

Multi-Level Denoising Filter Bank For Speech Signal

J. MUBEENA PARVEEN

Dept. Of Electronics & Communication Engineering,
SSN College Of Engineering,
Chennai, Tamil Nadu, India.
Mubeena227@Gmail.Com

R. RAJAVEL

Dept. Of Electronics & Communication Engineering,
SSN College Of Engineering,
Chennai, Tamil Nadu, INDIA.
Rajavelr@Ssn.Edu.In

Abstract— Denoising technique helps in speech enhancement. Denoising methods based on Spectral Subtraction, Minimum Mean Square Error produces musical noise distortion. This paper, proposes a new technique for musical noise reduction in speech signal. The noisy speech signal is first decomposed in to non-uniform sub bands using a gammatone perceptual filter model with Equivalent Rectangular Bandwidth Scale (ERB). The sub bands signals are divided into frames. Then, noise estimation is carried out over each frame; if noise exists in the particular frame then the denoising algorithm is applied. In this work two denoising algorithms are used they are Minimum Mean-Square Error Short-Time Spectral Amplitude (MMSE STSA) and Boll's Spectral Subtraction (Boll's SS) technique. Finally enhanced speech signal is obtained. The effectiveness of quality enhancement was measured using Perceptual Evaluation of speech Quality (PESQ) technique. Both MMSE STSA and Boll's spectral subtraction algorithms provide better results than the existing Spectral Subtraction and MMSE technique. However, quality enhancement in MMSE STSA is much better than Boll's SS.

Keywords— Gamma tone filter bank, Perceptual filter bank, Boll Spectral Subtraction (Boll SS), MMSE STSA, Denoising.

signal by subtracting the noisy spectral components from the original signal. However it introduces musical noise. Hence, designing speech enhancement algorithm without introducing any perceptible speech distortion is the main challenge. In this paper, the speech signal along with noise is processed through gammatone filter bank. The output of the gammatone filter bank is divided in to frames and windowed and converted in to frequency domain using FFT. Then continuous noise estimation is carried out over each frame and the required denoising method is applied. Finally IFFT and overlap add technique is carried out and the enhanced speech signal is obtained. In MMSE STSA, a priori SNR tracking obtained using a decision-directed method is more efficient than the previous MMSE method [4, 5]. Boll's spectral subtraction technique uses spectral averaging and residual noise reduction method which is more efficient than the previous method [6]. Both the methods, reduces the musical noise to a greater extent than the previous methods thereby improving the quality of the speech signal.

I. Introduction

In Speech Communication systems, when the speech signal is transmitted through the channel, noise gets added up along with the speech signal and it is very difficult to obtain completely noise free speech signal at the receiver. Hence denoising techniques such as Minimum Mean-Square Error Short-Time Spectral Amplitude (MMSE STSA) and Boll's Spectral Subtraction (Boll's SS) technique are required to improve the quality of the speech signal.

Filter bank employs decomposition of the signal in to its low frequency and high frequency components in the analysis block and then reconstructing it at the synthesis block [1]. However, a slight modification is required in developed systems to suppress undesired noise. An improved spectral subtraction method for reducing acoustic noise is obtained by using weighting filter based on psychoacoustic property to reduce the residual musical noise [2]. There are several speech enhancement algorithms, among them spectral subtraction algorithms is the oldest and most popular techniques due to its simplicity in implementation [3]. It improves the quality of the

II. Proposed method

A. Perceptual filter bank model for speech enhancement

Although most of the filtering techniques like wiener filtering filters the noise but it does not completely eliminate the musical noise. By using speech enhancement algorithm based on perceptual model residual noise can be eliminated to some extent. In this technique the degraded input waveform is decomposed in to non-uniform decomposition using gammatone filter bank [7]. Using this technique the lower frequency component can be efficiently obtained during the synthesis process. In this paper, the degraded input signal is decomposed using Gamma tone filter bank [8].

$$gt(t) = A t^{n-1} \exp(-2\pi b B c) \cos(2\pi f_c t + \phi) \quad (1)$$

Where A is the magnitude normalization parameter, n is the filter order, f_c is the centre frequency of filters, B is filters bandwidths, and $bB(f_c)$ represents the filter envelop. The frequency resolution of human hearing with broadband signals is expressed using Equivalent Rectangular Bandwidth (ERB) scale. The expression used to convert a frequency f in Hz in its value in ERB is

