Spectral Subtraction for Speech Enhancement in Modulation Domain

Muhammad T. Sadiq[1], Noman Shabbir[2] and Wlodek J. Kulesza[3]

[1] Dept of Electrical Engineering, Sharif College of Engineering & Technology (Affiliated with UET Lahore), Lahore, Pakistan

[2] Dept of Electrical Engineering, GC University, Lahore, Pakistan

[3] Dept of Electrical Engineering, Blekinge Institute of Technology, Karlskrona, Sweden

Abstract

In this paper, an enhancement method is proposed for a speech signal that is corrupted by background noise. The proposed method is based on the spectral subtraction technique. The conventional spectral subtraction method improves the quality of the signal by reducing noise but it introduces an annoying musical noise. To eliminate the musical noise, spectral subtraction method in modulation domain is proposed. A filter bank, the heart of modulation domain, splits the broadband signal into several sub-bands. Then Instantaneous Frequency (IF) carrier estimator technique is used to estimate the carrier and modulating signal of each sub-band. Implementing spectral subtraction on each modulating signal and by synthesizing signals at the output, we get the enhanced speech signal without musical noise. Both qualitative (listening tests) & quantitative tests were performed to determinate the intelligibility and quality of speech enhanced by the proposed method.

Keywords: *Speech Enhancement, Spectral Subtraction, Modulation domain, Noise Removal.*

1. Introduction

SPEECH signals are low frequency modulating signals which modulate high frequency carriers. During transmission, these signals are affected by Additive White Gaussian Noise (AWGN). The quality of the speech signal is deteriorated by the noise recorded in real time situations. The noise decreases the intelligibility and quality of the signal while increasing listener fatigue. To enhance the speech and suppress the noise, different speech enhancement methods such as adaptive gain equalizer, Wiener filtering and spectral subtraction are used. Experimental results show that spectral subtraction reduces the background noise effectively but it introduces a musical noise in the signal [3]. The musical noise is an offensive noise. In order to eliminate the musical noise, the spectral Subtraction method is implemented in modulation domain. Modulation domain is used to split a speech signal into carrier and modulating signals and an Instantaneous Frequency (IF) technique estimates these signals.

Our proposed method not only enhances the speech signal but it also provides better background noise reduction than conventional spectral subtraction method and does not introduce musical noise in the signal.

2. Survey of Related Works

A spectral subtraction algorithm [1] suppresses stationary noise from speech by subtracting the spectral noise bias calculated during non-speech activities; however it attenuates the musical noise left after subtraction. The speech enhancement cascaded algorithm increases the overall effectiveness of the speech enhancement system [3]. Meanwhile the drawback of this algorithm is that it introduces undesirable musical noise. Modulation domain is a new domain in signal processing, in which one separates the modulating and carrier signals of the speech and different techniques like smoothed Hilbert carrier estimator, Instantaneous Frequency (IF) carrier estimator and frequency reassignment carrier estimator are used to estimates these signals [2]. Modulation domain processing has grown in popularity finding applications in areas such as speech coding, speech and speaker recognition, objective speech intelligibility evaluation as well as speech enhancement [5].

3. Problem Statement and Main Contribution

Speech signal is contaminated by background noise that appears randomly in the signal from frame to frame. This result in a twinkling sounding noise called musical noise that can be quite annoying for the listener.

Our research question is how to implement spectral subtraction speech enhancement method for musical noise elimination in a modulation domain. We hypothesized that the spectral subtraction method can be implemented in modulation domain using IF technique which should eliminate the musical noise.

The main contribution of this paper is to model and implement spectral subtraction in modulation domain using MATLAB and then to validate the method for speech enhancement.

