ABSTRACT

Voice Activity Detection with Stochastic Resonance

by

İlker Özçelik

Stochastic resonance methods have recently been shown to improve the performance of certain suboptimal signal processing systems. In this thesis, we apply stochastic resonance to speech signal processing. In particular, we present a method to improve detection performance of a suboptimal voice activity detector without changing it. In our experiments, we use Sohn and Sung's voice activity detector as our reference detector. This detector was designed by using the Generalized Likelihood Ratio Test (GLRT) method under the assumptions that speech and noise signals are Gaussian random processes that are independent of each other and the Discrete Fourier Transform (DFT) coefficients of each process are asymptotically independent Gaussian random variables. However, recent studies have shown that Laplacian and Gamma distributions more accurately represent DFT coefficients of clean speech and noise signals. Therefore, Sohn and Sung's voice activity detector is suboptimal because of its underlying design assumptions and can be improved. In this study, in order to improve detection performance, the input signal of the detector is preprocessed using the bistable SR system used as an SR filter. Optimum SR filter parameters are obtained by using the deflection coefficient. Due to the high complexity of the signal of interest, the coefficients are found in an iterative manner. Experiments conducted on different input signals and training data sets showed that it is possible to get optimum parameters, even when only 20% of the input signal is used as training data, if it has enough information about the distribution of input signal. The detector performance of the system before and after the SR filter is compared using the Receiver Operating Curves (ROC). From the ROCs, we observe that our method improved detection performance up to 17.5% for lower false alarm rates and up to 4.5% for higher false alarm rates. Based on the simulation results, our method is an efficient method to improve detection performance of suboptimal voice activity detector.

Voice Activity Detection with Stochastic Resonance

By

İlker Özçelik

B. S. in Electronics and Communication Ed., Marmara University, Turkey, 2005

MASTER THESIS

Submitted in partial fulfillment of the requirements for the
degree of Master of Science in Electrical Engineering
in the Graduate School of Syracuse University

June 2010

Approved ______________________________
Advisor: Prof. Pramod K. Varshney

Date ______________________________

Approved ______________________________
Co-advisor: Prof. Hao Chen

Date ______________________________

Contents

Illustrations

Tables

Dedication

To my family.

Acknowledgments

I would first like to thank my advisor Professor Pramod K. Varshney for his invaluable guidance and support throughout my thesis process. He taught me the importance of patience and perseverance, as well as the value of a strong support system. With his acceptance into his research group, I found a strong community within which I could acclimate to a new academic environment and challenge myself to complete this thesis project. I am extremely grateful for these lessons and opportunities Professor Varshney has afforded me. Next, I would like to thank my co-advisor Professor Hao Chen. His knowledge and willingness to both listen and troubleshoot with me helped me stay motivated and positive throughout my thesis project. Additionally, I would like to thank my thesis defense committee members Professor Ruixin Niu, and Professor Utpal Roy.

I also wish to thank the many friends I made during my graduate study. Your willingness to listen to me, troubleshoot with me, and motivate me has greatly contributed to my success. I would like to especially thank, Engin Masazade, whose support, guidance, and friendship has been an unbelievable asset throughout this process; and will not be forgotten.

For making it possible for me to study abroad I would like to thank the Republic of Turkey Ministry of National Education for the financial support that will allow me to share my knowledge and learning experiences with future Turkish students for years to come. I would also like to thank the workers at the Education Attache in New York City for their patience and kindness while answering my many questions throughout my graduate study.

Finally and most importantly, I would like to thank my family. Thank you for your endless words of encouragement, and for always believing in me. Without your faith, love, and support none of this would have been possible.

Chapter 1

Introduction

1.1 Background

In today's developing world, in a quest to meet the needs of people, researchers study challenging "real life" problems everyday. Scientific knowledge plays a very important key role both in solving these problems and in establishing necessary background for future problems. However, these problems are not always easy to understand and there isn't always existing knowledge to provide a straightforward solution. Thus, both during the evolution of scientific knowledge and in the development of new technologies, assumptions and hypotheses are often used. It is very important to make reasonable and consistent assumptions while solving problems or designing systems, in order to get good results.

As previously stated, assumptions are a necessity in solving real life problems with limited knowledge. On the other hand, assumptions serve only as reasonable substitutes of complex phenomena. If they are not completely correct, systems designed based on these assumptions will be suboptimal. Only a complete understanding of these phenomena will lead us to optimum solutions. Therefore, until obtaining the optimum solutions, one could replace the systems with increasingly better ones which would arguably be a continual process. This, however, is not cost or time efficient, and is thus a less desirable option. This motivates the question, “ Is it possible to improve suboptimal system performance without making major changes?”

Recent research [2],[3],[4] shows that under suitable conditions, stochastic resonance (SR) can be a promising candidate to make performance improvements of suboptimal systems. Stochastic resonance is defined as “a nonlinear physical phenomenon in which the output signals of some nonlinear systems can be enhanced by adding suitable noise under certain conditions.” [2]. Specifically, stochastic resonance is a way to improve system performance under certain conditions by changing the input signal, instead of changing the system itself.

The term Stochastic Resonance was first used by Benzi *et al* in 1981 while exploring as to why ice ages in the Earth occur periodically. Two years later, Fauve *et al* reported that the SR phenomenon was observed in their Shcmitt trigger experiment [5]. Stochastic Resonance has been both observed and studied in many physical and biological nonlinear systems in the past three decades. Researchers have developed many explanations for this phenomenon. In the beginning, stochastic resonance was only observed in bistable systems. Next in the mid nineties, it was further investigated in neural and excitable systems. Then, the SR concept which was investigated only for weak periodic signals, was generalized to aperiodic signals. Different distributions of additive noise were also investigated besides the Gaussian distribution.

More recently, stochastic resonance has received a lot of attention in both communication and signal processing areas. Many image processing [6],[7], and weak signal detection problems [8],[9],[10],[11] have been investigated using stochastic resonance. In this thesis, we use stochastic resonance to improve detection performance of an existing voice activity detector, without any detector manipulation.

1.2 Motivation

Voice activity detection (VAD) plays a very important role in the areas of audio signal processing and communication. By distinguishing the presence or absence of the speech signal, noise statistics of the input signal can be obtained and used while designing robust speech recognition systems. Also communication channels can be used more efficiently by using speech absent time frames for sending other speech signals. Researchers have proposed many different types of voice activity detectors thus far. Earlier studies of voice activity detection focused on the various characteristic features of audio signals. Features such as energy, periodicity and zero crossing rate comprised the basis for the research studies. The focus of recent studies has shifted toward the statistical properties of signals [12],[13],[14],[15].

In many statistical VAD studies, the design of the detectors was based on the assumption that “Speech and noise signals were Gaussian random processes that were independent of one another and discrete fourier transform (DFT) coefficients of each processes were asymptotically independent Gaussian random variables” [12], [15], [16]. However, later investigations proved this to be wrong. Recently it has been reported that Laplacian and Gamma distributions more appropriately represent signal and noise DFT coefficient distributions [13]. It has been shown that detectors which are designed based on these distributions [13], [17] exhibit better performance. It is also important to note that, when considering the varying environmental conditions, one cannot assume a fixed distribution for the noise signal. Therefore, the proposed detectors are only optimum if the underlying assumptions are completely correct, which in general is not the case.

In order to get better detection performance, one can replace the existing detector with a better one that is matched better with the underlying statistics. However, sometimes replacing the system may not be easy or even possible. Therefore, as previously mentioned, stochastic resonance may be used to improve the performance of suboptimal systems under certain conditions. In our study, we use bistable SR model to improve Sohn and Sung's voice activity detector's [12] detection performance without making any changes in the detector.

1.3 Problem Definition

As we stated in the previous section, voice activity detectors are designed based on different signal and noise distribution assumptions. These assumptions are corrected with a better understanding of the signal of interest. In the beginning, it was assumed that Gaussian distribution can be used to represent DFT coefficient distributions of these signals, but recent research showed that these coefficients can be represented more accurately with Gamma and Laplacian distributions. It is obvious that these systems will give optimum performance only if the underlying assumptions are completely correct. However, we do not currently have a complete understanding of these signals. It should be noted that different environmental noises have different dominant frequency components, thus it is not correct to assume that different kinds of environmental noises have the same distribution. Therefore, it is possible to improve the performance of these suboptimal systems.

We propose a method to improve suboptimal voice activity detector performance with stochastic resonance. Instead of changing the detector's structure, we preprocess the input signal by using bistable stochastic resonance. The system is tested with different input data sets under different environmental noises and signal to noise ratios (SNR).

1.4 Voice Activity Detection

Environmental noise always affects speech processing systems. In order to overcome this problem and improve the system performance, noise statistic (σ^2) should be estimated. Voice activity detectors (VAD) help to estimate noise statistics by detecting noise only frames in the input signal. Voice activity detection is not limited to help estimate noise statistics, it can also be used to help Robust Speech Recognition, Discontinuous transmission and so on. Ramirez *et al* defined voice activity detection as,"a technique used in speech processing in which the presence or absence of human speech is detected"[18].

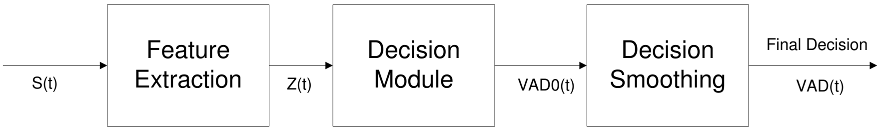

Figure 1.1 : Voice activity detector block diagram

Frame based detection is performed in voice activity detectors. In order to decide the presence or absence of the speech signal, the input signal is split into time frames where the signal is assumed to be stationary. Then, each frame is evaluated in the detector consecutively. A block diagram of VAD is shown in Fig. 1.1. A typical VAD consists of three blocks: feature extraction, decision and decision smoothing. In the feature extraction block, one or more input signal features are obtained and decision variables are generated. Then, in the decision block, these decision variables are evaluated by a decision rule. Here the presence or absence of the speech signal is decided. Finally, some of the errors which are made in the decision block are corrected in the decision smoothing block. If the detector decides that a speech signal is present it is called an *Active* frame, otherwise it is called an *Inactive* frame. The results of a three seconds speech signal voice activity detection can be seen in Fig.1.2.