J. Mubeena Parveen, R. Rajavel

$$\text{ERB}(f) = 21.4 \log(0.00437f_c + 1) \quad (2)$$

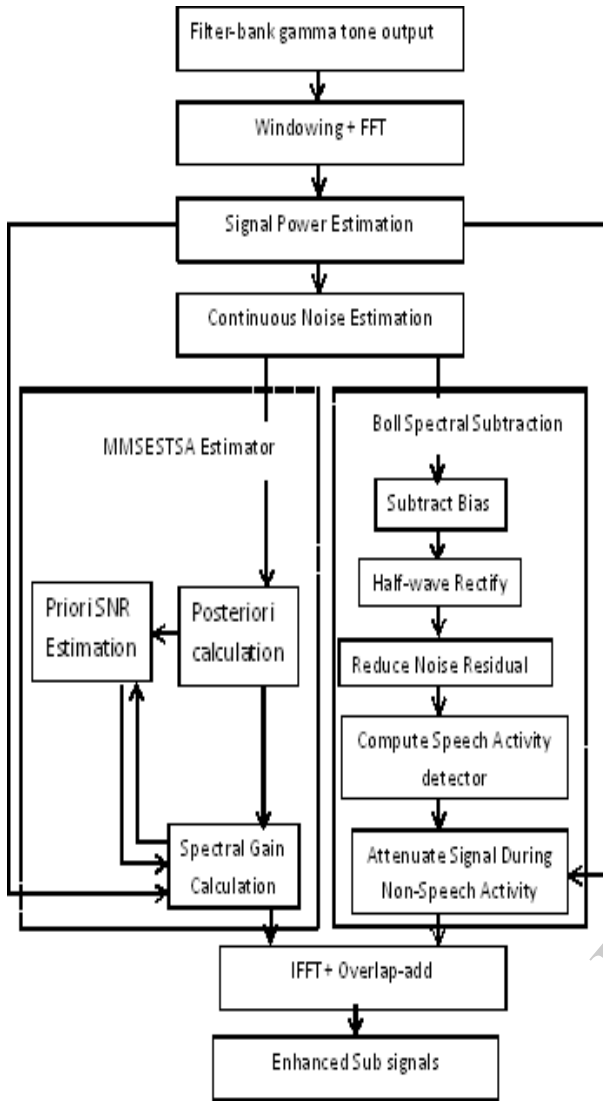


Fig.1. Denoising techniques using Method 1 – MMSE STSA, Method 2- Boll Spectral Subtraction

B. Multi Bands Perceptual Filter Bank

The processing is done by dividing the degraded waveform $y(n)$, into separate bands $Y_{k,gt}(n)$ based on gamma tone filter bank [7]. Using gammatone filter bank the lower frequency components can be efficiently obtained as the lower frequency components are divided in to more number of sub bands than the high frequency components. These bands are individually estimated using the required speech enhancement algorithm based on Boll's spectral subtraction or MMSE STSA technique. The analysis filter bank is made of 27- 4th order gamma tone filter [8]. The output of the k^h filter of the analysis gamma tone filter bank can be calculated using the

relation;

$$Y_{k,gt}(n) = y(n) * g_{t,k}(n) \quad (3)$$

The output of the gamma tone filter bank is divided in to frames and windowed using Hamming window and it is converted in to frequency domain using the FFT technique [5] and then the required speech enhancement algorithm is applied.

C. Perceptual generalized spectral subtraction techniques using Boll's method

The spectral subtraction estimator $\hat{S}(e^{j\omega})$, is calculated using the relation [6],

$$\hat{S}(e^{j\omega}) = \left[|X(e^{j\omega})| - \mu(e^{j\omega}) \right] e^{j\theta \times (e^{j\omega})} \quad (4)$$

Or

$$\hat{S}(e^{j\omega}) = H(e^{j\omega}) X(e^{j\omega}) \quad (5)$$

With

$$H(e^{j\omega}) = 1 - \frac{\mu(e^{j\omega})}{|X(e^{j\omega})|} \quad (6)$$

$$\mu(e^{j\omega}) = E \left\{ |N(e^{j\omega})| \right\} \quad (7)$$

Where $H(e^{j\omega})$ is the spectral subtraction filter, $\mu(e^{j\omega})$ is the average value taken during nonspeech activity,

The spectral error $\varepsilon(e^{j\omega})$, can be calculated as follows;

$$\varepsilon(e^{j\omega}) = \hat{S}(e^{j\omega}) - S(e^{j\omega}) = N(e^{j\omega}) - \mu(e^{j\omega}) e^{j\theta \times} \quad (8)$$

The spectral error can be reduced by magnitude averaging, half-wave rectification, residual noise reduction and by additional signal attenuation during nonspeech activity.