4. Problem Solution

4.1 Modeling of Spectral Subtraction in modulation domain

Our proposed model is depicted in Fig. 1. In this model, a noisy signal *x(n)* which is the combination of a clean speech signal *d(n)* and a noise signal *v(n)* is taken. A filter bank then divides the noisy speech signal into different sub-bands. The resultant signal is depicted as $x_N(n)$, where *N* is the number representing each sub-band. After that IF technique is used to estimate the carrier and modulating signal of each sub-band. The IF is the modification of the differential FM detector. A phase-only IF estimate $\alpha_N(n)$ is given by [2]:

$$\alpha_N(n) = \begin{cases} \left(\frac{Z_N(n)}{|Z_N(n)|}\right)^{\frac{1}{2}} & |Z_N(n)| > \epsilon \\ \alpha_N(n-1) & |Z_N(n)| \leq \epsilon \end{cases} \quad (1)$$

where $Z_N(n)$ is an un-normalized IF estimate, and ε is a small threshold that is used to reduce noise in the IF estimate. Smoothed IF estimates give an instantaneous phase estimate by the following equation, with the initial assumptions $W_N(-1)=1$ and $W_N(0)= \alpha_N(0)$:

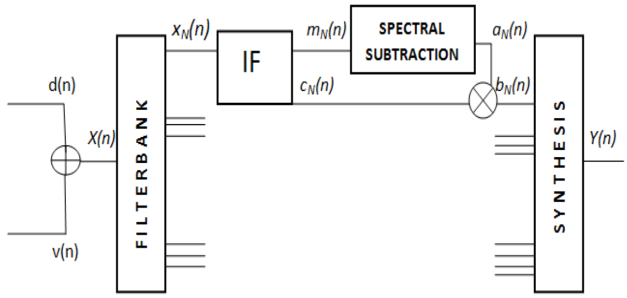

Fig 1: Block model of spectral subtraction in modulation domain.

$$W_N(n) = W_N(n-1)\alpha_N(n) \quad (2)$$

Once the instantaneous phase estimate $W_N(n)$ is calculated for sub-band $x_N(n)$, the carrier $c_N(n)$ is computed as:

$$c_N(n) = W_N(n) \quad (3)$$

and the modulating signal $m_N(n)$ can be found by demodulation of sub-band signal:

$$m_N(n) = x_N(n).\, c_N^*(n) \quad (4)$$

Now the spectral subtraction method is applied on the modulating signal $m_N(n)$. In spectral subtraction method, Voice Activity Detection (VAD) is used to determine the presence or absence of speech signal. Speech enhancer divides the input signal into different frames and then processes each frame. Let the signal of frame *k-th* be [4]:

$$m_{N,k}(n) = w(n) m_N(n + kD)\ for\ n = 0 \ldots L-1 \quad (5)$$

where *k* is the number of frame, *D* is the step size for each frame, *w(n)* is a window and *L* is the length of a frame.

The power spectral density of noise for frame *k* is calculated as [4]:

$$P_{vv,N,k}(\omega) = \begin{cases} \gamma P_{mm,N,k}(\omega) + (1-\gamma) P_{vv,N,k-1}(\omega) \\ \qquad \text{if } k \text{ is nonspeech frame} \\ P_{vv,N,k-1}(\omega) \quad \text{if } k \text{ is speech frame} \end{cases} \quad (6)$$

where $P_{vv,N,k-1}(\omega)$ is the estimated power spectral density of noise for frame *k-1*, $P_{mm,N,k}(\omega)$ is the estimated power spectral density of modulating signal for frame *k* and $0 < \gamma < 1$ is a forgetting-factor.

To enhance the signal $m_N(n)$, a gain function $G_{N,k}(\omega)$ is define as:

$$G_{N,k}(\omega) = 1 - \sqrt{\frac{P_{vv,N,k}(\omega)}{P_{mm,N,k}(\omega)}} \qquad (7)$$

The output signal for *k-th* frame is calculated as:

$$A_{N,k}(\omega) = M_{N,k}(\omega).G_{N,k}(\omega) \qquad (8)$$

where $M_{N,k}(w)$ is the spectrum of k^{th} frame of the modulating signal.

The output of spectral subtraction method $a_N(n)$ is obtained from $A_{N,k}(\omega)$ by using the overlap-add method. Each frame of $a_{N,k}(n)$ is calculated as the inverse Fast Fourier Transform (FFT) of $A_{N,k}(\omega)$. These frames are then rearranged after each other with respect to the frame length *L* and frame step size *D*. After that the signal $a_N(n)$ is calculated by adding the signal values where the frames overlap [4].

