Noise Reduction and Speech Enhancement

Basic methods and their improvements

Rong Xu
Department of EECS
University of Michigan
Ann Arbor, Michigan
rongx@umich.edu

***Abstract*—This paper is the final report of the EECS451 project. In this project, several audio enhancement methods are researched and implemented in MATLAB to reduce the noise in poorly recorded speech and reach a better signal-to-noise ratio. Improvement is applied to the basic spectral reduction algorithms. Issues of the trade off in this method and side effect included are also discussed.**

Keywords—Butterworth; spectral subtraction; MATLAB; audio enhancement;

I. Introduction

Sometimes people may want to record a motivating speech, a significant lecture suddenly happen, or listen to an old speech more clearly. However, professional audio recording devices are usually not accessible, which means mobile phones are always the only choice. Due to the undesirable noise of the environment and, poor recording quality of mobile phones and devices decades ago, noise reduction and speech enhancement is needed to get the satisfied records. In this project, some famous speeches with simulated noises are processed. Different basic filters like RLS filter and Butterworth filter is designed and applied to the original data. Then, spectral subtraction method is introduced and implemented in MATLAB and improvement is tried. In the end, the side effect is discussed.

II. Original data and FFT

A. 'I have a dream'

Several famous speeches is processed, but in this report I pick the most famous one, the 'I have a dream' speech given by Martin Luther King Jr.

B. Speech in time and frequency domain

Plot the original speech in time domain. Calculate its FFT and plot the magnitude in frequency domain.

III. Noise reduction

There are some common kinds of noise, such as periodic noise, pulse noise, wideband noise, lombard effect and etc. Due to the complexity of possible noise, The comparably simple one, white noise, is selected to simulate the noise in speech.

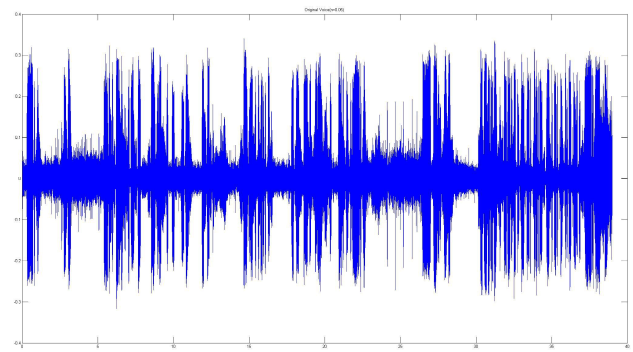

Fig. 1. Original speech in time domain.

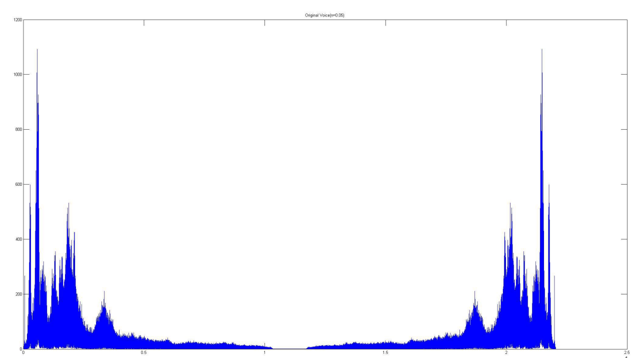

Fig. 2. Original speech in frequency domain.

A. White noise

white noise is a random signal with a constant power spectral density. The term is used, with this or similar meanings, in many scientific and technical disciplines, including physics, acoustic engineering, telecommunications, statistical forecasting, and many more. White noise refers to a statistical model for signals and signal sources, rather than to any specific signal.

B. Speech with white noise

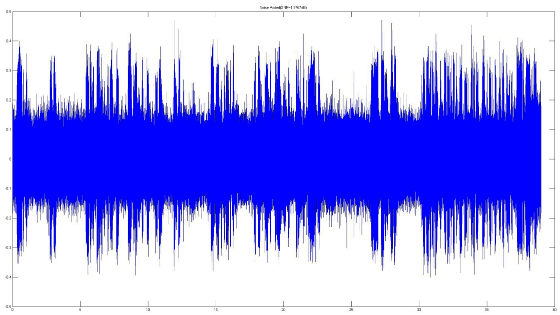

Fig. 3. Noisy speech in time domain, SNR=1.9767dB.