Using half-wave rectification the estimator becomes

$$\hat{S}(e^{j\omega}) = H_R(e^{j\omega}) X(e^{j\omega}) \quad (9)$$

Where

$$H_R(e^{j\omega}) = \frac{H(e^{j\omega}) + |H(e^{j\omega})|}{2} \quad (10)$$

The residual noise reduction scheme is given as,

$$|\hat{S}_i(e^{j\omega})| = |\hat{S}_i(e^{j\omega})|, \text{ for } |\hat{S}_i(e^{j\omega})| \geq \max |N_R(e^{j\omega})| \quad (11)$$

$$|\hat{S}_i(e^{j\omega})| = \min \left\{ |\hat{S}_j(e^{j\omega})|, j = i-1, i, i+1 \right\} \text{ for } |\hat{S}_i(e^{j\omega})| < \max |N_R(e^{j\omega})| \quad (12)$$

Where

$$\hat{S}_i(e^{jw}) = H_R(e^{jw}) X_i(e^{jw}) \quad (13)$$

And

$\max |N_R(e^{jw})| = \text{maximum value of noise residual}$
measured during nonspeech activity.
The absence of speech can be calculated by using,

$$T = 20 \log_{10} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{\hat{S}(e^{jw})}{\mu(e^{jw})} \right| dw \right] \quad (14)$$

For the value of T less than -12 dB, the frame is termed as no speech activity frame. The nonspeech activity including the output attenuation for the output spectral estimate is given as

$$\hat{S}(e^{jw}) = \begin{cases} \hat{S}(e^{jw}) & T \geq -12\text{dB} \\ cX(e^{jw}) & T \leq -12\text{dB} \end{cases} \quad (15)$$

Where $20 \log_{10} c = -30\text{dB}$

D. Minimum mean square error short time spectral amplitude method

The posteriori SNR estimate $\hat{\gamma}(f, \tau)$, is given by [4],

$$\hat{\gamma}(f, \tau) = \frac{|y_{DS}(f, \tau)|^2}{\lambda(f, \tau)} \quad (16)$$

Where $\lambda(f, \tau)$ is the power spectrum of the estimated noise given by,

$$\lambda(f, \tau) = \begin{cases} E\{z(f, \tau)^2\}^{\tau_{th}} & (\tau \leq \tau_{th}), \\ E\{z(f, \tau)^2\}^{\tau} & (\tau > \tau_{th}). \end{cases} \quad (17)$$

Where τ_{th} is a smoothing parameter for frame window and $E\{\cdot\}_a^B$ denotes an expectation operator from A to B.

The Priori SNR estimate is calculated as follows,

$$\hat{\varepsilon}(f, \tau) = \alpha \hat{\gamma}(f, \tau - 1) G^2(f, \tau - 1) + (1 - \alpha) P[\hat{\gamma}(f, \tau) - 1] \quad (0 \leq \alpha < 1), \quad (18)$$

α is the weighting factor of decision directed estimation, $G(f, \tau)$ is the spectral gain function and the operator $P[\cdot]$ is estimated as follows,

$$P[1] = \begin{cases} 1 & (1 \geq 0), \\ 0 & (\text{otherwise}). \end{cases} \quad (19)$$

$$G(f, \tau) = \Gamma(1.5) \frac{\sqrt{v(f, \tau)}}{\hat{\gamma}(f, \tau)} \exp\left(-\frac{v(f, \tau)}{2}\right) \cdot \left[(1 + v(f, \tau)) I_0\left(\frac{v(f, \tau)}{2}\right) + v(f, \tau) I_1\left(\frac{v(f, \tau)}{2}\right) \right] \quad (20)$$

$\Gamma(\cdot)$ is the gamma function and $I_0(\cdot)$ and $I_1(\cdot)$ denotes the modified Bessel functions of the zero and the first-order, respectively.

