IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 8, Issue 5 (Nov. - Dec. 2013), PP 10-16
www.iosrjournals.org

A New Technique for Perceptual Distortion Measure on A Spectro Temporal Auditory Model

A.Girish Kumar [1], B. Brahmareddy [2], Mr. S.A.Mansoor [3]
M.Tech [1], Professor [2], Asst. Professor [3],
Department of Electronics and Communication Engineering [1][2][3]
Vignana Bharathi Institute Of Technology, Aushapur, Ghatkesar, Ap.

Abstract: *Speech processing is used widely in every day's applications that most people take for granted, such as network wire lines, cellular telephony, telephony system and telephone answering machines. Due to its popularity and increasing of demand, engineers are trying various approaches of improving the process.*

One of the methods for improving is trying on different methods of filtering techniques. Thus, this instigates an introduction of a filtering technique known as Kalman filtering. In the early days, Kalman filtering was very popular in the research field of navigation because of its magnificent accurate estimation characteristic. Since then, electronic engineers manipulate its advantages to useful purpose in speech processing. Consequently, today it had become a popular filtering technique for estimating and resolving redundant errors containing in speech.

A speech pre-processing algorithm is presented to improve the speech intelligibility in noise for the near-end listener. The algorithm improves the intelligibility by optimally redistributing the speech energy over time and frequency for a perceptual distortion measure, which is based on a spectro-temporal auditory model. In contrast o spectral-only models, short-time information is taken into account. As a consequence, the algorithm is more sensitive to transient regions, which will therefore receive more amplification compared to stationary vowels. It is known from literature that changing the vowel-transient energy ratio is beneficial for improving speech intelligibility in noise. Objective intelligibility prediction results show that the proposed method has higher speech intelligibility in noise compared to other reference methods, without modifying the global speech energy.

I. Introduction:

Speech is a form of communication in every day life. It existed since human civilizations began and even till now, speech is applied to high technological telecommunication systems. As applications like cellular and satellite technology are getting popular among mankind, human beings tend to demand more advance technology and are in search of improved applications. For this reason, researchers are looking closely into the four generic attributes of speech coding. They are complexity, quality, bit rate and delay. Other issues like robustness to transmission errors, multistage encoding/decoding, and accommodation of non-voice signals such as in-band signaling and voice band modem data play an important role in coding of speech as well.

In order to understand these processes, both human and machine speech has to be studied carefully on the structures and functions of spoken language: how we produce and perceive it and how speech technology may assist us in communication. Therefore in this thesis, we will be looking more into speech processing with the aid of an interesting technology known as the Kalman Filter. Presently, this technique is not widely used in the field of signal processing, however it is a potential nominee to be considered. More details on Speech Processing and the Kalman filter will be explained in the later chapters of this thesis report.

II. Speech Processing

2.1Speech Production: Speech is produced when air is forced from the lungs through the vocal cords and along the vocal tract. The vocal tract extends from the opening in the vocal cords (called the glottis) to the mouth, and in an average man is about 17 cm long. It introduces short-term correlations (of the order of 1 ms) into the speech signal, and can be thought of as a filter with broad resonances called formants. The frequencies of these formants are controlled by varying the shape of the tract, for example by moving the position of the tongue. An important part of many speech codecs is the modeling of the vocal tract as a short term filter. As the shape of the vocal tract varies relatively slowly, the transfer function of its modeling filter needs to be updated only relatively infrequently (typically every 20 ms or so). The vocal tract filter is excited by air forced into it through the vocal cords. Speech sounds can be broken into three classes depending on their mode of excitation.

1. Voiced sounds
2. Unvoiced
3. Plosive sounds

2.2 Speech Processing:

The term speech processing basically refers to the scientific discipline concerning the analysis and processing of speech signals in order to achieve the best benefit in various practical scenarios. The field of speech processing is, at present, undergoing a rapid growth in terms of both performance and applications. This is stimulated by the advances being made in the field of microelectronics, computation and algorithm design. Nevertheless, speech processing still covers an extremely broad area, which relates to the following three engineering applications:

- **Speech Coding** and transmission that is mainly concerned with man-to man voice communication.
- **Speech Synthesis** which deals with machine-to-man communications.
- **Speech Recognition** relating to man-to machine communication.

