameters into multipath and direct signal. Besides, the autocorrelation function of multipath signals is subtracted from analog-to-digital (A/D) converter output of the received signal. In the decomposer process, it is assumed that the values of the first peak amplitude tap weight are the direct signal and the remainders are multipath signals. Fig. 5 presents an example where the direct signals refers to the first peak $k = l$ and the multipath signal amplitude denotes the remnants $l < k \leq K$. Suppose that the multipath channel has a decreasing power delay profile. The multipath signal parameter is adopted to calculate the correlation value. The correlation equation of estimated multipath signals with amplitude $\tilde{\alpha}_k$, delay $\tilde{\tau}_k$ and carrier phase $\tilde{\theta}_k$ is written as:

$$\Lambda_k(\tau) = \tilde{A}_k C(\tau - \tilde{\tau}_k) \cos(\tilde{\theta}_l - \tilde{\theta}_k), \quad (12)$$

where $C(\tau)$ denotes the autocorrelation function of the GPS PRN (Pseudo Random Noise) code signal.

$$E\{p[n]p[n-\tau]\} \quad (13)$$

Hence, the entire correlation value of the estimated multipath signal $\Lambda_k(\tau)$ is given by:

$$\Lambda_m(\tau) = \sum_{k=l+1}^K \Lambda_k(\tau) \quad (14)$$

2.4.4 Multipath removal

The entire correlation values of multipath signal Λ_m are subtracted from the correlation value of received signal Λ_r and the output of correlation value Λ_d is expressed as:

$$\Lambda_d(\tau) = \Lambda_r(\tau) - \Lambda_m(\tau) \quad (15)$$

In (12), the estimated correlation Λ_r of direct signal can be acquired using multi-correlator technique. Such a technique has been detailed in chapter 2.3.1. Multi-correlator technique can effectively estimate the correlation of direct signal and counteract multipath simultaneously. It can promote the convergence speed of Adaline-based filter.

The tracking error takes place in DLL and PLL due to the multipath effect. The effect primarily results from distortion of the correlation function receiving the IF signal, shown in Fig. 8(a), which illustrates the normalized correlation function with multipath effect. Fig. 8(a) presents that the symmetry is lost and the propagation delay is difficult to estimate. Thus, the range measurement accuracy is diminished. Nevertheless, the use of a subtractive method offers multipath mitigation in the tracking loop and the output $\Lambda_m(\tau)$ enables the tracking loops to track direct signal accurately.

The above processes: the estimate process, the correlation extractor and the cancellation method can counteract the multipath effects regarding the autocorrelation function of the received signal, since the tracking errors in DLL and PLL are not completely eliminated. Provided that the reference signal acquires the multipath error, the estimated parameters do not present accurately that of the real multipath. So as to obtain the ideal estimated parameters, the BP learning process is recursively employed. The use of multi-correlator technique can speed up BP learning process and enhance its performance.

3. Performance analysis and simulation results

In this section, computer simulations are performed to evaluate the performance of proposed method. To compare with other published methods in performance, the multipath tracking error envelopes in code and carrier phase for a multipath signal amplitude of half the LOS amplitude is denoted as $\alpha_0 = 1.0$ and $\alpha_1 = 0.5$. A GPS multipath model includes one direct signal and one delayed signal. Suppose that a high post signal to noise ratio (SNR) of 10 dB is located in this model. Simulation results are demonstrated in infinite bandwidth situation.

