Y Prasanthi, K Jyothi, GIET / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue4, July-August 2012, pp.2162-2165

Frequency Depended Spectral Subtraction For Speech Enhancement

Y Prasanthi, K Jyothi, GIET

Abstract:

The corruption of speech due to presence of additive background noise causes severe difficulties in various communications environments. This paper addresses the problem of reduction of additive background noise in speech. The proposed approach is a frequency dependent speech enhancement method based on the proven spectral subtraction method. Most implementations and variations of the basic spectral subtraction technique advocate subtraction of the noise spectrum estimate over the entire speech spectrum. However real world noise is mostly colored and does not affect the speech signal uniformly over the entire spectrum. In this paper, new spectral subtraction method proposed which takes into account the fact that colored noise affects the speech spectrum differently at various frequencies. This method provides a greater degree of flexibility and control on the noise subtraction levels that reduces artifacts in the enhanced speech, resulting in improved speech quality. This method outperforms the conventional spectral subtraction method with respect to speech quality and reduced musical noise.

1. INTRODUCTION

A major part of the interaction between humans takes place via speech communication.

Hence, research in speech and hearing sciences has been going on for centuries to understand the dynamics and processes involved in the production and perception of speech. The field of speech processing is essentially an application of signal processing techniques to acoustic signals using the knowledge offered by researchers in the field of hearing sciences. The explosive advances in recent years in the field of digital computing have provided a tremendous boost to the field of speech processing. Digital signal processing techniques are more sophisticated and advanced as compared to their analog counterparts. Ease and speed of representing, storing, retrieving and processing speech data has contributed to the development of efficient and effective speech processing techniques to address the issues related to speech. The presence of background noise in speech significantly reduces the intelligibility of speech. Degradation of speech severely affects the ability of person, whether impaired or normal hearing, To understand what the speaker is saying. Noise reduction or speech enhancement algorithms are used to suppress such background noise and improve the perceptual quality and intelligibility of speech. Even though speech is perceptible in a moderately noisy environment, many applications like mobile communications, speech recognition and aids for the hearing handicapped, to name a few, drive the effort to build more effective noise reduction algorithms for better performance. Over the years engineers have developed a variety of theoretical and relatively effective techniques to combat this issue. However, the problem of cleaning noisy speech still poses a challenge to the area of signal processing. Removing various types of noise is difficult due to the random nature of the noise and the inherent complexities of speech. Noise reduction techniques usually have a trade off between the amount of noise removal and speech distortions introduced due the processing of the speech signal. Complexity and ease of implementation of the noise reduction algorithms is also of concern in applications especially those related to portable devices such as mobile communications and digital hearing aids. The spectral subtraction method is a well-known noise reduction technique. Most implementations and variations of the basic technique advocate subtraction of the noise spectrum estimate over the entire speech spectrum. However, real world noise is mostly colored and does not affect the speech signal uniformly over the entire spectrum. In this paper, we propose a multi-band spectral subtraction approach that takes into account the fact that colored noise affects the speech spectrum differently at various frequencies. This method outperforms the standard power spectral subtraction method resulting in superior speech quality and largely reduced musical noise.

1.2 Short-term Spectral Amplitude Techniques

The short-term spectral amplitude (STSA) of speech has been exploited successfully in the development of various speech enhancement algorithms. The basic idea is to use the STSA of the noisy speech input and recover an estimate of the clean STSA by removing the part contributed by the additive noise. Expressed as

$$|\hat{S}(k)|^2 = |Y(k)|^2 - \alpha|\hat{D}(k)|^2 \quad (1)$$

Where S(k) and Y(k) are magnitude spectrum of clean speech and the noisy speech. Estimated noise magnitude $|\hat{D}(k)|$ is calculated during periods of speech, α is an over-subtraction factor.
A general representation of STSA technique is given in Figure 1.

