International Journal of Advanced Trends in Computer Science and Engineering

Speech Enhancement Using Novel Time-Frequency Analysis Techniques: A Survey on Comparison.

Jagadish S.Jakati[1], Shridhar S.Kuntoji[2].

[1]Assistant Professor, Department of Electronics & Communication Engg, S. G. Balekundri Institute of Technology, Belagavi, VTU Research Scholar Karanataka (State), India
[2] Professor, Department of Electronics & Communication Engg, Basaveshwar Engineering College Bagalkot, VTU Research Supervisor, Karanataka (State), INDIA
[1]jagadishjs30@gmail.com [2]shridhar.ece@gmail.com

ABSTRACT

Speech is an important essential function of human beings. Humans strive to recover data from the external world or to communicate with everyone. For many reasons, speech protrudes as the resourceful and suitable. Speech signal constitutes both vowels and consonants and consonants are more susceptible towards the noise than vowels as they have less energy. Hence when speech signal is communicated it is necessary to reduce the noise and this process is known as speech enhancement. In noisy environment listening task is very difficult at the destination, so reduction of noise is important research field in speech enhancement. In speech enhancement various filter banks are used to reduce the variety of noises. The proposed methods give better quality of speech and a minimized musical noise and background noises. Auditory filter bank, QMF banks, Warped Low Delay Filter-Banks (WLDFB), Unimodular Filter Banks (UMFB), Cascaded filter Banks, DFT modulated Filter Banks (DFTMFB) are the filter bank algorithms which very effective in noise reduction even in low delay and low SNR situations.

Key words: Speech enhancement, Filter Banks, QMF, FFT, DFT, ERB, WLDFB, SNR, PESQ score, ISD and MOS

1. INTRODUCTION

The quality of speech signal can be improved using speech enhancement technique. Enhancement of Speech signal is a technique which deals in processing of the noisy speech signals and aims in improving their intelligibility and quality or their proper decoding by machines. It is required that the noise signal is introduced in clean speech signal is to be minimized. This noise reduction technique finds applications in multidisciplinary areas such as, Communication Systems and Speech recognition Hearing aids [5] etc. The objectives of speech enhancement are: -
(a)To eliminate the noise associated with clean speech.

(b) To minimize the musical noise resulting due to transform domain processing.
(c) To increase the intelligibility and quality of speech signal.

The essential purpose of enhancement of speech is to minimize the noise in original speech signal and enhance the one or more intuitive facets of speech signal such as quality and/ or comprehensible. Therefore, increase speech intelligibility, the speech enhancement techniques are broadly classified into, single and multi channel enhancement techniques. In speech enhancement by single channel technique both clean speech and noises are present together. In multi channel speech enhancement technique both noise signal and clean speech are present separately [9].

The motivating factor for using filter banks in speech enhancement is, it is possible to control magnitude distortion and phase distortion resulting due to transform domain processing. These are not possible from conventional methods of speech enhancement [6].

This paper is organized as fallows; Section - II presents the detailed study of literature Survey, Section - III. Quantitative result analysis from literature survey. Section - IV gives the comparison of different speech enhancement techniques which are widely in use Lastly Section - V permits the conclusion and future scopes of this research area.

2. LITERATURE SURVEY

The various algorithms used for enhancement of noisy speech are as follows:

L.lin and E.Ambikairajah [1] study was to improve the noisy speech quality by removing the white and colored noise in each sub band channels.

The author has used the auditory filter bank structure. It consists of narrow band pass filters and placed according to equivalent rectangular bandwidth (ERB). The proposed

speech denoising system consists of analysis filter and synthesis filter. The input to auditotary filter bank is noisy speech signal corrupted by additive noise; this noisy speech signal is filtered by a bank of analysis filter spaced on equivalent react angular bandwidth scale, which have zero mean. Then denoising is achieved by multiplying the output of each filter by scaling factor and calculated signal and variances. This method can gives better quality speech denoising even in present of colored noise. This algorithm performs better in presence of both white noise and colored noise.

Ying Deng et.al.,[2] the author focused on design of the low delay uniform QMF bank with independent analysis and synthesis prototype filter. The author has proved that synthesis prototype has to be extended version of analysis prototype filter, in order to remarkable aliasing cancellation to occur in the less delay uniform QMF banks and to accomplish notable distortion cancellation in less delay non uniform filter bank, But this method shows slightly greater residual noise in segmented SNR levels as compared to Canon's bark extended wavelet packet decomposition.