2.3 Sampling:

The purpose of sampling is to transform an analog signal that is continuous in time to a sequence of samples discrete in time. The signals we use in the real world, such as our voices, are called "analog" signals. In order to process these signals in computers, most importantly it must be converted to "digital" form. While an analog signal is continuous in both time and amplitude, a digital signal is discrete in both time and amplitude. Since in this thesis, speech will be processed through a discrete Kalman filter, it is necessary for converting the speech signal from continuous time to discrete time, hence this process is described as sampling.

III. Wavelet Transformations

3.1 Wavelets:

It is well known to any scientist and engineer who work with a real world data that signals do not exist without noise, which may be negligible (i.e. high SNR) under certain conditions. However, there are many cases in which the noise corrupts the signals in a significant manner, and it must be removed from the data in order to proceed with further data analysis. The process of noise removal is generally referred to as signal denoising or simply denoising. Example of a noisy signal and its denoised version can be seen in below Figure. It can be seen that the noise adds high-frequency components to the original signal which is smooth.

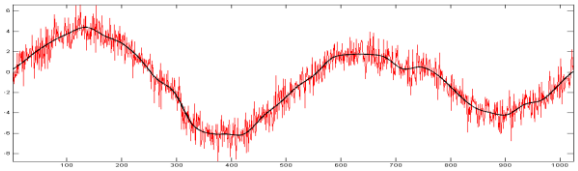

Figure 1. Noisy sine and its denoised version (solid line)

3.2 Mallet's algorithm:

In the case of DWT, assuming that the length of the signal satisfies $N = 2^J$ for some positive J, the transform can be computed efficiently, using Mallet's algorithm, which has a complexity of O (N). Essentially the algorithm is a fast hierarchical scheme for deriving the Required inner products using a set of consecutive low and high pass filters, followed by decimation. This results in a decomposition of the signal into different scales which can be considered as different frequency bands. The low-pass (LP) and high-pass (HP) filters used in this algorithm are determined according to the mother wavelet in use. The outputs of the LP filters are referred to as approximation coefficients and the outputs of the HP filters are referred to as detail coefficients. Demonstration of the process of 3-level decomposition of a signal can be seen in below Figure.

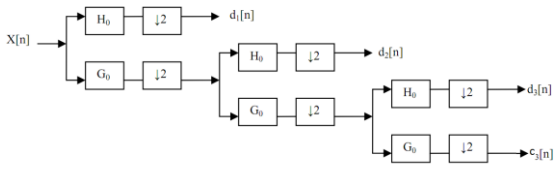

Figure6.Algorithmdemonstration 3-level decomposition of a signal.

IV. Kalman Filter

4.1Kalman Filter:

Theoretically, the Kalman Filter is an estimator for what is called the **"linear quadratic problem",** which focuses on estimating the instantaneous "state" of a linear dynamic system perturbed by white noise. Statistically, this estimator is optimal with respect to any quadratic function of estimation errors. In practice, this Kalman Filter is one of the greater discoveries in the history of statistical estimation theory and possibly the

greatest discovery in the twentieth century. It has enabled mankind to do many things that could not have been done without it, and it has become as indispensable as silicon in the makeup of many electronic systems

In a more dynamic approach, controlling of complex dynamic systems such as continuous manufacturing processes, aircraft, ships or spacecraft, are the most immediate applications of Kalman filter. In order to control a dynamic system, one needs to know what it is doing first. For these applications, it is not always possible or desirable to measure every variable that you want to control, and the Kalman filter provides a means for inferring the missing information from indirect (and noisy) measurements. Some amazing things that the Kalman filter can do is predicting the likely future courses of dynamic systems that people are not likely to control, such as the flow of rivers during flood, the trajectories of celestial bodies or the prices of traded commodities.

