A Review on Single Channel Speech Enhancement Methods

Brijesh Anilbhai Soni[1] Prof. Kinnar Vaghela[2]
[1]PG Student [2]Associate Professor
[1,2]Department of Electronics & Communication Engineering
[1,2]LDCE, Ahmedabad, Gujarat, India

Abstract— Speech is one of the fundamental means of communication. However noise free speech is never possible in the real world. It is always accompanied by the background noise and thus speech enhancement has been a long standing problem in signal processing .Speech enhancement algorithms are important components in many systems where speech plays a part, including telephony, hearing aids, voice over IP, and automatic speech recognizers. Speech enhancement is generally concerned with the problem of enhancing the quality of speech signals. Till date there have been various approaches for speech enhancement. In this paper, we provide the subjective description of this approaches.

Key words: Speech Enhancement, Spectral Subtraction, Noizeus, TIMIT Database

I. INTRODUCTION

In this paper, various techniques of suppression of acoustic noise has been discussed. The problem of noise suppression has received considerable attention, since it is relevant to many important applications like speech recognition and compression, restoration of archived files, speaker recognition etc. Our work targets the problem of speech enhancement in particular.

The organization of paper is as follows: Section II describes various speech enhancement techniques. Section III gives the brief description of various database available. And last section concludes the paper.

II. APPROACHES FOR SPEECH ENHANCEMENT

A. *Boll's Algorithm*

The firstly we compute the short-time Fourier transform (STFT) of the noisy signal using the fast Fourier transform (FFT) followed by windowing with the Hanning window. For this, we set the length of the window and the FFT to 256, with a shift in steps of 128 points. Next the noise is estimated in the unvoiced region through voice activity detection (VAD) technique [1]

$$Y(\omega) = X(\omega) + N(\omega) \qquad (1)$$

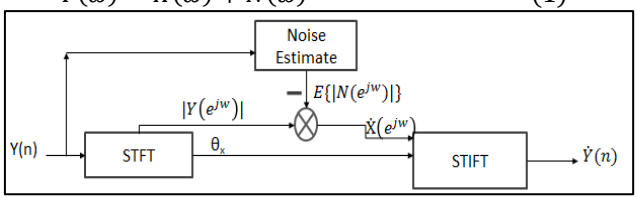

Fig. 1: Block diagram of spectral subtraction

After VAD, estimated noise spectra is subtracted from noisy speech signal followed by half wave rectification

$$|\hat{X}(\omega)| = |Y(\omega)| - E\{|N(\omega)|\} \qquad (2)$$

However two problems immediately appears: a clear narrowband of noise still remains in the spectrum, even if our estimate of noise is correct, and listening to the enhanced signal, we can notice an undesirable new noise appearing.

To mitigate the undesirable new noise, musical noise as claimed by berouti, Berouti prosed a better approach. As explained by Berouti [2], peaks and valleys exist in the noise spectrum, and once the estimate is subtracted, peaks remain as randomly occurring peaks, while valleys are set to zero. The peaks are "perceived as time varying tones which we refer to as musical noise."

B. *Berouti Algorithm*

Berouti generalizes the spectral subtraction technique by not only considering subtraction of amplitude spectra, but also power spectra [2]. Given P_X, P_y, P_n, the power spectra of the estimated clean signal, the noisy signal, and the noise (respectively), Berouti introduced two parameters in the spectral subtractor estimator, which is expressed as follows:

$$\hat{P}_x = (P_y^\gamma - \alpha P_n^\gamma)^{1/\gamma} \qquad (3)$$

The parameter' α', varying from 3 to 6 allows overestimating the power spectrum of noise and 'γ' raises the power of the power spectrum before subtraction [2].

However intelligibility of speech may suffer if overestimating factor is kept very high.

C. *Wavelet Domain Approach*

In this, spectral subtraction is applied to wavelet Approximations and details coefficients. A new parameter is introduced for spectral subtraction in unvoiced speech frames and the existing power factor in spectral Subtraction method is improved [3].

Discrete Wavelet Transform (DWT) is applied to the noisy signal frame, so the approximations and details coefficients are acquired. Also DWT is applied to the noise estimated from silence frames to acquire the estimated approximations and details coefficients of noise. In the next step, the GSS (Generalized spectral subtraction) algorithm proposed by Berouti et al. [2] has been improved and applied in parallel to both, approximations and details of noisy signal.

$$\text{if } |\hat{S}(\omega)|^\gamma > \beta|\hat{D}(\omega)|^\gamma \qquad (4)$$

$$|\hat{S}(\omega)|^\gamma = |\hat{Y}(\omega)|^\gamma - \eta.\alpha\,|\hat{D}(\omega)|^\gamma \qquad (5)$$

$$\text{else } |\hat{S}(\omega)|^\gamma = \beta|\hat{D}(\omega)|^\gamma \qquad (6)$$

Where $|\hat{S}(\omega)|^\gamma$ is the spectrum of enhanced approximations/details. Y (ω) is the spectrum of noisy approxiations/detail and $|\hat{D}(\omega)|^\gamma$ is the estimated noise approximations/detail

D. *Multitaper Spectral Estimation Approach*

In this approach, main idea is to reduce the variance by obtaining multiple independent estimates from the same sample. Taper is multiplied to the signal element wise providing a sub spectral signal. Since all tapers are orthogonal to all others, sub spectral signals provide statically independent estimate of spectrum thereby minimizing spectral leakage which exists in the finite length data set. Then final spectrum is obtained by averaging over all tapered spectra [4].

