Speech Enhancement using Critical Band Spectral Subtraction

L Singh[1] and S Sridharan[2]

Speech Research Laboratory
School of EESE
Queensland University of Technology
GPO Box 2434, Brisbane, QLD, 4001, Australia

[1]l.singh@qut.edu.au, [2]s.sridharan@qut.edu.au,
Tel: (+617) 3864 2113, Fax: (+617) 3864 1516

ABSTRACT

This paper proposes a new enhancement technique for the enhancement of broadband noise corrupted speech. The technique exploits the human auditory systems inability to distinguish between individual frequency components within critical frequency bands. Spectral subtraction is used and the spectrum is considered as critical frequency bands rather than individual frequency components. The proposed technique is compared with the existing spectral subtraction technique, using both subjective and objective speech assessment measures. Results are quoted and indicate that there is a significant increase in intelligibility and quality.

1.0 INTRODUCTION

Speech enhancement is a very difficult area of speech processing due to the fact that in many cases we are dealing with two non-stationary signals, of unknown distribution. We suggest a new technique that achieves improvements in intelligibility and quality by exploiting the deficiencies of the human auditory system and incorporating it into an existing spectral subtraction enhancement technique.

Critical band analysis has been used previously in speech enhancement (Virag 1995) for determining a noise masking threshold that was then used to modify a spectral noise estimate. The spectral noise estimate was modified so that only the audible components of the noise were suppressed. These approaches were extremely dependent on an initial clean speech estimate that was then used in the calculation of the noise masking threshold. Obtaining a good estimate of the clean speech can be difficult in low SNR cases. The technique proposed in this paper has been termed "critical band spectral subtraction". It still uses spectral subtraction for noise suppression but the noise that is suppressed has been calculated differently compared to previous techniques. The noise is estimated using a critical frequency band approach. Critical band analysis is employed because frequencies within the same critical band are perceived equally be the human auditory system (Ambikairajah 1997), (Virag 1995) and (Fletcher 1940). This knowledge is applied to the spectral noise estimate by forming a noise estimate for each critical band rather than for each individual frequency component. The noise estimate for each critical band is then used as the noise estimate for all the frequencies within the same critical band. This particular approach is effectively suppressing the noise spectral envelope rather than trying to suppress individual noise frequency components. We show that the proposed technique results in better performance when compared to previous work.

2.0 CRITICAL BAND SPECTRAL SUBTRACTION

The overall approach of the critical band spectral subtraction technique is still a frequency domain noise estimation and noise subtraction process as initially proposed by (Boll 1979). Previous broadband noise suppression techniques estimated the noise spectrum as individual frequency components. The proposed critical band spectral subtraction enhancement technique considers the frequency components of the noise spectrum as being grouped into critical frequency bands. Critical band analysis is used in the noise estimation process because the human auditory system can be modelled as a bank of bandpass filters know as critical band filters. The critical frequency bands for speech is shown in Table 1. In terms of perception, the critical

Critical Band No.	Frequency Range (Hz)
1	0 - 100
2	100 - 200
3	200 - 300
4	300 - 400
5	400 - 510
6	510 - 630
7	630 - 770
8	770 - 920
9	920 - 1080
10	1080 - 1270
11	1270 - 1480
12	1480 - 1720
13	1720 - 2000
14	2000 - 2320
15	2320 - 2700
16	2700 - 3150
17	3150 - 3700
18	3700 - 4400

Table 1: Critical Speech Frequency Bands

frequency bands can be considered as single entities within the speech spectrum, each consisting of a number of frequency components. Because of this grouping performed by the human auditory system, there is little perceived difference between noise frequency components within the same critical band. The noise estimate for each critical band is calculated by taking the average of the noise frequency components within the critical band. The spectral noise and the critical band noise estimates are shown in Figure 1.

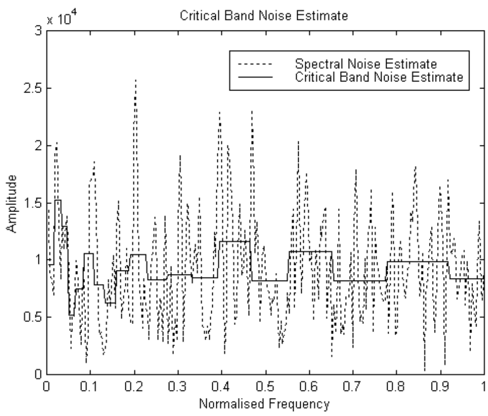

Figure 1: Critical Band Noise Estimate

As can be seen in Figure 1, the bandwiths of the critical bands increase with frequency. The critical band noise estimates are then used as the noise value for all frequency components within the same critical band. With the noise average found across all critical bands, it is then used in the noise suppression stage. Because of the averaging process it is sometimes necessary to use a larger subtraction weight factor. A diagram illustrating the critical band spectral subtraction process is shown in Figure 2.