From a practical standpoint, these are the perspectives that this section will present:

4.2 The Adaptive Kalman Filter Algorithm:

Many applications in speech enhancement are based on Kalman filtering algorithm. Most of those methods need to estimate the parameters of AR model at first, and then perform the noise suppression using Kalman filtering algorithm. In this process, the calculations of LPC (linear prediction coding) coefficient and inverse matrix greatly increase the computational complexity of the filtering algorithm. Although these methods can achieve a good filtering efficiency, the noise suppressed signal may deteriorate the quality of the speech signal dependent on estimation accuracy of the parameters of the AR model. Through equations [2] and [3] have been given a simple Kalman filtering algorithm without calculating LPC coefficient in the AR model, but the algorithm still contains a large number of redundant data and matrix inverse operations. In addition, the algorithm is non-adaptive.

The drawback of conventional Kalman filtering for speech enhancement must be overcome, for which we propose a fast adaptive algorithm of Kalman filtering. This algorithm only constantly updates the first value of state vector X (n), which eliminates the matrix operations and reduces the time complexity of the algorithm. Actually, it is difficult to know what environmental noise exactly is. This affects the application of the Kalman filtering algorithm. So there is a need for a real-time adaptive algorithm to estimate the ambient noise. We add the forgetting factor which has been mentioned by [4] and [5] to amend the estimation of environmental noise by the observation data automatically, so the algorithm can catch the real noise. Compared with the conventional Kalman filtering algorithm, the fast adaptive algorithm of Kalman filtering is more effective. This has been showed through the simulation results. At the same time, it reduced its running time without degrading quality of the speech signal. It also has good adaptability to improve the algorithm robustness

4.3 Improved Kalman Filtering Lagorithm:

4.3.1 Conventional Kalman Filtering Method:

Speech driven by white noise is All-pole linear output from the recursive process. Under the short-time stable supposition, a pure speech can establish L step AR model by

$$s(\mathrm{n}) = \sum_{i=1}^{L} a_i(\mathrm{n}) \times \mathrm{s}(\mathrm{n}-\mathrm{i}) + \omega(\mathrm{n}) \quad \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square\ \square$$

In (1), $a_i(\mathrm{n})$ is the LPC coefficient, $\square\ \square\ \omega(n)$ is the white Gaussian noise which the mean is zero and the variance is δ_M^2 . The speech signal $s(n)$ is degraded by an additive observation noise $v(n)$, in the real environment $\square\ \square$ Its mean is zero and its variance is $\delta_V^2\ \square$. This noise isn't related to $s(n)\ \square\ \square$ A noisy speech signal $y(n)$ is given by

$$y(n)\ \square\ \square s(n)\ \square v(n) \text{……..(2)}$$

In this paper, it is assumed that the variance $\delta_V^2\ \square$ is known, but in practice we need to estimate it by the "silent segment" included in the $y(n)$.

(1) and (2) can be expressed as the state equation and the observation equation which are given by

[State equation]

$x(n)\ \square\ \square F(n)x(n\square\ 1)\ \square G\square\ (n)$……..(3)

[Observation equation]

$y(n)\ \square\ \square Hx(n)\ \square v(n)$..... (4)

$F(n)\ \square$is the$\square\ \square L\square\ \square L$ transition matrix expressed as

$$F(n)=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ a_L(n) & a_{L-1}(n) & a_{L-2}(n) & \cdots & a_1(n) \end{bmatrix} \quad \ldots\ldots\ldots\ldots (5)$$

$$F=H=\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots (6)$$

G is the input vector and H is the observation vector. It is easy to see that the conventional Kalman filtering is using the LPC coefficient to estimate the observations of the speech signal. This part spends half the time of the whole algorithm.

In [2] the transition matrix *F* and the observation matrix *H* are modified. They has defined as It also has defined the *L*□ 1 state vector *X*(*n*) □ □[*s*(*n* □ □*L* □ 1) *s*(*n* □ □*L* □ □ 2)] , the*L*□ 1input vector *Q*(*n*) [*s*□*n*□ □ 0 □ □ 0]*T* □ □ , and the 1□ □*L* observation vector *R*(*n*) □ □ [*v*(*n*),...*v*(*n* □ □*L* □ □ 2)] . Finally, (3) and (4) can be rewritten into the matrix equations by

[State equation]

$$X(n) \square \square F \square \square X(n \square 1) \square Q(n) \ldots\ldots\ldots\ldots \square \square \square \square$$

[Observation equation]

$$Y(n) \square \square H \square \square X(n) \square \square R(n) \ldots\ldots\ldots\ldots\ldots\ldots \square \square \square$$

Then the recursion equation of Kalman filtering algorithm is given in Table. In this case the noise variance □ is δ_V^2 known.