Estimation errors in sub spectra will be approximately uncorrelated, which is a key to variance reduction.

The multitaper based spectral subtraction method provides a definite improvement over the conventional hamming window. The database used in this paper is Noizeus.

Spectral subtraction method fails for colored noise and this problem has been solved using proposed non-linear Multiband Spectral Subtraction algorithm.

The flow chart is as shown in the figure (2).

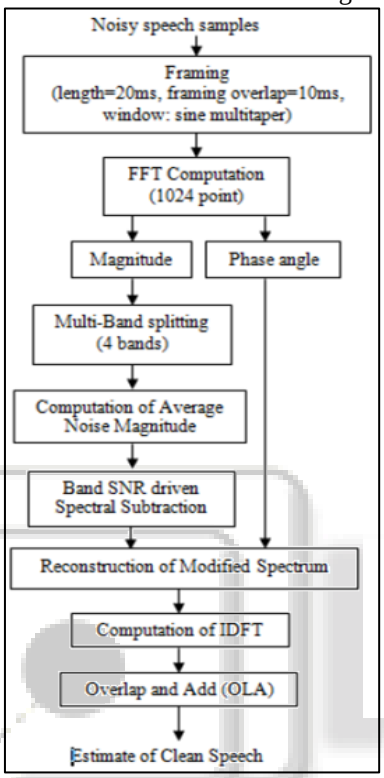

Fig. 2: Multitaper spectral estimation flow chart

E. Cepstral Smoothing Approach

Speech enhancement algorithms which modify short term spectral magnitudes of the noisy signal by means adaptive spectral gain functions are plagued by annoying spectral outliers which we call it as a musical noise. However cepstral smoothing can effectively prevent spectral peaks of short duration that may be perceived as musical noise. Also this approach preserves speech onsets, plosives, and quasi-stationary narrowband structures like voiced speech [6].

Authors propose the smoothing of the filter gain function in the cepstral domain to suppress the tendency of adaptive spectral filters to produce *musical noise*. The new method results in an effective smoothing of fine spectral variations that may be perceived as *musical noise*. At the same time, the spectral characteristics of speech are not affected.

They propose a temporal smoothing of the cestrum of the gain function G (k, l) in order to avoid a peaked shape of G (k, l) due to outliers in noise [6].

The spectral gain function is given by

$$G(k,l) = \frac{\xi_{n,k}}{1+\xi_{n,k}} \quad (7)$$

Where $\xi_{n,k}$ denotes the estimate of apriori SNR

$$\hat{\xi}_{n,k} = a\,\frac{\hat{X}^2_{n,k}(m-1)}{\hat{P}[K,(m-1)\ ]} + (1-a)[\gamma_{n,k}(m)-1\ ] \quad (8)$$

The smoothed version is approximated as

$$G^{ceps}_{smooth}(k,l) = (1-\beta)G^{ceps}(k,l) + \beta.G^{ceps}_{smooth}D_{n,k}(k,l-1) \quad (9)$$

Here cepstrum is divided as lower coefficients. As speech onsets and the spectral envelope of fricatives and plosives must not be distorted by the smoothing procedure, smoothing is not applied to low cepstral coefficients. This preserves the principal structure of the gain function in the case of speech presence.

However, Spectral peaks in $G(k,l)$ caused by outliers will be represented by higher cepstral coefficients, which will be smoothed out

F. Surround Noise Cancellation and Speech Enhancement using Sub Band Filtering and Spectral Subtraction

In this paper, algorithm based on Sub Band filtering, Spectral Subtraction, and Voice Activity Detector (VAD) for surround noise cancellation and speech enhancement is proposed.

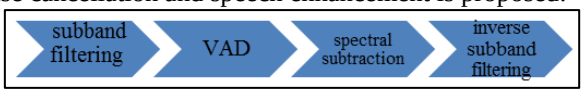

Fig. 3: Stages of surround noise cancellation approach

The first stage is sub band filtering, in which the given speech signal is filtered into smaller sub bands. Analysis stage performs the conversion of signal from time domain to frequency domain. The second stage is VAD, wherein presence of speech and noise alone signal is identified. Output of VAD is given to spectral subtraction, where the magnitude spectrum of the signal is either subtracted or attenuated depending on the presence of speech or noise. The final stage is to reconstruct the speech signal.

Simulation results show that the algorithm works efficiently in reducing the surround noise, which when mixed in a clean speech signal and this algorithm can also be tested for musical noise reduction [7].