3.1 Simulation parameter

The digital IF frequency of a GPS signal is $\omega / 2\pi = 1.25$ MHz and the sampling rate is 5 MHz. The delay chip of the multipath signal varies from 0 to 1.5 chips with the phase of 0 and π radians with regard to the direct signal. In conventional correlator simulations, code phase error and carrier phase error are computed with 1 chip early-late discriminator. The chip spacing of a narrow correlator is less than 1 chip. A spacing of 0.2 chips utilized to serve the discriminator functions. Two different narrow correlator discriminators are adopted in a strobe correlator and the chip spacing of the two narrow correlators can be adjusted to 0.1 and 0.2 chip. The same parameters are also adopted in both enhanced strobe and edge correlators. The E1/E2 tracker of the two correlators is located at $E1 = -0.55$ and $E2 = -0.45$ with 0.1 chip spacing (Irsigler, M. et al, 2003). The Adaline-based adaptive filter method with the parameter of tap delay $d_r = 0.01$ chip, 0.1 chip, 0.5 chip and its 5 delayed tap is employed as the input to the filter. The number of multi-correlator is set as five. The initial learning rate is 0.05, the number of training samples is 5000 at 1ms C/A code period and the weights are initialized to 1. The performance is assessed on a separate test set of 100ms samples measured at intervals of 1ms samples during the adaptive process.

3.2 Performance analysis and comparison

With regard to crucial multipath mitigation techniques of internal receiver, the multipath performance of these correlation techniques will be compared with each other, including the proposed method. Thus, the envelopes of all techniques described above are plotted into the same diagram to make a comprehensive comparison of multipath mitigation performance

Figs. 10-12 compare the error envelopes of the code phase and carrier phase for all of the multipath mitigation techniques. Simulation results show that the proposed method with multipath delay at the location of set correlator as $d_r = 0.5$ chip case has both the best multipath mitigation performance. Assume the location of correlator is not at multipath delay ($d_r = 0.1$ chip), it also presents good performance. The conventional PLL has a maximum 0.52 radians in carrier phase error. Therefore, the use of the conventional correlator can yield very large maximum multipath errors and reveals the worst mitigation performance. The same results hold true for both narrow and edge correlators. Note that since the narrow, the MEDLL, the edge and strobe correlators do not provide any carrier phase elimination, their sensitivity to multipath is almost the same as the 1-chip early-late correlator. Only slight differences can be observed due to differences in their code multipath mitigation.

These figures indicates that through the use of proposed method in combination with multi-correlator technique with a delay element $d_r = 0.5$ chip, both code and carrier phase errors are reduced in the range of delay from 0 through 1.5 chip. In contrast, through the adoption of the proposed method with a tap delay $d_r = 0.5$, the code and carrier phase error decrease significantly in the range of delay from 0 to 1.5 chip. The figure shows that the use of multi-correlator technique can effectively reduce code phase error. Nevertheless, for carrier phase error, its performance remains the same. The reason is because carrier phase error is not related to multi-correlator. In the case of the tap delay $d_r = 0.5$, multipath mitigation performance degrades in comparison with the case of $d_r = 0.1$. This is because to the accuracy of the estimated delay profile in the Adaline-based filter depends on the tap delay d_r . The smaller the d_r the better the performance of multipath mitigation. In the case of the

$d_t=0.5$ chip, the multipath mitigation performance degrades in code phase error and the carrier phase error also exceeds that of the conventional tracking loop. Though the use of a small tap delay can yield high performance in multipath mitigation, it also takes large computation cost to estimate delay profiles. However, the use of multi-correlator can save computation cost to estimate delay profiles.

