ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 212- 215 (2015)

Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati

http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp42.pdf

Multiband Spectral Subtraction for Electrolaryngeal Speech Processing

Prashant Kumbar[1], Dr. P. N. Kulkarni[2]

[1]Department of Electronics and Communication Engineering, Basaveshwar Engineering College Bagalkot(Autonomous), India, prashant.k.dce@gmail.com,

[2]Department of Electronics and Communication Engineering, Basaveshwar Engineering College Bagalkot(Autonomous), India, pnk_bewoor@yahoo.com

ABSTRACT

Multiband spectral subtraction approach is an useful noise reduction technique while dealing with the electrolaryngeal speech signals which most often contain colored noise. In this study we use multiband spectral subtraction algorithm for reducing colored noise produced by Electrolarynx device with different tweaking factors for each band of frequencies. The results obtained showed significant improvements in speech quality. The algorithm was evaluated using MOS test and PESQ test for speech quality.

Keywords: *Colored noise, Electrolarynx(EL), Multiband spectral subtraction, PESQ, Spectral subtraction.*

1. INTRODUCTION

Larynx is an important organ to produce sound in normal human beings. The persons who undergo laryngectomies due to some larynx related diseases have their larynx completely removed or in other case the persons who don't have larynx from birth will not be able to articulate due to disabled vocal function. Electrolarynx device can be a better substitute for regaining their vocal function partially and effectively.

Electrolarynx is an electronic, battery powered device. It can be fitted in the throat of the subject or can be used as a handheld device. Unfortunately, the Electrolarynx device suffers from some serious disadvantages such as directly radiated noise from the device itself, additive environmental noise and improper pitch adjustments.

The noise produced from Electrolarynx is colored in nature which affects the speech spectrum non-uniformly. Hence, the conventional spectral subtraction method proves improper because in normal spectral subtraction an estimated noise spectrum is subtracted from the entire speech spectrum which degrades the original speech. In this paper we propose an improved multiband spectral subtraction algorithm [1][2] which effectively reduces the colored noise in the speech produced by the electrolarynx device.

In section II we discuss the proposed method, section III presents the implementation details, and in section IV results are shown. Section V gives summary, future work and conclusion.

2. METHOD

The multiband spectral subtraction approach assumes that the additive noise is stationary and uncorrelated with the clean speech. The expression for noisy speech is given as,

$$y(n) = s(n) + d(n) \tag{1}$$

Where,

y(n) – Noisy speech

s(n) – Clean speech

d(n) – Noise

The power spectrum of noisy speech could be approximated as,

$$|Y(w)|^2 = |S(w)|^2 + |D(w)|^2 \tag{2}$$

Where $|Y(w)|^2$ is noisy speech short-time power spectrum, $|S(w)|^2$ is clean speech short-time power spectrum, $|D(w)|^2$ is noise power spectrum estimate

There are some varied spectral subtraction methods proposed by many others which took some extent of flexibility to modify the popular spectral

ISSN 2347 - 3983
International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 212- 215 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp42.pdf

subtraction method of Boll [3]. Some of the modified spectral subtraction algorithms are in [4][5][6].

Due the fact that the exact noise spectrum could not be calculated directly, its estimate D(k) is calculated during the silent periods. In the implementation proposed by Berouti *et.al.*[6], the estimate of the clean speech spectrum is obtained as:

$$|\hat{S}(k)|^2 = |Y(k)|^2 - \alpha|D(k)|^2 \tag{3}$$

Where α is the over-subtraction factor.

Taking into account that the colored noise affects the speech nonuniformly at different frequency bands. The estimate of clean speech at the i^{th} band is obtained as:

$$|\hat{S}_i(k)|^2 = |Y_i(k)|^2 - \alpha_i\delta_i|D_i(k)|^2 \tag{4}$$

where α_i and δ_i are the over-subtraction factor and tweaking factor of i^{th} band, that can be individually set for each frequency band to customize the noise removal properties.

The band specific over-subtraction factor α_i is a function of the segmental SNR_i of the 'i'th frequency band which subtracts an over-estimate of the noise in that particular frequency band. It is calculated as:

$$SNR_i(dB) = 10\log_{10} \frac{\sum_{k=b_i}^{e_i} |Y_i(k)|^2}{\sum_{k=b_i}^{e_i} |\hat{D}_i(k)|^2} \tag{5}$$

Using the SNR_i value calculated in Eq. (5) in consistent with Kamath et.al [5], α_i can be determined as:

$$\alpha_i = \begin{cases} 5 & SNRi < -5 \\ 4 - 3/20\,(SNRi) & -5 \le SNRi \le 20 \\ 1 & SNRi > 20 \end{cases} \tag{6}$$