Bastian Sauert et.al.,[3,5] proposed warped low delay filter bank using Discrete Fourier Transformer(DFT).In this method was used time and frequency dependent approach which intensifies far end speech signal to retrieve a target signal to noise ratio(SNR).This algorithm consists of analysis synthesis filter bank, implemented by overlap add method with non uniform filter bank equalizer.

In non uniform filter bank equalizer designed by all pass transformation, it changes both magnitude and phase response of filter. So this filter bank achieves the small signal delay by using analysis-synthesis filter bank in time domain filter, but low delay filter performed by moving average filter or auto regressive filter approximation, but it can achieve a delay of only few samples this approach applied to non uniform filter bank equalizer to calculate auto regressive co-efficient with delay feed bank loops. The proposed algorithm was evaluated the term of speech intelligibility index, apart from approximately calculation of spectral weights in bark scaled frequency band for the non uniform frequency resolution of human ear.

Asha Vijaykumar [4] author suggested novel unimodular filter bank algorithm which is the noisy input signal, processed by an M-channel one regular unimodular filter bank. The each M sub band channel output are scaled by scaling factor and computed with varying input signal, by using unimodular filter bank we have to reduce the sub band based noise suppression, after that we get speech and noise database and the database used for automatic speech recognition. The output of filter provides the effective noise

reduction as well as improve recognition rate at less delay speech signal. The demerit of this filter bank is produce highly non stationary noise.

Yu Cai et.al.,[7] author conveyed the sub band spectral subtraction algorithm based on over sampled DFT modulated filter bank .Here the input signal is filtered by k channel analysis filter and it is decimated by factor M, the decimated signal transforms into M sub band channels. These sub band signals are processed and interpolated by synthesis filter, these filter can remove the musical noise in speech enhancement and output of synthesis filter can sum together to get enhancement of speech signal, the enhanced speech signal gives better result rather than multiband method. But this filter bank also gives the same result as Kamath method in speech enhancement. The proposed algorithm may suppress the noise, to preserve relatively integrated speech component. This algorithm is also adaptable method that can acquaint into many audio applications.

Alexander Schase et.al.,[8] author presented cascaded filter bank algorithm which increases the frequency resolution without changing original filter bank system. The proposed algorithm, A time-domain signal is divided into oversampling samples which weighted by a analysis prototype low pass filter and these samples processed by sub band signal samples to get the frequency bins it depends on the sampling frequency of time domain signal. The frequency resolution in the low frequency range of the analysis system can be increased by decomposing sub band signals, if filter can cascaded it can reduce the spectral gain, but in this system gains of the a frequency domain wiener filter. The additional delay caused by cascaded system can be reduced by low delay filter bank.

Jiri Malek et.al.,[9] anticipated compare trained bank (L-CFB) and manually constructed bank(M-CFB) algorithms are used in real time scenario for noise free recording of mobile phones. This method described here scans the noise free recording, this data processed block by blocks it may overlap. In beginning stage the filter bank is empty. The algorithm computes target cancelation filter in the first block and it puts into the filter bank if output variance is smaller than threshold it detect the object. If variance is higher than, the algorithm is processed in the next block, this is repeated until a first target cancellation filter is added to the filter bank. Here each block is in parallel; filtered by all cancellation filters in the current filter bank and their output are compared. If cancellation filter bank yield the minimum output variance over first object, a new cancellation filter computed for the block of signal. The resulting filter bank is added to the filter bank. The resulting bank will henceforth be referred to as one pass target cancellation filter bank.

Kristian Timm Andersen et.al.,[10] author proposed Fast Fourier transformer algorithm is used to an adaptive time-frequency analysis/synthesis scheme for processing of audio less delay and less complexity. The adaptive time and frequency scheme is used for noise suppression gain factor by analysis/synthesis method.

The adaptive time-frequency can adopt its frequency resolution and develop the suitable time delay. The analysis filter is incepted into the frame of the DFT modulated filter bank and these filters are used to determine the gain factors while synthesis uses DFT filter bank [9]. The synthesis method uses an asymmetric window it is mixed with twiddle filter bank that maintains less delay of the reference filter bank but improves the frequency response. Additional delay can be reduced by applying the gain to signal. The complexity of the filter bank decreased in frequency by the FFT algorithm in the DFT filter bank and it is calculated by number of frequency band which selected in FFT algorithm.