Some of the most commonly used voice activity detection methods are based

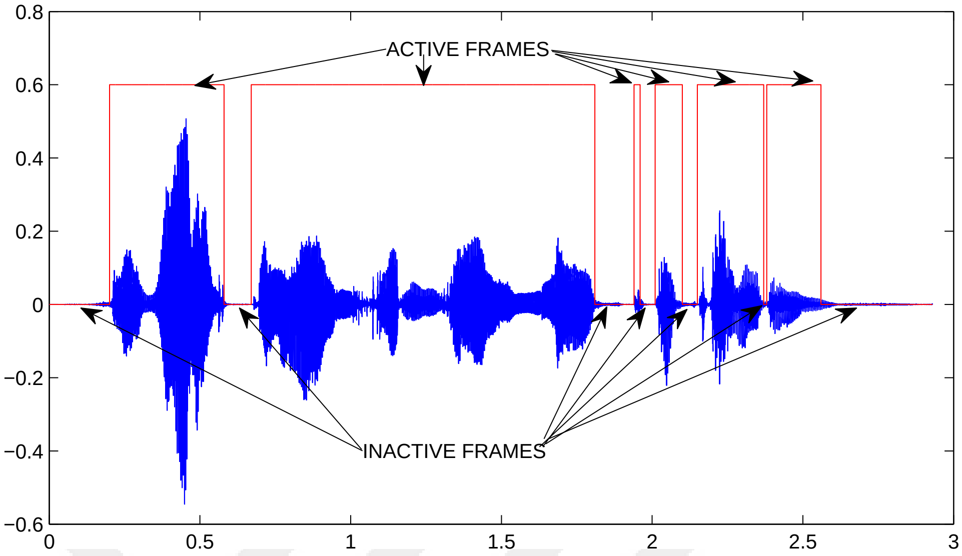

Figure 1.2 : Active and Inactive frames.

on[18]:

- Energy thresholds
- Pitch detection
- Spectrum analysis
- Zero crossing
- Periodicity
- Statistical based

Many different VAD algorithms have been proposed by using these methods. Some of the desirable aspects of a VAD algorithm can be listed as[14]:

- A good decision rule

- Adaptability to changing background
- Low computational complexity (for real time applications)

VAD algorithms can be divided in two broad categories: Time Domain VAD Algorithms and Frequency Domain VAD Algorithms.

1.4.1 Time Domain VAD Algorithms

In time domain VAD algorithms, decision variables of the detector are calculated in the time domain. Energy level and zero crossing rate are the most commonly used features while deciding the presence or absence of speech. These algorithms are relatively faster than frequency domain VAD algorithms. Some of these algorithms can be listed as: linear energy based, adaptive linear energy based and weak fricatives (zero crossing)

Linear Energy Based Detector

In a linear energy based detector, the energy of the frame is used as the decision variable when making the decision regarding the presence or absence of the speech signal. The energy level of the frame is compared to a threshold to decide whether it is an Active or an Inactive frame. If the background noise is not stationary, then the use of an adaptive threshold is necessary. The rule for updating the threshold can be found in [14] as,

$$tr_{(new)} = (1-p)tr_{(old)} + pE_{silence} \qquad (0 < p < 1) \tag{1.1}$$

Condition	“p” value
$\sigma_{RATIO} \geq 1.25$	0.25
$1.25 \geq \sigma_{RATIO} \geq 1.10$	0.20
$1.10 \geq \sigma_{RATIO} \geq 1.00$	0.15
$1.00 \geq \sigma_{RATIO}$	0.10

Table 1.1 : Adaptive linear energy based detector threshold update parameter “p” values

where $tr_{(new)}$ is the new threshold value, $tr_{(old)}$ is the previous threshold value and $E_{silence}$ is the energy of the most recent energy frame and p is the threshold update parameter. p is assumed to be a constant and preset while designing the system.

Adaptive Linear Energy Based Detector

This scheme works in a similar manner as a linear energy based detector except that the value of p is determined differently. In order to decide the new p value, most recent m silent frame energy values are buffered in an array. At every time step, the variance of the array is calculated and divided by the array variance of the previous time step.

$$\sigma_{RATIO} = \frac{\sigma_{new}}{\sigma_{old}} \tag{1.2}$$

By using the ratio which is given in Equation (1.2), the p value is chosen from Table 1.1.

Weak Fricatives (Zero Crossing) Detector

The energy levels of some sounds are very low and energy based detectors may not detect them. This algorithm combats this problem by using the zero crossing rate of the input signal. As a statistic, while the zero crossing rate of a 10ms frame lies between 5 and 10 for speech signals, this number increases to hundreds for noise only frames. Therefore, the detector makes its decision by comparing the zero crossing rate with a threshold. If it is above the threshold, the frame is decided as Inactive otherwise it is decided as an Active frame.

1.4.2 Frequency Domain VAD Algorithms

The second group of VAD algorithms are frequency domain VAD algorithms. In frequency domain VAD algorithms, decision variables of the detector are calculated in the frequency domain. Energy level and signal statistics are some of the most commonly used features while deciding on the presence or absence of speech. Although these algorithms are slower than time domain VAD algorithms, they give better detection performance. Some of these algorithms can be listed as: linear sub-band energy, spectral flatness and comprehensive VAD.

Linear Sub-Band Energy Detector

In the linear sub-band energy detector, the signal is divided into four frequency bands. Each of these frequency bands' energies are calculated in the frequency domain and compared to a threshold. Most of the energy in the voice signal tends to be at low frequencies. Therefore, if the low frequency sub-band and two out of the three high frequency sub-bands are active, the frame is decided to be Active.

Spectral Flatness

The algorithms discussed so far show good performance only in high SNR situations. If the input signal SNR decreases, system performance degrades significantly. The spectral flatness algorithm works well in low SNR also. In this detector, variance of the frame is compared to a threshold. High variance implies that speech is present. During the silent periods, the threshold value is updated in a similar manner as previously given in Equation (1.1).

Comprehensive VAD

Comprehensive VAD is a combination of all the previous algorithms. A decision flow chart of the algorithm is shown in Fig. 1.3.

First, the input signal is tested with an energy based detector. If this detector decides Active, the frame is decided as an active frame. If the energy based detector decides Inactive, a zero crossing detector is used to double check the result. If it is decided as Inactive again, the frame is decided to be Inactive. However, if the zero crossing detector decides Active, then the final decision is made after testing spectral flatness.

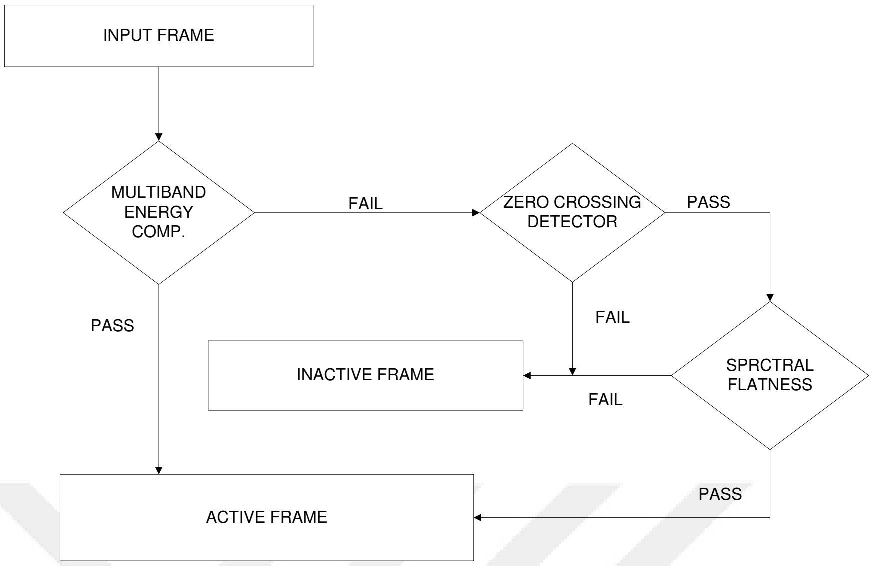

Figure 1.3 : Comprehensive Voice activity detector flowchart

1.4.3 Performance Evaluation for VAD

Performance evaluation of voice activity detection involves a very detailed procedure. There are many performance evaluation parameters that have been defined for this purpose. These parameters can be grouped as objective and subjective parameters. There are five objective evaluation parameters. These include [19]:

- Front End Clipping (FEC)
- Mid Speech Clipping (MSC)
- Over Hang (OVER)
- Noise Detected as Speech (NDS)

- Correct

As it can be seen in the Fig.1.4, when the speech signal starts, if the detector does not decide that speech is present then front end clipping occurs. Another objective evaluation parameter, mid speech clipping, can be defined as the misclassification of the signal in the middle of the speech present frames. Over Hang is the result of the detector continuing to decide speech is present after the speech signal has ended. Noise Detected as Speech happens during the silent period, when noise is interpreted as speech. The final objective evaluation parameter is correct, when the correct decision is made by the VAD algorithm.

The first two parameters, front-end clipping and mid-speech clipping, can be considered as misses in the context of detection theory, third and the fourth ones, over hang and noise detected as speech, are false alarms. The final parameter, correct, can be considered to be detections and can be used in plotting receiver operating curves.

In some instances, the clipping of speech is not audible and does not have an effect on the detected signal quality. Thus, subjective parameters are also very important for performance evaluation in voice activity detection. However, these parameters are not easy to utilize. For subjective performance evaluation, the detected signal is listened to by a group of people and graded for certain parameters. These parameters include:

- Quality

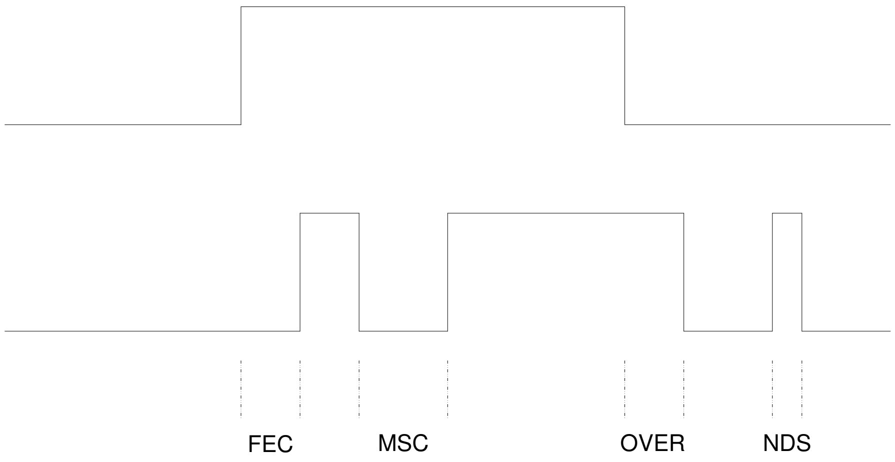

Figure 1.4 : Objective voice activity detection performance evaluation parameters.