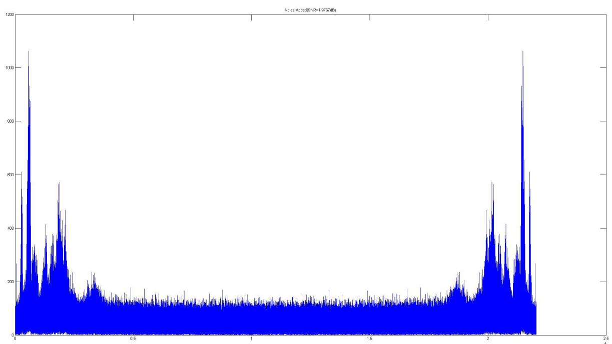

Fig. 4. Noisy speech in frequency domain, SNR=1.9767dB.

C. Butterworth Filter

Early in this project, RLS, LMS and Butterworth filters are used to implement band-pass filter to do basic noise reduction. However, only the Butterworth filter has desirable result in noise suppression.

The Butterworth filter is a type of signal processing filter designed to have as flat a frequency response as possible in the passband. It is also referred to as a maximally flat magnitude filter. In MATLAB, use buttord() and buttap() to setup a lowpass butterworth filter. After observing the frequency representation of the speech and trying, select the most efficient combination of passband corner frequency and stopband corner frequency.

```
wp=0.15*pi; ws=0.32*pi;
Fs=8000;
T=1/Fs;
Ap=(2/T)*tan(wp/2);
Ar=(2/T)*tan(ws/2);
rp=0.6; rs=8; as=50;
ripple=10^(-rp/20);
attn=10^(-rs/20);
[n,wn]=buttord(Ap,Ar,rp,rs,'s');
[Z,p,k]=buttap(n);
[b,a]=zp2tf(Z,p,k);
[bt,at]=lp2lp(b,a,wn);
[b,a]=bilinear(bt,at,Fs);
```

Following is the filtered speech given in progress report 2.

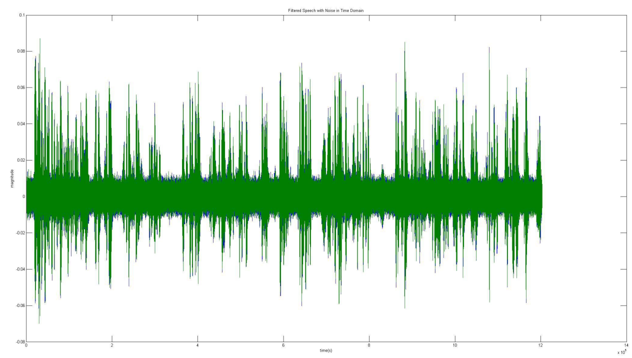

Fig. 5. Butterworth filtered speech in time domain.

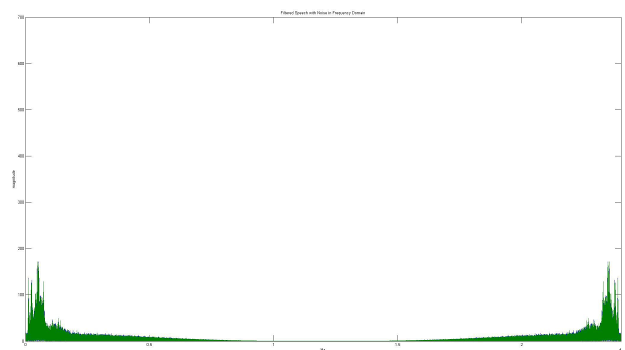

Fig. 6. Butterworth filtered speech in frequency domain.