Pinki, Sahil Gupta [11.] author proposed Spectral Subtraction method. This method operates on the frequency domain; the input signal may be presented as the addition of the speech signal and noise signal. In this method, first we have to estimates spectrum of the background musical noise and eliminating the noise spectrum from input speech signal. To perform the frequency domain processing, the continuous time signal cab be split into overlapping pieces is called frames. These frames can operate; the frames can reassemble to create a non-discrete output signal. To avoid the spectral effects, we have to multiply the window function before processing the fast Fourier transformer and again after evaluating inverse fast Fourier transformer. In this technique the speech spectrum is divided into number of frequency bands and spectral subtraction is operated independently and this algorithm can readjust the over subtraction factor in every frequency band, to estimate the clean speech magnitude spectrum. This algorithm removes the additive noise and remnant noise using spectral subtractive type algorithm.

3. QUANTITATIVE RESULT ANALYSIS FROM LITERATURE SURVEY.

In these section quantitative results of various earlier speech enhancement algorithms is presented.

The researcher L. Lin et.al.,[1] have shown that there is a significant improvement in segmental signal to noise ratio (SNR) for various SNR levels with pink noise as background. The details are shown in table 1.

Table 1: Segmental SNR improvement with respect to initial SNR with pink noise.

Sl. No.	Input segmental SNR(dB)	Improve mental in segmental SNR(dB)
01	-4	9
02	-3	8.5
03	-2	8
04	-1	7.5
05	0	7
06	1	6.5
07	2	6
08	3	5.5
09	4	5

Above results shows that improvement in segmental SNR decreases as SNR of noisy speech signal increases.

The researchers Ying Deng et.al.,[2] have disported that there is pregnant improvement in segmental SNR for various SNR levels with different noise like White Gaussian noise, Cockpit noise, factory noise and traffic noise as background. The details are shown in table 2.

Table 2: Segmental SNR improvements with respect to initial SNR with different background noises.

Sl. No.	Type of Noise	Input segmental SNR(DB)	Improvement in segmental SNR(DB)
01	White Gaussian Noise	-5	8.5
		0	7.5
		5	5.5
02	Cockpit Noise	-5	6.5
		0	5.5
		5	3.5
03	Factory Noise	-5	6.5
		0	4.5
		5	3.0
04	Traffic Noise	-5	9.8
		0	8.2
		5	6.0

Above results shows that improvement in segmental SNR decreases as SNR of noisy speech signal increases which is in conformity with theory.

The authors Bastian Saurt et.al.,[3] have conveyed that there is a considerable improvement in mean speech intelligibility index (SII) as a function of input speech signal SNR using time domain frequency independent processing and frequency dependent processing using different filter banks. The details are shown in table 3.

Table 3: Comparison of mean SII in various frequency bands using filter banks.

Sl. No	Type of method	Type of filter bank	Input segmental SNR(DB)	Improve mental in segmental SNR(DB)
01	Time domain frequency independent	Uniform ASFB	-20	0.55
			-10	0.75
			0	0.85
			10	0.87
			20	0.88
		Warped FBE	-20	0.62
			-10	0.78
			0	0.82
			10	0.85
			20	0.86
02	Time domain frequency dependent	Uniform ASFB	-20	0.60
			-10	0.75
			0	0.85
			10	0.86
			20	0.88
		Warped FBE	-20	0.62
			-10	0.75
			0	0.82
			10	0.86
			20	0.88

Above results shows that improvement in segmental SNR decreases as SNR of noisy speech signal increases which is in conformity with theory.

The researcher Asha Vijaykumar has presented that [4] there is outstanding improvement SNR in different noisy environments for various SNR using Gammatone filter bank and unimodular filter bank. The details are shown in table 4.

Table 4: Noisy type versus initial SNR using filter banks.

Sl. No	Filter Bank	Noise type	Input SNR(dB)			
			0	5	10	15
			Word Recognition Rate			
01	GFB	White	33.82	34.86	42.20	44.95
02	UMFB	White	38.53	48.62	56.88	69.72
03	GFB	Babble	14.68	27.52	45.87	62.39
04	UMFB	Babble	33.94	49.54	69.72	78.90

The performance of word recognition increase as input SNR increases using unimodular filter bank.

The author Yu Cai et.al.,[5] have shown that there is significant improvement in output segmental SNR for different input SNR levels in different noise environments using DFT modulated filter bank. The details are shown in table 5.

Table 5: Comparison of Itakura-Saito distance resulting from different algorithms with proposed algorithm (DFT-MFB).