The signal $b_N(n)$ is the product of $a_N(n)$ and the carrier signal, $c_N(n)$. Similarly, this whole process is applied on each sub-band and after synthesizing signals at output we achieve enhanced speech signal $Y(n)$.

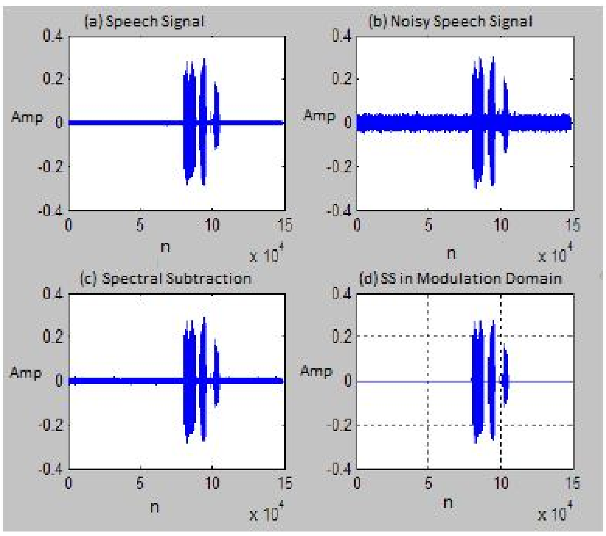

Fig 2: The enhancement effects.

4.2 Model Implementation and Results

Our model is implemented in MATLAB. For evaluation purposes a male's speech "The world is beautiful", recorded with little noise and the sampling frequency of 16 kHz is used. In Fig. 2(a), the original speech signal is depicted. The noise presented in the signal is not much so there is no need for enhancement at this stage. AWGN (with SNR=0dB) is added in the signal to make the scenario more practical. In Fig. 2(b), the speech signal contaminated with noise is represented. Then conventional spectral subtraction method is applied on this noisy speech signal and the output signal is shown in Fig. 2(c). Now our proposed spectral subtraction speech enhancement method in modulation domain is applied on the noisy speech signal and the result is shown in Fig. 2(d).

The validation of our model is firstly done by listening tests. Seven listeners participated in this experiment. The listening tests were conducted in a quiet room. Also in Fig. 2(c), the enhanced signal with conventional spectral subtraction is shown. It can be seen from the figure that it still suffers from undesirable musical noise. Musical noise is still present in the signal because modulating signal was not detected by using IF technique. But in our case, IF is used to detect modulating signal and then spectral subtraction method is applied in modulation domain. After listening tests, different listeners confirmed that our proposed method in modulation domain completely eliminates musical noise from the signal. This can also be verified by the signal represented in Fig. 2(d).

For quantitative purpose, the performance is measured through the Signal to Noise Ratio (SNR) improvement. The results were taken by varying number of sub bands. Numbers of sub-bands used are 2, 4, 8 and 16. In Fig. 3, SNR values for each band with conventional spectral subtraction method are shown. Improved values of SNR with our proposed method are depicted in Fig. 4. It can be seen from these figures that SNR value is around 12 dB for conventional spectral subtraction method where as spectral subtraction method in modulation domain provides SNR around 14 dB. There is an improvement of 2 dB in each sub-band's SNR.