- Comprehension difficulty
- Audibility of clipping

To summarize, voice activity detection is the process of discriminating between speech and silent periods for a given input speech signal. As can be understood from the definition, it is a detection problem. Many solutions to this problem have been derived by using different features of the input signal, such as energy, zero crossing and statistics. Linear energy based, zero crossing and comprehensive VAD are some examples of these detectors.

1.5 Stochastic Resonance

Noise is always considered to be one of the major problems that causes performance degradation in systems. Different techniques have been developed to remove noise and improve system performance. In signal processing and communications, filters and signal processing algorithms are used to denoise the signal. However, studies performed over the last few decades show that under certain conditions adding noise may in fact improve system performance. This phenomenon is called Stochastic Resonance (SR). As defined earlier in Section 1.1, Stochastic Resonance is a nonlinear phenomenon in which output performance of some nonlinear systems is improved by adding certain amount of noise under certain conditions. After its first use by Benzi *et al* in 1981, this counter-intuitive phenomenon has received lots of interest and has been studied for many physical and biological nonlinear systems.

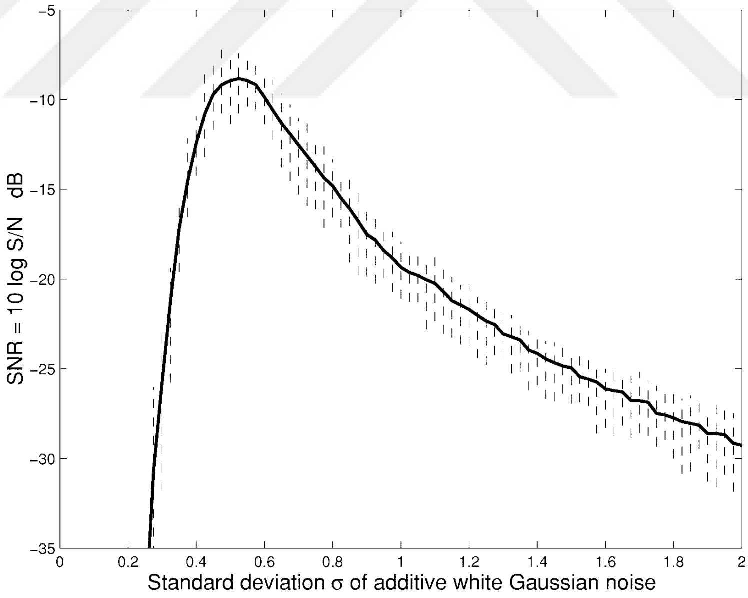

Figure 1.5 : Typical SR characteristic [1].

From the perspective of engineering systems, stochastic resonance has very appealing advantages. Typically to improve performance, a system would need to be either changed or replaced. However, instituting a change or replacement is not always a possibility. For example, in biological and some mechanical systems, the systems are permanent or are very expensive to change. In such instances, stochastic resonance can sometimes offer the ability to improve system performance by adding a random or a deterministic signal to the input signal [20]. This phenomenon offers engineers a lot of flexibility. Instead of redesigning the systems for variable environmental conditions, stochastic resonance makes it possible to tune existing systems for varying conditions.

A typical SR characteristic is shown in Fig. 1.5. As can be seen from the figure, by increasing the variance of additive noise; the output performance of the system at first increases to reach the highest performance value, with addition of the optimum amount of noise. However, the output performance of the system degrades gradually with further increase of the noise variance σ.

In the past three decades, many theoretical studies have been carried out and different models have been established for stochastic resonance. Some of the most used theoretical models are: double-well SR, two state SR, single-well SR and non-dynamical threshold[5]. In our study, we design our system by using the double-well SR model which is discussed in some details next.

1.5.1 Double-well SR

Double well is the most studied SR model in the literature. The most appealing aspect of this model is that it has fewer parameters to estimate when compared to other SR models. It is represented in terms of the following equation:

$$\begin{aligned}\dot{x} &= -\frac{\partial}{\partial x}U(x,t) + s(t) + n(t) \\ &= ax - bx^3 + s(t) + n(t) \qquad (1.3)\end{aligned}$$

where $U(x,t)$ is the unforced quartic potential defined as $U(x,t) = -(a/2)x^2 + (b/4)x^4$ with $a > 0$ and $b > 0$, s is the input signal and n is white Gaussian noise with zero mean and variance σ^2. It has two fixed stable points at $\pm\sqrt{a/b}$ and one fixed metastable point at $x = 0$. The first two points are minima and $x = 0$ is the local maximum of the potential. Sometimes signal and noise are also added to the potential function to obtain a forced form of quartic potential: $U(x,t) = -(a/2)x^2 + (b/4)x^4 + x(t)[s(t) + n(t)]$.

This model is generally explained by the analogy of the movement of a particle, under the influence of a periodic signal and random noise, between two minima of a bistable potential. In the literature, it is common to assume that this potential is purely symmetric [5]. Assume that there is a particle in one of the minimum of the bistable potential which is given in Fig.1.6 and assume that this potential is modulated with a weak periodic signal $\epsilon \sin \omega t$ in an antisymmetric way. Without noise, even if the periodic signal makes the particle's minima shallow, the particle would be stuck in its place forever. The addition of some noise would help the

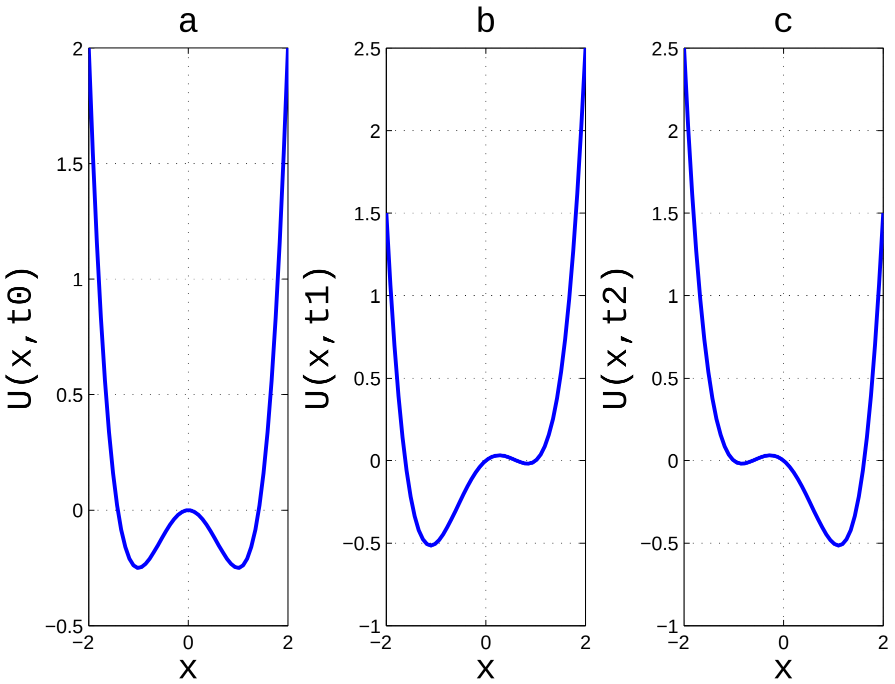

Figure 1.6 : Forced Quartic potential: $U(x,t) = -(a/2)x^2 + (b/4)x^4 + \sin 2\pi t$ (a) Potential $U(x,t)$ when forcing signal at $t = 0$, (b) Potential $U(x,t)$ when forcing signal at $t = 1/4$, (c) $U(x,t)$ when forcing signal at 3/4.

particle to jump from one minima to another when its minima becomes shallow. At the optimum noise level, the particle would synchronize with the weak periodic signal and jump between the minima twice per modulation period. The result of this is that a spike will be observed at the frequency of weak periodic signal in the frequency spectrum. However, under the influence of very strong noise, this particle would move from one minima to another regardless of the weak periodic signal and the spike in the frequency spectrum would gradually disappear. Although, in reality there is no particle which would jump between minima, this analogy helps us to understand how double-well SR improves the detection of the weak periodic signal.

1.6 Stochastic Resonance Effect in Signal Detection

As is true in many nonlinear systems, stochastic resonance plays a very important role in improving the performance of signal detection applications. Many researchers have investigated methods to maximize output SNR of the system with SR to improve system performance. Although, SNR is an important performance parameter for systems, and an accepted classic SR characteristic is the improvement in output SNR with the addition of noise; it is not possible to optimize SNR without complete prior information of the signal. Additionally, in non-Gaussian noise cases, SNR is not always directly related to detection probability.

Many special cases of signal detection with SR have been reported, such as weak periodic signal detection[8], dc signal detection in Gaussian Mixture noise[4]. Instead of focusing on a more specific case, Chen *et al.* explain Stochastic resonance effect with a more general, two hypothesis detection problem[2]. This can be summarized as follows.

Consider a two hypothesis detection problem. The input signal is N dimensional and represented by $x \epsilon R^N$. Decision is to be made between hypothesis 1 (H_1) and hypothesis 0 (H_0) which are given as:

$$\begin{aligned} H_0 :&\ \ p_x(x, H_0) = p_0(x) \\ H_1 :&\ \ p_x(x, H_1) = p_1(x) \qquad (1.4) \end{aligned}$$

where $p_0(x)$ and $p_1(x)$ are pdfs under H_0 and H_1 respectively. The test to decide between H_0 and H_1 can be characterized by a critical function $\phi(x)$ where $0 \leq \phi(x) \leq$

1. Generally, this function is represented with a test statistic $T(x)$ which is a function of the input signal and the decision is made by comparing it with a threshold η.

$$T(x) \gtrless_{H_0}^{H_1} \eta \tag{1.5}$$

The corresponding critical function can be defined as:

$$\phi_T(x) = \begin{cases} 1 : T(x) > \eta \\ \alpha : T(x) = \eta \\ 0 : T(x) < \eta \end{cases} \tag{1.6}$$

where $0 \leq \alpha \leq 1$.

