



Modified Wiener Filter Based Speech Restoration with Autocorrelation of Isolated Kannada Word Speech Recognition

Puttaswamy Gowda Vanajakshi^{1*} Murugavelu Mathivanan¹ Thangamuthu Senthil Kumaran¹

¹*ACS College of Engineering, Visveswaraya Technological University, India*

* Corresponding author's Email: vanaja_gowda@rediffmail.com

Abstract: There is a big challenge in speech recognition system due to variability of the spoken languages and also speech signal is degraded due the environmental noise. Therefore speech recognition system requires pre-processing, which plays vital role to restore speech signal more effectively. Many challenges are there to restore the speech from different noisy environment. In this paper, proposed method is divided into two parts. First part is Kannada speech restoration, the modified wiener filter is proposed to restore the Kannada speech signal with very good speech quality. The experimental results are verified for different types of noise by varying input Signal to Noise Ratio (SNR) from -20dB to 5 decibels (dB). Second part is Autocorrelation based Kannada speech recognition system. Autocorrelation technique gives better output SNR for degraded input SNR with -20dB. The Autocorrelation is the simple method to recognize the isolated Kannada word in speech signal. This proposed technique results 100% recognition rate for male, female, with different accents.

Keywords: Restoration, Wiener filter, Auto correlation, Speech recognition, Speech quality.

1. Introduction

Speech is one of the supreme efficient methods of communication to transfer the information from source to destination users. For distant communication from source to destination undergo with different environments.

This different environment degrades the information varying speech signal. So, the quality and intelligibility of speech restoration is very challenging for different types of environment. With these backgrounds intelligence will be added to the speech signal. To achieve perception of speech quality is still real challenges for researchers.

Probability distribution over sequences of words that matches the sounds with word sequences which distinguishes between words that sounds similar. The applications of speech recognition are general purpose dictation, military application, healthcare, handheld devices for speech to speech translation. The healthcare Automatic Speech Recognition (ASR) aimed at doctors so as to build patient records mechanically. In vehicles speech devices

take drivers input and display route map etc., help for disable the systems.

Deep Neural Networks (DNN's) are used to determine the mapping function among the noisy and clean speech. The DNN regression model training comprises of unsupervised retraining portion and supervised fine tuning portion. The proposed DNN [1] is also adopted with non-stationary different types of noise to smoothing the speech, Minimum Mean Square Error (MMSE) optimize DNN are used to smoothing the speech quality. DNN based speech enhancement is more effective for different environments, experimentally verified for different Standard data base set (TIMIT set) with an average Perceptual evaluation of speech quality (PESQ) is improving for different training dataset.

The paper [2] proposed features of noise and clean in cepstrum domain. A set of polynomial with respect to the utterance of polynomial are used to approximate the features of clean and noise in the cepstrum domain. Maximum Likelihood (ML) estimation of regression polynomials are used to

remove the noise in the cepstrum domain. Depending on the environment polynomials are adopted. In this paper two Maximum Likelihood linear regression (MLLR) schemes are tested with respect to the adoption of data for different noise levels. This method most effective even for high mismatched cases.

Two stage algorithms are used for reducing the combining influence of additive noise and reverberation [3]. These can be done sequentially using deep neural networks and also incorporate the phase information during training. This proposed algorithm improves the speech intelligibility and also quality. Experimental results verified for different noise input SNR from -5dB to 5dB. The output achieved significantly better signal output SNR with average PESQ is 2.238.

HMM-GMM (Hidden Markov model-Gaussian mixture model) baseline for Kannada speech recognition based model. Kannada phonic recognition improves the recognition rate. Training speech corpus using HTK (Hidden Markova model tool kit), ANN are initialized with random weight and training is done using back propagation algorithm. DNN [4] has two stages one is unsupervised free training and another one is fine tanning stage, Kannada continuous speech corpus are recorded in real mode, lecture mode and conversation mode. It is independent of the phone modelling on three baselines and on different modes of corpus. Acoustic modelling based on HMM-DNN shows 7 to 8% improvement.

Deep neural network is used to enhance the signal acoustic model. Signal restoration based on multichannel beam forming with acoustic model using neural network [5]. This neural network learns the target speaker direction with different environment. To achieve good performance first layer of neural network separate the multichannel filtering operation from a single channel in frequency domain. Filtering coefficients are updated by considering the previous inputs. This paper nearly improves 5% compare to the existing system.

