Spectral Subtraction Abstract

Basic Spectral Subtraction Algorithm

We look at the spectral subtraction method of speech enhancement. Speech Enhancement means reducing noise level from a noisy speech as much as possible while retaining the intelligibility of the speech. The basic spectral subtraction method consists of subtracting the estimated noise magnitude spectrum from the noisy speech spectrum to obtain the clean speech spectrum. From this clean speech spectrum, we can obtain the clean speech time waveform by taking the inverse Fourier transform and by using the original phase. The estimated noise spectrum is obtained by considering the noisy speech in silent intervals, i.e. in intervals where speech is not present, or where there is non-speech activity. The noise at such places is averaged to obtain the estimated noise spectrum. If after subtraction, the value of obtained spectrum is negative, then we make those values equal to zero. This basic algorithm with further residual noise reduction is discussed in Boll [1]. There are various limitations of this method which are discussed in [2]. These limitations are mostly phase, cross-term and magnitude errors and are more pronounced when the noise level reaches 0 dB and below. To overcome these errors and for other improvements, there have been proposed many modifications of the spectral subtraction algorithm: -

1. Berouti et al. [3]

In this modification, they tried to reduce the 'musical noise'. ('Musical noise is perceptually annoying due to the presence of random frequency components). This was done by subtracting an 'overestimate' of the noise spectrum from the noisy speech. This 'overestimate' was more than the original estimate by a factor $\propto$ which is greater than 1 and usually taken between 3 to 6 and it is made to vary with the SNR (dB) to get the desired noise reduction and intelligibility. Also, in this method, we introduce a spectral floor which is little above zero and not zero. In the original method, we had reduced all the negative results (obtained after subtraction of spectrums) to zero, but in this case, we make them equal to the spectral floor the position of which is described by a factor β which is usually taken between 0.005 to 0.1. With the results it shows, this method was successful in reducing the 'musical noise' as well as some 'background noise' and intelligibility did not suffer for most SNRs. This is the paper that I would be implementing in MATLAB.

2. Geometric Spectral Subtraction [4]

In the conventional spectral subtraction method, the cross terms between signal and noise spectra are considered zero because the signal and noise are uncorrelated. Although this assumption is valid, and our algorithm will work well for very high or low SNR, this is not valid for SNRs close to 0 dB and this is where most of the speech enhancement algorithms work. This assumption, then, is unsafe to work with and can cause considerable estimation errors. So, a modification was proposed which takes into account the cross terms and the phase differences between speech and noise. In this modification, the analysis is done after representing the noisy speech signal as a sum of clean speech and noise, geometrically in the complex plane.

3. Spectral Subtraction based on Minimum Statistics [5]

In this modification, the noise estimation part of the original spectral subtraction algorithm is modified. In the original algorithm, noise is estimated during the non-speech activity. Due to this, tracking of varying noise levels is slow and confined to periods of non-speech activity. In this modification, the valleys of the noisy speech spectrum are used to estimate noise spectrum based on minimum statistics. The noisy speech spectrum is analysed in windows of appropriate duration. This modification eliminates the need for speech activity detection with comparable computational complexity.

4. Spectral Subtraction in the short-time modulation domain [6]

In this modification of the original method, we proceed in the modulation domain, i.e. we apply spectral subtraction in the modulation domain as opposed to acoustic domain in the original algorithm. The original method uses AMS (Analysis-Modification-Synthesis) technique. This speech enhancement method processes each frequency component of the acoustic magnetic spectra, obtained during the analysis stage of the acoustic AMS procedure frame-wise across time using a secondary (modulation) [7]. Thus, the AMS framework is extended to include modulation processing. Various subjective and objective tests show that the method is successful in reducing 'musical noise' and also it leads to improved speech quality for various input SNRs.

5. Speech Enhancement based on the Masking properties of Human Auditory System [8]

This method introduces human perception into the speech enhancement algorithm. In our auditory system, there is a masking threshold. All the noises below this threshold are inaudible. So, the subtraction parameters ($\propto$, β) discussed in [3] are chosen such that the residual noise is below this masking threshold and thus, inaudible. The major step in this method is to obtain the noise masking threshold. It is done after estimating the noise spectrum. The noise masking threshold can be obtained from clean speech signal, but we only have the noisy speech and especially for high frequencies, its values differ slightly. This method is very useful when the SNRs are very low, i.e. when noise levels are high as compared to signal (<10 dB SNRs).

6. Multiband Spectral Subtraction [9]

This method is used when the noise present in the noisy speech is mainly colored noise. And this is the case in many real-life situations. The noise, if it is colored effects different frequencies differently as opposed to white noise which effects all the frequencies in the same way. This method is a modification of [3] and in this, the over subtraction parameter ($\propto$) and the spectral floor parameter (β) are calculated independently for multiple bands. It is shown that this method produced better results than the basic spectral subtraction method and the Non-linear spectral subtraction method.

7. Extended Spectral Subtraction [10]

This modification is related to the noise spectrum estimation. Traditionally, noise spectrum is estimated using a VAD (Voice Activity Detector) and it is done in periods of non-speech activity. In this modification, a Voice Activity detector is not required. Here, we use a Wiener Filter to estimate noise and it can be done even when speech is present. And thus, the noise spectrum can be continuously updated. This method is a combination of Wiener Filtering and Spectral Subtraction Principles.

Performance of Spectral Subtraction Algorithms

Many objective tests comparing the enhanced signal with the original signal have been performed and various subjective tests have also been performed like formal listening tests. The intelligibility and quality of the enhanced speech is measured. It is expected that the spectral subtraction algorithms should improve both the quality and intelligibility of speech, but it has been found that it is not the case. Most of the tests and studies confirm that the algorithms improve speech quality, but not speech intelligibility. In fact, in some cases the intelligibility decreases. This is because, although the spectral subtraction algorithm does an excellent job in removing the noise, but also, we lose low energy information, which may lead to a decrease in intelligibility.

BIBLIOGRAPHY OF PAPERS

[1] Boll, S., 1979. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on acoustics, speech, and signal processing*, *27*(2), pp.113-120.

[2] Evans, N.W., Mason, J.S., Liu, W.M. and Fauve, B., 2006, May. An assessment on the fundamental limitations of spectral subtraction. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on* (Vol. 1, pp. I-I). IEEE.

[3] Berouti, M., Schwartz, R. and Makhoul, J., 1979, April. Enhancement of speech corrupted by acoustic noise. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'79.* (Vol. 4, pp. 208-211). IEEE.

[4] Lu, Y. and Loizou, P.C., 2008. A geometric approach to spectral subtraction. *Speech communication*, *50*(6), pp.453-466.

[5] Martin, R., 1994. Spectral subtraction based on minimum statistics. *power*, *6*, p.8.

[6] Paliwal, K., Wójcicki, K. and Schwerin, B., 2010. Single-channel speech enhancement using spectral subtraction in the short-time modulation domain. *Speech communication*, *52*(5), pp.450-475.

[7] Goel, P. and Garg, A., 2011. Review of Spectral Subtraction Techniques for Speech Enhancement 1.

[8] Virag, N., 1999. Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Transactions on speech and audio processing*, *7*(2), pp.126-137.

[9] Kamath, S. and Loizou, P., 2002, May. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. In *ICASSP* (Vol. 4, pp. 44164-44164).

[10] Sovka, P., Pollak, P. and Kybic, J., 1996, September. Extended spectral subtraction. In *European Signal Processing Conference, 1996. EUSIPCO 1996. 8th* (pp. 1-4). IEEE.

[11] Loizou, P.C., 2013. *Speech enhancement: theory and practice*. CRC press.