Sl. No.	Noise Type	Input SNR(dB)	Algorithms		
			Berouti	Kamath	DFT-MFB
01	Station	0	9.85	2.32	2.18
		5	6.94	1.80	1.83
		10	6.62	1.25	1.04
02	Car	0	6.75	3.09	2.90
		5	4.95	2.03	2,02
		10	3.84	1.77	1.80
03	Babble	0	8.99	2.66	2.64
		5	3.58	2.02	1.93
		10	2.59	1.34	1.31
04	Airport	0	5.60	2.54	2.51
		5	4.46	1.79	1.79
		10	4.14	1.38	1.38

Itakura Saito distance decreases as SNR increases and it is theoretically true.

The researcher Alexander Schasse et.al.,[6] have presented perceptual evaluation of speech quality score and Itakura-Satio distance for various input SNR levels with different filter banks. The details are as shown in table 6.

Table 6: Comparison of PESQ score and ISD with respect to different input SNR levels from different filter banks.

Sl. No	Type of Filter bank	Input SNR(DB)	PESQ Score	Itakura-Satio distance
01	Low frequency resolution	-5	0	3.5
		0	0	3.4
		5	0	3.2
02	High frequency resolution	-5	0.15	3.2
		0	0.1	3.15
		5	0	3.1
03	Second order filter bank with linear phase FIR Filter	-5	0.2	3.2
		0	0.19	3.15
		5	0.17	2.95
04	Second order filter bank with linear phase IIR Filter	-5	0.15	3.2
		0	0.1	3.15
		5	0.9	3.1

There is no significant improvement in PESQ score and ISD from the work.

The author pinki et.all [8] shown the SNR in a given frame is not constant rather it is function of frequency. The details are shown in table 7.

Table 7: SNR value as a function of frame number and frequency bands.

Sl. No.	Frequency band(HZ)	Frame number	Segmental SNR(dB)
01	60Hz-1KHz	84	6
		86	10
		88	13
		90	4
		92	15
		94	5
02	1KHz-2KHz	84	4
		86	5
		88	7
		90	3
		92	15
		94	5

		84	2
		86	5
03	2KHz-3KHz	88	3
		90	2
		92	3
		94	3
04	3KHz-4KHz	84	2
		86	3
		88	2
		90	1
		92	3
		94	2

The above result shows that noise is colored. SNR of a given segment does not remain constant throughout the frame.

4. OBSERVATION OF DIFFERENT ALGORITHMS FOR SPEECH ENHANCEMENT WITH VARIOUSDATABASES

Table 8: Comparative analyses along with the future scope of various algorithms used in speech enhancement

Author	Objective	Algorithm	Result	Future Scope
L.Lin et.al., [1]	To improve speech quality by reducing musical noise in speech enhancement using auditory filter bank.	Modified Auditory filter bank Algorithm	To achieve Speech denoising by scaling factor of the each filter, calculated from each channel in time domain.	The denoising technique integrated to an auditory filter based speech recognition under noise condition using FFT algorithm.
Ying Deng et.al.,[2]	A Quadrature Mirror Filter (QMF) design algorithm that significantly suppress delay by relaxing the linear phase constraints.	Oversampling critical-band wavelet packet decomposition and Recursive Averaging Method.	QMF bank achieved linear phase constraints and it provides better balance between delay and system performance.	Significantly to Reduce the noise in speech enhancement.
Bastain Sauert et.al.,[3]	To remove the background noise in speech enhancement by using Non-unifrom filter bank with Bark scaled Frequency Band.	Uniform DFT bank Algorithm	The proposed algorithm has evaluated the term of speech intelligibility index	The warped time domain filter of filter bank equalizer reduces the signal delay in speech intelligibility.
Asha vijaykumar[4]	The effective noise reduction at low frequencies in speech signal.	Unimodular Filter bank and Sub band algorithm.	The One-regular Uni-Modular filter bank noise reduction is improved in quality of speech enhancement	The TIV One-regular Uni-modular filter bank are replaced with time varying one regular UMFB, was remove non stationary noise.
Yu cai et.al.,[5]	To enhance the speech quality by reducing the additive background noise using DFT Modulator filter banks.	Sub band spectral Subtraction method based on oversampling DFT Modulator filter bank.	This algorithm can remove the music noise in speech Enhancement	This algorithm can be used in audio applications.
Alexander schase et.al.,[6]	To improve the frequency resolution for speech enhancement techniques in Hearing aids.	Cascaded filter bank algorithm	The cascaded filter bank can enhance the frequency resolution without changing original filter bank system.	The additional delay will be reduced by using low delay filter bank system such as equalizer filter bank.