This algorithm abrogates the computation of the LPC coefficient. The number of calls for the filtering equations is equal to the number of sampling point's n of the speech signals, so the algorithm's time complexity is O (Ln).

4.3.2 Improved Filtering Method:

By the recursive equations of the conventional Kalman filtering algorithm, we can find that (5) - (13) contain a large number of matrix operations. Especially, the inverse matrix operations lead to an increase in the algorithm's complexity. We can greatly reduce the complexity of the algorithm, if we can reduce the dimension of matrix or eliminate matrix operations. In Table I, we can find that (11) constantly pans down the value of state vector X(n) and then (12) constantly updates the first value *s*(*n*) of X(n). However, during the whole filtering process, only the value of *s*(*n*) is useful. So we can use the calculation of *s*(*n*) instead of the calculation of the vector in order to avoid the matrix inversion. Furthermore, computational complexity of the algorithm can be reduced to O(n).

The recursive equations of the improved filtering method are shown in Table II.

TABLE-I

THE CONVENTIONAL METHOD PROCEDURE

□ Initialization]

$$X(0/0)=0, P(0/0)=I$$

$$R_v(n)=\delta_V^2, G=[10.....0]$$

$$R_s(n)[i,j]=\begin{cases} E(Y(n)\times Y(n))-\delta_V^2 & G,j=1) \\ 0 & (others) \end{cases}$$

[Iteration]

$$P(n/n-1)=F\times P(n-1/n-1)\times F^T \div G\times R_s(n)\times G^T \ldots\ldots\ldots\ldots(9)$$

$$K(n)=P(n/n-1)\times G^T/(G\times P(n/n-1)\times G^T \div R_V(n)) \ldots\ldots\ldots\ldots(10)$$

$$X(n/n-1)=F\times X(n-1/n-1) \ldots\ldots\ldots\ldots\ldots\ldots(11)$$

$$X(n/n)=X(n/n-1)+K\times(y(n)-G\times X(n/n-1)) \ldots.(12)$$

$$P(n/n)=(I-K(n)\times G)\times P(n/n-1)) \ldots\ldots\ldots\ldots(13)$$

4.3.3 The Adaptive Filtering Algorithm:

As the noise changes with the surrounding environment, it is required that the estimation of noise is constantly updated. From which we can get a more accurate expression of noise. Here we further improve the Kalman filter algorithm, so that it can adapt to any changes in environmental noise and become a fast adaptive Kalman

TABLE-II

THE IMPROVED FILTERING METHOD PROCEDURE

$$S(0) = 0, \mathrm{R}_V = \delta_V^2,$$

$$R_s(\mathrm{n}) = \mathrm{E}(\mathrm{y}(\mathrm{n}) \times \mathrm{y}(\mathrm{n})) - \delta_V^2$$

[Iteration]

$$K(\mathrm{n}) = \mathrm{R}_s(\mathrm{n}) / (\mathrm{R}_s(\mathrm{n}) + \mathrm{R}_V) \quad \text{.....(14)}$$

$$\mathrm{S}(\mathrm{n}) = \mathrm{K}(\mathrm{n}) \times y(\mathrm{n}) \quad \text{.....(15)}$$

filtering algorithm for speech enhancement.

The capability of constant updating of the estimation of background noise is the key for the fast adaptive algorithm. We can set a reasonable threshold to determine whether the current speech frame is noise or not. It consists of two steps: one is updating the variance of environmental noise $\mathrm{R}_V(\mathrm{n})$ and the other is updating the threshold U.

1) Updating the variance of environmental noise by

$$\mathrm{R}_V(\mathrm{n}) = (1-\mathrm{d}) \times \mathrm{R}_V(\mathrm{n}) + \mathrm{d} \times \mathrm{R}_U(\mathrm{n}) \quad \text{.....(16)}$$

In (16), d is the loss factor that can limit the length of the filtering memory, and enhance the role of new observations under the current estimates. According to [4] its formula is

$$\mathrm{d} = \frac{1-b}{1-b^{t+1}} \quad \text{.....(17)}$$

(b is a constant between 0.95 and 0.99. In this paper, it is 0.99)

Before the implementation of (16), we will use the variance of the current speech frame $\mathrm{R}_U(\mathrm{n})$ to compare with the Threshold U which has been updated in the previous iteration. If $\mathrm{R}_U(\mathrm{n})$ is less than or equal to U, the current speech frame can be considered as noise, and then the algorithm will re-estimate the noise variance.