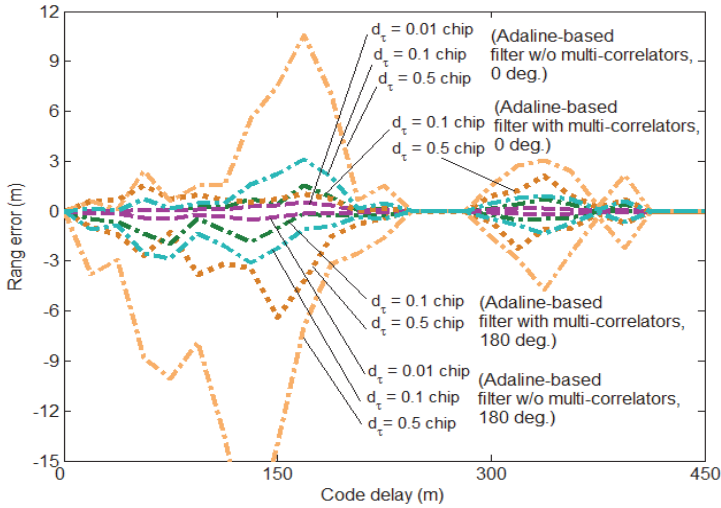


Fig. 10. Code-phase error simulation results of proposed method. ($\alpha_0 = 1.0$, $\alpha_1 = 0.5$, $\tau_0 = 0$ chip, $\tau_1 = 0 \sim 1.5$ chip, $\theta_0 = 0^\circ$, ($\theta_1 = 0^\circ, 180^\circ$); delay element $d_t = 0.01$ chip, 0.1 chip, and 0.5 chip with and without multi-correlators.)

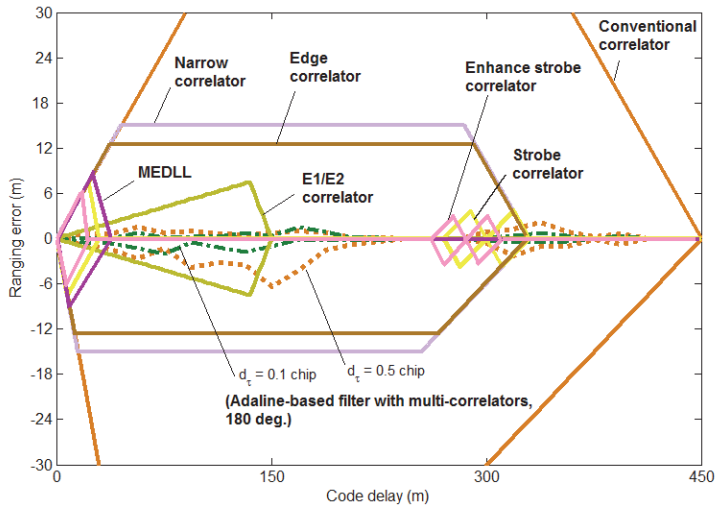


Fig. 11. Code-phase error simulation results of existing methods. ($\alpha_0 = 1.0$, $\alpha_1 = 0.5$, $\tau_0 = 0$ chip, $\tau_1 = 0 \sim 1.5$ chip, $\theta_0 = 0^\circ$, ($\theta_1 = 0^\circ, 180^\circ$)).

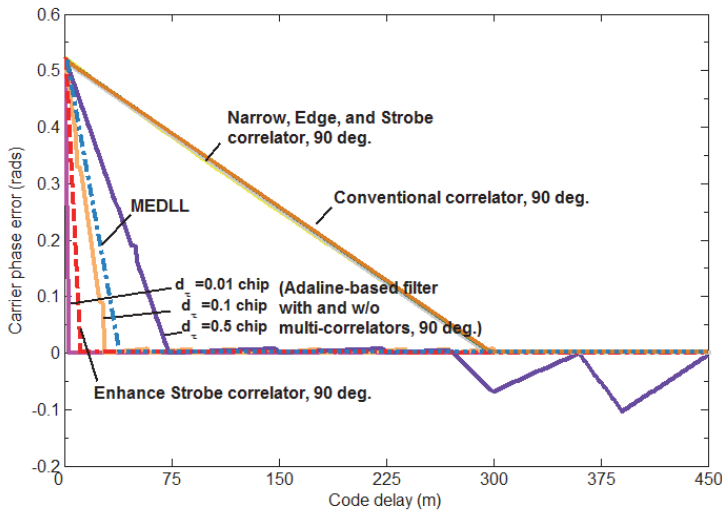


Fig. 12. Carrier phase error simulation results. ($\alpha_0 = 1.0$, $\alpha_1 = 0.5$, $\tau_0 = 0$ chip, $\tau_1 = 0 \sim 1.5$ chip, $\theta_0 = 0^\circ$, ($\theta_1 = 0^\circ, 180^\circ$)); delay element $d_\tau = 0.01$ chip, 0.1 chip, and 0.5 chip with and without multi-correlators.)