Each target speaker discriminated by considering Deep Belief Networks and Deep Neural Networks [6]. The speaker recognition, the target speaker enrolls one sample or multiple samples. The main principle of this paper is to end the Oracle Probabilistic Linear Discriminant Analysis (PLDA) scoring system with no proper label background

The proposed technique [7] suppressed the background without losing information contained in the speech. It uses both frequency blind signal suppression (FDBSS) and Posterior Wiener filter processing before applying the wiener filter, it

estimate the depressed background noise using FDBSS and then applied to Wiener filter. This is very complex and more computational time is required.

The paper [8] uses wavelet filter banks are used to decompose the speech into number of sub bands. The output of each sub band is given to the basic wiener filter to control the noise.

The paper [9] pools an amended Wiener filter through new model (Posterior Union mode) PUM. The paper proposed the filter out the broadband noise and PUM estimates merely fractional band noise. PUM is the post processing of speech enhancement. This adds complexity to the system.

The paper [10] uses noise reduction by using Wiener filter based speech de-noising algorithm to find errors in the target speech. It gives the quantitative performance behavior of the Wiener filter with respect to noise reduction.

The paper [11] model the speech signal processing as an Autoregressive using Kalman filter and it focuses on noise reduction that utilizes past and future form of estimation of the current state of the speech. This method uses recursive to minimize mean square error. Because of recursive it adds more complexity to the system

In this paper proposed method is divided into two stages first stage is the (preprocessing) restoration of the Kannada speech signal using modified wiener filter which is a simple and efficient way to restore the speech with different types of real world noisy environment. By calculating the significant parameters like variance, and of the clean and noise signals in the short term Fourier transform (STFT) domain are very much useful to restore the speech signal with different types of noisy environment. Finally required estimation of the speech signal output achieved with improved sound quality by varying input SNR (dB) using inverse STFT.

The second stage is, to recognize the isolated Kannada word based on the autocorrelation. This is very simple and efficient recognition system with less computation compared to the latest methods. The latest methods use different feature extracting techniques to extract significant features of speech signal as a database in speech recognition system and adaptive threshold setting is challenging to recognize the target speech in the database. The proposed method used isolated Kannada word speech signal and threshold setting is simple to recognize the target speech in the database with less recognition error rate.

2. Restoration of speech

Restoration of speech is one of the most important pre-processing stages in speech processing system. Maintaining the quality and intelligibility of speech in an everyday environs which contains the background interference in additive noise form significantly reduces the systems performance. Due to degradation of speech, the restoring the desired speech signal still poses a considerable challenge. Many researchers have been developed a different technique to restore speech from the degraded speech due to innumerable amount of complexities present in the speech.

2.1 Wiener filter

Wiener Filter plays an important role in speech restoration with assume that speech is stationary in process. The aim of this filter is to diminish average squared distance in the middle of filter output and an anticipated speech. However, the filter coefficients on behalf of each block of P number speech samples are calculated at the moment it adopt itself within the blocks to the average speech characteristics [12]. The input-output filter relation is given in Eq. (1).

$$\hat{x}(m) = \sum_{k=0}^{P-1} w_k y(m-k) \quad (1)$$

$$\hat{x}(m) = W^T Y \quad (2)$$

Eq. (2) contains Y is speech signal with P number of samples and W is the P number of filter coefficients.

The wiener filter objective stands to find Least-mean Square Error (LSE) among filter output and desired speech. The wiener filter coefficients acquired via minimizing average squared error function $E[e(m)]$ pertaining to filter coefficients vector The MSE is given in Eq. (3).

$$E[e^2(m)] = E[(x(m) - \hat{x}(m))^2] \quad (3)$$

The MMSE Wiener filter is obtained by

$$R_{yy}W = r_{yx} \quad (4)$$

Eq. (4) Where $R_{yy} = E[y(m)y^T(m)]$ is input signal autocorrelation matrix and r_{yx} is the input cross correlation matrix and $W = R_{yy}^{-1}r_{yx}$ filter output.