Jiri Malek et.al.,[7]	Preparing a bank of target cancellation filter to reduce the noise present in the adaptive target.	Trained bank[L-CFB] and Constructed bank[M-CFB] algorithm is used	This algorithm is used in real time scenario for noise free recording in mobile phones.	L-CFB and M-CFB ability to reduce the unwanted noise in mobile phones.
Kristian Timm Andersen et.al.,[8]	Adaptive analysis scheme to adopt the input signal in both time and frequency domain a complex filter on a DFT modulated filter bank.	Fast Fourier transformer algorithm is used	The adaptive time-frequency analysis/synthesis filter bank processing of audio with less delay and less computational complexity in real time system.	The assessment of the adaptive TF analysis method how the delay is essential in the analysis impacts the application of the gain factor.
Pinki,Sahil Gupta	Speech enhancement using spectral subtraction.	Spectral Subtraction algorithm is used	In this paper, comparison and simulation of different spectral subtractive technique can suppress the additive noise.	This algorithm can be used in audio applications to reduce the additive noise.

5. CONCLUSION

In speech enhancement various filter banks are used, such as Warped Low Delay Filter-Banks (WLDFB), Unimodular Filter Banks (UMFB), Cascaded filter Banks, DFT modulated Filter Banks (DFTMFB) to reduce non stationary noise for low delay signal in speech intelligibility. This paper is a firsthand document based on the literature study of few techniques, which delineate the challenges, associated with related work, objective, algorithm with possible outcome in various publications. We can conclude from this literature survey that increment in segmental SNR, PESQ score, ISD and MOS values decreases as input noisy speech SNR increases. It is also observe that noise is colored that is SNR level is frequency dependent. The key inference is noise level should not be treated as constant value in a given time frame, within that time frame it is function of frequency. Finally it is worth to note that processing of noisy speech for its enhancement should be done on time-frequency grids.

ACKNOWLEDGEMENT

REFERENCES

[1] Thomas F. Quatieri, **"Discrete-Time Speech Signal Processing: Principles and Practices"**, Prentice Hall, 2001.

[2] L.Lin and E Ambikairajah**"Speech Deoising on an auditory Filter Bank"**The IEEE ICPS'02 Proceeding-7803-7488-6/02@2002 IEEE

[3] Ying Deng Member IEEE,John Mathew,Fellow IEEE and Behrouz-Borojeny Senoir member IEEE**"Low Delay Nonunifrom Pseudo-QMF Bank with application to Speech Enhancement"**,IEEE Transaction on signal processing Vol.55,No-5,May 2007

[4] Bastian Sauret,Heinrich W Lolmann and Peter Vary**"Near End Listening Enhancement by means of Warped low delay Filter-Bank"**The ITG-Fachtagung spachommunication 8-10 october 2008 in Aachen

[5] Asha Vijaykumar**"Noise Supression Using One Regular Unimodular Filter Bank"** The ICME@2009-1-4244-4291-1/09 IEEE

[6] Dr.(Mrs)S.D Apte,Shridhar **"Speech Enhancement in Hearing Aids using Conjugate Symmetry property of short Time Fourier transform"** International Journal of Recent Trends in Engineering,Academy publishers,finand,Nov 2009,Vol 2,No-16 pp346-351

[7] Dr.(Mrs)S.D.Apte.Shridhar **"Speech Enhancement in Hearing Aids using Conjugate symmetry of DFT and SNR-Perception Models"** International Conference on futuristic computer application,IISC,Banglore,20 March2010,PP 43-5
https://doi.org/10.5120/58-650

[8] Alexander Schase, Rainer Martin,Wolgang Sogel,Thomas Pilrim,Henninig Puder **"Efficent Implementation of single Channel noise reduction for hearing Aids using a cascaded filter "**The ITG-Fachberichiy236-Spachommunication -28.09.2012 in Braunschweing@ VDE VERLAG GMBH Berlin offenbanch

[9] Yu cai and Chaohuan Hou **"Subband spectral subtraction speech enhancement based on the DFT Modulated Filter Banks"**

[10] Jiri Malek david Botka,Zbynek Koldoysky and Sharon Gannot "**Method to learn bank of filter steering nulls towards potential position of a target sources"** The HSCMS 978-4799-31095/14 @IEEE 2014.

[11] Pinki Sahil Gupta**"Speech Enhancement using spectral subtraction type algoritms:A Survey on Comparison"** The IJCS Volume 04 Issue 10-oct-2015,page no.1487-14878.