The tweaking factor δ_i provides an additional degree of control over each frequency band [5]. The values of δ_i were determined empirically in consistent with Sheng Li [4], as:

$$\delta_i = \begin{cases} 1 & 60Hz \le fi \le 300Hz \\ 1.3 & 0.3kHz \le fi \le 1kHz \\ 1.6 & 1kHz \le fi \le 2kHz \\ 1.8 & 2kHz \le fi \le 3kHz \\ 1.3 & 3kHz \le fi \le 5kHz \end{cases} \tag{7}$$

3. IMPLEMENTATION

Hamming window of length 20ms, with an overlap of 10ms between each frame was applied to the speech signal. Then the Fast Fourier Transform (FFT) of windowed speech is calculated and smoothed as per [7] and noisy spectrum is extracted with a posterior and a priori SNR values which are calculated as in equation (5). Empirically determined δ_i values are applied. The noise spectrum is subtracted from corrupted speech spectrum.

The enhanced speech spectrum in each frame is combined together and inverse Fourier transform is applied to obtain enhanced clean speech spectrum. Standard overlap-add technique [5] is used to obtain enhanced speech.

4. RESULTS

The speech signal is arbitrarily divided into five frequency bands as 0~300Hz (Band 1), 0.3k~1kHz (Band 2), 1k~2kHz (Band 3), 2k~3kHz (Band 4), 3k~5kHz (Band 5). Spectral subtraction is performed over each band with different α_i and δ_i values as in equations (6) and (7) respectively. The results are shown below.

The speech signal used in this experiment is "|A||I||U|" i.e. only vowel sounds are used in this experiment.

Figure 1 shows the spectrogram of clean speech recorded from a normal person. Figure 2 show the spectrogram of noised speech and figure 3 shows combined enhanced speech.

ISSN 2347 - 3983
International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 212- 215 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp42.pdf

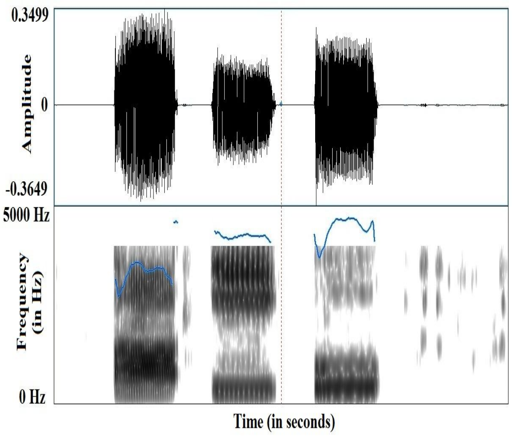

Figure 1: Spectrogram of clean speech “|a||i||u|”, sampling frequency 8kHz and 16 bit quantization. Duration is 3s

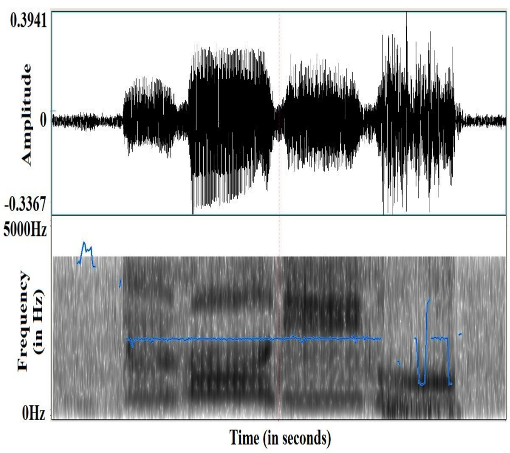

Figure 2: Spectrogram of noised speech “|a||i||u|”, sampling frequency 8kHz and 16 bit quantization. Duration is 3s.

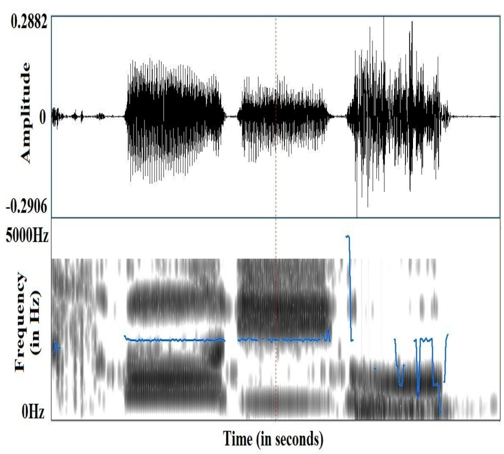

Figure 3: Spectrogram of enhanced speech “|a||i||u|”, sampling frequency 8kHz and 16 bit quantization. Duration is 3s.