2.2 Wiener filter for additive noise reduction

Consider a noisy speech signal in Eq. (5) where $y(m)$ is the added version of clean speech $x(m)$ with noise $n(m)$.

$$y(m) = x(m) + n(m) \quad (5)$$

By assuming signal and noise are uncorrelated. The Eq. (6) is the sum of the autocorrelation matrix R_{xx} of the signal and the autocorrelation matrix R_{nn} of the noise is the autocorrelation matrix of noisy speech signal

$$R_{yy} = R_{xx} + R_{nn} \quad (6)$$

So, the Wiener filter coefficient matrix is defined in Eq. (7). And the input auto correlation matrix r_{xx} .

$$W = (R_{xx} + R_{nn})^{-1} r_{xx} \quad (7)$$

2.3 Modified Wiener filter

Consider a noisy speech signal $x(n)$ in Eq. (8) is

$$x(n) = s(n) + N(n) \quad (8)$$

Where $s(n)$ the clean speech is $N(n)$ is the noisy signal. The signal is divided into a number of frames by windowing technique. The windowing function is $W(n) = \sin\left(\frac{l}{\pi L}\right)$, $1 \leq l \leq L$ Where L is the length of the window function. The short term Fourier transform (STFT) of the clean speech signal $s(m)$, $N(m)$ is the noisy signal and $x(n)$ is noisy speech signal given in Eq. (9), (10), and (11), respectively.

$$S(k) = \sum_{m=1}^L s(m) w_m^{mk}, \quad \text{Where } 1 \leq k \leq L \quad (9)$$

$$N(k) = \sum_{m=1}^L n(m) w_m^{mk}, \quad \text{Where } 1 \leq k \leq L \quad (10)$$

$$X(k) = \sum_{m=1}^L x(m) w_m^{mk}, \quad \text{Where } 1 \leq k \leq L \quad (11)$$

Where $w_m = e^{\frac{-j2\pi}{L}}$

Then the Variance Vector of $s(m)$ and $N(m)$ calculated in STFT domain is $V_S = |S(k)|^2$, $V_N = |N(k)|^2$.

By setting some of the parameters like μ and λ vectors in STFT domain are

$$\mu = \frac{V_S}{(V_S + V_N)X}, \quad \lambda = \frac{1}{V_S} + \frac{1}{V_N}$$

Intermediate estimation of desired signal is calculated in Eq. (12) by taking the inverse STFT of μ .

$$s_e(m) = \frac{1}{L} \sum_{k=1}^L \mu(k) w_m^{-mk} \quad (12)$$

The output of Modified Wiener Filter is $\hat{x}(m)$ is obtained by setting another the parameter $\mathbf{b} = \mu \times \lambda$. The final estimation of the restored speech signal is reconstructed in Eq. (13).

$$\hat{x}(m) = iSTFT(\mathbf{b}) - (iSTFT(\lambda) * s_e(m)) \quad (13)$$

3. Speech recognition

Speech recognition is useful in multimedia applications. It allows to search and index of recorded audio and video data. The linguistic model to be needed to distinguish between different languages in the speech recognition system. Still forming of linguistic model is very much typical for researches to achieve less word error rate.

3.1 Autocorrelation

Autocorrelation is the correlation of a signal with same signal of delayed version [15]. The output of the autocorrelation is symmetric about the middle point $N/2$, where N is the length of the signal. Because of even symmetric nature in Eq. (15), the output co-efficient value is maximum at $N/2$ middle point of the signal. The discrete time auto correlation of speech signal and its time lag k of speech signal as in Eq.(14). The cross correlation of two different signal output will not give maximum value at $N/2$ middle point of the signal. The proposed Kannada speech recognition system uses both autocorrelation and cross correlation techniques to recognize the target speech signal in the database. The threshold setting carried by exploiting the even symmetric property of autocorrelation with its maximum value at $N/2$ middle point of the speech signal.

$$R_{yy}(k) = \sum_{n=1}^N y(n) \bar{y}(n-k) \quad (14)$$

where $y(n)$ is the speech signal of N number of samples and $\bar{y}(n-k)$ is the delayed version $y(n)$.

$$R_{yy}(n) = R_{yy}(-n) \quad (15)$$

4. Proposed model algorithm

Proposed method is divided into two parts first part of the system is restoration of Kannada speech and second part is isolated Kannada Word Recognition.