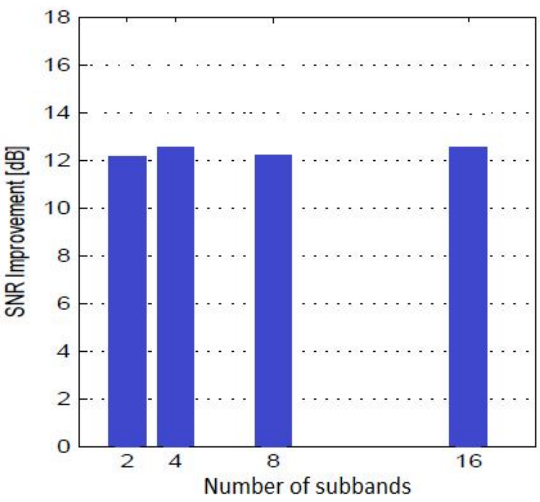

Fig 3: SNR without improvement

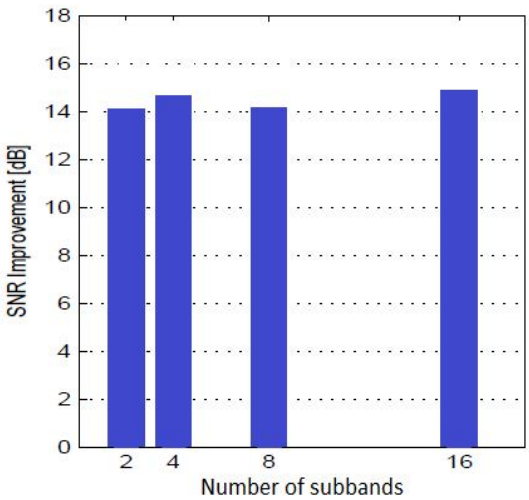

Fig 4: SNR after Improvement

4. Conclusions

In this paper, we proposed spectral subtraction method in modulation domain to enhance noisy speech with AWGN. Fig. 2(d) demonstrates that the modulation domain processing is useful when IF is used for the detection of modulating signal. Also the problem of musical noise is removed by using this technique.

The problem observed in the results is that a little amount of background noise is still present which requires to be improved. Though this noise does not effect much the intelligibility of speech but can be further improved.

References

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, pp. 113–120, 1997.

[2] R. Ishaq, "Adaptive gain equalizer and modulation frequency domain for noise reduction," Master's thesis, Dept. Signal Processing, BTH, Karlskrona, Sweden, 2010.

[3] T. K. Kumar, "Optimum cascaded design for speech enhancement using kalman filters", *WASET*, vol. 41, pp. 355-359, May 2008.

[4] T. S. Gunawan, "Audio compression and speech enhancement using temporal masking models," Ph.D. dissertation, Dept. Elect. Eng., New South Wales Univ., Sydney, NSW, 2007.

[5] L. Atlas, P. Clark, and S. M. Schimmel, "Modulation toolbox for matlab,"Available: http://isdl.ee.washington.edu/projects/modulationtoolbox/

Muhammad Tariq Sadiq was born in Lahore, Pakistan, in 1985. He recieved his BS Electrical Engineering degree from COMSATS Institute of Information Technology (CIIT), Lahore, Pakistan, in 2009 and MS Electrical Engineering degree from Blekinge Institute of Technology, Sweden in 2011. He is currently working as a Lecturer & Project Manager in the Dept. of Electrical Engineering., Sharif College of Engineering & Technology (Affiliated with UET Lahore) Lahore, Pakistan. His research interests are in the field of cognitive radio, speech processing and wireless networks.

Noman Shabbir was born in Lahore, Pakistan in 1985. He got his BS Computer Engineering degree from COMSATS Institute of Information Technology (CIIT), Lahore, Pakistan, in 2007 and MS Electrical Engineering (Radio Comm.) degree from Blekinge Institute of Technology, Sweden in 2009. His project on Unmanned Arial Vehicles (UAV) got 3rd position in a National competition in 2006. He is currently working as a lecturer in the Dept of Electrical Engineering., GC University, Lahore, Pakistan. His research interests are in the field of wireless networks and computer networks.

Wlodek J. Kulesza received the M.Sc. and the Ph.D. degrees from Lodz University of Technology, Poland, and a docent degree from Linkoping University, Sweden. In 2001 he became Professor in Measurement Science at the University of Kalmar, Sweden. Since 2005 he has held a Professor position at the School of Engineering in the Blekinge Institute of Technology (BTH), Karlskrona, Sweden. His current research interests are multi-sensor systems and wireless sensor networks. IEEE member since 1995.