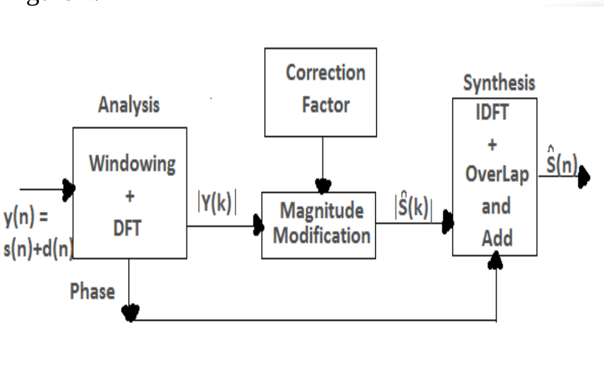

Figure 1: Diagrammatic Representation of the short-time Spectral Magnitude Enhancement System

The input to the system is the noise-corrupted signal y(n) . While there are many methods for the analysis-synthesis processing, the Short-term Fourier Transform (STFT) of the signal with OverLap and Add (OLA) is the most commonly used method. Spectral subtraction is a well-known noise reduction method based on the STSA estimation technique. The basic power spectral subtraction technique, as proposed by Boll, is popular due to its simple underlying concept and its effectiveness in enhancing speech degraded by additive noise but drawback of the power spectral subtraction technique is residual noise.

2. FREQUENCY DEPENDED SPECTRAL SUBTRACTION

Along with the actual noise suppression operation, some pre-processing methods are needed to achieving good speech quality. In frequency depended spectral subtraction, the speech spectrum is divided into non-overlapping bands, and spectral subtraction is performed independently in each band. The fact that colored noise affects the speech spectrum differently at various frequencies.

$$|\hat{S}_i(k)|^2 = |Y_i(k)|^2 - \alpha_i\delta_i|\hat{D}_i(k)|^2 \quad b_i \le k \le e_i \quad (2)$$

Where bi and ei are beginning and ending frequency bins of the ith frequency band.δi is tweaking factor that can be individually set for each band to customize the noise removal properties.The negative values in the enhanced spectrum were floored to the noisy spectrum as

$$|\hat{S}_i(k)|^2 = \begin{cases} |\hat{S}_i(k)|^2 & |\hat{S}_i(k)|^2 > 0 \\ \beta|Y_i(k)|^2 & else \end{cases} \quad (3)$$

Where The spectral floor parameter was set to β=0.001.

3. IMPLEMENTATION AND PERFORMANCE EVALUTION

3.1 Objective Measures for Performance Evaluation

In the evaluation of speech enhancement algorithm, it is necessary to identify the similarities and differences in perceived quality and subjectively measured intelligibility. Speech quality is an indicator of the naturalness of the processed speech signal. Intelligibility of speech signals is a measure of the amount of speech information present in the signal that is responsible for conveying what that speaker is saying. The interrelationship between perceived speech and intelligibility is not clearly understood. Performance evaluation tests can be done by subjective quality measures or objective quality measures. While subjective measures provide a broad measure of performance since a large difference in quality is necessary to be distinguishable to the listener. Hence, it becomes difficult to get a reliable measure of changes due to algorithm parameters. Objective measures, on the other hand, provide a measure that can be easily implemented and reliably reproduced.

It is necessary to conduct off-line simulations to check the validity and feasibility of an algorithm before it can be implemented on a real-time system. Implementation on a workstation permits modifications and changes to the algorithm without constraints of time, memory or computational power. The simulations were carried out on an using Matlab, a technical computing software. The speech signal is first Hamming windowed using a 20-ms window and a 10-ms overlap between frames. The windowed speech frame is then analyzed using the Fast Fourier Transform (FFT). Smoothing of the magnitude spectrum was found to reduce the variance of the speech spectrum and contribute to the enhancement in speech quality. A weighted spectral average is taken over preceding and succeeding frames of data as given by Equation.