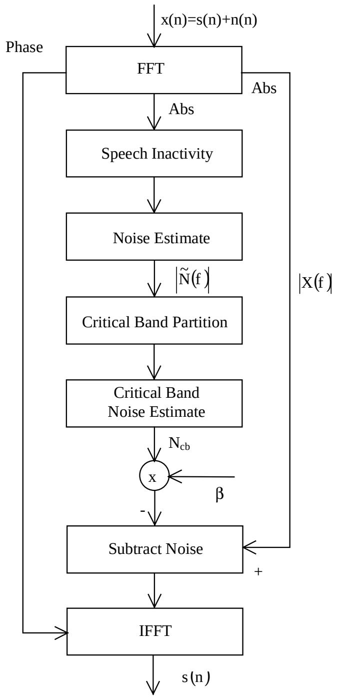

Figure 2: Critical Band Spectral Subtraction System

In Figure 2, $x(n)$ is the broadband noise corrupted speech signal, $\left|\tilde{N}(f)\right|$ is the magnitude spectrum of the noise estimate, $\left|X(f)\right|$ is the magnitude spectrum of the noisy speech signal, N_{cb} is the critical band noise estimate, β is the subtraction weight and $\tilde{s}(n)$ is the estimate clean speech signal. The critical band spectral subtraction method can be summarised by (1) and (2).

$$N_{cb}(i) = \overline{\left|\tilde{N}(f_{low}, i), \ldots, \tilde{N}(f_{high}, i)\right|} \qquad (1)$$

$$\left|\tilde{S}(f, i)\right| = \left|X(f, i)\right| - \beta * N_{cb}(i) \qquad (2)$$

where i is the critical band index (for i = 1, 2, 3, …, 18), $N_{cb}(i)$ is the noise estimate for all frequency components within critical band i, $\tilde{N}(f_{low}, i)$ is the lower frequency index for critical band i, $\tilde{N}(f_{high}, i)$ is the upper frequency index for critical band i, $\left|\tilde{S}(f, i)\right|$ is the magnitude of the estimated clean speech frequency components in critical band i, $\left|X(f, i)\right|$ is the magnitude of the noisy speech frequency components in critical band i and β is the subtraction weight applied to the subtracted noise.

The main advantage that the critical band spectral subtraction technique offer is a reduction in the error that occurs due to estimating individual noise frequency components, through the use of averaging in critical frequency bands.

3.0 RESULTS

The critical band method for noise estimation was incorporated into a standard spectral subtraction technique, as shown in Figure 2. The standard technique was also used for comparison purposes. In evaluating the performance of the critical band spectral subtraction technique, a speech signal was corrupted by various levels of additive white gaussian noise (AWGN). The noisy signal was then enhanced using the proposed critical band spectral subtraction technique and standard spectral subtraction technique. The enhanced signal was then assessed using objective measures (SNR_{seg}) and subjective measures (MOS). The SNR_{seg} was calculated using (3).

$$SNR_{seg} = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\sum_{n=Nm}^{Nm+N-1}\left(\frac{s(n)^2}{\left[s(n)-\tilde{s}(n)\right]^2}\right) \quad (3)$$

where M are the number of frames or segments, N is the length of the frames, $s(n)$ is the original clean speech signal and $\tilde{s}(n)$ is the enhanced speech signal or estimate of the original speech signal. The objective results obtained in evaluating the performance of the critical band spectral subtraction technique are shown in Figure 3. It can be seen from the objective results in Figure 3, that there appears to be very little difference in the output SNR_{seg} of the critical band spectral subtraction technique and the standard spectral subtraction technique. These objective results suggest that the enhancement techniques perform similarly. However MOS give a better indication of the proposed techniques performance. The MOS testing was