Therefore, the probability of detection, P_D is given as:

$$P_D = \int_{R^N} \phi(x) p_0(x) dx \tag{1.7}$$

and the probability of false alarm P_{FA} is given as:

$$P_{FA} = \int_{R^N} \phi(x) p_1(x) dx \tag{1.8}$$

It is well known that in statistical detection, the Neyman-Pearson test is the most powerful test in the sense of maximizing P_D for given fixed P_{FA} value. In order to find the test statistic, it is necessary to have complete knowledge of pdfs under H_0 and H_1. However, in practical applications, this information is not always available. Additionally, sometimes the form of test statistic is too complicated and/or input statistic may vary with time. Considering these circumstances, there are two widely used approaches utilized to improve a suboptimal detector's detection performance.

One can either change the system parameters or find the optimum noise and add it to the input signal.

While it is known that, detection performance can be improved by adding noise if it is statistically dependent on the existing noise and/or the pdf of the true hypothesis[4]; it is not always possible to find dependent noise because prior information about the signal is not available. To combat this problem Kay [4] showed that under certain conditions, it is possible to improve detection performance by also adding independent noise. Another important point that should be emphasized is that optimum SR noise does not just mean the amount of the noise, it is also required to determine appropriate noise pdf.

Chen *et al*[2],[3] explained the stochastic resonance effect in signal detection in a more general way. They showed that the optimal SR noise is “a proper randomization

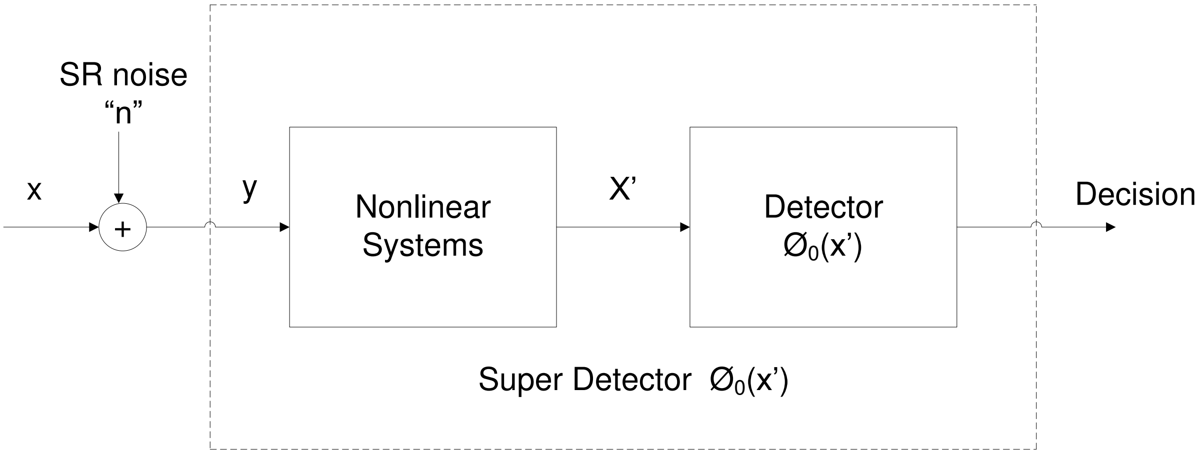

Figure 1.7 : Illustration of super detector.

of no more than two discrete signals”[2] and they proposed a theoretical approach to obtain optimum SR noise pdf for a given p_0 and p_1 distributions.

They also generalized their results to different SR detectors described in the literature, e.g., bistable systems and defined a ”super detector” which is demonstrated in Fig. 1.7. In the figure, x represents input signal array, nonlinear system block can represent bistable systems and x' is the output of the nonlinear system. Using this framework, the SR detectors proposed in the literature are also incorporated in their theories.

1.7 The Layout of the Thesis

Organization of the thesis is summarized as follows. In Chapter 1, we presented the motivation behind the thesis as well as the problem definition. This chapter also contained the literature review which explored existing research and studies in voice activity detection, stochastic resonance and stochastic resonance effect in signal detection. Chapter 2 explains both the voice activity detector which is used in the experiments as well as the SR method, which is used to improve detection performance. Simulation details and significant results are presented in Chapter 3. Finally, Chapter 4, briefly discusses highlights of the thesis and the thesis is concluded with suggestions for future work.

Chapter 2

Voice Activity Detection Using Stochastic Resonance

This chapter presents voice activity detection using stochastic resonance. Our goal is to improve voice activity detection performance, without changing the existing detector. The system block diagram is presented in Fig. 2.1 and details of each system block are provided in the following sub-sections.

Our system consists of two main blocks: Voice Activity Detector (VAD) and Stochastic Resonance filter (SR filter).

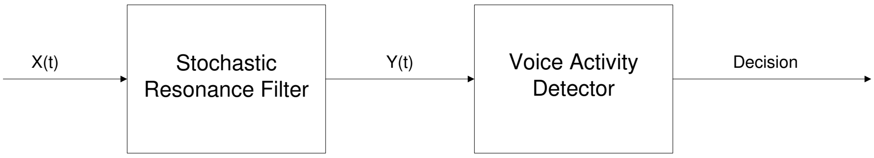

Figure 2.1 : System block diagram

2.1 Voice Activity Detection Block

In the voice activity detection block, we use Sohn and Sung's statistical based voice activity detector as our reference detector [12]. This detector is assumed to perform frame based detection and during each frame, the signal is assumed to be stationary. Also, it is assumed that the first 110ms of the input signal is noise. The voice activity detection block can be split into two parts, detector and noise statistics estimator, as shown in the Fig.2.2.

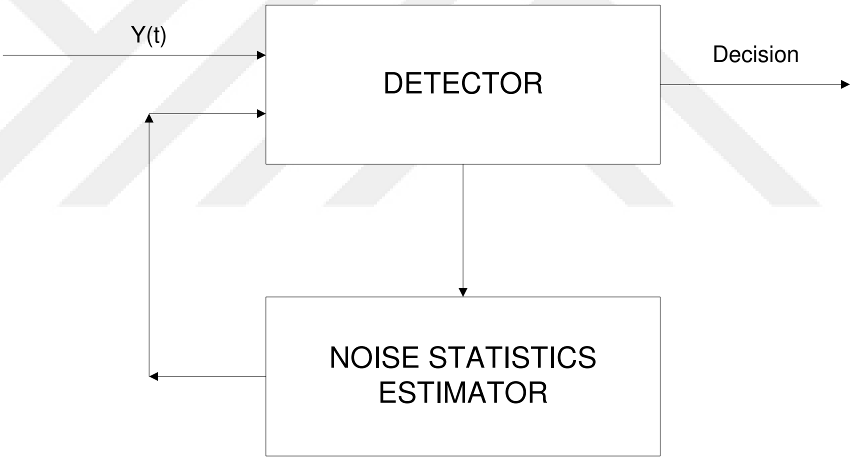

Figure 2.2 : Reference voice activity detector block diagram

2.1.1 Detector

In the reference detector, as a statistical model, speech and noise signals are assumed to be Gaussian random processes that are independent of each other, and discrete fourier transform (DFT) coefficients of these processes are assumed asymptotically independent Gaussian variables. In this statistical model, noise and speech variances are given by [12]:

$$\begin{aligned} \lambda_N(k) &= S_N(2\Pi k/M) \\ \lambda_S(k) &= S_S(2\Pi k/M) \end{aligned} \tag{2.1}$$

where $S_N(w)$ and $S_S(w)$ are true power spectra of noise and speech signals, respectively, and the variance of incoming signal Y_k is given by:

$$\sigma_Y^2(k) = \lambda_S(k) + \lambda_N(k) \tag{2.2}$$

Two hypotheses of the voice activity detection problem can be defined as;

$$\begin{aligned} H_0 &: \text{ speech absent} : \mathbf{Y} = \mathbf{N} \\ H_1 &: \text{ speech present} : \mathbf{Y} = \mathbf{S} + \mathbf{N} \end{aligned} \tag{2.3}$$

To be able to solve this problem with statistical methods, probability density functions $p(\mathbf{X}|H_0)$ and $p(\mathbf{X}|H_1)$ of hypothesis 0 and hypothesis 1 respectively should be obtained. For M dimensional DFT coefficient vector, the joint probability density functions conditioned by H_0, H_1 are given by [12]:

$$p(\mathbf{X}|H_0) = \prod_{k=0}^{M-1} \frac{1}{\pi\lambda_N(k)} \exp\left\{-\frac{|X_k|^2}{\lambda_N(k)}\right\} \tag{2.4}$$

$$p(\mathbf{X}|H_1) = \prod_{k=0}^{M-1} \frac{1}{\pi[\lambda_S(k) + \lambda_N(k)]} \exp\left\{-\frac{|X_k|^2}{\lambda_S(k) + \lambda_N(k)}\right\} \tag{2.5}$$

This detector is designed using one of the most widely employed methods for composite hypothesis testing [12], generalized likelihood ratio test.

As previously mentioned, we assume that the in first 110 ms of the signal voice is absent and contains only noise. By using this information, the initial noise statistics are obtained via a noise statistics estimator. This will be explained in detail in the following sub-section.

Although the probability density functions under both hypotheses and the noise variance are assumed to be known, in order to design this detector, speech variance should be estimated. In a generalized likelihood ratio test, $\hat{\lambda}_S(k)$ can be replaced by its maximum likelihood estimate $\{\hat{\lambda}_S(k) : k = 0, ..., M-1\}$, where $\hat{\lambda}_S(k)$ is the maximum likelihood estimate of the speech variance which is obtained by:

$$\hat{\lambda}_S(k) = |X_k|^2 - \lambda_N(k) \tag{2.6}$$

Substituting (2.6) in (2.5), the decision rule can be obtained as follows:

$$\begin{aligned} \Lambda &= \frac{1}{M}\frac{p(\mathbf{X}|\hat{\theta}, H_1)}{p(\mathbf{X}|H_0)} \\ &= \frac{1}{M}\sum_{k=0}^{M-1}\left\{\frac{|X_k|^2}{\lambda_N(k)} - \log\frac{|X_k|^2}{\lambda_N(k)} - 1\right\} \gtrless_{H_0}^{H_1} \eta \end{aligned} \tag{2.7}$$

It can be seen from (2.7) that $\frac{1}{M}\sum_{k=0}^{M-1}\left\{\log\frac{|X_k|^2}{\lambda_N(k)}\right\}$ term is dominant only when $\frac{|X_k|^2}{\lambda_N(k)}$ is much smaller than one, which is not true in either H_0 or H_1 cases.