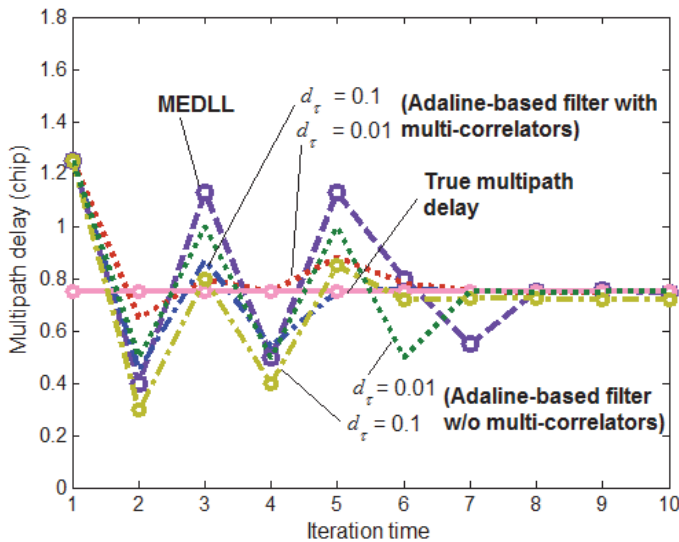


Fig. 13. Delay estimated by MEDLL and proposed method with and without multi-correlators.

Note that Fig. 11 reveals that every DLL structure lacks of performance enhancement for short delay multipath signals. Nevertheless, the proposed method with multi-correlators can perform better in short delay and medium-to-long-delay multipath environment.

Suppose a given application involves short delay and medium-to-long-delay multipath, then the best correlation techniques such as the enhanced strobe correlator will not outperform the proposed method of this paper.

To accomplish estimated performance of proposed method, the desired correct of multipath delay profiles are set $\alpha_0 = 1.0$, $\alpha_1 = 0.5$, $\tau_0 = 0$, $\tau_1 = 0.75$, $\theta_0 = 0^\circ$ and $\theta_1 = 0^\circ$. The delay element number is five and the number of multi-correlator is set as five. An estimated multipath delay versus the true multipath delay curve for three considered algorithms, the MEDLL and the Adaline-based filter with and without multiple correlators, is shown in Fig. 13. Note that the proposed method with and without multi-correlator technique of $d_r = 0.01$ has the faster convergence rate than the MEDLL. The Adaline-based filter without multi-correlator technique is rapid in convergence rate with $d_r = 0.1$. However, it suffers from a steady state error 0.03 chip in delayed estimation. Nevertheless, the use of five multi-correlators with shift 0.1 chip, the error approximates zero.

Several concessions exist in these architectures such as: noise performance, code versus carrier performance, a priori information needed as input, short delay performance and hardware/software complexity. These factors are compared in Table 1. This performance comparison is on the basis of the published methods and simulation results of this paper. The research analysis is shown in the following:

Concerning the noise mitigation performance, when SNR = -10dB, the simulation result presents that the narrow correlator and proposed method with multi-correlators are the best in performance with the code tracking error of about 0.034 chip and 0.04 chip, respectively. The proposed method without multi-correlators in this paper is medium in performance with the tracking error of around 0.05~0.1 chip, which equals the medium noise performance of the edge and E1/E2 correlator. In contrast, the conventional correlator, strobe, enhanced strobe correlator and the MEDLL are inferior in noise performance, with the tracking error around 0.2 chip.