IV. SPECTRAL SUBTRACTION

If it is assumed that the signal is distorted by a wide-band, stationary, additive noise, the noise estimate is the same during the analysis and the restoration and the phase is the same in the original and restored signal, we can use the basic spectral subtraction, a simple and effective method of noise reduction. In this method, an average signal spectrum and average noise spectrum are estimated in parts of the recording and subtracted from each other, so that average signal-to-noise ratio (SNR) is improved. However, a pretty annoying side-effect called 'musical noise' will be included in the output.

A. Basic theory

Due to the assumptions before, we can get the additive module of the noisy speech:

$$y[n]=s[n]+d[n] \tag{1}$$

After processed with window function:

$$y_w[n] = s_w[n] + d_w[n] \tag{2}$$

Do DTFT in both side:

$$|Y_w(\omega)|^2 = |S_w(\omega)|^2 + |D_w(\omega)|^2 + S_w(\omega)D_w^*(\omega) + S_w^*(\omega)D_w \tag{3}$$

$|S_w(\omega)|^2$ and $|D_w(\omega)|^2$ are the spectral of original speech and noisy speech. Out goal is to find the estimation of $|S_w(\omega)|$.

Because we can not get the accurate value of terms in equation (3), we usually use its average energy to approximate. Due to the independence of s[n] and d[n], we have:

$$E[|S_w(\omega)D_w^*(\omega)|] = E[|S_w^*(\omega)D_w(\omega)|] = 0$$

Then we get the estimation of $|S_w(\omega)|^2$:

$$|\hat{S}_w(\omega)|^2 = |Y_w(\omega)|^2 - E[|D_w(\omega)|^2] \tag{4}$$

$$|\hat{S}_w(\omega)| = |Y_w(\omega)| - E[|D_w(\omega)|] \tag{5}$$

It is known that human's ear is not sensitive to the value of phase, so we can assume that the phase does not change during spectral subtraction. Then the estimated spectral and sequence in time domain is:

$$\hat{S}_w(\omega) = |\hat{S}_w(\omega)| \exp(jY_w(\omega)) \tag{6}$$

$$\hat{s}_w(n) = F^{-1}[\hat{S}_w(\omega)] \tag{7}$$

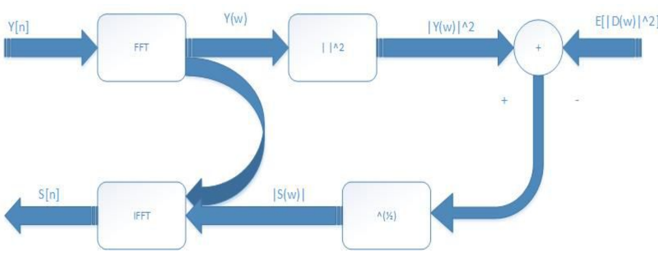

Fig. 7. Diagram of basic spectral subtraction method.

Spectral subtraction can also be done by multiplication:

$$|\hat{S}(\omega)| = G(\omega) \bullet |Y(\omega)| \tag{8}$$

In which:

$$G(\omega) = \sqrt{1 - \frac{|\hat{D}(\omega)|^2}{|Y(\omega)|^2}} \tag{9}$$

When $|\hat{D}(\omega)|^2 > |Y(\omega)|^2$, we define $G(\omega) = 0$ to guarantee that:

$$0 \le G(\omega) \le 1 \tag{10}$$

In equation (9), notice that $G(\omega)$ is determined by SNR:

$$SNR(\omega) = \frac{|Y(\omega)|^2}{|\hat{D}(\omega)|^2} \tag{11}$$

The basic spectral subtraction algorithm is implemented in MATLAB code to process the 'I have a dream' speech with white noise. Following is the enhanced speech output in both time and frequency domain:

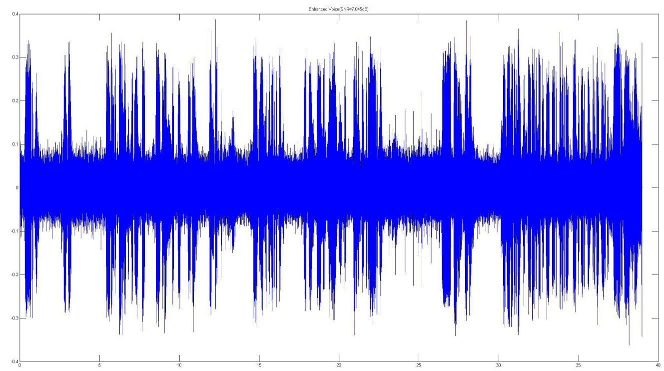

Fig. 8. Enhanced speech in time domain, SNR=7.045dB.

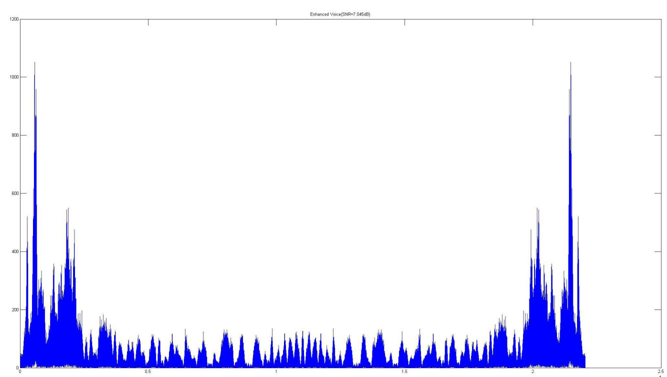

Fig. 9. Enhanced speech in frequency domain, SNR=7.045dB.

Compared to the frequency representation of noisy speech given in part III, the high frequency noise is largely reduced but leave lots of little peaks. Although the enhanced speech sounds more clearly than the noisy one, it included strong 'musical noise' mentioned before. In the following part, an alternative improvement is applied to the basic spectral subtraction method.

B. Improved spectral subtraction

We can modify the equation (5) to a more general form as :

$$|\hat{S}_w(\omega)|^a = |Y_w(\omega)|^a - b \cdot E[|D_w(\omega)|^a] \tag{12}$$

In this method, we can modify 2 parameters a and b. This is also a important way to reduce the musical noise. During my implementation, this method is proved more effective than the basic spectral subtraction.

When a=1, b=1, it is the basic spectral subtraction. When b>1, it is the over subtraction form; if a =2, b=1, we can get the power spectrum subtraction form. The influence of parameter b is that with larger b, the noise reduction is more significant but more information of effective speech is reduced, which means the distortion of original speech is more serious. However, with smaller b, the distortion is lighter but the effect of noise reduction is poor. In fact, the choice of parameter b is a trade off between distortion and noise reduction.

For signal has low signal-to-noise ratio, the variance of noise is large, we can choose larger b; as for signal has high SNR, b can be smaller.

In this project, the parameter a is fixed at 2 and different parameter b is tried to find a better trade off than the basic spectral subtraction. Following is the output in time and frequency domain when b equals to 6:

This time, the enhanced speech sounds more clearly, the volume of noise is less than the previews enhanced speech. Just as discussed before, the distortion becomes a bigger issue, but it is also the most acceptable trade off during experiment. As we can see in the figure, SNR is increased further compared to the basic method. Furthermore, ‘musical noise’ included by the improved spectral subtraction method is suppressed but still not negligible. Although it is much smaller than the white noise added, it is pretty annoying.

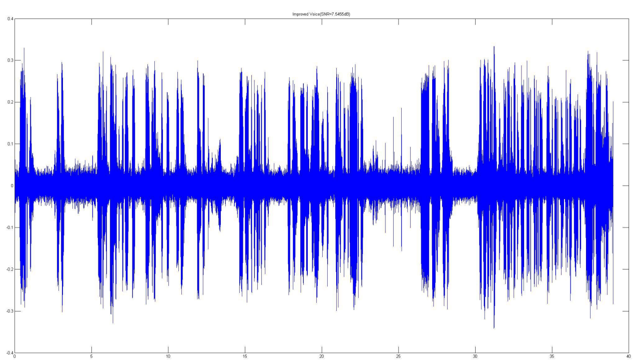

Fig. 10. Improved speech in time domain, SNR=7.5455.

