

An Improved Algorithm for Noise Reduction in Hearing Aids

Yueqin Feng*, Guichen Tang, Ruiyu Liang, Qingyun Wang

School of Communication Engineering, Nanjing Institute of Technology, Jiangsu, Nanjing, 211167, China
fengyueqin@njit.edu.cn

The noise reduction algorithm is a vital algorithm in hearing aids. To improve performances, an improved algorithm is proposed. For the basic perception filter, the masking tone will be audible after filtering. To solve this problem, one weighted perception filter algorithm for the simulation of the auditory mechanism is designed. Based on the basic perception filter, the priori SNR is firstly estimated by the direct decision method and then revised by the optimal smoothing factor method. The optimal value is obtained based on minimum mean square error criterion, and the gain of weighted perception filter is computed. Experiment results show that the proposed method can effectively eliminate the residual noise, reduce the interference, and improve the speech comprehension of hearing-impaired patients under the noisy environment.

1. Introduction

How to improve the performance of hearing aids in the noisy environment is an important research direction. Many noise reduction methods have been proposed after decades of research. For single channel noise reduction algorithm, the problem needs to be solved is how to estimate desired speech signal from the signals corrupted by noise. The earliest method is the spectrum subtraction algorithm proposed by Boll in 1979 Boll (1979), which is computationally simple, but will bring out music noise. Afterwards, many improved algorithms have been proposed by researchers, such as the adaptive gain average spectrum subtraction algorithms proposed by Gaustafsson Gustafsson, Nordholm and Claesson (2001) reported, the Minimum Mean Square Error (MMSE) spectrum subtraction algorithms proposed Sim (1998) reported, etc. These algorithms have achieved a certain effect. Another common single channel noise reduction method is the Wiener filter method Wiener (1949) reported, which takes the Minimum Mean-Square Error between estimated signal and real signal as the optimum criterion. Then Lim and Oppenheim proposed an iterative Wiener filter method Lim (1979) reported. In 1984, Ephraim et al proposed an optimal amplitude spectral estimation method based on MMSE Ephraim (1984) reported. Because the human ear sound perception of loudness is in direct proportion to the spectrum amplitude, it is more appropriate to use logarithmic spectrum. Thus Ephraim and Malah proposed a MMSE estimation method based on logarithmic spectrum Ephraim and Malah (1985) reported. In addition, there are other improved forms of MMSE, such as MMSE estimation based on Laplace distribution proposed by Chen and Loizou, optimal modified logarithmic spectrum amplitude estimation proposed by Cohen, etc. Cohen and Berdugo (2001) reported.

Because the human sound subjective perception is the result from the joint effect of human psychology and physiology, many studies for this research have been done. Now, the speech enhancement based on auditory masking model has been a hot research area. Virag et al. combined the auditory masking of human ear auditory system and the spectrum reduction algorithm to reduce the speech noise Virag (1999) reported. Loizou and Alam proposed a noise reduction method based on perceptual filter, by which the speech signal with noise is filtered by designing a perceptual filter Udrea (2008), Vizireanu and Ciochina reported. However, because of the variety of noise type and application occasions, no one is able to suppress all the noise. In addition, although the perceptual filter method performs better than the traditional noise reduction methods, it still has residual audible noise. The ultimate cause is: after the noise reduction by perceptual method, the speech element is reduced, which leads to the reduction of original masking threshold, so the inaudible noise element which is masked and not processed exceeds the masking threshold and become audible.

Therefore, a perceptual filter algorithm based on the mechanism of auditory masking is proposed. The algorithm designs the weighting factor of the perceptual filter based on the principle of MMSE, so the estimation of prior SNR is modified, and the optimum gain function is achieved. For three noise types with different SNR,

three noise reduction algorithms are compared. The subjective and objective marker shows that the noise reduction performance of this algorithm is outstanding and the speech quality is good.

2. The principle of auditory masking

The speech enhancement algorithm based on auditory masking does not need to eliminate all the noise, but just satisfies the condition that the residual noise is not perceived by human, which reduce the distortion of speech and improve the hearing comfort of human ear.

Set $y(n)$ to be the speech signal with noise, $x(n)$ to be the pure speech signal, $d(n)$ to be the additive noise signal, and suppose $x(n)$ and $d(n)$ are not relevant, then the speech signal with noise is:

$$y(n) = x(n) + d(n) \quad (1)$$

After frame separation and Fourier transform, we get

$$Y(m, k) = X(m, k) + D(m, k) \quad (2)$$

Where, $Y(m, k)$ is the amplitude spectrum of speech with noise, $X(m, k)$ is the amplitude spectrum of the pure speech, $D(m, k)$ is the amplitude spectrum of noise, m is the number of frame, k is the discrete frequency.