Restoration

- Initially, limit the Kannada speech signal dynamic range to avoid clipping of the signal output of the system. The sampling frequency is 8000 samples per second.
- The clean speech signal is added with different types of noise to by varying input SNR in dB.
- The degraded speech signal is alienated into number of frames, every single frame size length is of 1024 samples with overlapping is 512 samples.
- Compute absolute value of STFT coefficients for each frame. Then compute the variance, μ and λ of degraded speech and noise in STFT domain.
- Intermediate estimation of desired signal is calculated using Eq. (12) by taking the inverse STFT of μ .
- The final estimation of the restored speech is reconstructed using Eq.(13)
- To restore the speech signal by taking inverse STFT of each frame.
- Finally combine all frames to sound the speech signal.
- Calculate output SNR for different input SNR and the output of restored signal is given to the recognition system.

Recognition

- Kannada data base is created nearly for 500 different speech signals and also similar speech signal of male and female of different accents.
- Take the test speech signal, which is to be recognized.

- Calculate the autocorrelation between the test speech signals with other sound signal in the data base.
- Set the threshold value by calculating the autocorrelation using Eq (14) between the test speech signals with itself and find the maximum value from autocorrelation output signal.
- This maximum value of autocorrelation output signal is used as a threshold value to recognize the Kannada word.
- To recognize the target signal present in the data base by comparison between the maximum values of all cross correlation of speech signals with threshold value.
- Finally if maximum values of all cross correlation speech signals value greater than or equal to threshold value will be treated as recognized test speech signal otherwise not recognized.

5. Sound quality measures

In general, using mathematical formula via clean and degraded speech the objective speech quality measures are calculated. Some objective and subjective quality measures [13] are relating towards common percipience whereas others are more similar pertaining to recorded speech quality of every word being understandable. The Perceptual Evaluation of Speech Quality (PESQ) [14,15] Log Likelihood Ratio (LLR), Weight Spectral Scope (WSS), Signal Distortion (SD), Background Distortion (BD), Overall Quality (OQ).

5.1 Signal to noise ratio (SNR)

It is extensively utilized objective quality measure using the following mathematical equation.

$$(SNR)_{dB} = 10 \times \log_{10} \left(\frac{\sum_{n=1}^N x^2(n)}{\sum_{n=1}^N (x(n) - \bar{x}(n))^2} \right) \quad (16)$$

5.2 Weighted spectral slope

It is a spectral distance measure based on spectra comparison of clean and noisy speech samples and Eq. (17) shows one of the measures of WSS.

$$d_{WSS} = \frac{1}{M} \sum_{m=0}^{M-1} \left(\frac{\sum_{j=1}^K W(j,m) (S_c(j,m) - S_d(j,m))^2}{\sum_{j=1}^K W(j,m)} \right) \quad (17)$$

The Speech Production Model demonstrated via calculating linear coefficients. The clean speech LPC and distorted LPC can be calculated. The LLR measures the distance among clean and noisy speech LPC vector and the measures calculated as follows:

$$d_{LLR}(a_d, a_c) = \log \left(\frac{a_d R_c a_d^T}{a_c R_c a_c^T} \right) \quad (18)$$

6. Results and analysis

The algorithm and experimental results are simulated using Mat Lab R2013a tool. Experimental observations are made by taking 500 isolated Kannada words of different pitch frequencies of different accent with different environmental noise as an input to the preprocessing system. By observing the spectrogram of the clean, noise and addition of clean with noise signals indicates the presence of noise in different colors.

The first stage of preprocessing input and output wave forms are generated and experimentally verified with large vocabulary of input Kannada speech signal of different words with different accents.

The Fig. 1 shows that the clean speech of male and its spectrogram with -5dB normally distributed pseudo random noise is added. The experimental results show that for a -5 dB noise, the input speech completely degraded almost zero PESQ. The output enhanced speech signals and its spectrogram with better PESQ.

For varying input (SNR) from -20dB to 5dB with different types of noise for male and female shown in Table 1. The output Enhanced Speech SNR (dB) is increasing with varying input (SNR) and gives better output signals.

Table 2 and 3 show the standard speech quality measures the output speech for normally distributed Pseudo random noise with varying input (SNR) in dB for male and female speech signals and output measures are LLR, WSS, PESQ, SD, BD and OQ. In this speech quality measures are better for varying input SNR.

