The Effect of Sampling Frequency on a FFT Based Spectral Estimator

Saeed Ayat

The Effect of Sampling Frequency on a FFT Based Spectral Estimator

Saeed Ayat

Payame Noor University, Najafabad, Iran.
dr.ayat@pnu.ac.ir

Abstract:

This paper reviews the effect of sampling frequency on a FFT-based spectral estimator. In signal processing applications usually a fix window size is used for obtaining the current frame spectral.
For an application like speech enhancement this accuracy of this estimation has a great influence in the quality of the system, because listener feeling is very important in this subject.
In our proposed method we divided the well-known spectral subtraction method in two phases. Then by using different frame sizes that we used in these two phases the overall quality of the system has increased in different sampling frequencies.

1. Introduction

One of the first methods introduced for speech enhancement is spectral subtraction. Till now, different versions of spectral subtraction have been proposed to increase the performance of this method, for example [1, 2, 3].
Despite of its high noise removal, it can cause an annoying noise called musical noise and hence it can reduce overall quality. Musical noise is produced because, we don't have the needed spectra exactly, so we have to use their estimations.
In signal processing applications usually a fix window size is used for obtaining the current frame spectral. As we know if the frame length is L the frequency resolution in Fourier spectral analysis is Fs/L. For example if F_s=11025Hz and L=256 then Fs/L is 43Hz and this resolution may not be enough for speech signal.
As we know a clean speech signal consists of some sections that have speech and some others that have no speech and we call them silences.
In a noisy speech signal these silence sections have only noise and other sections have noisy speech signals. If the noise is stationary we can estimate its spectrum in the noise sections.
In spectral subtraction method, after framing the noisy speech signal we use a silence detector or a voice activity detector for separating noisy speech frames and noise frames.
After that with applying, FFT we have the spectrum of each frame. By calculating the average of the noise frames spectra we have estimation for noise spectra. Now with subtracting this estimation of noise spectrum from the spectrum of each noisy speech frames we can achieve enhanced speech signal.
The paper is organized as follows: In section 2 we have a review on spectral subtraction method. In section 3 we proposed our method and in section 4 we present the simulation results.

2. Spectral Subtraction

There are many different versions for spectral subtraction. In a generalized spectral subtraction [4] we have:

$$\left|\hat{S}(w)\right| = \max\left\{\left(\left|S(w)\right|^{\alpha} - \beta\left|N(w)\right|^{\alpha}\right)^{\frac{1}{\alpha}}, \gamma\left|N(w)\right|\right\} \quad (1)$$

Where $\left|S(w)\right|, \left|N(w)\right|$ and $\left|\hat{S}(w)\right|$ are magnitude spectrum of noisy speech, estimation of noise and enhanced speech. β is the oversubtraction factor and γ is spectral floor. Both β and γ are adjusted to improve the quality of enhanced speech.
By the assuming that the noise is stationary, a good estimation can be resulted by computing the average of the noise in silence frames spectra. We called such average $\left|\overline{W}(w)\right|$.
In presence of nonstationary noises, an adaptation technique can be used. Given an initial value $\left|\overline{W_0}(w)\right|$, if the current frame is silence, $\left|\overline{W_m}(w)\right|$ is updated using this equation:

$$\left|\overline{W_m}(w)\right| = (1-f)\left|\overline{W}_{m-1}(w)\right| + f\left|Y_m(w)\right| \quad (2)$$

In this formula $\left|Y_m(w)\right|$ is the spectrum of current silence frame and f is a coefficient called forgetting factor. This factor is changed depending on the noise changing rate.
The main problem of spectral subtraction method is the production of musical noise. Musical noise is produced because we don't have the exact spectrum of the noise signal.

3. Proposed Method

In our method that estimates the spectrum better than the basic averaging method, after separating speech and silence frames in the noisy signal with a basic analysis frame, we can increase the analysis frame length until it covers all the current silence frames. As in periodogram estimator technique the accuracy improves by increasing the number of signal samples, by using this adaptive analysis frame length we can have a better spectral estimation for noise and noisy signal and so the system can produce a better enhanced signal with less musical noise.

As we know if the frame length is L the frequency resolution in Fourier spectral analysis is Fs/L. For example if F_s=11025Hz and L=256 then Fs/L is 43Hz and this resolution may not be enough for speech signal.