In term of the GPS mobile applications, high precision is required even at the expense of slightly increased complexity. The best options are the enhanced strobe correlator and the Adaline-based filter. The proposed method has the best performance in multipath mitigation. Nevertheless, its hardware complexity, such as the number of required multiplications per delay estimate is on the order of $O[N_{iter}(Kd_r)^3]$ where N_{iter} is the number of filter iterations and Kd_r is an estimate of the maximum delay spread of the channel in the samples. The high complexity of this method is principally due to the matrix inversion operations. However, in short-delay and medium-to-long delay multipath environments, the number of delay samples Kd_r are smaller. Thus, the complexity of the Adaline-based filter is not very high. The enhanced strobe correlator has lower complexity, on the order of $O[(Kd_r)^2]$, but it does not perform as well as Adaline-based filter. With regard to design perspective, the best tradeoff between accuracy and complexity should be determined based on estimated maximum delay spread of the channel.

Concerning the conventional receiver design for civilian application, the lowest complexity solutions of the 1-chip E-L correlator and the narrow correlator, appear to be the best choice. What is more, complexity is the top priority and is emphasized more than performance in the design of a receiver provided that no significant degradation in performance occurs. All of the conventional, strobe and narrow correlator designs have least medium performance and reduced complexity in multipath scenarios. Hence, they are viable options for a low

complexity receiver. In comparison, even though the edge, the E1/E2 and the MEDLL designs are higher in cost, they are better than the conventional, narrow, and strobe correlators in performance.

In fact, there are inherent limitations in almost every technique. Note that the combined features of proposed method prevails over those of other techniques. In addition, the condition of short-delay and medium-to-long-delay multipath renders the effect of hardware complexity in Adaline-based filter insignificant. Consequently, the proposed method is a well-suited and well-balanced application in multipath mitigation.

	Conventional correlator	Narrow Strobe	Enhanced strobe	Edge	E1/E2	MEDLL	Adaline-based filter		
							Without multi-correlators	With multi-correlators	
Code multipath performance	×	△	△	○	△	○	○	✓	✓
Carrier multipath performance	×	×	×	△	×	×	○	△ (Count on number d_r)	✓
Short-delay multipath performance	×	×	×	△	×	○	○	✓	✓
Medium-to-long-delay multipath performance	×	×	○	○	△	✓	✓	△	✓
A priori information	Yes (Coarse delay)	Yes (Coarse delay)	Yes (Coarse delay)	Yes (Coarse delay)	Yes (Coarse delay)	Yes (Coarse delay)	Reference function	None	None
Noise Performance (SNR= -10dB)	×	○	×	×	△	△	△	△	○
	(above 0.2 chip error)	(0.034 chip error)	(0.2-0.25 chip error)	(below 0.2 chip error)	(0.054 chip error)	(0.04-0.06 chip error)	(below 0.18 chip error)	(0.05-0.1 chip error)	(0.03-0.05 chip error)
Hardware complexity	☺	☺	☺	☺	☺	☺	☹	☹ (Count on number of iteration)	☹
Software complexity	☺	☺	☺	☹	☺	☺	☺	☺	☺

✓: Best ○: Good △: Fair ×: Poor

Table 1. Comparative performance of multipath mitigation techniques.

4. Conclusion

Multipath is the primary error source in high-precision-based GNSS applications and is also a significant error source in non-differential applications. Various receiver designs have been on the market and claim various multipath mitigation functions. Most of these techniques can be characterized either as discriminator function shaping or correlation function shaping. In this

paper, an Adaline-based filter with multi-correlators method is adopted in multipath mitigation for GNSS application. A simplified direct plus multipath signal model is employed in this simulation. This approach enhances the performance of code phase and carrier phase errors compared with other published methods. Simulation results demonstrates that the proposed method is a viable and effective solution to increase the positioning accuracy for GNSS navigation in the presence of short-delay and medium-to-long-delay multipath environment.

5. Acknowledge

The author would like to thank the National Science Council of Taiwan (R.O.C.) for their support of this work under grant NSC 99-2221-E-020-036.

