CHAPTER 4

PARAMETRIC REPRESENTATION BY PREPROCESSING

The research work attempts to authorize the voice pathology disorder from the various input speech signal by preprocessing performances. The vital role of preprocessing is considered as Noise Removal and Silence Removal. This helps to find the acoustic parameters by Electro Glotto Graph (EGG) with the help of Wiener Filter and DWT Filters. The Wiener Filter and DWT Filters generate the Electro Glotto Graph (EGG) of each input speech signal. For this concern, this research proposed the Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDWT) measure to calculate the enhanced speech signal from the affected raw input speech signal. The preprocessing progression is labeled as a denoising estimation in voice pathology disorder. Perhaps the Wiener filter minimizes the Mean Square Error between the estimated random process and the desired process by Linear Time-Invariant (LTI) filtering. Similarly, the Discrete Wavelet Transform (DWT) produced the discrete wavelets sampled over Fourier transforms. The present work combines both the "Wiener Filter Minimization" as well as "Discrete Wavelets Sample" called Hybrid Wiener Filter Discrete Wavelet Transform (HWFDWT).

Voice Denoising is a vital issue for pathology voice prediction. This vital issue is rectified by various denoising processes like Endpoint Detection, Pre-emphasis, Framing, Windowing, Noise Removal, and Silence removal. This research scrutinizes the input voice signals from other noises and silence. Many researchers tried to remove the Noise and Silence from the original voice. Previously, the speech and noise model was proposed by Wilson et al. (2008) but it displayed incorrect results for some infected voice's noisy conditions. To overcome this problem, the Wiener Filter and Discrete Wavelet Transforms are combined for preprocessing called Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDWT). With the help of Linear Time-Invariant (LTI) filtering and Voice Enhancing Threshold (VETh), the Voice Denoising is carried out by the Wiener filter. The input noisy voice signals always have a high "Mean Square Error". This occurrence is also called as voice noise spectrum. In this case, the infected voice's noisy conditions were removed by thresholding the "Mean Square Error". The Mean Square Error of noise signals is relatively large compared to silence signals. For this, a smaller coefficient is considered as the zero-

Mean Square Error. So it is the finest elimination of Silence signal while protecting the Original input signal. From this consideration, the Wiener filter minimizes the Mean Square Error between the input noisy voice signal and output Denoised Voice. The main goal of the HWFDWT is to estimate the silence signal using related noisy voice signal. The Mean Square Error is calculated for each input signal. The Mean Square Error is insignificant for some input signal called silence filtered signal. The statistical estimation is carried out to all signal filtering. Meanwhile, systematic pitch period and approximation pitch periods were calculated based on the acoustic indices' types. With respect to the acoustic indices, Electro Glotto Graph (EGG) is plotted for voice samples. This EGG is plotted for Constant voice signal glottal period denoised voice signal strength. In preprocessing analysis of input noisy voice signal, the silence and noises were removed with respect to the Electro Glotto Graph (EGG) plotted. In such a way, the input noisy voices are taken from the two different test sets such as www.stimmdatenbank.coli.uni-saarland.de and another private dataset from Department of Pathology, Karpagam Faculty of Medical Sciences and Research, Coimbatore, subject to the analysis terms. The Laryngitis acoustic, laryngoceles acoustic, dysphonia acoustic, diplophonia acoustic and chorditis acoustic analysis is also carried out on these two test sets of dataset signals. Subsequently, these preprocessed voice signals are applied to various selection and extraction techniques used for voice signal analysis. The preprocessed output is calculated based on the Signal to Noise ratio, calculated for the sampled data of both normal and pathological persons; SNR graph is plotted for different filters. After analytical examination, the SNR values show that the proposed Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDWT) outperforms well in the noise reduction process.

4.1 Vital Role of Preprocessing

Preprocessing is considered as the initial processing phase to smoothen the input taken from the two different databases for several disorders of speech such as Laryngitis Acoustic, Laryngoceles Acoustic, Dysphonia Acoustic, Diplophonia Acoustic and Chorditis. Acoustic speech analysis is considered by expert reputed pathological doctors. Preprocessing Acoustic Parameters such as Signal Energy, Pitch, Silence removal, Noise Removal, and Mel filtering must be identified as the male normal speech and male pathological speech as well as female normal speech and female pathological speech. This differentiation in speech is evaluated by using thresholding the "Mean Square Error". Preprocessing is to make way to find a pathological voice by analyzing for the input signal of different voice from the database, namely, Saarbruecken Voice Database and private dataset from Department of Pathology, Karpagam Faculty of Medical Sciences and Research, Coimbatore to recognize various pathological voice disorders.

Preprocessing is an exact phenomenal to observe the pathological voice disorder mainly with respect to the Silence removal and Noise Removal by extracting the necessary parameters on the voice's noisy conditions.

4.1.1 Preprocessing Phase

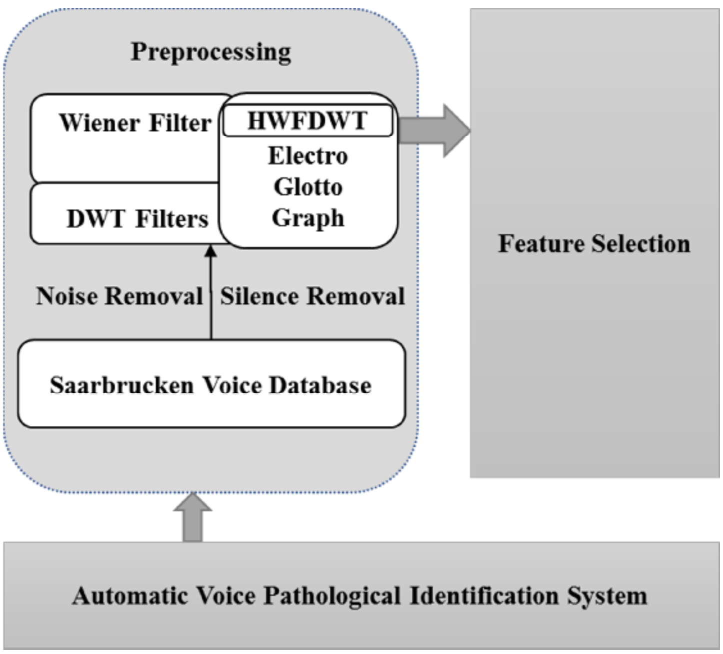


Figure 4.1 Steps in Preprocessing

Figure 4.1 depicts the steps involved the preprocessing phase of the research work as per the vital role of preprocessing, the pathological speech is preprocessed confidently to identify the acoustic features and its purposes. For this preprocessing, the focus is on the following:-

- Parameters Diagnostician,
- Identification Pathological Voice Difference,

- Denoising Estimation,
- Voice analysis by Thresholding,
- Mean Square Error calculation, and
- SNR analytical examination values.

In preprocessing, various filters are used for voice enhancement and recognition with respect to the acoustic factors of pathological voice.

4.1.2 Requirements for Preprocessing

- To generate a parametric representation of input voice signal by preprocessing using the proposed Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDWT).
- Parametric representation by preprocessing helps to find the acoustic parameters using Electro Glotto Graph (EGG) with the help of Wiener Filter and DWT Filters.
- Preprocessing representation overcome infected voice noisy problem. This research forms the Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDWT). This HWFDWT is formed by the combination of Wiener Filter and Discrete Wavelet Transforms. Linear Time-Invariant (LTI) filtering and Voice Enhancing Threshold (VETh) are also used to voice denoising.
- Scrutinize the acoustic factors and voice features in a male/female with the help of comparison between silence signals with the normal signal as well as abnormal signal.

4.2 Voice Database Specification

Voice Database is an assortment of vocal sound recordings from various persons. A single vocal sound recording Database comprises the Recording of the vowels [a,e,i,o,u] with respect to pitch. Generally, there are various forms like the low pitch, normal pitch and high pitch. The same vowels [a,e,i,o,u] also differ with respect to the falling- rising-pitch. The sentence also differs from the actual sentences. For example, "Good, how are you?" may be heard like “Guten, weight es Ihnen?” From this consideration, the voice input and the EGG speech signal have been listed in the individual categorized files. Depending on the user input text-file the desired output voice signal was downloaded from the database. From this downloaded file and recorded file, the quality of the recording was not the up to the requirement, because of the characteristic of low pitch, normal pitch, high pitch, and falling- rising-pitch.