$v(f, \tau)$ is given by,

$$v(f, \tau) = \frac{\hat{\varepsilon}(f, \tau)}{1 + \hat{\varepsilon}(f, \tau)} \hat{\gamma}(f, \tau) \quad (21)$$

The noise reduction is estimated using,

$$y_{PROP}(f, \tau) = G(f, \tau) y_{DS}(f, \tau) \quad (22)$$

III. simulation

A. MATLAB Simulation

Simulation is done using MATLAB. It is used due to its high performance in technical computing.

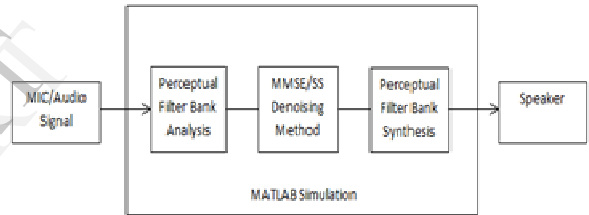


Fig. 2. Block Diagram of Denoising Algorithm

It allows computation and programming in a user friendly environment. In Perceptual Filter bank Analysis, the noisy speech signal is decomposed in to non-uniform sub bands and the required denoising method is used and then the enhanced speech signal is synthesized using Perceptual Filter bank Synthesis.

B. Speech quality and measure

Perceptual Evaluation method is an objective measure used for speech quality measurement. This method is used in order to evaluate the quality of the speech signal. This method captures the received audio stream (degraded input) and compares it with the original signal and simulated using MATLAB thus predicting the quality of the signal [9]. PESQ range corresponding to 0.5 indicates poor quality of signal and the values reaching towards 4.5 indicates efficient quality of signal.

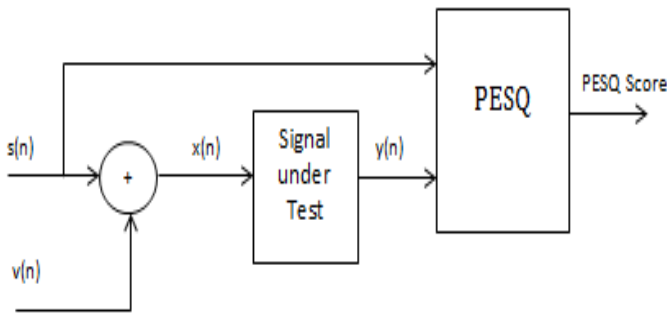


Fig. 3. Block Diagram for Perceptual Evaluation of Speech Quality

IV. RESULTS AND DISCUSSION

Method 1: Speech denoising algorithm based on Minimum Mean Square Error Short-Time Spectral Amplitude method with gamma tone filter bank reduces musical noise more efficiently than the previous MMSE method using gammatone filter bank. Using this method the PESQ value 2.38 is obtained which is better than the previous MMSE method which provides the PESQ value 2.25.

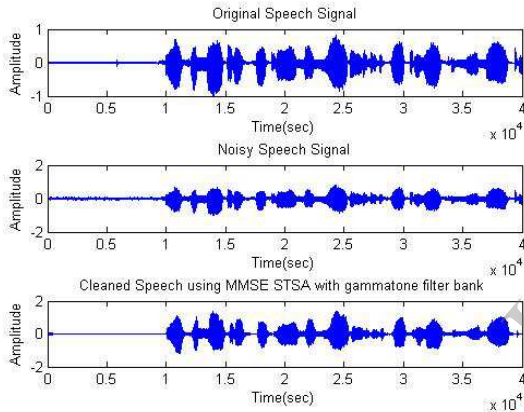


Fig. 4. MMSE STSA with ERB Perceptual filter bank (Gammatone)

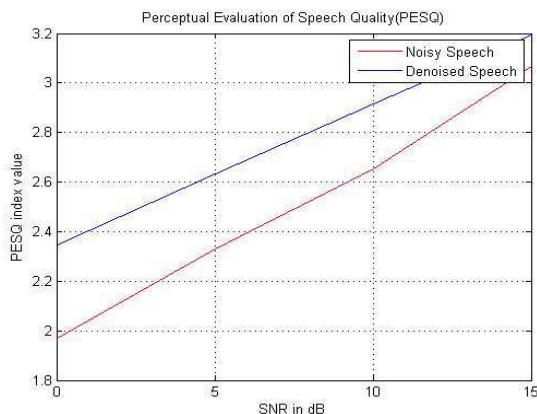


Fig .5. Perceptual Evaluation of speech Quality (PESQ) for MMSE STSA method

Method 2: Speech denoising algorithm based on Boll Spectral Subtraction method with gamma tone filter bank provides better result than the previous spectral subtraction method. This method provides the PESQ value 2.3 which is better than the PESQ value 2.2 of the previous Spectral Subtraction method.