Therefore, the test statistic can be simplified to:

$$\Lambda = \frac{1}{M} \sum_{k=0}^{M-1} \left\{ \frac{|X_k|^2}{\lambda_N(k)} \right\} \gtrless_{H_0}^{H_1} \eta + 1 \tag{2.8}$$

So, the decision statistic of the voice activity detector becomes an average of the subband SNR values.

2.1.2 Noise Statistics Estimator

As stated before, the initial noise statistics are assumed to be known and stationary. But in the real world, this statistic is not stationary. Even if the noise statistics of the environment changes very slowly, this will cause a gradual performance degradation of the system. In order to overcome this problem, noise statistics estimation and update processes should be performed periodically. This can be achieved by updating the noise statistics during noise only frames. Hence, to be able to detect noise only frames, a more conservative second detector should be used. However, in highly non-stationary noisy environments or very high speech rate situations, noise statistics will be outdated rapidly. Contrary to this approach, in our reference detector, the soft decision information of the decision rule is used to track and update noise statistics in both speech absent and present frames.

The optimum estimation of variance of noise Fourier coefficient $\hat{\lambda}_N(k)$ in terms of minimum mean square error is given by [12]:

$$\begin{aligned} \hat{\lambda}_N(k) &= E(\lambda_N(k)|X_k) \\ &= E(\lambda_N(k)|H_0)P(H_0|X_k) + E(\lambda_N(k)|H_1)P(H_1|X_k) \end{aligned} \tag{2.9}$$

By using Bayes rule, $P(H_0|X_k)$ and $P(H_1|X_k)$ can be derived as:

$$P(H_0|X_k) = \frac{1}{1+\varepsilon\Lambda(k)} \quad \text{and} \quad P(H_1|X_k) = \frac{\varepsilon\Lambda(k)}{1+\varepsilon\Lambda(k)} \tag{2.10}$$

where $\varepsilon = \frac{P(H_1)}{P(H_0)}$ and $\Lambda(k) = \frac{p(X_k|H_1)}{p(X_k|H_0)}$ where the generalized likelihood ratio that is derived in (2.8) can be used.

Now (2.10) can be substituted in (2.9);

$$E(\lambda_N(k)|X_k) = \frac{1}{1+\varepsilon\Lambda(k)}E(\lambda_N(k)|H_0) + \frac{\varepsilon\Lambda(k)}{1+\varepsilon\Lambda(k)}E(\lambda_N(k)|H_1) \tag{2.11}$$

In (2.11), the current frame measurement $|X_k^n|$ can be used instead of $E(\lambda_N(k)|H_0)$ while speech is absent. However, if speech is present $|X_k^n|$ should not be used while updating the noise statistic estimate $\hat{\lambda}^n(k)$ in the n^{th} frame. Instead, $\hat{\lambda}^{n-1}(k)$ can be used for $E(\lambda_N(k)|H_1)$ while speech signal is present.

After rearranging (2.11), noise statistics can be estimated as:

$$\lambda_N^n(k) = \frac{1}{1+\varepsilon\Lambda^n}|X_k^n|^2 + \frac{\varepsilon\Lambda^n}{1+\varepsilon\Lambda^n}\hat{\lambda}^{n-1}(k) \tag{2.12}$$

In (2.12), ε can be explained as a parameter which determines the convergence speed of the adaptation system [12]. In our experiments, we fixed this value as 1.

Noise update operation is performed in each frame after detection. With this algorithm, noise variance could be updated during both speech absent and present frames. This algorithm is also used to obtain initial noise statistics estimation at the system start up.

2.2 Stochastic Resonance Filter Block

In the stochastic resonance filter block, a non-linear transformation is applied to the incoming signal sequence to improve the detection performance. Our system uses, one of the most commonly used SR models in the literature, namely the Double Well SR. Also, another important feature of this model is that the performance of the system can be improved by finding Double-well coefficients a and b for a certain noise instead of dealing with relatively difficult problem, finding optimum additive noise. This model is defined as [1],[21],[6]:

$$\dot{x}(t) = -V'(x) + s(t) + \xi(t) \tag{2.13}$$

where $s(t)$ is incoming signal, $\xi(t)$ is additional noise and $V(x)$ is symmetric bistable potential which is $V(x) = -\frac{a}{2}x^2 + \frac{b}{4}x^4$. If we write (2.13) in an expanded form:

$$\dot{x}(t) = ax(t) - bx^3(t) + s(t) + \xi(t) \tag{2.14}$$

In order to implement this model, coefficients a, b and additive noise $\xi(t)$ in Equation (2.14) should be determined. This problem can be solved by either fixing a and b coefficients to be constant and finding optimum noise and its distribution, or by using a constant noise level and searching for optimum values of a and b coefficients. Due to the fact that our signal is very complex and finding suitable noise distribution is difficult; we chose to find optimum values of a and b coefficients for a fixed noise. Considering that our signal has already had its own noise component, we didn't use additional noise.

In a discrete computer environment, Double-well SR can be implemented with the stochastic version of Euler's method (Euler Maruyama scheme) which is defined in [22] as :

$$x_{m+1} = x_m + \triangle T(ax_m - bx_m^3 + s_m) \tag{2.15}$$

where $\triangle T$ is the time step which is defined as the difference between two consecutive samples. In our system, it is defined as $\frac{1}{\text{sampling frequency}}$, but if the difference between two consecutive samples is changed for some reason (e.g: sub-sampling) then this ratio changes. Furthermore, the initial condition of the transformation x_0 is chosen as 1 for convenience while finding the coefficients a and b. In each frame, this procedure is applied to each sample in the frame in an iterative way, and the frames are transformed separately.

It is known that, while having limited information about the input signal, such as just first and second moments under each hypothesis, the linear receiver can be found by maximizing the deflection coefficient[23]. In our system, we chose the SR filter coefficients "a" and "b" which resulted in maximum deflection coefficient after the transformation of the training data. In order to calculate deflection coefficient D_0 and D_1 arrays are generated for every a and b coefficient set by using Equation

(2.16).

$$D_0\{p|a,b\} = \textstyle\sum_{k=0}^{L} S_p[\frac{2\pi k}{M} + \frac{2\pi f_l}{f_s}|a,b]$$
$$D_1\{q|a,b\} = \textstyle\sum_{k=0}^{L} S_q[\frac{2\pi k}{M} + \frac{2\pi f_l}{f_s}|a,b] \tag{2.16}$$

where L is defined as $\{\frac{M(f_h-f_l)}{f_s}\} - 1$, and $D_0\{p|a,b\}$ and $D_1\{q|a,b\}$ are power sums of the requested frequency band of transformed frame with coefficients a and b for H_0 and H_1 respectively, f_h is highest frequency of the frequency band, f_l is the lowest frequency of the frequency band and f_s is the sampling frequency. After generating these arrays, their mean values and variances are calculated. Finally deflection coefficient is calculated by using Equation (2.17).

$$\mathcal{D}(a,b) = \frac{(E[D_1|a,b] - E[D_0|a,b])^2}{var[D_0|a,b] + var[D_1|a,b]} \tag{2.17}$$

where $\mathcal{D}$ is deflection coefficient of the signal for given a and b coefficients.

Flowchart of the algorithm used to find SR filter coefficients is given in Fig. 2.3. This procedure can easily be summarized as follows. The input training signal with a given SNR value and frame state information is transformed by the SR filter and each frame's power spectral density (PSD) is calculated. Then, for both H_0 and H_1, each frame's desired frequency band power is added and the respective D_0 and D_1 arrays are generated. In our applications, considering the speech signal frequency band, we choose this interval as 300 Hz to 3KHz. Next, the mean and variances of D_0 and D_1

arrays array are calculated.

It is obvious that if the difference between the means of the two hypotheses is large and variances are small, then the detection performance improves. Taking this into account we calculated the deflection coefficient $\mathcal{D}$ for different combinations of a and b values around our initial points. Finally, we chose the a and b values which maximized the deflection coefficient and used these values as SR Filter coefficients to improve detection performance. The performance of the system is evaluated in the next chapter.