4.2.1 Saarbruecken Voice Database

Saarbruecken Voice Database contains 2000 person's voice database with various pitch levels. It is a downloadable database with respect to the syntax rules creation and deletion. This rule will work based on the gender and age groups. For example, if the syntax is 25, the files of speakers who are 25 years old are downloaded. If the syntax is 26-30, the files of speakers who are at least 26 years old and 30 at the most are downloaded. This downloaded file is diagnosed with respect to the pathologies such as obligatory pathologies and excluded pathologies. Those obligatory pathological and excluded pathological are sorted with respect to the user input condition specified on the sorting stage. From this condition, the recording sessions were grouped by the speaker. The retrieved recording section might be containing the sex of the speaker, the age of the speaker, the diagnosed pathologies status, the remarks about diagnosis, and also comments about the recording session. After the status of the retrieved recording, the section has been saved, the Saarbruecken voice database might be worked based on automatic reset, approval without changes and saved unchanged. Finally, the speech and EGG downloaded signals listed recordings file should be the original NSP and EGG file format and.WAV format.

This research uses The Saarbruecken Voice Database (SVD) initiated for the general purpose in healthy/pathological voice classification. In this Saarbruecken Voice Database, each voice has vowels /a/, /i/ and /u/. Each vowel has High tones, Low tones and Neutral tones with a total of 9 voice sectors. These voice sectors include the parameters jitter, shimmer, and HNR. The number of samples taken from Saarbruecken Dataset is provided in Table 4.1.

Table 4.1- Saarbrücken Voice Database

Saarbruecken Dataset	**http://www.stimmdatenbank.coli.uni-saarland.de/help_en.php4**
No. of Samples	**2000 samples; Male-1000; Female-1000;**
Recording Collections	**vowels /a/, /i/, /u/ and Phrases**
Minimum Duration	**2 seconds**
Frequency	**50 kHz**
Resolution	**16-bit**
Voice Disorder Types	**Laryngitis, laryngoceles, dysphonia, diplophonia, and chorditis**
Age	**30-35**

The Saarbruecken Voice Database is downloaded from the http://www.stimmdatenbank.coli.uni-saarland.de, which is a collection of more than 2000 person's voice recordings and EGG signals. It contains recordings of 1000 males (687 healthy ones and 313 patients) and 1000 females (727 healthy ones and 273 patients) with different voice pathologies. These pathologies have the Vowels /i/, /a/, /u/ produced at Low pitch, Normal Pitch and High pitch with a rising pitch - falling pitch. The input samples of vowels are 2 seconds long, sampled at 50 kHz with a 16-bit resolution which is recorded in the same environment of 30- 35 age group people.

4.2.2 Private Real-time Database

The private real-time dataset contains the 80 samples voice database with various pitch levels. It is a manually collected database with the support of the Karpagam Academy of Higher Education (KAHE), Coimbatore, Tamil Nadu, India and Pathology Department from Karpagam Nursing College. Karpagam Academy of Higher Education is an educational institution located in Coimbatore Tamil Nadu, India. It was established under Section 3 of UGC Act 1956 and is approved by the Ministry of Human Resource and Development, Government of India. This data set will work based on the gender and age groups. A collection of 80 person's voice recordings are collected in order to perform the testing, it contains recordings of 50 males (Normal – 20; Pathology - 30) and 50 females (Normal – 20; Pathology - 30) with different voice pathologies. This pathology has the Vowels /a/, /e/, /i/, /o/, /u/ produced at Low pitch, Normal Pitch and High Pitch also with a rising pitch - falling pitch. The input samples are samples of vowels are 2 seconds long, sampled at 50 kHz with a 16-bit resolution which is recorded in the same environment of 30- 35 age group people. The number of samples taken from Real-time Dataset is presented in Table 4.2.

Table 4.2- Private Real-Time Database

Real-time Dataset	Department of Pathology, Karpagam Faculty of Medical Sciences and Research, Coimbatore.
No. of Samples	80 samples; Male-20; Female-30;
Recording Collections	vowels /a/,/e/, /i/ /o/, /u/ and Phrases
Minimum Duration	2 seconds
Frequency	50 kHz
Resolution	16-bit
Voice Types	Normal – 20 pathology - 30
Age	30-35 years

4.3 Acoustic Parameters

Many researchers analyzed the acoustic parameters with respect to the acoustic analysis such as Mel filtering, Jitter, Shimmer and Pitch for Silence Removal and Noise Removal. This acoustic analysis gives the successive result of pathological voice quality from the vocal fold vibration. This study concentrates on the voice quality checking and removal of Silence and Noise due to the breathiness, hoarseness and roughness that cause voice pathology. So, the researcher has considered the various perturbations to find the difference between a normal speech signal and the pathological speech signal. An attempt is also made to find many voice disorders. Some of the cases may fail to measure the Mel filtering, Jitter, Shimmer, and Pitch. This reflects on the prediction of vocal quality. In this research, The Sensitivity, Selectivity, and Accuracy may be failing to predict the exact result of the acoustic analysis. The Sensitivity, Selectivity, and Accuracy are the most common prediction measures for the Mel filtering, Jitter, Shimmer and Pitch calculation. These factors might be varied depending on the fundamental frequency of the input voice signal.

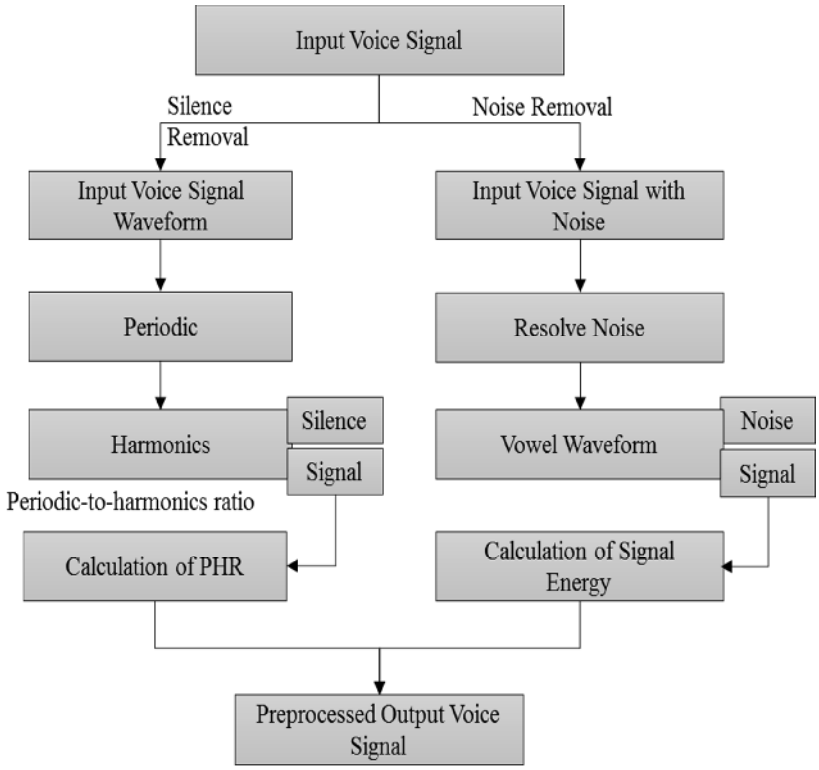


Figure 4.2. Preprocessed Voice Signal

From this acoustic parameter with respect to the acoustic analysis, the input voice signal is applied to the two major processes such as Periodic-to-harmonics ratio measurement and Calculation of Signal Energy. It does not affect the input voice signal, because, the input voice signal is in the form of a glottal waveform. Generally, the glottal waveform does not allow the glottal noise inside the Periodic-to-harmonics ratio and Signal Energy measurement. This is highlighted in the above Figure 4.2. and, both remove the Silence and Noise of pathological voice samples.

In this acoustic analysis, the input voice signals are applied to the Periodic-to-Harmonics Ratio (PHR). The PHR is calculated based on the Input Voice Signal Waveform using equation 4.1. Here, the minimum 25 glottal cycles are taken for estimation of the silence voice prediction. This will be calculated based on the following formulae.

$$PHR = \frac{1}{Avg\left\lfloor \max_{Periodic} glottalnoise \right| - Avg\left\lfloor \max_{harmonics} glottalnoise \right|}$$

$$\text{G}lottal\ \text{N}oise = averagedwaveform - individualglottalwaveforms$$

(4.1)

This glottal noise calculation, measuring the PHR measures, also attempts to resolve the vowel waveform into signal and noise components. It calculates the signal energy of this glottal noise signal.