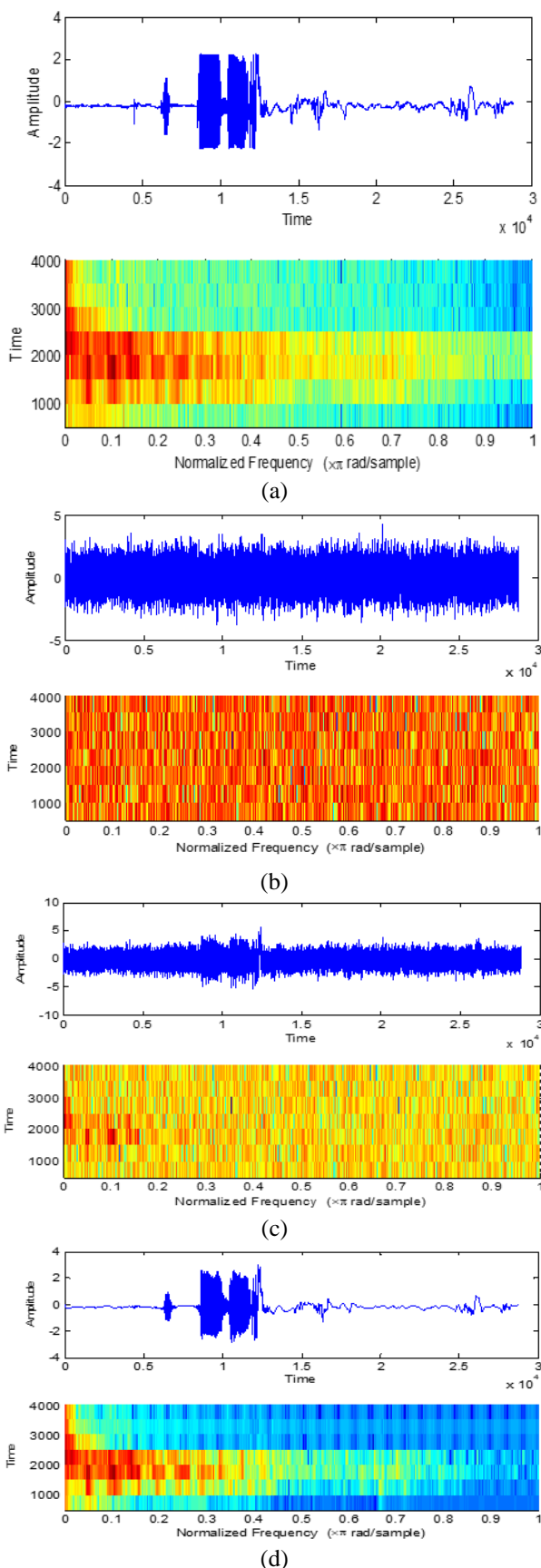


Figure.1 Kannada speech restoration input and output speech signals: (a) Kannada speech clean speech, (b) -5dB Normally distributed pseudo random noise, (c) clean + noise, and (d) enhanced speech signals.

Table 1. Input/output SNR for different types of noise

| Noise type | Input SNR (dB) | Enhanced Speech SNR (dB) | |
|------------------|----------------|--------------------------|-------|
| | | Female | Male |
| Street Noise | -20 | 5.02 | 6.12 |
| | -15 | 6.51 | 7.20 |
| | -10 | 8.40 | 8.67 |
| | -5 | 10.74 | 10.64 |
| | 0 | 13.54 | 13.05 |
| | 5 | 16.56 | 15.72 |
| Bubble Noise | -20 | 2.99 | 2.94 |
| | -15 | 4.66 | 4.41 |
| | -10 | 6.74 | 6.36 |
| | -5 | 9.18 | 8.85 |
| | 0 | 12.02 | 11.69 |
| | 5 | 15.22 | 14.83 |
| Exhibition Noise | -20 | 5.79 | 7.10 |
| | -15 | 7.72 | 8.48 |
| | -10 | 9.92 | 10.18 |
| | -5 | 12.27 | 12.24 |
| | 0 | 14.83 | 14.67 |
| | 5 | 17.62 | 15.32 |
| Airport Noise | -20 | 6.23 | 5.92 |
| | -15 | 8.41 | 8.15 |
| | -10 | 10.83 | 9.69 |
| | -5 | 13.48 | 11.02 |
| | 0 | 16.42 | 13.03 |
| | 5 | 19.17 | 14.91 |