In our method we first apply a SAD algorithm with L=256 and L/2=128 points overlap to detect the silence frames. Now we can increase the analysis frame length until it covers all the current silence frames. By this method we have larger window length and hence better frequency resolution. If we have several silence areas with the new frame length, the average of them is the overall noise spectrum.

By applying such method we have better noise spectrum estimation with less musical noises.

In section 4 we give experimental results that confirm this improvement clearly.

4. Simulation Results

In this section we explain our simulation. The speech signal that used for these tests was chosen from TIMIT data base and was pronounced with a female speaker. Then this sentence converted to different sampling frequencies by cool edit software. All these sentences degraded by additive Gaussian white noise, so we can have the noisy signal in required SNR, here 5dB.

For evaluating our method we calculate SNR improvement as below.

If $s(n)$ is the clean speech, y(n) the noisy, $\hat{s}(n)$ the enhanced signal and w(n) the noise then we have:

$$y(n) = s(n) + w(n) \tag{3}$$

and the SNR improvement is computed as follows[5]:

$$SNR_{imp} = SNR_{out} - SNR_{in} \tag{4}$$

In which SNR_{in} and SNR_{out} are the SNRs for noisy and enhanced:

$$SNR_{in} = 10\log_{10}\frac{\sum s^2(n)}{\sum (y(n)-s(n))^2} \tag{5}$$

$$SNR_{out} = 10\log_{10}\frac{\sum s^2(n)}{\sum (\hat{s}(n)-s(n))^2} \tag{6}$$

In this experiment a listener listens to the enhanced signal and increases β until the musical noise appeares in the enhanced signal. At this point, β and SNR improvement is recorded. This is done for SNR equal to 5dB and different frame lengths with 256, 512, 1024 and 2048 samples.

This test was evaluated for different sampling frequencies equal to 8000Hz, 11025 Hz and 16000Hz. α is fixed to 1.0 and γ to 0.0. Note that the frame length is 256 in silence detection step.

Tables 1 to 3 show the results for β and SNR improvement at the appearance of musical noise in the enhanced signal for tested SNRs.

L	256	512	1024	2048
SNR_{imp}	0.8	1.0	1.4	1.44
β	0.1	0.15	0.25	0.45

Table 1: β and SNR improvement at the start of musical noise (Fs=8000Hz)

L	256	512	1024	2048
SNR_{imp}	1.0	1.8	2.3	3.1
β	0.15	0.3	0.5	0.9

Table 2: β and SNR improvement at the start of musical noise (Fs=11025Hz)

L	256	512	1024	2048
SNR_{imp}	1.2	2.3	3.1	3.8
β	0.2	0.5	0.7	1.0

Table 3: β and SNR improvement at the start of musical noise (Fs=16000Hz)

As we can see the SNR improvement is better for longer frame lengths in all different sampling frequency rates. This show that the musical noise arises from inaccurate noise estimation and reduces as the frame length increases, and this result is true for different sampling frequencies.

So with a greater frame length, we can choose a greater β without production of musical noise and by increasing it we can have less noise in the enhanced signal and then achieve more SNR improvement, too.

5. Conclusions

In this paper we studied the effect of sampling frequency on a FFT-based spectral estimator. We also proposed an improved spectral subtraction method by increasing the accuracy of spectral estimator.

This adaptive estimator can give better spectral estimation by increasing the analysis frame length that achieves in silence regions.

In this method for separating silence frames we use a basic analyzing frame and for estimation the spectrum we use an adaptive frame length that can increase until it covers all current silence region. By this method we could have a better spectral estimation for noise and noisy signal and so the system can produce a better enhanced signal with less musical noise.

References:

[1] S. F. Boll, "Suppression of Acoustic Noise in Speech using Spectral Subtraction," IEEE Trans. Acoustics, Speech and Signal processing, vol. ASSP-27, No. 2, pp. 113-120, 1977.

[2] H. Hu, F. Kuo, H. Wang, "Supplementary Schemes to Spectral Subtraction for Speech Enhancement," Speech Communication, 2002.

[3] H. Gustafsson, S. Nordholm, "Speech Subtraction using Reduced Delay Convolution and Adaptive Averaging", *IEEE Trans. Speech and Audio Processing*, vol. 9, No. 8, pp. 799-807, 2001.

[4] J. S. Lim, A. V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech", Proceedings of the IEEE, vol. 67, 1972.

[5] S. Ayat, "Enhanced Human-Computer Speech Interface Using Wavelet Computing", IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2008.