4.3.1. Acoustic Analysis

From the analyzed voice signal, only the standard duration voice signal segments of the input waveform of continued vowel (a) (e) (i) (o) (u) of 25 samples were sent for acoustical analysis (Muhammad, G., et al 2011). Each sample has an average of 30 sec to 35 sec of standard duration voice signal segments. The initial and end standard duration voice signal segments were cut off, i.e. some cases of ~0.25s may be Noise/Silence and final ~0.25s maybe Silence/Noise). The subsequent standard duration voice signal segments (25 Sec) are taken for the Silence/Noise analysis. i.e. SNR ratio calculation. From this consideration, the sustained vowels|(a)| |(e)| |(i)| |(o)| |(u)| were discarded. This elimination reflecting the starting and the end of the voice signal did not influence the preprocessing result. Simultaneously, the acoustic Input Voice Signal Waveform and Input Voice Signal with Noise data were measured for the fundamental frequency with a fixed 100- Hz sampling frequency. Meanwhile, the mean of 25 samples was calculated for each of the Acoustic Parameters such as Silence removal and Noise Removal.

Preprocessing the pathological voices, so many acoustic factors are established based on the vocal cords data. These factors are analyzed based on the fundamental frequency in noisy environments. For this noisy environment, the feasibility of pathological voices is classified with respect to the Periodic-to-Harmonics Ratio (PHR) measures. However, the Periodic-to-Harmonics Ratio (PHR) measures consider the signal energy and fundamental frequency of the input voice signal. To analyze the signal energy and Mel filtering, the averaged Glottal Noise is calculated from the glottal waveforms. It is used to preprocess the non-stationary signals and track the higher order energy of harmonic oscillators. In this research, the Periodic-to-Harmonics Ratio (PHR) measures are used to efficiently preprocess the input voices in-terms of silence and noise removal on Saarbruecken Voice Database (SVD) and Real-Time data set collection. For finding the preprocessing, the commonly SNR analytical examination values were implemented.

a. Noise Removal

The air breath out over the vocal tract is comparatively low with respect to the loud voice breathe out power. The mean velocity is often very minimal at the stage of breath out over the vocal tract. This is called Noise. In preprocessing, some of the voice signals have the partial closure of the glottis. This affects the amplitudes of the discontinued voice signal at the vocal tract (Parsa, V., & Jamieson, D. G. 2000). The lower amplitudes of discontinued voice signals reduced by increasing the breath out- breathe in the vocal tract amplitudes of the discontinued voice signal. This causes the improvement of voice signal energy, and so increasing signal energy might cause Noise Removal. This ratio might be maintained till all noise is removed from the discontinued voice signal samples at vocal tract. This method is called the normalized noise energy.

b. Signal Energy

The signal energy calculation is a predominant one in acoustic voice analysis approaches. This signal energy calculation is used to measure a vocal function of fundamental frequency or affected pathology.

Signal energy calculation is an important fact in acoustic methodology, which has the reversal of being from non-noise signal, it requires clear input voice signal strategies and being conceivable on a recording system, yet since the system has numerous pathological applications for the early detection and differential finding of many pathologies, similarly with respect to the quantitative examination of the vocal limit of patients encountering such treatment as therapeutic strategy and voice treatment. The Signal energy calculation is also used to recognize the voice disorders from the general speech communication processes. Also, it makes a significant contribution to noise removal. The measures reported so far are primarily associated with the following acoustic characteristics of the voice signal:

In the acoustic examination of the infected voice, which is of fundamental frequencies is a choice of acoustic assessments which reflect the vocal limit and physiological behaviors of the voice (Moran, R. J., et al 2006). Since preprocessing reported might be useful in the revelation of infected pathology. Diverse undertakings have been made to search for significant acoustic measures. The measures of voice end to end are mainly associated with the fundamental pitch period, a signal with vocal noise, waveform variations, frequency variations, signal transition and so on. These are the acoustic examination over the acoustic characteristics of the input voice signal.

c. Silence removal

Silence removal is normal in voice pathology procedures on removing the initial voice of input signal with the quantity of ± 20 ms of the vowel. So, the maximum level of silence might be removed on the initial voice of the input signal and at the end of the voice signal. The fundamental frequency was calculated with respect to the input vowel. This frequency calculation was used in the acoustic examination of pathological voices.

d. Pitch

The preprocessing valuation is used to evaluate the input voice signal in the glottal waveform. This process is called as Voice Signal Pitch Marking (VS-PM).

The Pitch extraction technique has the following procedures. The steps are discussed below and the flow diagram presentation of Voice Signal Pitch Marking (VS-PM) is shown in Figure 4.3.

Step 1: In the Earlier stage of pre-processing the input voice signal is considered as the Glottal Closure Instant (GCI). Sometimes this glottal closure is secured into the Glottis, in such a way the Glottis may be continuing or discontinuing in the input voice signal. At this stage, the continuing signal is already preprocessed in the Acoustic analysis of silence/noise removal process. Other than the continuing signal, the irregular input voice signals are also processed under the Pitch extraction technique with respect to the abrupt closure of the vocal folds and Passage Interruption.

Step 2: To find the residue signal from the resulting discontinues input voice signal. This signal may apply to the inverse filtering by the LPC filters. The result causes the speech signal to traces the glottal flow in the glottal closure instant (GCI). This glottal flow is called as the Glottal Flow Approximation (GFA). The GFA having a discontinuity in the flow produces large negative peaks (Muhammad, G., & Melhem, M. 2014).

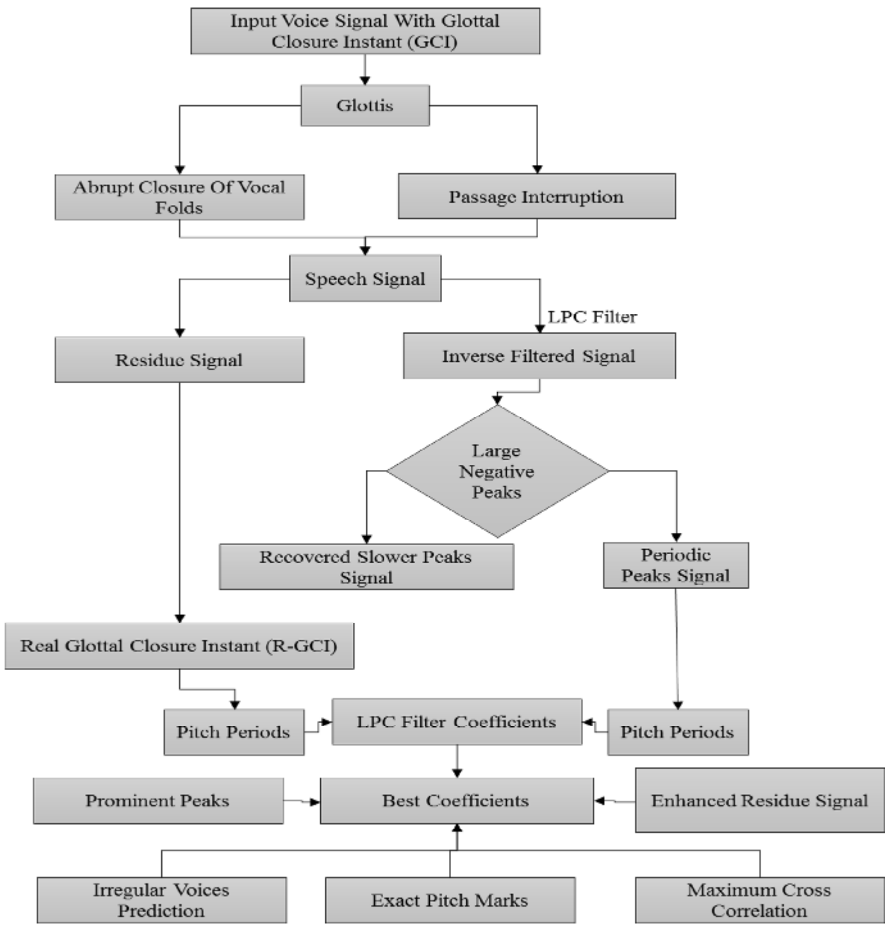


Figure.4.3. Voice Signal Pitch Marking (VS-PM)

Step 3: In large negative peak, the recovered signal is always audible compared to the glottal waveform signal. Because of peaks recover at the stage of vocal fold's closing/opening process. The closing/opening process might be calculated with respect to the fundamental frequency of voice vibration periodic peaks.

Step 4: A general process used to glottal closure instant detection is the Real Glottal Closure Instant (R-GCI) implementation. This process is works based on the Voice input data which is taken from the SVD data set as well as Real-Time data set.