In this project, $\mathrm{R}_U(\mathrm{n})$ can't replace $\mathrm{R}_V(\mathrm{n})$ directly, because we do not know the exact variance of background noise. In order to reduce the error, we used.

2) Updating the threshold by

$$U = (1-\mathrm{d}) \times U + \mathrm{d} \times \mathrm{R}_U(\mathrm{n}) \quad \text{.....(18)}$$

In (15), d is used again to reduce the error. However, there will be a large error when the noise is large, because the updating threshold U is not restricted by the limitation $\mathrm{R}_U(\mathrm{n}) \le U$ □ .It is only affected by$\mathrm{R}_U(\mathrm{n})$. So, we must add another limitation before implementation (18). In order to rule out the speech frames which their SNR (Signal-to-noise rate) is high enough, it is defined that δ_r^2 □ is the variance of pure speech signals,δ_x^2 □ is the variance of the input noise speech signals, and δ_V^2 □ is the variance of background noise. We calculate two SNRs and compare between them. According to [6], one for the current speech frames is

$$SNR_1(\mathrm{n}) = 10 \times \log_{10}\left(\frac{\delta_r^2(\mathrm{n}) - \delta_V^2(\mathrm{n})}{\delta_V^2(\mathrm{n})}\right) \quad \text{..... (19)}$$

Another for the whole speech signal is

$$SNR_0(\mathrm{n}) = 10 \times \log_{10}\left(\frac{\delta_r^2 - \delta_V^2(\mathrm{n})}{\delta_V^2(\mathrm{n})}\right) \ldots\ldots\ldots (20)$$

In (19) and (20), n is the number of speech frames, and δ_V^2 has been updated in order to achieve a higher accuracy. The speech frame is noise when $SNR_1(\mathrm{n})$ is less than or equal to, $SNR_0(\mathrm{n})$ or $SNR_0(\mathrm{n})$ is less than zero, and then these frames will be follow the second limitation $(\mathrm{R}_U(\mathrm{n}) \le U)$. However, if $SNR_1(\mathrm{n})$ is larger than $SNR_0(\mathrm{n})$, the noise estimation will be attenuated to avoid damaging the speech signals. According to [7], this attenuation can be expressed as

$$R_V(n) = R_V(\mathrm{n})/1.2 \ldots\ldots\ldots\ldots\ldots (21)$$

V. Simulation Results

Simulated Results of Kalman filter: The reason for employing the discrete Kalman filter is due to its very accurate estimation capability. The simulated results presented in this section will prove the above statement. The following results that are about to be displayed indicate the estimates of 5 Kalman coefficients of a 5th order Kalman filter. The following results are obtained by setting 5 Kalman coefficients, -0.8, 0.2, -0.6, 0.7 and –0.4. This is followed by an input of random generated noise In the previous section, it has been proven that the Kalman filter estimates the correct coefficients according to the initialized values set. Another test was conducted to prove that the discrete Kalman filter is capable of adapting to random changes in the coefficients at different iterations. By doing so, it will prove that the Kalman filter can be successfully applied to time varying signals such as speech. In this case, a 5th order Kalman filter will have 5 groups of coefficients, however each group of coefficients have different values set at different iterations

Results of Speech Samples: In this section, 3 different speech samples will be presented, namely all samples will employ a order Kalman filter to reconstruct their output speech. With all 3 speeches, the initiate settings of Process Noise Covariance, Q = 1 x 10-3 and Measurement Noise Covariance, R = 0.1 are used. The input and output reconstructed speech signals are given in the following figures. As you can see, with the current settings of Q and R stated above, speeches are able to reconstructed accurately after several hundred iterations. On the other hand, s680 is unable to be reconstructed satisfactorily. In order for s680 to be constructed satisfactorily, we try tuning the parameters of Q and R.