G. STFT phase improvement for single channel speech enhancement

In single channel short-time Fourier transform (STFT) based speech enhancement algorithms only the amplitude of the noisy speech signal is improved, but its phase is kept unchanged.

In [8], phase of voiced clean speech is blindly reconstructed in the STFT-domain and the proposed algorithm leads to frequency weighted SNR improvements of up to 1.8 dB.

At each sample n the noisy speech y (n) is given by an additive superposition of speech s (n) and noise v (n), i.e. y (n) = s(n)+ v(n)

$$Y(k,l) = \sum_{n=0}^{N-1} y(lL+n)w(n)e^{-j\Omega n} \quad (10)$$

$$Y(k,l) = s(k,l) + V(k,l) \quad (11)$$

$$Y(k,l) = \left|Y(k,l)e^{j\emptyset(k,l)}\right| \quad (12)$$

Further this paper is divided into two halves. In the first half, phase is reconstructed along time and in the second it is reconstructed along frequency.

Timit database is used for implementation. This paper concludes that beside the standalone application, its combination with STFT-amplitude enhancement algorithms can provide promising results [8].

III. DATABASE

We need standard speech samples for uniquely measuring various speech parameters. Few database for speech enhancement are discussed as below.

A. Noizeus Database

A noisy speech corpus (NOIZEUS) is developed to facilitate comparison of speech enhancement algorithms.

The database contains 30 IEEE sentences (produced by three male and three female speakers) corrupted by eight different real-world noises at different SNRs. The noise was taken from the AURORA database and includes suburban train noise, babble, car, exhibition hall, restaurant, airport, street and train-station noise [9].

The sentences were originally sampled at 25 kHz and then down sampled to 8 kHz. Noise signal are taken from Aurora database. A noise of the same length as the speech signal is randomly cut out of the noise recordings, appropriately scaled to the desired SNR level and added to the filtered clean speech signal.

B. Timit Database

The TIMIT corpus of speech is designed to provide speech data for acoustic-phonetic studies and for the development and evaluation of automatic speech recognition systems and enhancement algorithms.

TIMIT contains broadband recordings of 630 speakers of eight major dialects of American English, each reading ten phonetically rich sentences. The TIMIT corpus includes phonetic and word transcriptions as well as a 16-bit, 16kHz speech waveform file for each utterance.

C. Spear Database

Spear stands for Speech Enhancement and Assessment Resource. It has the following contents

- Noisy Speech Recordings- Recorded speech and recorded noise: acoustically combined and re-recorded. Various noise sources and different SNR levels.
- Lombard Speech - Live speech in a noisy environment. Recorded noise sources are used to allow extraction of a cleaned speech reference.
- Monaural Recordings - Two recorded speech signals: acoustically combined and re-recorded.

IV. CONCLUSION

From this review we conclude that there are various approaches for speech enhancement. However, spectral subtraction remains one of the most simple and basic method, but it inherits musical noise. Most other methods do not introduce such musical noise but they come at the cost of complexity of implementation.

Also till date all most all approaches were ignoring the phase for speech enhancement as per [5]. However according to [8], phase has got importance in reconstruction of speech signal and same has been incorporated for speech enhancement with very promising result.

REFERENCES

[1] Boll, S.F. Suppression of Acoustic Noise in Speechusing Spectral Subtraction. IEEE Transactions on Acoustics, Speech, and Signal Processing (27), pp. 113-120, 1979

[2] Berouti, M., Schwartz, R., and Makoul J, *Enhancement of speech corrupted by additive noise.* IEEE Transactions on Acoustics, Speech, and Signal Processing, pp. 208-211, 1979

[3] Malihe hassani, M. R. Karami mollaei,Speech Enhancement Based on Spectral Subtraction in Wavelet Domain, 2011 IEEE 7th International Colloquium on Signal Processing and its Applications

[4] Supriya.P.Sarvade, Dr.Shridhar.K, Varun.P.Sarvade Multi-Band Spectral Subtraction for Speech Enhancement Using Sine Multitaper IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 6, Issue 6, Ver. II (Nov. - Dec. 2016), PP 70-76

[5] D. L.Wang and J. S. Lim, "The unimportance of phase in speechenhancement," IEEE Trans. Acoust., Speech, Signal Process.,no. 4, pp. 679–681, 1982.

[6] Colin Breithaupt,Timo Gerkmann, and Rainer Martin Senior Member, Cepstral smoothing of spectral filter gains for speech enhancement without musical noise, Institute of Communications Acoustics (IKA) Ruhr-University Bochum, 44780 Bochum, Germany.

[7] Surround noise cancellation and speech enhancement using Sub Band Filtering and Spectral Subtraction, Vignesh Ganesan and Sangeetha Manoharan, Indian Journal of Science and Technology, Vol 8(33), December 2015.

[8] Stft phase immprovement for single chanel speech enhancement, Martin Krawczyk and Timo Gerkmann, Speech Signal Processing Group, Institute of Physics, University of Oldenburg, Germany

[9] Noizeus database available on http://ecs.utdallas.edu/loizou/speech/noizeus