Step 5: The implemented result has the LPC Filter Coefficients for detecting the sustained vowels from the input data set. At this stage, the input voice data set may be modified with respect to the glottal closure instants. The modification changes may analyze at the stage of approximation of the real GCIs. This approximation result produces the Best Coefficients of each glottal closure instants.

Step 5: The Best Coefficients of each glottal closure instants reflect the following result of pitch preprocessing:

- Prominent Peaks,
- Enhanced Residue Signal,
- Irregular Voices Prediction,
- Exact Pitch Marks, and
- Maximum Cross-Correlation.

Algorithm for Pitch extraction techniques

Read: whole SVD and private body dataset, amplitude sequences,
autocorrelation pitch detection (APD)
Input: Vowel interval (VI), pitch period,
Compute: average pitch period (APP)
Split (VI)
Compute: accuracy (APP)
APD = short time autocorrelation (preprocessed signal)
Normalization (APD) by median
Check: if pitch period > short time autocorrelation
Return: average pitch period
Else: Silence Voice
Check: if pitch period < short time autocorrelation
Return: peak locations
Else: Silence Voice

Figure 4.4. Algorithm for Pitch Extraction

The algorithm for pitch extraction is given in Figure 4.4, this will be executed until the end of the Autocorrelation Pitch Detection (APD) interval is reached. This helps to detect the accurate pitch period and amplitude of the negative side of the signal.

Table 4.3. Description of Preprocessing Datasets

Preprocessing	**Requirements**	
Database	**SVD**	**Private**
No. of Voices (Pathological + Healthy)	**2000 samples; Male-1000; Female-1000**	**80 samples; Male-20; Female-30**
No. of Voices used for Training	**Male-1000 Female-1000**	**Male-50 Female-50**
Pathological + Healthy	**(687 - healthy and 313 - patients) (727 - healthy and 273- patients)**	**(Normal – 20 and Pathology - 30) (Normal – 20 and Pathology - 30)**
Accuracy (%) over Preprocessing before Silence Removal	**77.90**	**84.69**
Accuracy(%) over Preprocessing after Silence removal	**96.5 - 95.50**	**97.6 - 98.50**
Accuracy(%) over Preprocessing before Noise Removal	**82.37**	**88.35**
Accuracy(%) over Preprocessing after Noise Removal	**95.20**	**97.56**

The Acoustic Parameter evaluation depends on the fundamental frequency alone and it is somewhat difficult to find the Voice pathology. But in case of preprocessing, the combination of Fundamental frequencies makes the evaluation quite easy to compare between normal and pathology voices. The Acoustic parameter Signal Energy was used along with the pitch value to extract the speech signals alone present in the input signals.

The pitch extraction measurements of the input voice signal are calculated for both the databases used in this research. The accuracy measures are considered over preprocessing before and after Silence Removal and also for Preprocessing before and after Noise Removal. The performance of the system was evaluated by using the measurements is given in the above Table 4.3.

4.4 Electro Glotto Graph (EGG)

Input voice signals may be imperfect by dissimilar forms of additive noises that diminish the excellence and clarity in the voice. Therefore, Electro Glotto Graph (EGG) based noise reduction processes are planned to get rid of the additive noise from a speech signal. Noise reduction methodology is employed to considerably decrease noise and intrusion and, hence it progresses the perceived excellence and clarity of a speech signal. The matter of noise reduction has attracted a substantial quantity of attention over the past many decades. Wiener filter could be a measure of the foremost elementary noise reduction approaches, which have been described in various forms and adopted during a kind of applications.

The choice between using and not using a noise reduction technique may have a significant impact on the functioning of these systems. Noise reduction is a very challenging and complex problem according to the input voice quality. First of all, the nature and the characteristics of the noise signal vary significantly from application to application, and moreover vary in time. It is therefore very difficult, and impossible, to develop a versatile algorithm that works in diversified environments. Secondly, the objective of a noise reduction (Bhuta, T., et al 2004) system is heavily dependent on the specific context and application. Speech signals play an important role in conveying messages. Due to this reason, the information of the speech should be preserved. The quality of a speech signal can be degraded by interfering noise in a communication channel.

The quality and clarity of the speech signal are degraded in the input voice signal, it will be difficult to recognize and categorize the speech signals. Consequently, it is aimed to reduce the noise in such a way that the speech would be able to convey the information properly. Wiener filter could be a classical noise reduction methodology that is widely used for removing noises from the speech signal. The Wiener filter methodology supported frequency warp has been planned for noise reduction method. The discontinuities in the frequency of the speech signals are rectified by the FIR filter, which is helpful in noise reduction of speech signals and yields higher speech perceived quality. And the Wiener filter will make changes to the speech signal, in order to perform noise reduction and speech distortion. By shaping a speech-distortion guide to assess the degree to which the speech signals are irregular and two noise-reduction factors to work out the quantity of noise being attenuated, this research provides the quantitative presentation performance of the Wiener filter within the framework of noise reduction.

Wavelet Transform (WT) could be a powerful tool for removal of noise from varied signals. Combining Wavelet Transform (WT) with alternative noise reducing techniques might lead to a reduction of noise. Almost like WT, Singular decomposition (SVD) is additionally an efficient noise reduction tool. The Wiener filter and Wavelet transform are applied in combination for the method called Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDWT) which is precisely reducing the noise. In case, there are multiple mic sensors, the multiple observations of the speech signal which might be used for noise reduction with less or perhaps no speech distortion.

Noise reduction is used in voluminous series of applications such as hands-free mobile phones, teleconferencing, in-car cabin communication, etc. For signal and image processing the Wavelet Transform (WT) is a powerful tool. In various research fields like signal processing, image compression, pattern recognition, etc. Continuous Wavelet Transform (CWT) gives reliable and detailed time-scale information compared to classical Short Time Fourier Transform (STFT). Wiener filters are used for development of speech signal which is surrounded with background noise. The procedure of speech enrichment for noise reduction aims at reducing the power of preservative noise by Wiener filtering. By combing the wavelet transform and Wiener filter, the noise is reduced by the proposed algorithm.

4.4.1 Wiener filter and DWT Filters

The Wiener filter and DWT Filters signal preprocessing process is shown in Figure 4.5. The Wiener filter will filter and estimate of a desired target preprocessed method by Linear Time-Invariant (LTI) filtering of Wiener Filter method, it would filter the eminent input voice silence signal and the noise signal. The Wiener filter minimizes the mean square value between the input voice signal and pathology voice detection method. The goal of the Wiener filter is to find a mathematical solution for unknown signal processing with the silence and noise signal discrimination. This noise signal will be applied to the Wiener filter, and also, performs filtering the silence signal to provide the noise removed output. The Wiener filters are often used to filter the noise from the infected signal and also produce a preprocessed voice signal.

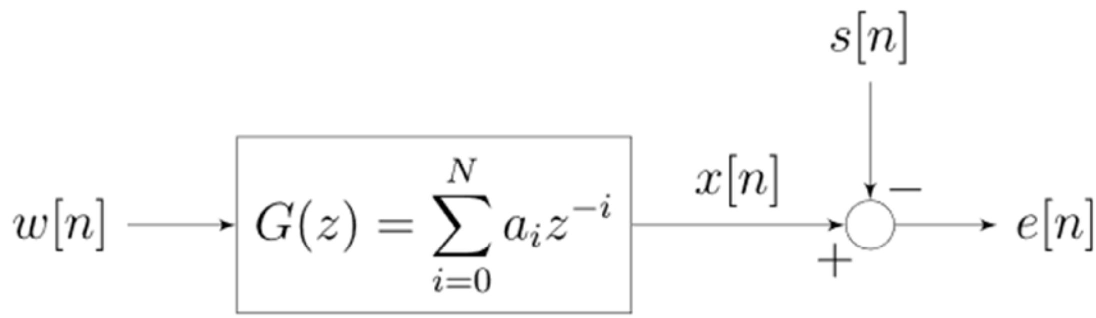


Figure 4.5. Block Diagram of Wiener filter

a. Wavelet Transformer

A wavelet transformer is a mathematical relation, which helps in the digital signal processing and voice signal compression. In the signal processing, wavelets create potential to recover weak signals from noise. A wavelet could be a wave-like oscillation with associate amplitude that begins at zero, increases, and decreases back to zero. It will generally be unreal as a "brief oscillation" like one recorded by a measuring instrument or monitor. Generally, wavelets are on purpose crafted to possess specific properties that create them helpful for the signal processor. Employing a "reverse, shift, multiply and integrate" technique known as convolution, wavelets may be combined with notable parts of a broken signal to extract information from the unknown parts. In arithmetic, a riffle series could be an illustration of a square-integrable (real- or complex-valued) performs by an explicit orthonormal series generated by a wavelet is presented in Figure 4.6.