Table 2. The standard speech quality measures of SD, OQ and PESQ versus Input (SNR)

| Speech signal | Input SNR (dB) | SD | OQ | PESQ |
|---------------|----------------|-------|-------|-------|
| Female | -5 | 4.596 | 3.195 | 2.986 |
| | 0 | 4.338 | 3.786 | 3.199 |
| | 5 | 3.410 | 3.945 | 3.331 |
| Male | -5 | 4.272 | 2.939 | 2.724 |
| | 0 | 3.837 | 3.342 | 2.860 |
| | 5 | 3.215 | 3.607 | 2.970 |

Table 3. The standard speech quality measures of LLR, WSS and PESQ versus Input (SNR)

| Speech signal | Input SNR | LLR | WSS | PESQ |
|---------------|-----------|-------|--------|-------|
| Female | -5 | 0.339 | 10.383 | 2.986 |
| | 0 | 0.513 | 17.201 | 3.199 |
| | 5 | 1.195 | 27.175 | 3.331 |
| Male | -5 | 0.403 | 18.298 | 2.724 |
| | 0 | 0.721 | 26.483 | 2.860 |
| | 5 | 1.161 | 36.077 | 2.970 |

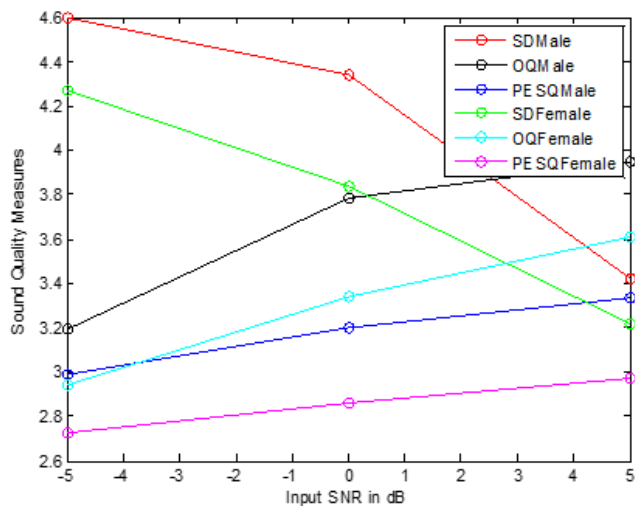


Figure.2 The graph of SD, PESQ and OQ verses Input (SNR) dB for male and female

Table.4 Comparison table for different filters

| Input SNR dB | Wiener Filter Output SNR dB | Improved Wiener Filter Output SNR dB | Modified Wiener Filter Output SNR dB |
|--------------|-----------------------------|--------------------------------------|--------------------------------------|
| -10 | 1.91 | 0.17 | 10.69 |
| -5 | 2.67 | 4.11 | 13.47 |
| 0 | 7.08 | 8.48 | 16.27 |
| 5 | 11.05 | 13.24 | 19.23 |
| 10 | 15.53 | 17.28 | 21.73 |

The Fig.2, shows that the SD, PESQ and OQ Verses input (SNR) dB for female and male respectively. The graphs shows that SD is decreasing with increasing input (SNR)dB for both male and female and PESQ, OQ, LLR,WSS increasing with increasing input (SNR)dB for both male and female. Here the overall speech quality increasing.

Table 4 shows the comparison for Wiener filter [9]. The proposed modified wiener filter gives the good performance compare to basic Wiener filter and Improved Wiener filter. The PESQ of speech signal is effectively improved.

Table 4 shows normally distributed pseudo random noise of varying input SNR in dB with clean speech and the corresponding output SNR in dB of restored speech. The proposed Modified Wiener Filter gives better sound quality than the other methods because this method estimates the statistical parameters in STFT domain. In STFT domain, the lower frequencies of speech signal effects with less noise than the higher frequencies. So proposed method is more effective.