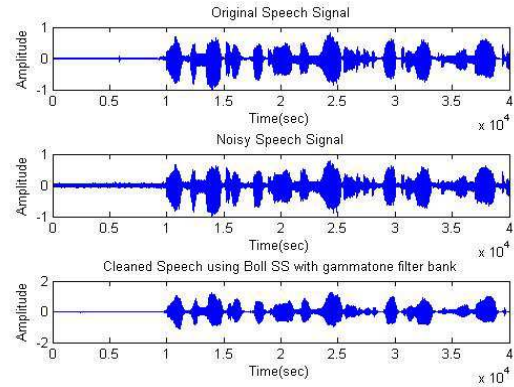


Fig. 6. Boll's Spectral Subtraction with ERB Perceptual filter bank (Gammatone)

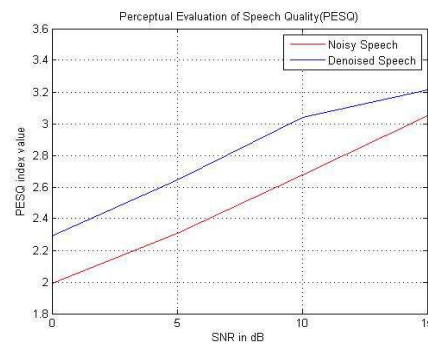


Fig .7. Perceptual Evaluation of speech Quality (PESQ) for Boll's Spectral Subtraction method

V. Conclusion

This paper proposes a new technique for musical noise reduction in speech signal based on MMSE STSA with gammatone filter bank and Boll's Spectral Subtraction with gammatone filter bank. The noisy speech is decomposed in to nonuniform sub bands using gammatone filter banks that are manipulated in each nonlinear block with the Boll's spectral subtraction process and the MMSE STSA technique and simulated in MATLAB. PESQ was calculated for each combination and Experimental results obtained were found to be better than the previous existing method. It was observed from the above methods that noisy signal was efficiently denoised and original speech signal reconstructed by MMSE STSA the gammatone perceptual filter bank model was found to provide better results than Boll's Spectral Subtraction in terms of better PESQ.

J. Mubeena Parveen, R. Rajavel

References

- [1] Muhammad Amir Shafiq and Saqib Ejaz, "Real Time Implementation Of Multi-Level Perfect Signal Reconstruction Filter Bank," *International Journal of Engineering & Technology*, Vol. 10, No. 04, pp.40-47, 2010.
- [2] R. M. Udrea, N. D. Vizireanu, and S. Ciochina, "An improved spectral subtraction method for speech enhancement using a perceptual weighting filter," *Digital Signal Processing*, vol. 18, no. 4, pp. 581–587, 2008.
- [3] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pp. 208–211, April 1979.
- [4] Ryoi Okamoto, Yu Takahashi, Hiroshi Saruwatari, Kiyohiro Shikano, "MMSE STSA Estimator With Nonstationary Noise Estimation Based On ICA For High-Quality Speech Enhancement," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ')*, pp. 4778-4781, 2010.
- [5] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error-log-spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 443–445, 1985.
- [6] Steven F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Transactions On Acoustics, Speech And Signal Processing*, Vol. Assp-27, No. 2, pp. 113-120, April 1979.
- [7] Novlene Zoghalmi and Zied Lachiri, "Application of Perceptual Filtering Models to Noisy Speech Signals Enhancement," *Hindawi Publishing Corporation Journal of Electrical and Computer Engineering*, pp.1-12, 2012.
- [8] V. Hohmann, "Frequency analysis and synthesis using a Gammatone filterbank," *Acta Acustica United with Acustica*, vol. 88, no. 3, pp. 433–442, 2002.
- [9] S. S.V Sumanth kotta and B.K . kommineni, "Acoustic beamforming for hearing using multi micro phone array by designing graphical user interface", *Blekinge Institute of technology, Sweden*, pp.1-72, 2012.

IJERT