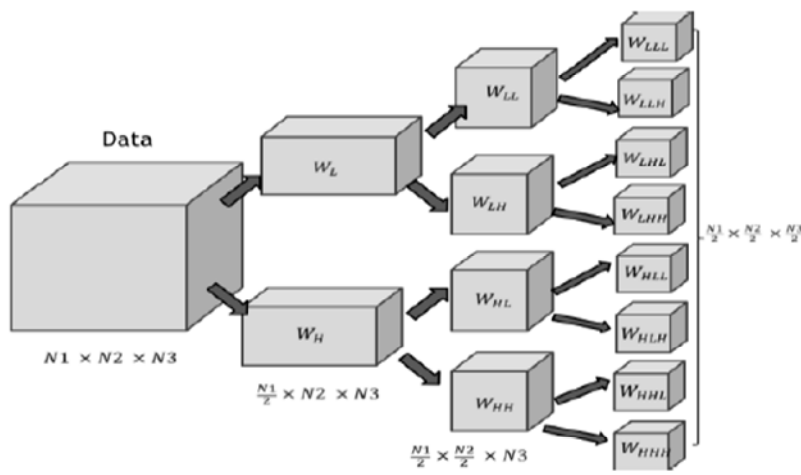


Figure 4.6.Wavelet Transformer System Design

The speech enhancement methods are centered at suppression of the background noise. However, it is very difficult to suppress the noise when it is combined with speech as there is a chance of information loss. This can be rectified by estimating the background noise in the speech signal. In the enhancement process, initially, the voiced and unvoiced section of the speech signal is detected. It is easy to estimate the noise during the pauses (Silence Section) in speech, and noise estimation in the voice section is done by fundamental frequency. Here, the wave primarily based methodology for the speech improvement is supported by hybrid Wiener filter with slight modification in the rule. This hybrid Wiener filter methodology begins from the interference of speech signal, then preprocessing with completely different window functions on the same length of the block. Further, frequency spectral analysis is carried out using wave filters. Hence, during this methodology, the hybrid Wiener filtering is performed on the wave speech coefficients for noise estimation and suppression.

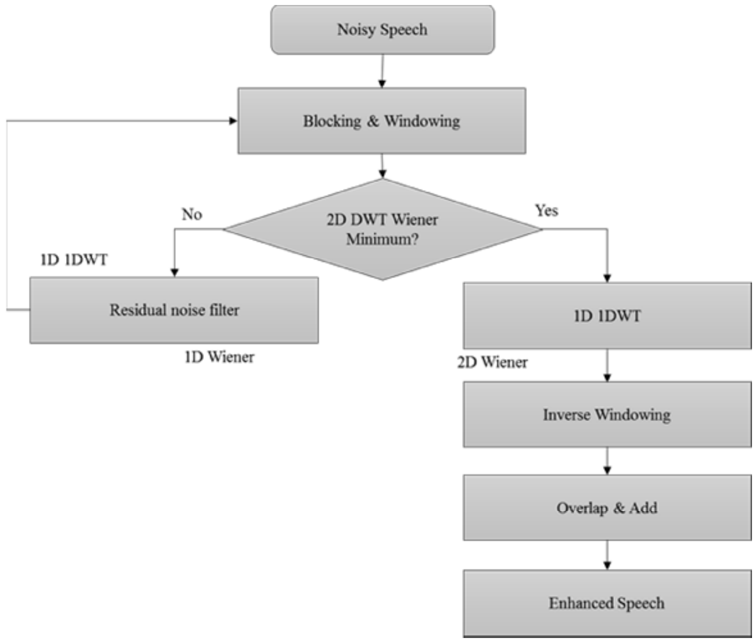


Figure 4.7. Hybrid Wiener Filter Discrete Wavelet Transforms

As seen in Figure 4.7, both the Wiener and wavelet transform filters are combined together as Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDWT) so that the resulting output produced will be specific. Here, the noisy speech signals are given as input. Then the speech signals are divided into blocks, and windowing is done by taking a small subset of a larger dataset and the samples are taken. To those signals one-dimensional WT and Wiener filters are applied, the minimum speech values are considered to perform noise filtering. In the proposed method WT along with Double density, a wavelet is applied to get the resulting output close to the original signals. Finally, WT and Inverse windowing are performed to get enhanced speech as output.

4.4.2. Hybrid Wiener Filter Discrete Wavelet Transform (HWFDWT)

Discrete wavelet transform is widely used in signal de-noising. The DWT is understood as signal decomposition in an exceedingly set of freelance, spatially directed frequency channels. A signal experiences two complementary filters and emerges as two signals, approximation, and details. This is known as decomposition. The parts are assembled back to the first signal without loss of data. This method is named as reconstruction. The Gaussian noise can nearly be averaged into low-frequency wavelet coefficients.

All digital signals contain some degree of noise. The proposed de-noising algorithm attempts to remove the Gaussian noise from a signal. Ideally, the resulting denoised signal will not contain more noise. De-noising of natural images corrupted by Gaussian noise using wavelet techniques is very effective, because of its ability to capture the energy of a signal in little energy transform values. The methodology of the discrete wavelet transform based image de-noising has the following three steps:

- Transform the noisy image into the frequency domain by discrete wavelet transform,
- Apply Wiener filter on each sub-band, by using the local window, and
- Perform inverse discrete wavelet transform to obtain the de-noised image.

The Wiener filter is used to remove Gaussian noise from a corrupted signal based on statistics estimated from a local neighborhood of each speech (Wilson. K et al 2008). This filter depends on the noise power (i.e. noise variance in a corrupted signal). Where the variance is large, the filter performs little smoothing. Where the variance is small, the filter performs more smoothing; this filter often produces better results than other filtering used for speech enhancement (Silva, D. G., et al 2009). In this work, the filter with a local window applied on the DWT to remove Gaussian noise from each sub-band is used.

4.4.3. Denoising Estimation and SNR Analytical Examination

A speech distortion index is defined to measure the degree to which the speech signal is deformed. The noise reduction to quantify the amount of noise being attenuated, analytically examined the performance behavior of the Wiener filter and wavelet transformer for noise reduction technique is calculated. In sample 1(SVD), the SNR value for the Wiener filter algorithm is 5.78 dB, for the DWT it is 6.23 dB and for the proposed system (HWFDWT) it is 7.33 dB for de-noising signals. In sample 2(Real-Time Dataset), the SNR value for the Wiener filter is 6.45 dB, for the DWT it is 7.21dB and for the proposed system (HWFDWT) it is 8.4 dB for feature extraction of noising signals. Since comparing from the existing system, the proposed algorithm is efficient and is used, the SNR value is significant for denoising the signal using the proposed algorithms, which is used in the speech distortion.

a. Signal to Noise Ratio Calculation

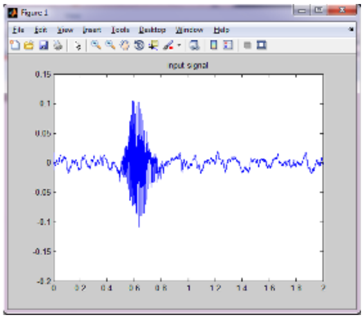
Signal to Noise ratio is calculated for the sampled data of both normal and pathological persons. SNR graph is plotted for different filters. After analytical examination, the SNR values shows that the proposed Hybrid Wiener Filter Discrete Wavelet Transform (HWFDWT) outperforms well in the noise reduction process.

Equation 4.2 shows the calculation of SNR. In this Equation, the SNR values of normal and pathological sustained vices were maximum Silence dB and minimum Noise dB, respectively are calculated based on the voice signal input. Table 4.4.shows the SNR values of normal and pathological running voices Saarbruecken Dataset and Real-Time private Dataset.