Let $\hat{X}(m, k) = G(m, k) \cdot Y(m, k)$ is the speech signal enhanced by the perceptual filter, $G(m, k)$ is the gain function of perceptual filter, then the error spectrum between enhanced speech and pure speech is:

$$\begin{aligned} \varepsilon(m, k) &= \hat{X}(m, k) - X(m, k) \\ &= (G(m, k) - 1) \cdot X(m, k) + G(m, k) \cdot D(m, k) \\ &= \varepsilon_x(m, k) + \varepsilon_D(m, k) \end{aligned} \quad (3)$$

Form formula (3) we can know that distortion of speech signal and distortion of noise signal is opposite with the variation trend of gain function $G(m, k)$, and it is impossible to make both down simultaneously. Therefore, an ideal gain function needs to make a reasonable compromise between distortion of speech signal and distortion of noise signal. The purpose of perceptual filter is not to eliminate the residual noise completely, but keep the residual noise below the hearing threshold T by using the auditory masking of human ear and can not be heard by human ear, and make the distortion of speech signal minimum. As:

$$\begin{cases} \min \varepsilon_x^2(m, k) \\ \text{constraint condition: } \varepsilon_D^2(m, k) \leq T(m, k) \end{cases} \Rightarrow \begin{cases} \min \{(1 - G(m, k))^2 |X(m, k)|^2\} \\ \text{constraint condition: } G^2(m, k) |D(m, k)|^2 \leq T(m, k) \end{cases} \quad (4)$$

Bring in a Lagrange factor $\mu(m, k)$, and $\mu(m, k) \geq 0$. Let Lagrange cost function be:

$$\begin{aligned} L &= \varepsilon_x^2(m, k) + \mu(m, k) [\varepsilon_D^2(m, k) - T(m, k)] \\ &= (G(m, k) - 1)^2 |X(m, k)|^2 + \mu(m, k) [G^2(m, k) |D(m, k)|^2 - T(m, k)] \end{aligned} \quad (5)$$

Let $\frac{\partial L}{\partial G} = 0$, then get:

$$\begin{aligned} G(m, k) &= \frac{|X(m, k)|^2}{|X(m, k)|^2 + \mu(m, k) |D(m, k)|^2} \\ &= \frac{\xi(m, k)}{\xi(m, k) + \mu(m, k)} \end{aligned} \quad (6)$$

Where, $\xi(m, k)$ denotes the SNR of the m th frame. Put the above formula into the constraint condition $G^2(m, k) |D(m, k)|^2 \leq T(m, k)$, and $\mu(m, k) \geq 0$, then get:

$$\mu(m, k) \geq \max \left(\sqrt{\frac{|D(m, k)|^2}{T(m, k)}} - 1, 0 \right) \cdot \xi(m, k) \quad (7)$$

As long as $\mu(m, k)$ satisfies formula (7), the residual noise can be masked. Put formula (6) into $\varepsilon_x^2(m, k)$, then

$$\begin{aligned} \varepsilon_x^2(m, k) &= (1 - G(m, k))^2 \cdot |X(m, k)|^2 \\ &= \left(1 - \frac{\xi(m, k)}{\xi(m, k) + \mu(m, k)} \right)^2 \cdot |X(m, k)|^2 \end{aligned} \quad (8)$$

It can be inferred from formula (8) that with the increase of $\mu(m, k)$, $\varepsilon_x^2(m, k)$ increases, namely the speech distortion increases. Therefore, in order to reduce the distortion of speech, take the minimum value of $\mu(m, k)$,

the gain function can be obtained:

$$G(m,k) = \frac{1}{1 + \max\left(\sqrt{\frac{|D(m,k)|^2}{T(m,k)}} - 1, 0\right)} = \min\left(\sqrt{\frac{T(m,k)}{|D(m,k)|^2}}, 1\right) \quad (9)$$

The gain function in formula (9) is obtained by keeping the speech distortion minimum while keeping the residual noise below the masking threshold.

3. Improvement strategy

When the power spectrum $|D(m,k)|^2$ of the k th discrete frequency point of the m th frame of noise is higher than the corresponding masking threshold $T(m,k)$, we can obtain $G(m,k) < 1$, namely the k th discrete frequency point of the m th frame of speech with noise is suppressed; and for the frequency point of $|D(m,k)|^2 \leq T(m,k)$, $G(m,k) = 1$, which is output directly without any reduction process, namely only the noise above the masking threshold is filtered, and the noise below the masking threshold still exist. So the MAN phenomenon is generated. Based on the consideration of real-time and low-power of digital hearing aids, usually a Wiener filter based on prior SNR is used to suppress, because the Wiener filter is computationally simple and the residual noise of Wiener filter is similar with white noise to make the auditory perception good. Here, a weighting factor W is introduced and defined as:

$$W(m,k) = \begin{cases} \frac{\xi(m,k)}{1 + \xi(m,k)}, & T_{abs}(m,k) < |D(m,k)|^2 \leq T(m,k) \\ 1, & \text{others} \end{cases} \quad (10)$$

Where, $T_{abs}(m,k)$ is the absolute hearing threshold, $\xi(m,k)$ is prior SNR. When noise power spectrum satisfied $T_{abs}(m,k) < |D(m,k)|^2 \leq T(m,k)$, suppress the speech with noise based on Wiener filter, so the MAN phenomenon is reduced and the speech quality is improved.