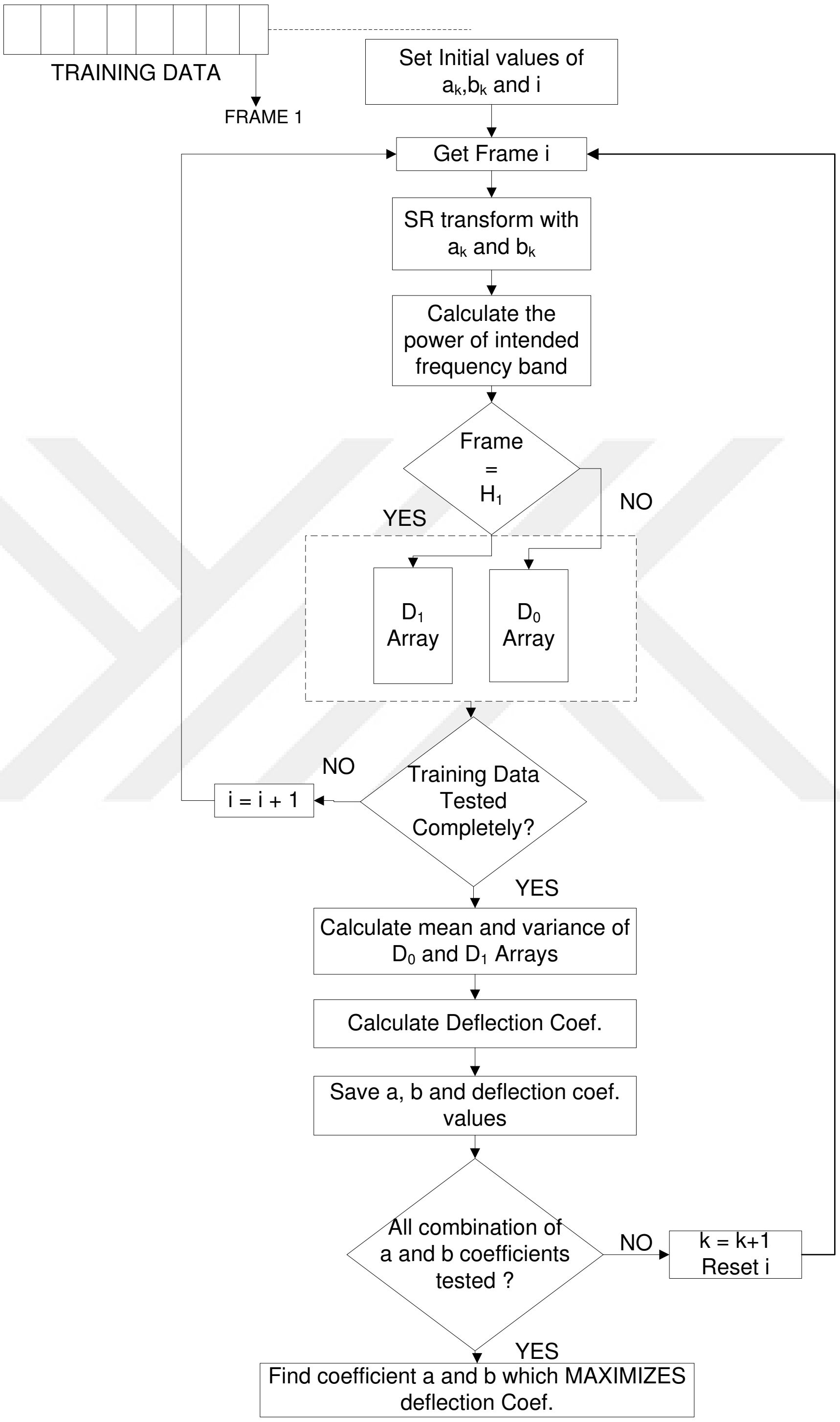

Figure 2.3 : Flowchart of the exhaustive algorithm used to find SR Filter coefficients.

Chapter 3

Simulation Results

Our simulations were performed on MATLAB v7.6 platform. The test data was obtained from a noisy speech corpus, NOIZEUS which is available online at Speech Processing Lab web page at University of Texas at Dallas [24]. Clean speech signals and eight different real world noise corrupted forms of the same data at different SNRs were taken from NOIZEUS. These real world noise signals were babble (crowd of people), exhibition hall, restaurant, street, airport, train, car and train station. Speech data contains 30 IEEE sentences which are phonetically-balanced sentences with relatively low word-context predictability that are produced by 3 male and 3 female speakers. By using corrupted and clean forms of speech signals, new speech signals at desired intermediate SNRs are generated for each of the eight different noise type.

In order to evaluate system performance, we investigate the probability of false alarm (P_F) and the probability of detection (P_D) of the system both with and without stochastic resonance. Here P_D means the probability of detecting a speech signal while speech is present and P_F means the probability of deciding speech is present during a silent period. To be able to obtain P_D and P_F values, we manually labeled speech signal at every 10ms time frame and recorded our reference decisions.

For our simulations, we used J. Sohn and W. Sung 's detector [12] as our reference detector which was described in detail in Section 2. During the experiments, we tested the performance of the Sohn and Sung detector using the original input signal. Then without making any structural change in the detector, we evaluated its performance using the new input signal which is preprocessed with a stochastic resonance filter.

In the stochastic resonance filter, we found the parameters a and b that optimize the deflection coefficient in an iterative manner. For an iterative approach, the initial values for parameters are very important. Therefore, to be able to find a better initial value, we calculated and plotted the deflection coefficients for different values of a and b. We repeated the same procedure many times for different SNR values, different noise types and different input signals. With small differences, we obtained similar results. One sample result is shown in Fig. 3.1. Also two dimensional illustration of the same results for value of a is fixed to 1 and value of b is fixed to 1 can be seen in Fig.3.2 and Fig.3.3 respectively. One of the most important observation that we made in Fig.3.1 was that the peaks were lined up mostly along two parallel lines which are $(-8000, 1)$ to $(1, 8000)$ and $(1, 1)$ to $(8000, 8000)$ and the global maximum is either around $(1, 1)$ or around $(1, 8000)$ depending on the SNR values. In order to avoid finding local optima, parameters a and b were searched around 1 and 8000, as the initial points, using a greedy approach. Another important observation from this figure is that, one of the initial points is actually our sampling frequency. We repeated the same experiment for a data set whose sampling frequency is $16KHz$ and

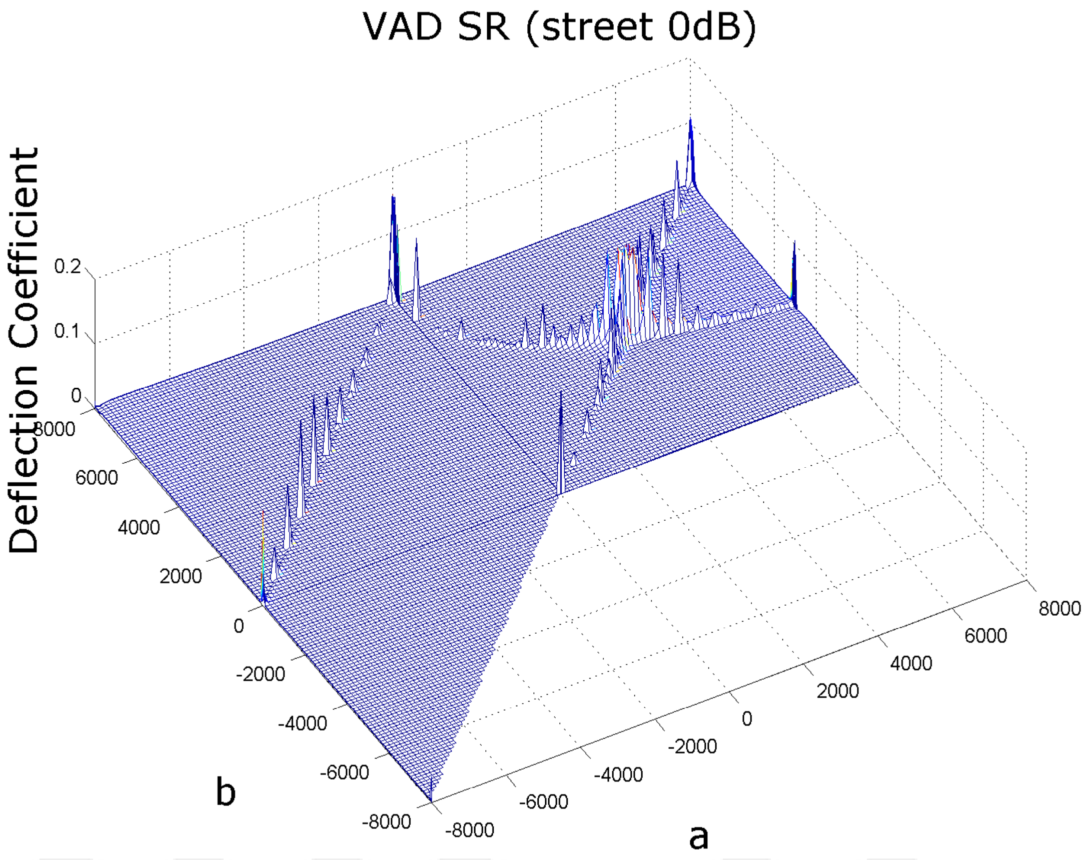

Figure 3.1 : Deflection coefficients for different values of a and b.

$Coef./SNR$	-5dB	-2dB	0dB	2dB	5dB	10dB
a	0.758	-0.758	-1.263	-1.263	-2.273	-2.278
b	0.758	-0.758	-1.263	-1.263	-2.273	-2.278

Table 3.1 : SR Filter coefficients for street noise at different SNRs

peaks were observed around $(1,1)$ and $(1,16K)$ points.

Considering these two initial points, we obtained coefficient a and coefficient b

$Coef./NoiseType$	Airport	Car	Exhibition Hall	Station	Street	Train
a	0.758	-3.778	-4.798	-4.293	-1.263	-1.263
b	0.758	-3.778	-4.798	-4.293	-1.263	-1.263

Table 3.2 : SR Filter coefficients for different noise types at 0dB SNR

in an iterative manner. At first, we fixed b at 1 and looked for a around values 1 and 8000. Then, by using the obtained value of a the value of b was searched. This procedure was repeated until a and b values converge at a point. Most of the time the system converged to the optimum value after 3 iterations. Some of the a and b coefficients obtained for different values of SNR and different types of noise are presented in Table 3 and Table 3 respectively.

In order to obtain the values of a and b, we used either complete data or some portion of it as the training data set.

3.1 Training with complete data

Although training with complete data is not a realistic way to solve this problem, it gave us some insight about the possible upper bound on the performance improvement by using our method. In this case, we assume that we already know the frame state information of the complete data set. After each transformation of the data with distinct a and b pairs, we separate the H_0 and H_1 frames to calculate the deflection

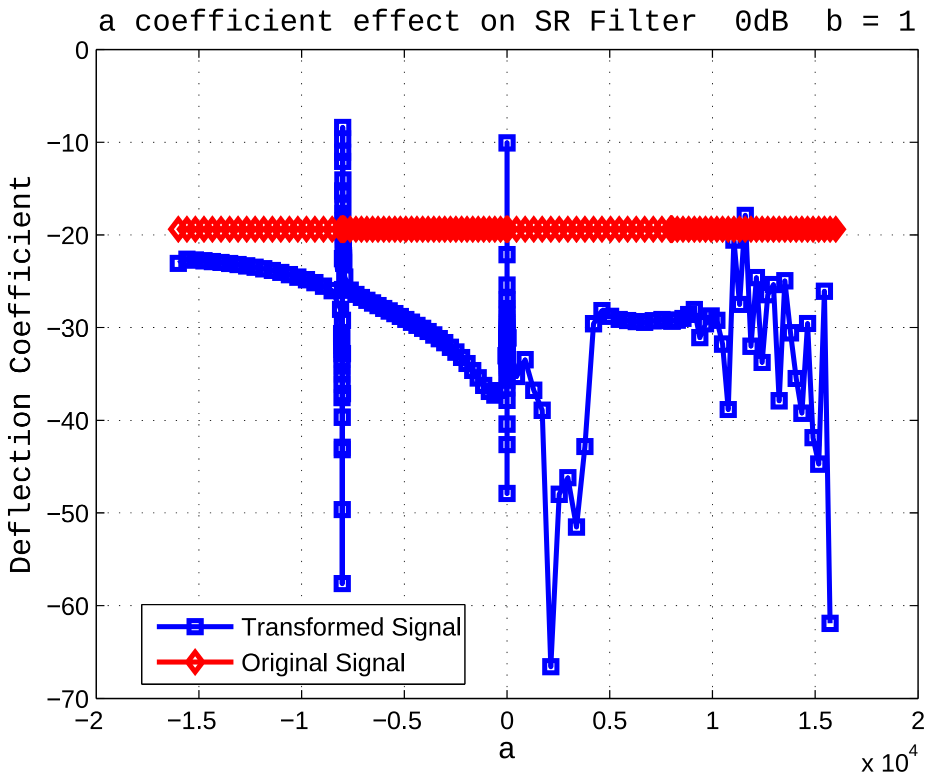

Figure 3.2 : Change of deflection coefficient with different values of a.

coefficient.