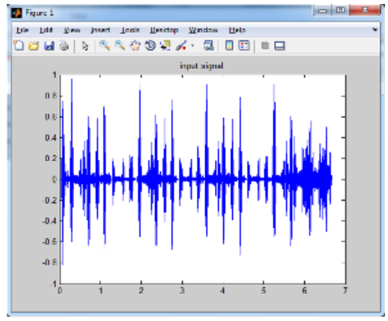
$$SNR = 20 \cdot log10\left(\frac{S}{N}\right) \tag{4.2}$$

Table 4.4. SNR values - Saarbruecken Dataset and Real-Time Dataset

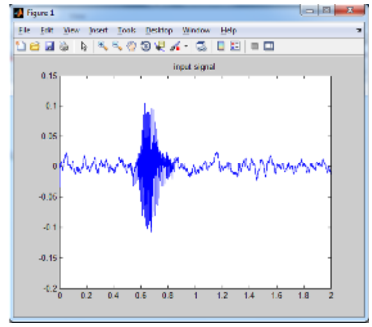
Input Speech Signals	Saarbruecken Dataset			Real-Time Private Dataset		
	DWT (dB)	Wiener (dB)	HWFDWT (dB)	DWT (dB)	Wiener (dB)	HWFDWT (dB)
Sample 1	6.23	5.78	7.33	7.21	6.45	8.4
Sample 2	6.45	5.84	7.12	7.56	5.89	8.12
Sample 3	6.48	5.69	6.96	6.89	5.68	8.32
Sample 4	6.24	5.71	7.14	6.78	6.32	8.47
Sample 5	6.84	5.65	6.98	5.87	5.21	7.58
Sample 6	6.52	5.76	7.26	6.24	5.67	7.68
Sample 7	6.89	6.14	7.56	6.52	5.68	7.21
Sample 8	6.25	5.97	7.23	6.41	5.47	8.14
Sample 9	6.47	5.68	6.87	6.48	5.32	8.31



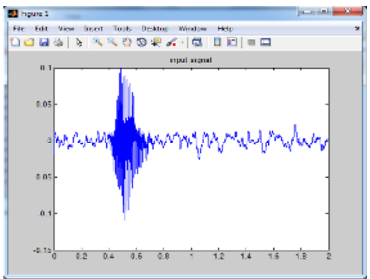
(a) Input Speech Signal Sample 1



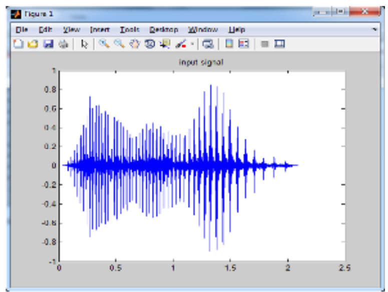
(b) Input Speech Signal Sample 2



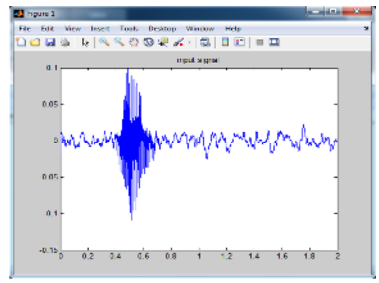
(c) Input Speech Signal Sample 3



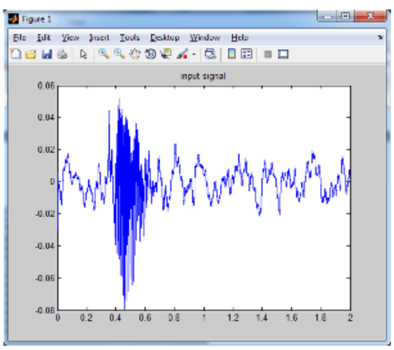
(d) Input Speech Signal Sample 4



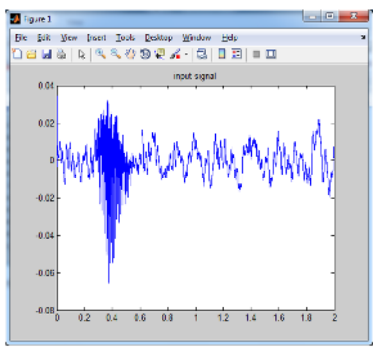
(e) Input Speech Signal Sample 5



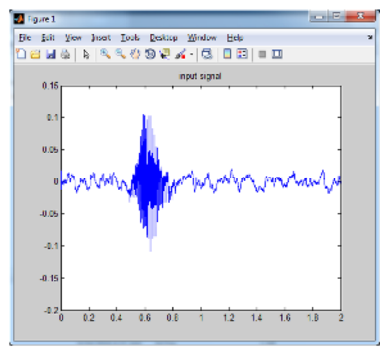
(f) Input Speech Signal Sample 6



(g) Input Speech Signal Sample 7



(h) Input Speech Signal Sample 8



(i) Input Speech Signal Sample 9

Figure 4.8 (a) to (i). Input Voice Samples for SVD

Figures 4.8 present the input voice samples for SVD with different glottal noise. The input of the SVD analysis of preprocessing from the SNR from normal and pathological voices with the removal of different glottal noise, as well as running voices with silence are shown in Figure 4.8. The median values of normal and pathological groups for sustained vowels were 7.934694 and 15.869388 respectively. The median preprocessed values of normal and pathological sustained voice were found by the respective median values of the input voice.

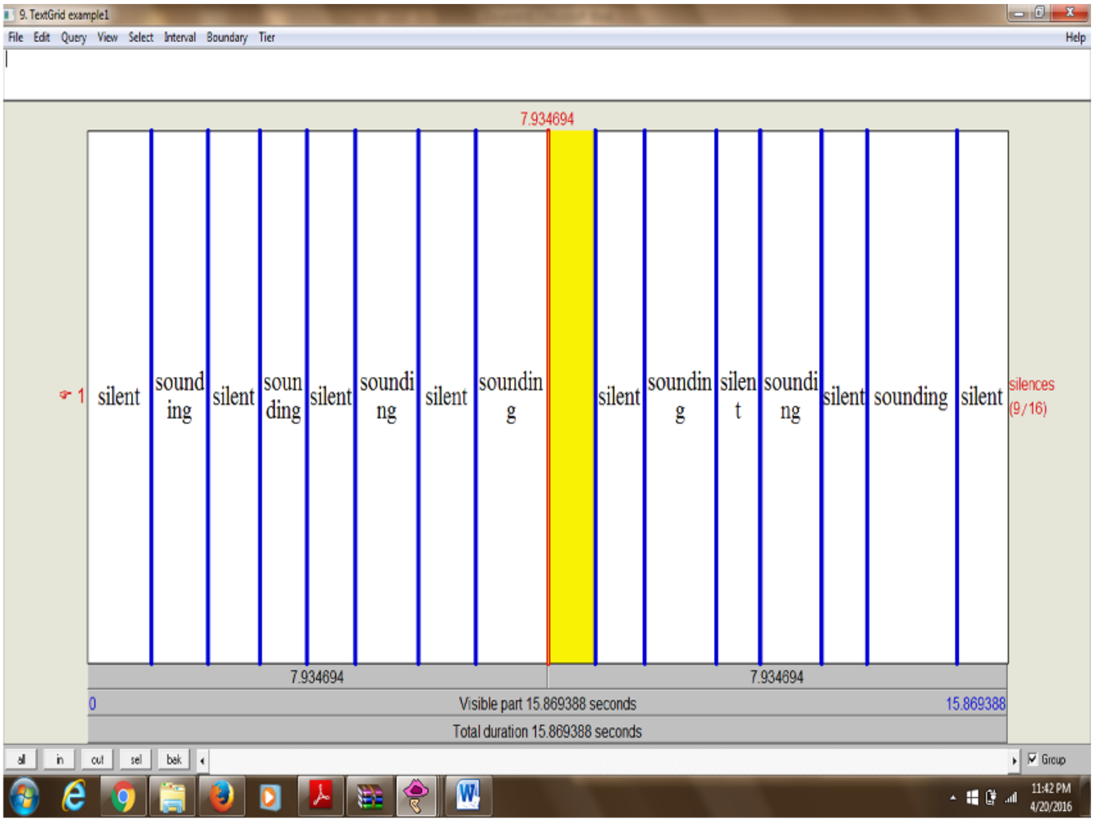
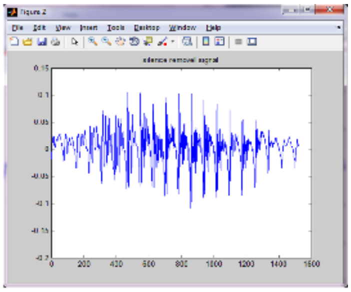
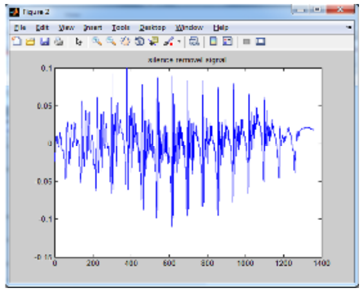


Figure 4.9. Significance level of Silence and Voice

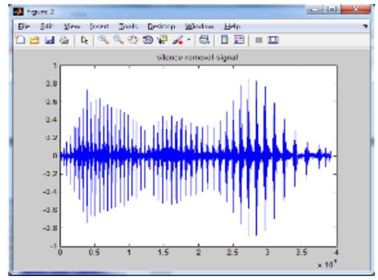
Figure 4.9. depicts the significance level in determining if the discriminatory performance of two silence and voice acoustic variables are significantly different. For example, there was no significant difference in silence areas of the input voice signal, likewise, the indicating voices provide higher SNR discriminatory value.