6. References

- Chang, C. L., & Juang, J. C. (2008). An Adaptive Multipath Mitigation Filter for GNSS Applications," *EURASIP Journal on Advances in Signal Processing*, Vol. 2008, Article ID 214815, 10 pages
- Garin, L., van Diggelen, F., & Rousseau, J. (1996). Strobe and Edge Correlator Multipath Rejection for Code, in *Proc. ION GPS-96*, pp. 657-664, Kansas City, MO, September 17-20
- Garin, L. and Rousseau, J. (1997). Enhanced Strobe Correlator Multipath Rejection for Code and Carrier, in *Proc. ION GPS-97*, pp. 559-568, Kansas City, MO, September 16-19
- Hagerman, L. L. (1973). Effects of Multipath on Coherent and Non-coherent PRN Ranging Receiver, *Aerospace Report No. TOR-0073 (3020-03)-3*, Development Planning Division, The Aerospace Corporation
- Haykin, S. (1986) *Adaptive Filter Theory*, ISBN 013-0484-34-2, Prentice Hall, USA.
- Irsigler, M., & Eissfeller, B. (2003). Comparison of multipath mitigation techniques with consideration of future signal structures, in *Proc. ION-GPS/GNSS*, pp. 2584-2592, Portland, OR, USA, September 9-12
- Jacobs, R. A. (1988). *Increased Rates of Convergence Through Learning Rate Adaptation*, *Neural Networks*, Vol. 1, pp. 295-307
- Mattos, P. (1996). Multipath Elimination for the Low-Cost Consumer GPS, in *Proc. ION GPS-96*, pp. 665-671, Kansas City, September 1996.
- Rumelhart, D. E., Hinton, G.E., & Williams, R. J. (1986). Learning Internal Representations by Error Propagation, *Parallel Distributed Processing*, Vol. 1, MIT Press, Cambridge, MA, USA, pp. 318-362
- Schalkoff, R. J. (1997). *Artificial Neural Networks*, ISBN 007-1155-54-6, McGraw-Hill.
- Sleewaegen, J.-M., & Boon, F. (2001). Mitigating Short-delay Multipath: A Promising New Technique, pp. 204-213, in *Proc. ION GPS 2001*, Salt Lake City, UT, USA
- Stone, J., & Chansarkar, M. (2004). Anti-multipath Triangulation (AMT) for Positioning in Dense Urban Environments, in *Proc. ION GNSS*, pp. 1165-1168, Long Beach, CA, September 21-24
- van Dierendonck, J., Fenton, P., & Ford, T. (Fall 1992). Theory and Performance of Narrow Correlator Spacing in GPS Receiver, *Navigation: Journal of the Institute of Navigation*, Vol. 39, No. 3, pp.265-283

- van Dierendonck, A. J., & Braasch, M. S. (1997). Evaluation of GNSS Receiver Correlation Processing Techniques for Multipath and Noise Mitigation, in *Proc. ION-NTM*, pp. 207–215, Santa Monica, CA, USA, January 14–16
- van Nee, R. D. J. (1992). The Multipath Estimating Delay Lock Loop, in *Proc. IEEE Second Symposium on Spread Spectrum Techniques and Applications*, pp. 39–42
- van Nee, R. D. J., Sierveld, J., Fenton, P. C., & Townsend, B. R. (1994). The Multipath Estimating Delay Lock Loop: Approaching Theoretical Accuracy Limits, in *Proc. IEEE PLANS*, pp. 246–251, Las Vegas, Nev, USA, April 11–15
- Widrow, B., & Hoff, M. E. (1960) Adaptive Switch Circuits. In *IRE WESCON Convention Record*, pp. 96–104, New York, USA
- Widrow, B., & Lehr, M. A. (1990). 30 Years of Adaptive Neural Networks: Perceptron, Madaline, and BP., *Proc. IEEE*, pp. 1550–1560