As can be seen from the Fig. 3.4, after finding a and b coefficients a significant performance improvement is observed. Although this transformation can not help the detection performance for very low false alarm rates, it yields a significant improvement, up to 17.5%, between 0.15 to 0.5 false alarm rates. On the other hand, for higher false alarm rates the improvement decreases, up to 4.85%, as expected. Also, the potential for improvement depends on the incoming signal and detector design assumptions. As mentioned earlier, to be able to get a performance improvement, the

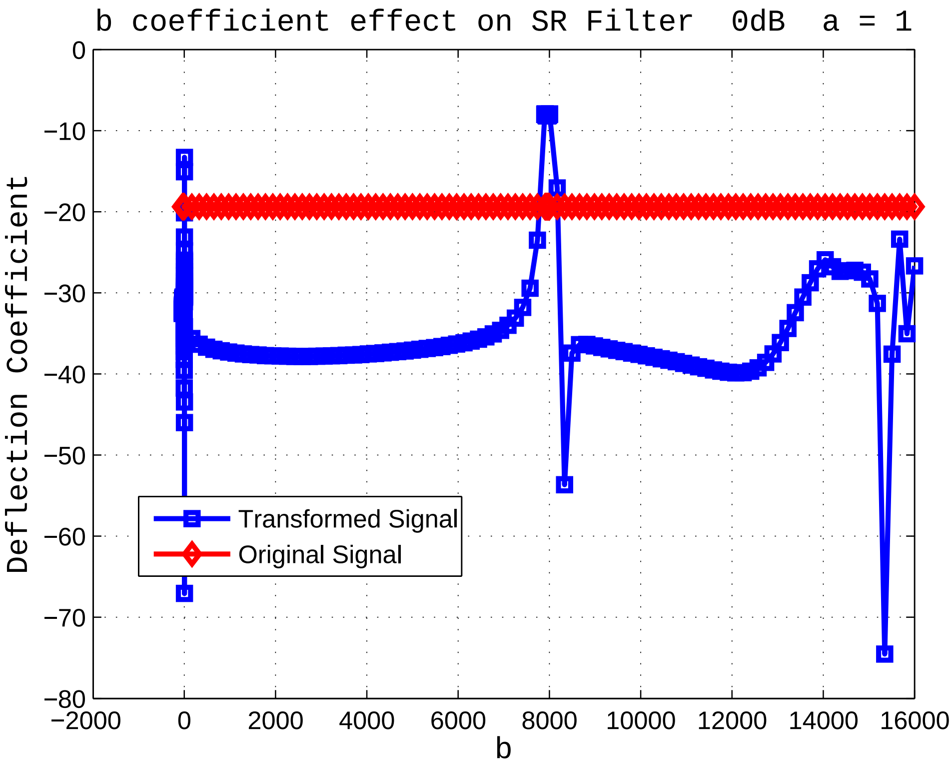

Figure 3.3 : Change of deflection coefficient with different values of b.

system should be suboptimal. Therefore, if the incoming signal distribution is close to detector design assumption, the improvement will decrease.

3.2 Training with partial data

For this case we assume that, we only have limited information about the input signal. In our experiment, we selected a percentage of input data as our training data. Then, we generate the H_0 and H_1 arrays in the same way as was explained in Section 3.1 and calculated the deflection coefficient. We repeated this procedure for different percentage of training data sets. Even though we reduced the training data percentage down to 20%, it can be seen from Fig. 3.5 that we found a and b values which gave us similar performance improvement as was obtained while training with a complete input data set. On the other hand, reducing the training data percentage caused a lower probability of convergence of coefficients a and b. Specifically, when we used less input data as training data, the coefficients couldn't converge at a point. Therefore, if the coefficients do not converge at a point after certain number of iterations, we stopped the itteration and repeated it with a different training data set. The most important inference we can make here is that, if our training data has enough statistical information about the input signal, even 20% training data is enough to obtain SR filter coefficients.

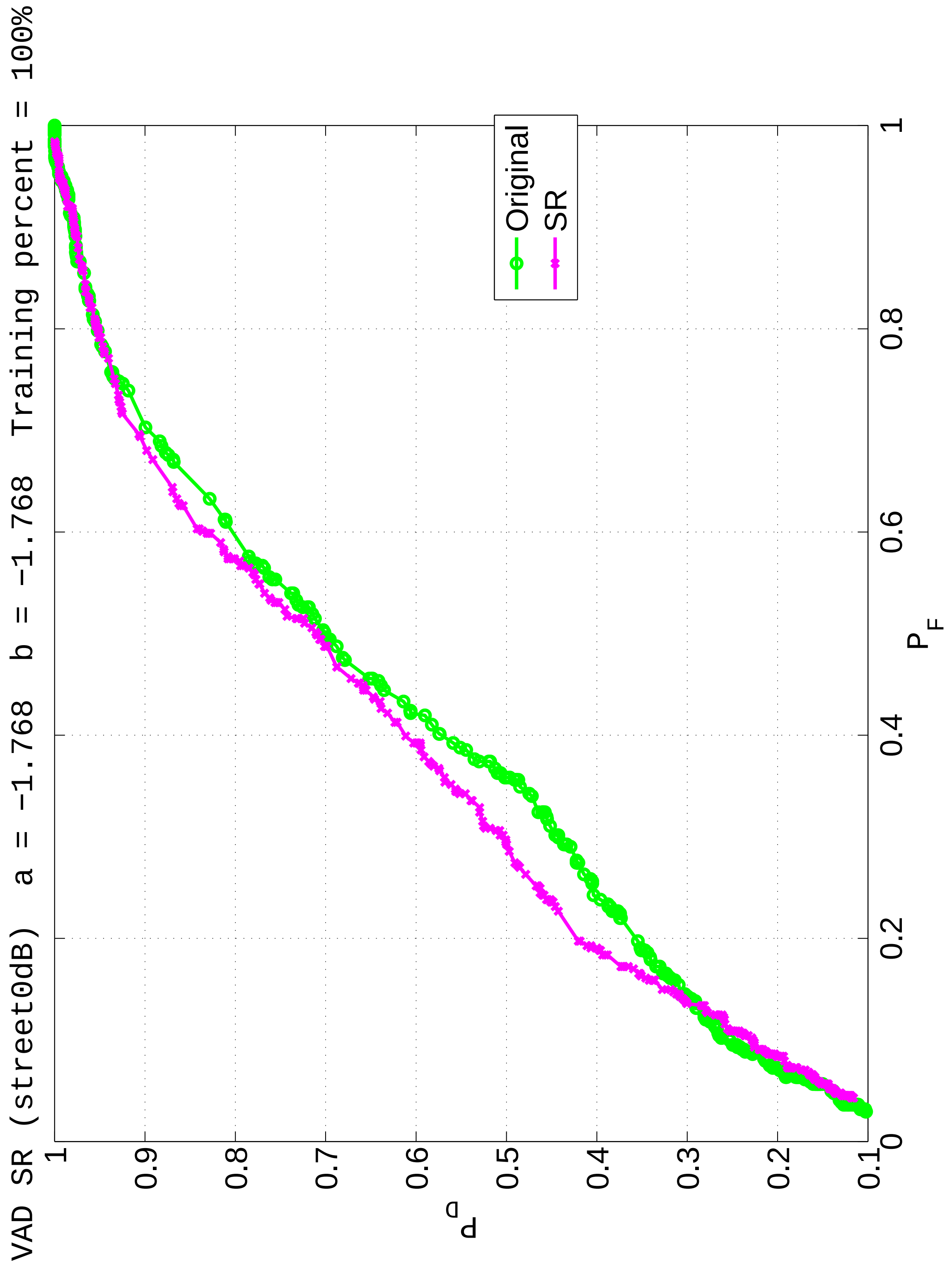

Figure 3.4 : Performance improvement after training with complete input data.