$$\overline{Y_j(k)} = \sum_{i=-M}^{M} W_i Y_{j-i}(k) \quad (4)$$

Where j is the frame index 0<Wi<1. The averaging is done over M preceding and succeeding frames of speech. The number of frames M is limited to 2 to prevent smearing of the speech spectral content. The filter weights Wi were empirically determined and set to W= [0.09, 0.25, 0.32, 0.25, 0.09]

$$\delta_i = \begin{cases} 1 & f_i \le 1\text{kHz} \\ 2.5 & 1\text{kHz} < f_i \le \frac{Fs}{2} - 2\text{kHz} \\ 1.5 & f_i > \frac{Fs}{2} - 2\text{kHz} \end{cases} \quad (5)$$

Where fi is the upper frequency of the ith band, and Fs is the sampling frequency in Hz. The motivation for using smaller δi values for the low frequency bands is to minimize speech distortion, since most of the speech energy is present in the lower frequencies. Relaxed subtraction is also used for the high frequency bands. Subtraction is performed over each band as indicated in equation and the negative values are rectified using the spectral floor. A small amount of the original noisy spectrum can be introduced back into the enhanced spectrum to mask any remaining musical noise. In this implementation, 5% of the original noisy spectrum within each band is combined, and the enhanced signal is obtained by taking the IFFT of the enhanced spectrum using the phase of the original noisy spectrum. Finally, the standard overlap-and-add method is used to obtain the enhanced signal.

3.2 Subjective Evaluation of Speech Intelligibility

Subjective tests are conducted by having some human subjects listen to the prepared test speech files and evaluate based on some criteria. Intelligibility test were carried out at the Callier Center for Communication Disorders/UTD on seven hearing-impaired subjects with severe to profound hearing loss. Speech enhanced by the MBSS algorithm with four linearly spaced frequency bands was evaluated against the noisy speech. The sentences were corrupted using speech-shaped noise at 0 dB SNR. The noise-corrupted sentences were played in a random order through speakers in a sound insulated booth. The subjects were asked to repeat the sentence they heard. Intelligibility was measured in terms of percentage of words correct. Figure 2 gives the bar plots of the intelligibility scores achieved by each subject and the average score for both the test conditions. The results obtained by objective evaluation are an indicator of the best speech quality that can be obtained by the different configurations of the algorithm. The proposed multi-band spectral subtraction approach(with number of bands>3) performed better than the PSS approach for both SNRs.

4. RESULTS

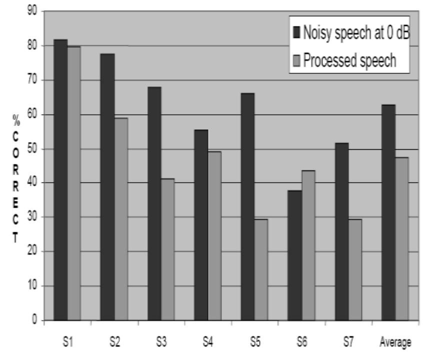

Figure 2: Intelligibility Test Results for Seven Subjects scored on Percentage Words Correct

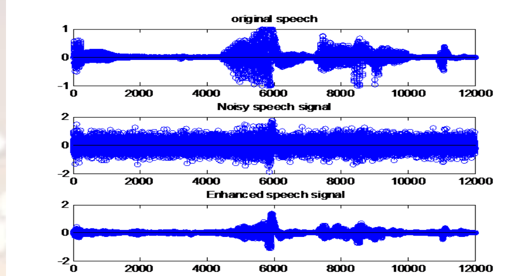

Figure 3: signal representation of the sentence "I am David". The top representation is the original signal, the middle representation is corrupted signal, and the bottom representation is the enhanced signal obtained by the frequency depended spectral subtraction method

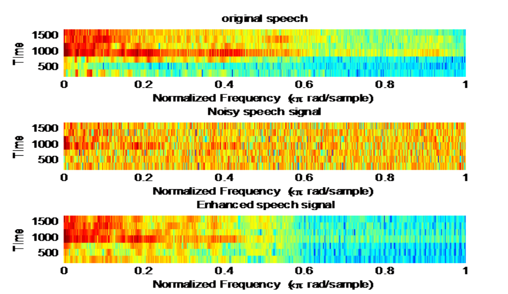

Figure 4: Spectrogram of the sentence "I am David". The top spectrogram is the original signal, the middle spectrogram is the corrupted signal, the bottom spectrogram is the enhanced signal obtained by frequency depended spectral subtraction method.