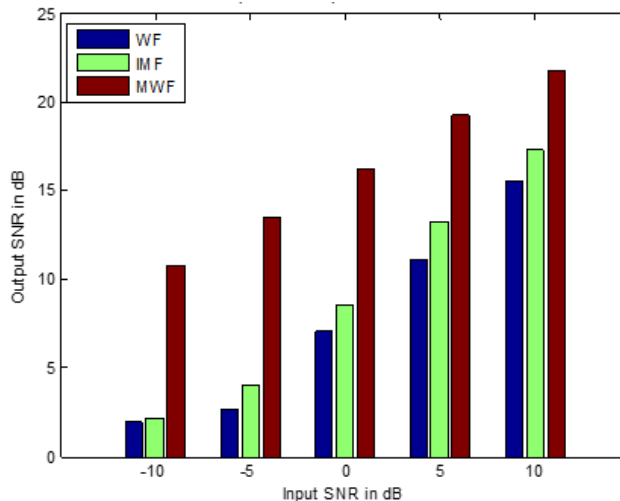


Figure. 3 The graph of Wiener Filter (WF), Improved Wiener Filter (IMF), Modified Wiener Filter (MWF) Verses Input (SNR) in dB

Table 5. Recognition status with nine different words and/or digits of male and female speech signals. Recognition status ‘0’ represents not recognized words and / or digits and ‘1’ represents recognized words and/or digits in the Isolated Kannada speech signals database. Recognition tested nearly 500 words with 10 digits from 0-9 different accents. Tabulated only for nine and /or digits of male and female with different accents. Experimentally proposed recognition system status as follows:

Case1: if the data base contains the target speech signals, 100% recognition.

Case2: if the data base does not contains the target speech signals, 100% not recognition.

In Table 6, Recognition status of Kannada words are used experimentally with starting same letter of different words of different sounds are tabulated. These different kanndda words are generated using UNICODE for Kannada letters.

For example the Kannada word ಅವಳು (Avalu), the Unicode of this word is 0C85 (ಅ), 0CB5 (ವ), 0CB3 (ಳ), 0CC1 (ು). The word ಅವಳು is created by appending Unicode’s 0C850CB50CB30CC1. Recognition status ‘0’ represents not recognized words and ‘1’ represents recognized words.

Table 5. Recognition status with nine different words and/or digits of male and female speech signals

| Threshold | Female/Male Speech Signals of maximum correlated value with target speech signal | | | | | | | | |
|------------------|--|--------|--------|--------|--------|--------|--------|--------|--------|
| Test Signal | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Female 343.88 | 230.68 | 193.93 | 343.88 | 197.25 | 218.17 | 221.57 | 192.73 | 219.66 | 343.88 |
| Recognition | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| Female 410.22 | 206.16 | 190.17 | 410.22 | 410.22 | 136.48 | 219.28 | 193.70 | 410.22 | 410.22 |
| Recognition | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 |
| Male 489.54 | 136.81 | 129.60 | 489.54 | 137.24 | 489.54 | 285.41 | 223.77 | 198.44 | 489.54 |
| Recognition | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| Male 232.90 | 200.34 | 191.03 | 122.15 | 205.31 | 122.15 | 212.35 | 191.72 | 204.78 | 122.15 |
| Recognition | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 6. Recognition status of different Kannada words

| Test kannada word | kannada word database | | | | | | | |
|-------------------------|-----------------------|------------------|-----------------|----------------|------------------|-------------------------|------------------|------------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| ಅವಳು (Avalu) | ಅರಸ Arasa | ಅವನು Avanu | ಅವಳು Avalu | ಅವರು Avaru | ಅರಮನೆ Aramane | ಅವತಾರ Avathara | ಅಮರ Amara | ಅನುಪಮ Anupama |
| Recognition | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| ಕಮಲ (Kamala) | ಕರು Karu | ಕರುಳು Karulu | ಕಮಲ Kamala | ಕಮಲ Kamala | ಕವನ Kavana | ಕದನ Kadana | ಕಟಕ Kataka | ಕಮಲ Kamala |
| Recognition | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| ಟಂಕಣ (Tankana) | ಟಗರು Tagaru | ಟಂಕ Tanka | ಟಂಕಣ Tankana | ಟವಣೆ Tavane | ಟಂಕಣ Tankana | ಟಕಾಯಿ ಸು Takayisu | ಟಪಾಕಿ Tapaki | ಟಮಟೆ Tamate |
| Recognition | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| ವಿನುಜ (Vinuja) | ವನಜ Vanaja | ವಿನುತ Vinutha | ವಿನಯ Vinaya | ವರಣ Varana | ವರಿಸು Varisu | ವರಹ Varaha | ವಾರಿಜ Vaarija | ವಿಜಯ Vijaya |
| Recognition | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

7. Conclusion

The experimental outcomes shows, the algorithms used in this paper are very simple and more effective in the different scenario. To restore Kannada speech signal using modified Wiener filter results better restoration with good speech quality for -5dB input SNR. Output SNR and PESQ are significant improved for different types of noise. Isolated Kannada words are used experimentally with starting same letter of different words of different sounds of male, female. The proposed method gives 100% recognition rate for male, female, with different accents.