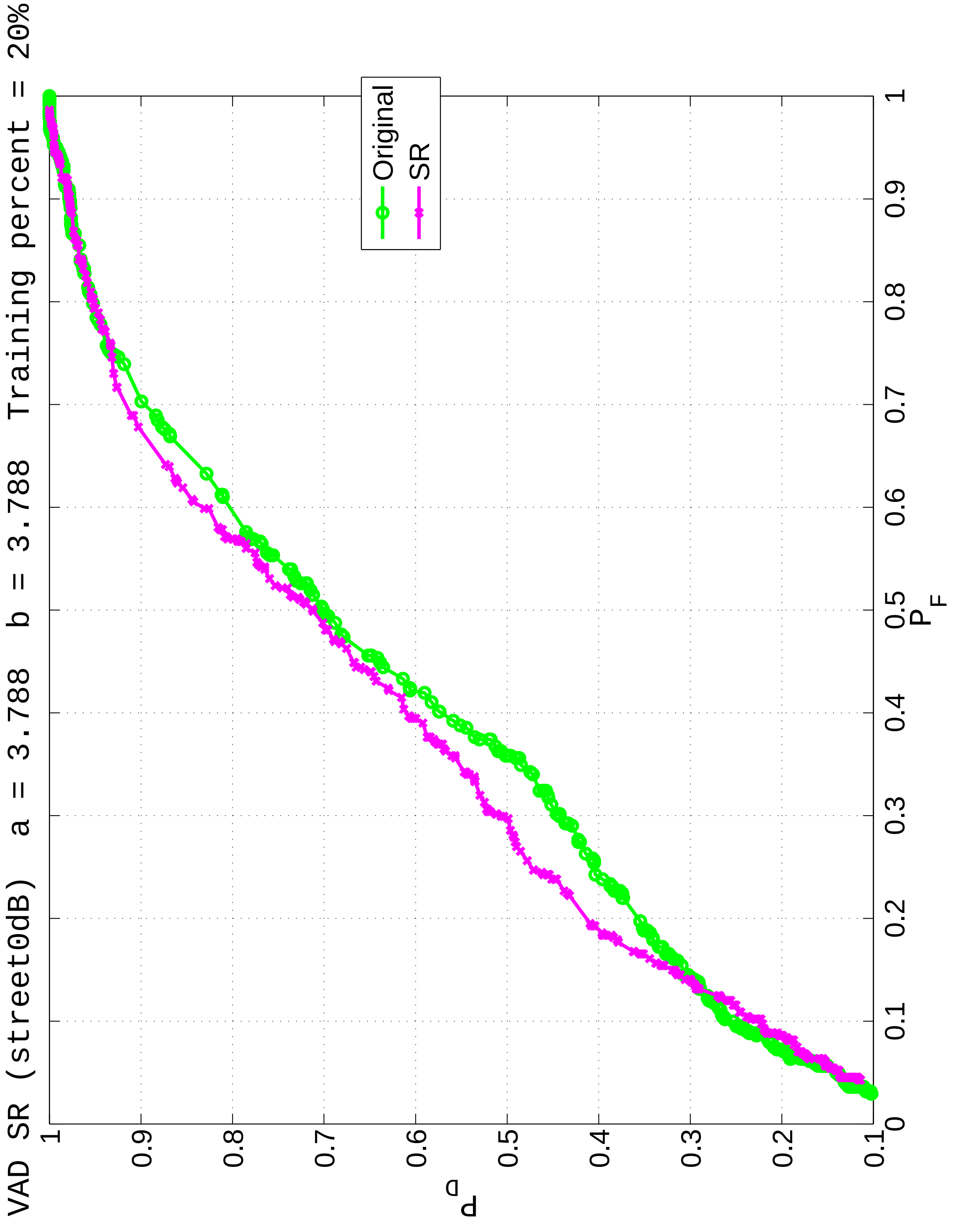

Figure 3.5 : Performance improvement after training with partial input data .

Chapter 4

Conclusion

4.1 Summary and Conclusions

Assumptions have a very important role in evolution of modern science and technology. Researchers make assumptions either to fulfill inadequate information about a phenomenon or to express complicated problems in a simpler form. Optimum systems are designed by using these assumptions under some constraints. It is important to realize that these systems will be suboptimal, if the underlying assumptions are not completely correct. This gives us the inference that it is possible to obtain better performance with better assumptions. However, "Is it necessary to redesign the whole system to improve its performance?" Recent studies [2],[3],[4] showed that, performance of the suboptimal systems can be improved by using stochastic resonance. In this thesis, we proposed a method to improve detection performance of a voice activity detector with stochastic resonance without making changes to the detector.

Experimental results showed that the proposed method improved detection performance up to 17.5% for low false alarm rates and 4.85% for high false alarm rates. In our experiments, in order to find the SR coefficients a and b, we used deflection coefficient as a measure. In the first part of the experiments we used complete data as our training data. Although it is not a realistic approach, the results gave us some insight about possible improvement rates of detection. Then, we used different fractions of input signal as our training signal. During the experiments we picked training frames randomly. Experimental results showed that, it is possible to find SR coefficient even by using 20% of input signal as training signal.

Therefore, based on the simulation results, our method is an efficient method to improve detection performance of suboptimal voice activity detector. Also, it is possible to obtain SR coefficients by using just a small portion of the input signal as a training signal if the training data properly represents the input signal.

4.2 Suggestions for Future Research

In this study, we investigated the stochastic resonance affect on a suboptimal voice activity detector. In our simulations, we observed up to 17.5% detection improvement by using a stochastic filter before the reference detector. In our experiments, because the signal of interest was a speech signal, we used a 300Hz to 3000Hz frequency band while finding optimum SR filter coefficients. Our method can be extended to different frequency bands for other multi tone signal detection problems.

In our study, in order to find SR filter coefficients, we used deflection coefficient as a measure. As the optimum SR filter coefficient set, we chose the coefficients which gave the highest deflection coefficient. We determined these coefficients by using the greedy search method. The greedy search approach restricted our method to offline applications. In order to find optimum coefficients, a faster algorithm can be used, such as a genetic algorithm. Then, the system can be used online also.

Finally, in our experiments we observed that it is possible to obtain SR coefficients by using a small amount of input signal as our training data. It was also examined that decreasing the training data percentage caused a decrease in the probability of convergence for SR coefficients to one point. Therefore, further investigation of the sufficiency conditions for training data is also necessary.

Bibliography

[1] S. Mitaim and B. Kosko, “Adaptive stochastic resonance,” *Proceedings of the IEEE*, vol. 86, pp. 2152 –2183, Nov 1998.

[2] H. Chen, P. Varshney, S. Kay, and J. Michels, “Theory of the stochastic resonance effect in signal detection: Part I;fixed detectors,” *Signal Processing, IEEE Transactions on*, vol. 55, pp. 3172 –3184, July 2007.

[3] H. Chen and P. Varshney, “Theory of the stochastic resonance effect in signal detection ;part II: Variable detectors,” *Signal Processing, IEEE Transactions on*, vol. 56, pp. 5031 –5041, Oct. 2008.

[4] S. Kay, “Can detectability be improved by adding noise?,” *Signal Processing Letters, IEEE*, vol. 7, pp. 8 –10, Jan 2000.

[5] K. Wiesenfeld and F. Jaramillo, “Minireview of stochastic resonance,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 8, no. 3, pp. 539–548, 1998.

[6] Q. Ye, H. Huang, and C. Zhang, “Image enhancement using stochastic resonance [sonar image processing applications],” in *Image Processing, 2004. ICIP ’04. 2004 International Conference on*, vol. 1, pp. 263 – 266 Vol. 1, Oct. 2004.

[7] R. Jha, P. Biswas, and B. Chatterji, “Enhancement of digital images using stochastic resonance,” in *TENCON 2005 2005 IEEE Region 10*, pp. 1 –6, 21-24 2005.

[8] K. Zheng, H. Li, S. M. Djouadi, and J. Wang, “Spectrum sensing in low snr regime via stochastic resonance,” Tech. Rep. arXiv:0906.0739, Jun 2009. Comments: 5 pages, 9 figures, submitted to Asilomar 2009.

[9] A. Asdi and A. Tewfik, “Detection of weak signals using adaptive stochastic resonance,” in *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, vol. 2, pp. 1332 –1335 vol.2, 9-12 1995.

[10] J. Wang, Q. Xiao, and X. Li, “The high-frequency weak signal detection based on stochastic resonance,” in *Test and Measurement, 2009. ICTM ’09. International Conference on*, vol. 1, pp. 431 –434, 5-6 2009.

[11] Z. Hou, J. Yang, Y. Wang, and K. Wang, “Weak signal detection based on stochastic resonance combining with genetic algorithm,” in *Communication Systems, 2008. ICCS 2008. 11th IEEE Singapore International Conference on*, pp. 484 –488, 19-21 2008.

[12] J. Sohn and W. Sung, “A voice activity detector employing soft decision based noise spectrum adaptation,” in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 1, pp. 365 –368 vol.1, May 1998.

[13] J.-H. Chang, N. S. Kim, and S. Mitra, “Voice activity detection based on multiple statistical models,” *Signal Processing, IEEE Transactions on*, vol. 54, pp. 1965 – 1976, June 2006.

[14] R. Venkatesha Prasad, A. Sangwan, H. Jamadagni, M. Chiranth, R. Sah, and V. Gaurav, “Comparison of voice activity detection algorithms for VoIP,” in *Computers and Communications, 2002. Proceedings. ISCC 2002. Seventh International Symposium on*, pp. 530 – 535, 2002.

[15] D. K. Kim, K. W. Jang, and J.-H. Chang, “A new statistical voice activity detection based on ump test,” *Signal Processing Letters, IEEE*, vol. 14, pp. 891 –894, Nov. 2007.

[16] J. Sohn, N. S. Kim, and W. Sung, “A statistical model-based voice activity detection,” *Signal Processing Letters, IEEE*, vol. 6, pp. 1 –3, Jan 1999.

[17] J. W. Shin, J.-H. Chang, S. Barbara, H. S. Yun, and N. S. Kim, “Voice activity detection based on generalized gamma distribution,” in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP ’05). IEEE International Conference on*, vol. 1, pp. 781 – 784, 18-23, 2005.

[18] J. Ramirez, J. M. Gorriz, and J. C. Segura, *Robust Speech Recognition and Understanding*, pp. 1–22. I-Tech Education and Publishing, 2007.

[19] A. Davis, S. Nordholm, and R. Togneri, “Statistical voice activity detection using low-variance spectrum estimation and an adaptive threshold,” *Audio, Speech,*

and Language Processing, IEEE Transactions on, vol. 14, pp. 412 – 424, March 2006.

[20] H. Chen, "Constructive role of noise in signal processing @ONLINE," May 2009.

[21] Q. Ye, H. Huang, X. He, and C. Zhang, "A study on the parameters of bistable stochastic resonance systems and adaptive stochastic resonance," in *Robotics, Intelligent Systems and Signal Processing, 2003. Proceedings. 2003 IEEE International Conference on*, vol. 1, pp. 484 – 488 vol.1, Oct. 2003.

[22] D. J. Higham., "An algorithmic introduction to numerical simulation of stochastic differential equations," *SIAM Review*, vol. 43, no. 3, pp. 525–546, 2001.

[23] B. Picinbono, "On deflection as a performance criterion in detection," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 31, pp. 1072 –1081, Jul 1995.

[24] Y. Hu and P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech Communication*, vol. 49, no. 7-8, pp. 588 – 601, 2007. Speech Enhancement.

VITA

NAME OF AUTHOR : İlker Özçelik

PLACE OF BIRTH : Eskişehir, Turkey

DATE OF BIRTH : February 24, 1981

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:
Marmara University, İstanbul, Turkey

DEGREES AWARDED:
Bachelor of Science in Electronics and Communication Education, 2005, Marmara University

AWARDS AND HONORS:
Full graduate education scholarship from Republic of Turkey, Ministry of National Education, 2007