Mean PESQ scores are calculated as per [8] for objective analysis of enhanced speech quality at different SNR values (-6db, -3db, 0db, 3db, 6db) and is given in table 1.

Perceptual assessment test was conducted based on the results of improvement in the noisy speech over ten subjects, they were asked to rate the improvement in the perceived quality on (0-5) with ‘0’ as the worst and 5 as the best for speech at different SNR values. The ratings given by 10 subjects are given in the Table 2 with mean and

SNR in db	PESQ Score Processed speech Vs Unprocessed speech (EL Speech)	PESQ Score Processed speech Vs normal speech
-6	1.17	2.54
-3	1.16	2.49
0	1.13	1.73
3	1.17	1.68
6	1.22	1.7
At Infinite SNR	2.86	2

standard deviation.

Table 1: PESQ scores for objective analysis of enhanced speech quality.

In Table 1, column 2 gives PESQ scores of processed speech versus unprocessed speech and column 3 depicts PESQ scores of clean speech versus processed speech. These PESQ scores are calculated using Philipos C. Loizou method hence for detailed explanation of PESQ calculation algorithm refer [8].

5. CONCLUSION

Experimental results show that the multi-band spectral subtraction approach enhances the noised speech effectively. We believe that this effective noise reduction is because multiband spectral subtraction algorithm reduces the colored or musical noise and the spectral degradation of clean speech can be avoided. We found that 5 frequency

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 212- 215 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp42.pdf

bands are sufficient to obtain good results. Subjective analysis was done by perceptual assessment tests over 10 listeners and objective analysis was carried out by writing MATLAB code for PESQ analysis based on [8].

Table 2: Perceptual Assessment test for speech quality using Mean Opinion Score (MOS) (Ratings out of 5)

Subject	**Unprocessed speech at different SNR in dB**						**Processed speech at different SNR in dB**					
	∞	**-6**	**-3**	**0**	**3**	**6**	**∞**	**-6**	**-3**	**0**	**3**	**6**
1	3	1.5	2	1	2.5	1.5	3.5	2	2.5	1.5	3.5	3
2	3.5	1.5	2	1	2	1.5	3	1.5	2.5	1	4	3
3	4	1.5	2.5	1	2	2	4	2	2.5	1	4	3
4	4	2	2	1	2.5	1.5	4	2	3	1.5	4	3
5	4	1.5	2	1	2	1.5	3.5	2	3	1	4	3.5
6	3.5	2	2.5	1	2.5	1.5	4	2	3.5	1	4	3
7	4	1.5	2.5	1.5	2.5	2	4	2.5	3	1	3	2.5
8	4	1	2	1	2.5	1.5	4.5	2	3	1	4.5	3
9	4	1.5	2	1.5	2	2	4	2	3	1.5	4	3
10	3.5	1.5	2	1	2	2	4.5	2	2.5	1	4.5	3.5
Mean	**3.47**	**1.43**	**1.99**	**1.02**	**2.09**	**1.54**	**3.54**	**1.84**	**2.65**	**1.07**	**3.59**	**2.77**
Standard Deviation	**0.98**	**0.46**	**0.57**	**0.34**	**0.61**	**0.44**	**0.96**	**0.53**	**0.77**	**0.37**	**0.97**	**0.72**

6. REFERENCES

1. L. R. Rabiner, R.W. Schafer. ***Digital Processing of Speech Signals***, second edition, 2005.
2. Thomas. F. Quatieri. ***Discrete-Time Speech Signal Processing Principles and Practice***, Third edition, 2007.
3. S.F. Boll, **Suppression of acoustic noise in speech using spectral subtraction**, *IEEE Trans. Acoust., Speech, Signal Process., vol.27, pp. 113-120, Apr. 1979.*
4. Sheng Li, MingXi Wan, and SuPin Wang. **Multi-Band Spectral Subtraction Method for Electrolarynx Speech Enhancement Algorithms,** *2009, 2, 550-564;doi:10.3390/a2010550.www.mdpi.com/journal/algorithms*
5. Sunil D. Kamath and Philipos C. Loizou, **A Multiband Spectral Subtraction Method for Enhancing Speech Corrupted by Colored Noise**, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 2002.*
6. Berouti. M, Schwartz R, Makhoul J. **Enhancement of speech corrupted by acoustic noise**. *IEEE International Conference on Acoustics, Speech, and Signal Processing 1979, 208-211*.
7. L. Arslan, A. McCree and V. Viswanathan, **New methods for adaptive noise suppression**, *ICASSP, vol.1, pp. 812-815, May 1995*.
8. Yi Hu and Philipos C. Loizou, **Perceptual evaluation of speech quality (PESQ), andobjective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs**. *ITU-T Recommendation P. 862. ITU (2000).*