5. CONCLUSIONS

The work in this paper addressed the problem of enhancing speech in noisy conditions. A multi-band spectral subtraction method, based on the direct estimation of the short-term spectral amplitude of speech and the non-uniform effect of noise on speech. The results establish the superiority of the proposed method over the conventional spectral subtraction method with respect to speech quality of the enhanced signal and reduced residual noise.

The major contributions of this paper are

(a) Development of a multi-band speech enhancement strategy based on the spectral subtraction method. Speech processed by the new algorithm shows reduced levels of residual noise and good speech quality.

(b) Proposed a band subtraction factor that provides greater control over the subtraction process in each band and can be tweaked to minimize speech distortion.

(c) Evaluation of various pre-processing strategies for improving the output speech quality. It was shown that spectral smoothing and weighted spectral averaging of the input speech spectrum helped preserve the speech content and improved speech quality.

Reference/Bibliography

[1] L. Arslan, A. McCree and V. Viswanathan, " New methods for adaptive noise suppression," ICASSP, vol.1, pp. 812-815, May 1995.

[2] M. Berouti, R. Schwartz and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," Proc. IEEE Int. Conf. on Acoust., Speech, Signal Procs., pp. 208-211, Apr. 1979.

[3] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process., vol.27, pp. 113-120, Apr. 1979.

[4] Y. Cheng and D. O'Shaughnessy, "Speech enhancement based conceptually on auditory evidence," ICASSP, vol.2, pp. 961-964, Apr. 1991.

[5] J. Deller Jr., J. Hansen and J. Proakis, "Discrete-Time Processing of Speech Signals", NY: IEEE Press, 2000.

[6] Y. Ephraim, "Statistical-model-based speech enhancement systems," Proc. IEEE,80, No.10, pp. 1526-1555, Oct.1992.

[7] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-term spectral amplitude estimator," IEEE Trans. on Acoust.,Speech,Signal Proc., vol.ASSP-32, No.6, pp. 1109-1121, Dec.1984.

[8] Y. Ephraim and H. Van Trees, "A signal subspace approach for speech enhancement," IEEE Trans. Speech Audio Procs., vol. 3, pp. 251-266, Jul. 1995.

[9] Z. Goh, K.Tan and T. Tan, "Postprocessing method for suppressing musical noise generated by spectral subtraction," IEEE Trans. Speech Audio Procs., vol. 6, pp. 287-292, May 1998.

[10] J. Hansen and B. Pellom, "An effective quality evaluation protocol for speech enhancements algorithms," Inter. Conf. on Spoken Language Processing, vol.7, pp. 2819-2822, Sydney, Australia, Dec.1998.

[11] C. He and G. Zweig, "Adaptive two-band spectral subtraction with multi-window spectral estimation," ICASSP, vol.2, pp. 793-796, 1999.

[12] Y. Hu, M. Bhatnagar and P. Loizou, "A cross-correlation technique for enhancing speech corrupted with correlated noise," ICASSP, vol. 1, pp. 673-676, 2001.

[13] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," submitted to ICASSP 2002.

[14] P. Kasthuri, "Multichannel speech enhancement for a digital programmable hearing aid," Master thesis, University of New Mexico, 1999.

[15] W. Kim, S. Kang and H. Ko, "Spectral subtraction based on phonetic dependency and masking effects," Proc. IEEE Vis. Image Signal Procs., vol. 147, No. 5, Oct. 2000.

[16] H. Levitt, "Noise reduction in hearing aids: An overview", Journal of Rehabilitation Research and Development, vol. 38,No.1 January/February 2001.

[17] J. Lim and A. Oppenheim, "All-pole modeling of degraded speech," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 26, No. 3, pp. 197-210, June 1978.

[18] J. Lim and A. Oppenheim, "Enhancement and bandwidth compression of noisy speech," Proc. IEEE, vol. 67, No. 12, pp. 221-239, Dec. 1979