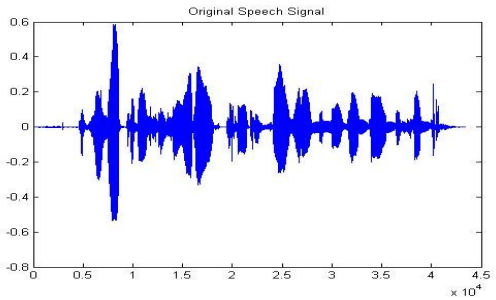

Figure 8. Original Speech Signal

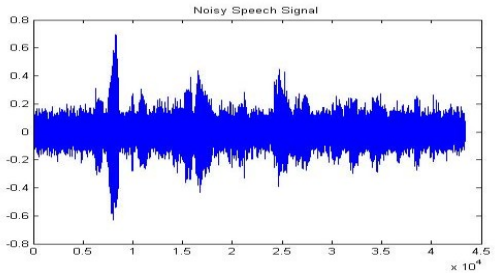

Figure 9. Noisy Speech Signal

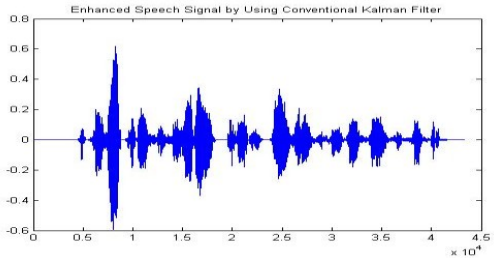

Figure 10. Enhanced Speech Signal by Using Conventional Kalman Filter

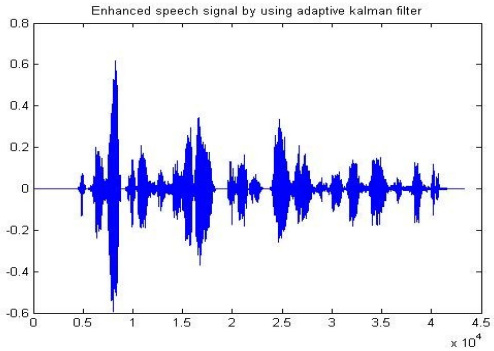

Figure 11. Enhanced Speech Signal by Using Adaptive Kalman Filter

As we can see from the speech signal in Figures above**,** the output speech is not well reconstructed. However if we zoom in to the first few iterations, in this case taking the first 1000 iterations, the output speech has a similar structure to the input signal. The cause of this occurrence is the estimation noise of the Kalman filter. For which the magnitude of the speech signal is comparative high to its input. Due to the scale of the magnitude, it is unable to clearly determine the output speech signal.

Simulated Values:

Algorithm	SNR value	Speech Integebility
DWT	6.9767	0.7034
KALMAN FILTER	8.2773	0.7983

VI. Conclusion:

In this project, an implementation of employing Kalman filtering to speech processing had been developed. As has been previously mentioned, the purpose of this approach is to reconstruct an output speech signal by making use of the accurate estimating ability of the Kalman filter. True enough, simulated results from the previous chapter had proven that the Kalman filter indeed has the ability to estimate accurately. Furthermore, the results have also shown that Kalman filter could be tuned to provide optimal performance.

This test is of necessity for the reason that different signals are bound to be similar but not identical. By and large, this thesis has been quite successful in terms of achieving the objectives. Consequently, perception on signal processing and Kalman filter had also been treasured throughout the process. Most importantly, the skill in time management applied during the research of this project had been developed.

References

Reference Paper

[1]. Robust adaptive kalman filtering based speech enhancement algorithm by Marcel Gabrea IEEE-2004.

Bibliography

[2]. S. Crisafulli, J.D. Mills, and R.R Bitmead, "Kalman Filtering Techniques in Speech
[3]. Coding". In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing,
[4]. S. Saito and K. Nakata, "Digitization", Fundamental of Speech Processing"
[5]. C. R. Watkins, "Practical Kalman Filtering in Signal Coding", New Techniques in Signal Coding,
[6]. M.S. Grewal and A.P. Andrews, Kalman Filtering Theory and Practice Using MATLAB.