The future work for restoration with maintaining good quality is required very simple algorithms with less complexity is a still challenging for researchers and significantly speech recognition

with less complexity and more recognition rate for continuous speech recognition to be challenging to significantly improve recognition rate for kannada speech signal with different accents.

References

- [1] Y. Xu, J. Du, L.R. Dai, and C.H. Lee, "A Regression Approach to Speech Enhancement Based on Deep Neural Networks", *IEEE/ACM Transactions On Audio, Speech And Language Processing*, Vol. 23, No. 1, pp.7-19, 2015.
- [2] X. Cui and A. Alwan, "Noise Robust Speech Recognition Using Feature Compensation Based on Polynomial Regression of Utterance SNR", *IEEE Transactions on Speech and Audio Processing*, Vol. 15, No. 6, pp. 1161-1172, 2005.

- [3] Y. Zhao, Z.Q. Wang, and D. Wang, "A Two-Stage Algorithm for Noisy and Reverberant Speech Enhancement", In: *Proc. of the International Conference on Acoustics, Speech and Signal Processing*, pp.5580-5584, 2017.
- [4] R. Pradeep and K.S. Rao, "Deep Neural Networks for Kannada Phoneme Recognition", In: *Proc. of the Ninth International Conference on Contemporary Computing*, pp.1-6, 2016.
- [5] T.N. Sainath and R.J. Weiss, K.W. Wilson, B. Li, A. Narayanan, E. Variiani, M. Bacchiani, I. Shafran, A. Senior, K. Chin, A. Misra, and C. Kim, "Multichannel Signal Processing With Deep Neural Networks for Automatic Speech Recognition", *IEEE/ACM Transactions on Audio, Speech, And Language Processing*, Vol. 25, No. 5, pp. 965-979, 2017.
- [6] O. Ghahabi and J. Hernando, "Deep Learning Backend For Single And Multisession i-Vector Speaker Recognition", *IEEE/ACM Transactions on Audio, Speech, And Language Processing*, Vol. 25, No. 4, pp. 807-817, 2017.
- [7] J. Even, H. Saruwatari and K. Shikano, "Enhanced Wiener Post-Processing Based On Partial Projection Back Of the Blind Signal Separation Noise Estimate", In: *Proc. of the 17th European Signal Processing Conference*, pp. 807-817, 2009.
- [8] A. Biswas, P.K. Sahu, A. Bhowmick, and M. Chandra, "Acoustic Feature Extraction using ERB Like Wavelet Sub-band Perceptual Wiener Filtering for Noisy Speech", In: *Proc. of the Annual IEEE India Conference*, pp.1-6, 2014.
- [9] L. Xuanmin, G. Yue, X. Peng, and D. Chao, "An anti- noise speech recognition model based on improved Wiener filter and PUM", In: *Proc. of the Sixth International Conf. on Instrumentation and Measurement, Computer, Communication and Control*, 2016.
- [10] S. Shekar and D.J. Ravi, "Denoising of a Speech Signal using Wiener Filter", In: *Proc. of the International Conf. on Current Trends in Engineering, Science and Technology*, 2017.
- [11] A.P. Nair and S. Krishnan, "MFCC Based Noise Reduction in ASR Using Kalman Filtering", In: *Proc. of the International Conference on Advances in Signal Processing*, 2016.
- [12] S.V. Vaseghi, "Advanced Digital Signal Processing and Noise Reduction", *Second Edition. John Wiley & Sons*. 2000.
- [13] Y. Hu and P.C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement", *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 16, No.1, pp. 229-238, 2008.
- [14] ITU-T, "Perceptual Evaluation of Quality (PESQ), An Objective Method for End-to End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs", *International Telecommunication Union*, p.862, 2001.
- [15] L.R. Rabiner and B.H. Juang, "Fundamentals of Speech Recognition", *1st Edition, Prentice Hall*, Upper Saddle River, 1993.