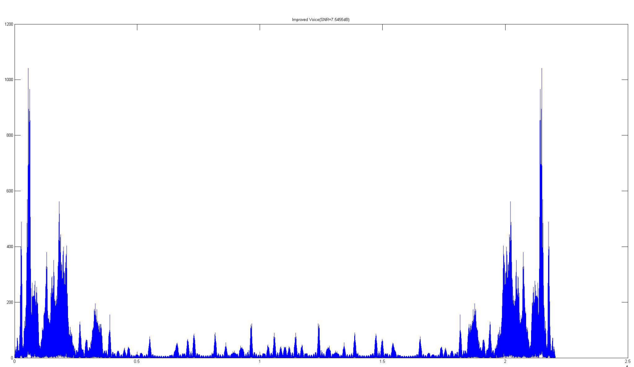

Fig. 11. Improved speech in frequency domain, SNR=7.5455.

C. Musical noise

a) Drawback of spectral subtraction: Listening to the enhanced speech, it is easy to notice that the speech is accompanied by perceptually annoying noise, namely musical noise. The musical noise consists of some noticeable random rapid-changing tones in the background of speech.

Musical noise is a perceptual phenomena that occurs when isolated peaks are left in a spectrum after processing with a spectral subtraction type algorithm. In the enhanced speech spectral, these isolated peaks are obvious in the high frequency part. In speech absent sections of a signal, these isolated components sound like musical tones to our ears. In speech present sections, it produces an audible ’warble’ of the speech.

b) Possible solutions:

- Parametric Spectral Subtraction. Just discussed before, this method can suppress musical noise but how to select the appropriate parameters is the tricky part.
- Musical Noise Filtering. When musical noise is still present after spectral smoothing and parametric tuning, we can turn to post-processing solutions for further musical noise reduction.In this model, speech presence or absence needs to be estimated for the current frame, and based on this estimate a window is derived to filter the gain function thereby reducing musical noise.

SUMMARY AND FUTURE WORK

In this project, DSP tools from class like time domain analysis, Fast Fourier Transform and window functions are used in speech analysis. Fdatool and spectral subtraction algorithms are what I learned outside of classes.

Different filters are applied to original speech to reduce noise. A popular audio enhancement algorithm, spectral subtraction, is discussed and implemented in MATLAB. A alternative improvement is completed to further reach a better enhance effect. Also, the drawback and possible solutions are discussed.

After complete these, solutions like post-processing filtering can be implemented to further reduce the musical noise.

REFERENCES

[1] Selesnick, Ivan W., and C. Sidney Burrus. "Generalized digital Butterworth filter design." Signal Processing, IEEE Transactions on 46.6 (1998): 1688-1694.

[2] Boll, Steven. "Suppression of acoustic noise in speech using spectral subtraction." Acoustics, Speech and Signal Processing, IEEE Transactions on 27.2 (1979): 113-120.

[3] Goh, Zenton, Kah-Chye Tan, and B. T. G. Tan. "Postprocessing method for suppressing musical noise generated by spectral subtraction." IEEE Transactions on Speech and Audio Processing 6.3 (1998): 287-292.

[4] Sim, Boh Lim, et al. "A parametric formulation of the generalized spectral subtraction method." Speech and Audio Processing, IEEE Transactions on 6.4 (1998): 328-337.

[5] Ogata, S., and Tetsuya Shimamura. "Reinforced spectral subtraction method to enhance speech signal." TENCON 2001. Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology. Vol. 1. IEEE, 2001.

[6] Udrea, Radu M., and S. Ciochina. "Speech enhancement using spectral over-subtraction and residual noise reduction." Signals, Circuits and Systems, 2003. SCS 2003. International Symposium on. Vol. 1. IEEE, 2003.