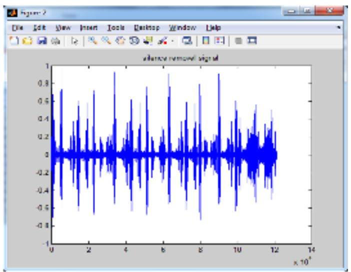
(a) Silence Removal Signal Sample 1



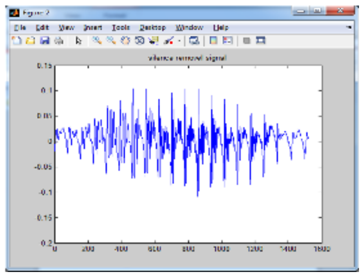
(b) Silence Removal Signal Sample 2



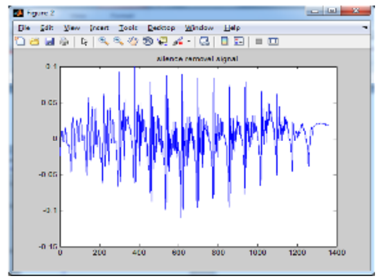
(c) Silence Removal Signal Sample 3



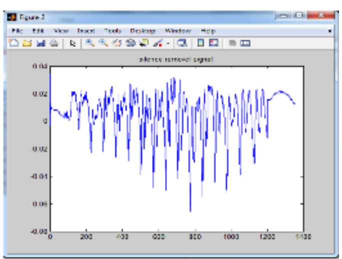
(d) Silence Removal Signal Sample 4



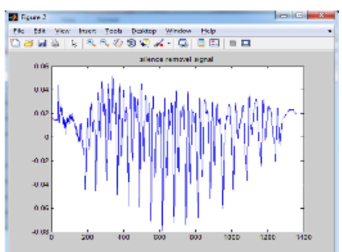
(e) Silence Removal Signal Sample 5



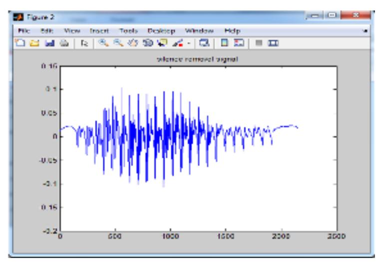
(f) Silence Removal Signal Sample 6



(g) Silence Removal Signal Sample 7



(h) Silence Removal Signal Sample 8



(i) Silence Removal Signal Sample 9

Figure 4.10. (a) to (i). Silence Removal for Input Voice Signals

Figures 4.10 shows the distributions of the signals after silence removal for the normal and pathological input voice signals. Results of silence removal of input voice signal measurements for normal and pathological sustained input voices, as well as running voices are calculated. For input pathological voices, the median values of normal and pathological groups relatively vary depending on the SNR value. Meanwhile, all the SNR varied with respect to the values of normal and pathological samples in the input voice data.

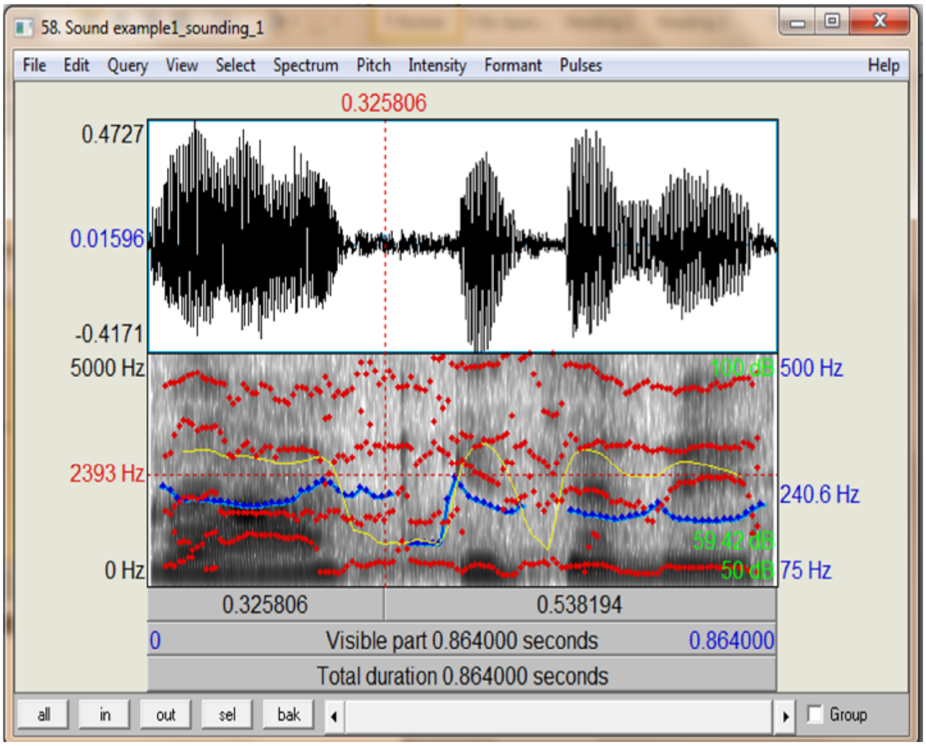
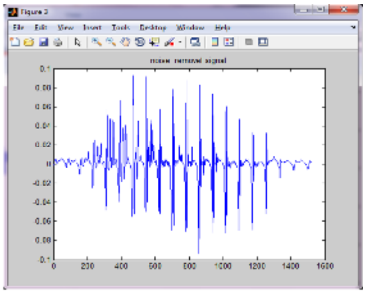


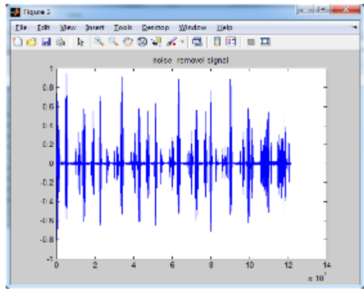
Figure 4.11. Noise Waveform for Input Voice Signals

Figure 4.11. shows the distribution of the input noise waveform for normal and pathological input performance of the signal parameter is similar to silence removal process with the normal and pathological preprocessing.

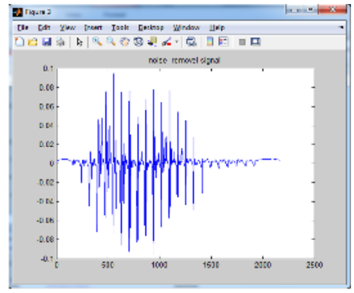
As per Figure 4.12., the noise removal of the input voice signal of running voice samples (1- 9) for both normal and pathological voices was higher than for those of silence removed voice samples. The median value of the SNR of silence removed voice groups was lower than that of noise removed voice. These results were expected due to the high degree of variations presented in silence removed voice samples compared to noise removed voice