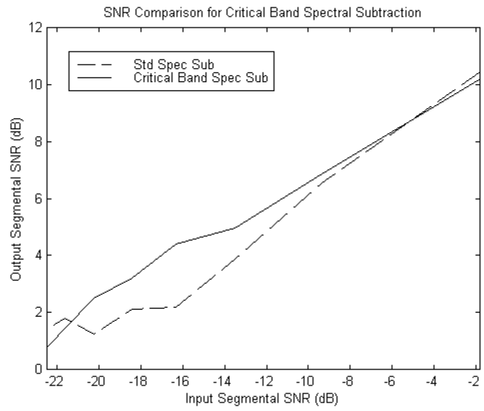

Figure 3: Critical Band Spectral Subtraction SNR Comparison

performed in a controlled environment using headphones as the listening device. Six listeners (five native english speakers) were asked to rate the various enhanced speech signals using a five point rating scale. The subjective results obtained in evaluating the performance of the critical band spectral subtraction technique are shown in Figure 4.

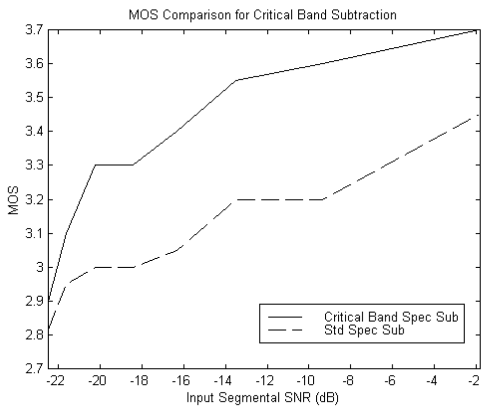

Figure 4: Critical Band Spectral Subtraction MOS Comparison

It can be clearly seen from the subjective results in Figure 4 that the improvement in the MOS is approximately 0.4 of a point over the existing spectral subtraction technique when using a five point MOS scale. The MOS indicate that the critical band spectral subtraction technique produces enhanced speech that is significantly better in quality. The performance of the critical band spectral subtraction technique was also tested on noise corrupted speech signals received over a HF radio. In this case it is only possible to quote the MOS since calculation of SNR_{seg} requires the original clean speech signal. The original received speech signals were also rated to show the relative improvement in the MOS. The subjective results obtained are shown in Table 2.

Type of Enhancement	MOS
Original Received Signal	2.0
Standard Spectral Subtraction	2.4
Critical Band Spectral Subtraction	2.9

Table 2: Critical Band Spectral Subtraction MOS Comparison for HF Radio Signals

As can be seen from the subjective results in Table 2 that were obtained using actual HF radio signals, the improvement over standard spectral subtraction is approximately 0.5 of a point when using a five point MOS scale. Listening tests also indicate that the perceived differences are a reduction in the residual broadband noise due to incorrect noise estimation (which is usually noticeable during speech segment) and a reduction of musical noise which accompanies standard spectral subtraction.

4.0 CONCLUSIONS

In this paper we have proposed a new speech enhancement technique that incorporates knowledge of the human auditory system into an existing spectral subtraction technique. Both objective and subjective results were used to assess the performance of the proposed critical band spectral subtraction technique. The objective results indicate that there is little difference between the SNR_{seg} of the proposed critical band spectral subtraction and the standard spectral subtraction technique. However subjective results show that the proposed technique produces enhanced speech that is better in quality overall, with less residual broadband noise and less introduced musical noise.

In low SNR cases it is often difficult to estimate the interfering noise. Usually when suppressing the noise in a noisy speech signal, residual artefacts are introduced such as musical noise or there is some residual original noise. By exploiting the deficiencies of the human auditory system and applying this to the noise estimate within critical frequency bands through noise averaging, we have gained an increase in performance when compared with the standard spectral subtraction technique.

5.0 REFERENCES

1. Ambikairajah E, Davis A G and Wong W T K, “Auditory Masking and MPEG-1 Audio Compression”, Electronics and Communication Engineering Journal, Volume 9, Number 4, pp 165-175, August 1997.

2. Boll S F, “Suppression of Acoustic Noise in Speech using Spectral Subtraction”, IEEE Transactions on Acoustics, Speech and Signal Processing, Volume 27, Number 2, pp 113-120, April 1979.

3. Fletcher N, “Auditory Patterns”, Revised Modern Physics, Volume 1, Number 2, pp 47-65, 1940.

4. Virag N, “Speech Enhancement Based on Masking Properties of the Auditory System”, Proceedings of the International Conference on Acoustics, Speech and Signal Processing 1995 (ICASSP-95), pp 796-799, 1995.