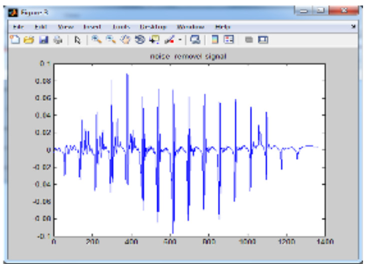
(a) Noise Removal Signal Sample 1



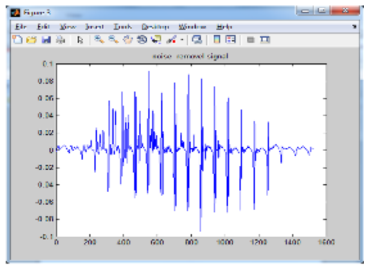
(b) Noise Removal Signal Sample 2



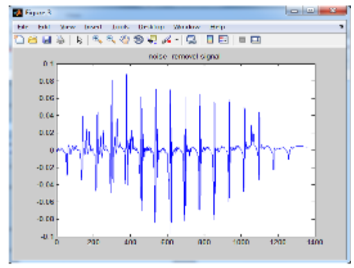
(c) Noise Removal Signal Sample 3



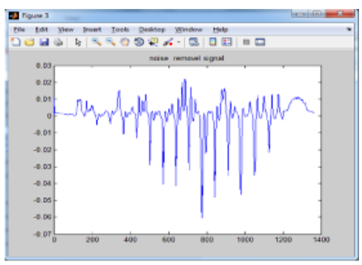
(d) Noise Removal Signal Sample 4



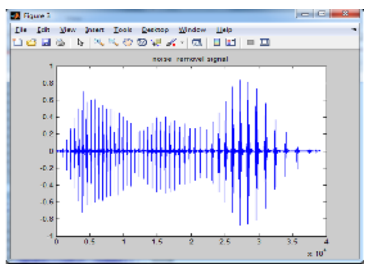
(e) Noise Removal Signal Sample 5



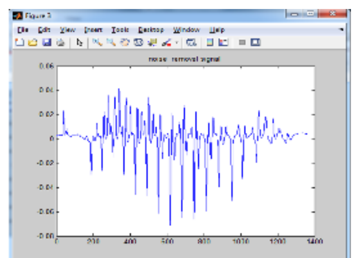
(f) Noise Removal Signal Sample 6



(g) Noise Removal Signal Sample 7



(h) Noise Removal Signal Sample 8



(i) Noise Removal Signal Sample 9

Figure 4.12 (a) to (i) Noise Removal for Input Voice Signals

Figure 4.12. shows the distribution of the Noise Removal Signal Samples for normal and pathological voices using continuous speech. The noise removal voice has performed better than the original input signal on speech samples but lowers to the feature

extracted signal parameter. The noises are removed from the input voice signals and SNR ratio is calculated for the samples.

Figure 4.13 shows the amount of the SNR values of SVD dataset with respect to the silence and noise removal for the normal and pathological input signal. The performance of the SNR ratio is measured.

A. SNR Values for SVD Dataset

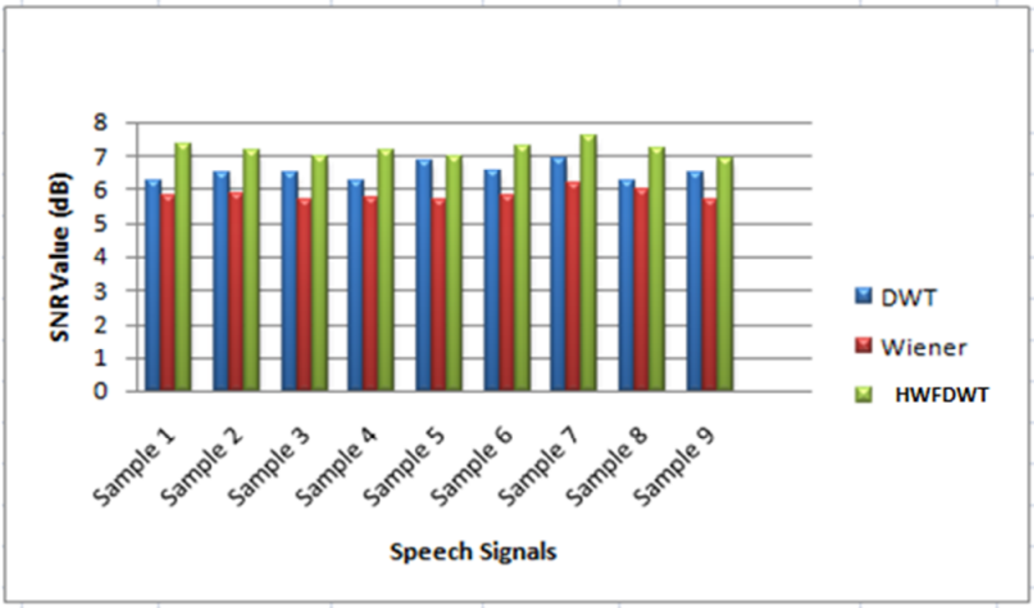


Figure 4.13. SNR Values of SVD Dataset

Figure 4.13 displays the SNR Values for SVD Dataset for normal and pathological voice signal with respect to the distribution of noise removal and silence removed signal. The SNR ratio is calculated with respect to all samples. The samples values are compared to the Wiener filter, Discrete Wavelet Transforms and Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDWT), the preprocessing characteristic for SVD Dataset gives the efficient SNR rate in HWFDWT for all samples. All those values are corresponding to the SNR of normal voices and pathological voices.

B. SNR value for Private Real-time dataset

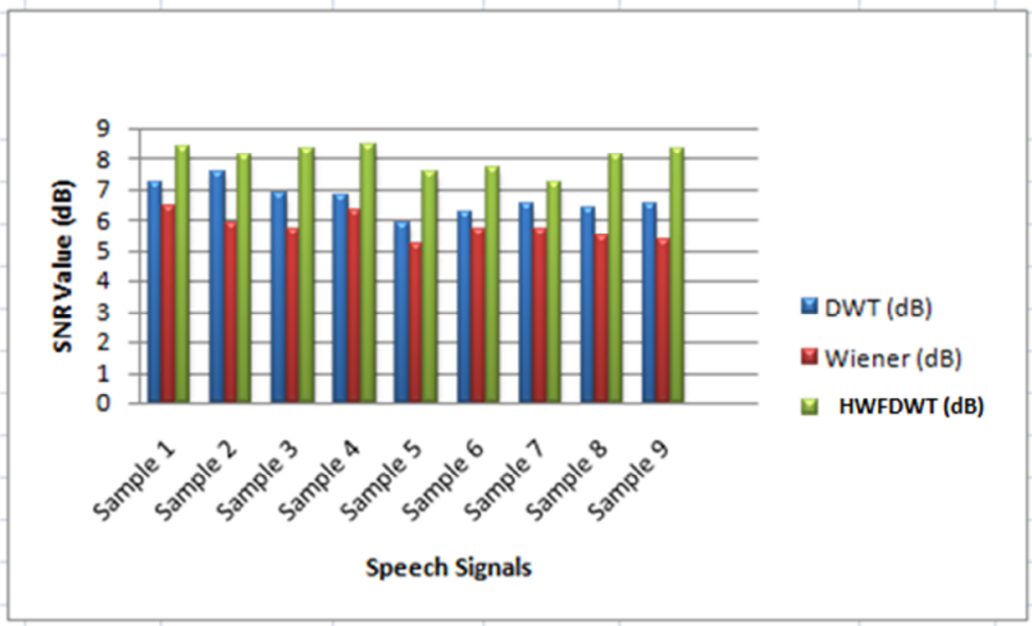


Figure 4.14 SNR Values of Real-Time Dataset

Figure 4.14 displays the SNR Values of the Private Real-Time Dataset with respect to the distribution of noise removal and silence removal signals. The SNR ratio is calculated with respect to all samples. The samples values are compared to the Wiener filter, discrete wavelet transforms and Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDWT). The preprocessing characteristic of Private Real-time Dataset gives the efficient SNR rate in HWFDWT for all samples. All those values are corresponding to the SNR of normal voices and pathological voices.

4.5 Chapter Summary

Finally, it is understood that the preprocessing phase is an essential procedure in identifying pathology disorder from the various affected input speech signal by preprocessing techniques. The Noise and Silence were removed by using Electro Glotto Graph (EGG) with the help of hybrid Wiener Filter and DWT Filters. The preprocessing progression is labeled as a denoising estimation in voice pathology disorder. Perhaps the Wiener filter minimizes the Mean Square Error between the estimated random process and the desired process by Linear Time-Invariant (LTI) filtering. The design of the Hybrid

Wiener Filter Discrete Wavelet Transforms (HWFDWT) algorithm was formed for Voice Denoising in pathology voice prediction. Simultaneously, the EGG plotted for Constant voice signal glottal period denoised voice signal strength analysis is performed. Finally, the preprocessing analysis of input noisy voice signal, the silence and noise were removed subject to the Electro glotto graph (EGG). In such a way, the preprocessed output result was plotted based on the two different test sets such as Saarbruecken database and Real-time dataset subject to the analysis terms. The analytical examination of the preprocessed output voice signals clearly shows that the proposed Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDWT) outperforms well than the traditional Wavelet Transform and Weiner Filtration methods. The next step of the proposed CSOMFCC with optimal feature extraction, which is described in the next chapter, Chapter 5 (**Feature Selection and Extraction**).