Usually a decision-directed method is used to estimate the prior SNR $\xi(m,k)$.

$$\hat{\xi}(m,k) = \alpha \cdot \frac{|\hat{X}(m-1,k)|^2}{E\left[|\hat{D}(m-1,k)|^2\right]} + (1-\alpha) \cdot \max(\gamma(m,k) - 1, 0), \quad 0 \leq \alpha < 1 \quad (11)$$

Where, $|\hat{X}(m-1,k)|^2$ denotes the estimated pure speech spectrum of the $(m-1)$ th frame, $E\left[|\hat{D}(m-1,k)|^2\right]$ denotes the estimated noise spectrum of the $(m-1)$ th frame, α is the smoothing factor, $\gamma(m,k)$ is the posterior SNR.

The value of smoothing factor α is very important to estimate the prior SNR. Usually, it is constant close to 1. Literature Cappe (1994) reported denotes that the residual music noise is lower when the value of α is closer to 1, but it will lead to the transient distortion of final signal increase. In order to balance both aspects, the value of α is usually set to be a constant between 0.95-0.99. However, the fixed value of α is not able to handle the situation of the abrupt change of posterior SNR. Therefore, in order to solve the above problem, an optimal smoothing factor method is used, where α is no longer a fixed value, but a smoothing coefficient varying with time.

Rewrite the original prior SNR smoothing formula (11) to the formula effected by time-varying smoothing factor:

$$\hat{\xi}(m,k) = \alpha(m,k) \cdot \bar{\xi}(m-1,k) + [1 - \alpha(m,k)] \cdot \max(\gamma(m,k) - 1, 0) \quad (12)$$

$$\text{Where } \bar{\xi}(m-1,k) = \frac{|\hat{X}(m-1,k)|^2}{E\left[|\hat{D}(m-1,k)|^2\right]}.$$

In order to reduce the error between estimated value and real value, a MMSE criterion is used to update $\alpha(m,k)$.

Mean square error (MSE) is defined as:

$$J_\alpha = E\left\{\left[\hat{\xi}(m,k) - \xi(m,k)\right]^2 \mid \bar{\xi}(m-1,k)\right\} \quad (13)$$

Substitute fomula (13) in fomula (14) and simplify, we can obtain:

$$J_\alpha = \alpha^2(m,k) \left[\bar{\xi}(m-1,k) - \xi(m,k) \right]^2 + [1 - \alpha(m,k)]^2 \cdot [\xi(m,k) + 1]^2 \quad (14)$$

Compute the partial derivative of J_α , and let $\frac{\partial J_\alpha}{\partial \alpha(m,k)} = 0$, the optimal value of $\alpha(m,k)$ can be obtained as:

$$\alpha^{opt}(m,k) = \frac{1}{1 + \left[\frac{\xi(m,k) - \bar{\xi}(m-1,k)}{\xi(m,k) + 1} \right]^2} \quad (15)$$

In practical application, because $\xi(m,k)$ is unknown, we can substitute $\bar{\xi}(m,k)$ in the above formula with $\max(\gamma(m,k) - 1, 0)$. In order to prevent the deadlock of $\alpha(m,k) = 1$, the optimal smoothing factor $\alpha^{opt}(m,k)$ is constrained below a max value α_{max} , for example $\alpha_{max} = 0.998$. Equally, in order to ensure the smoothing performance under low posterior SNR, also set a lower limit value α_{min} of smoothing factor, for example, 0.3. The final gain function of perceptual weighting filter can be denoted as:

$$G^{WPF}(m,k) = G^{PF}(m,k) \cdot W(m,k) \quad (16)$$

4. Experimental simulation and analysis

In order to measure noise reduction performance, the test is carried out under different noisy environment. Four noise types such as White noise, Pink noise, Speech Babble noise and HF channel noise in NOISE-92 noise database are applied in this experiment. The speech signals with noise at different SNR (-5dB, 0dB, 5dB, 10dB, 15dB) are obtained by superposition, and MOS method among subjective assessment methods as well as segSNR (segmental SNR) and WSS (Weighted-Slope Spectral distance) among objective assessment are used to analyze the performances of three methods. Both the pure speech and noise in the experiment are re-sampled as 16KHz, and a Hamming window is selected, the frame length is 512 sample points, the overlap between the frames is 50%. Wiener Filter (WF) method and Perceptual Filter (PF) method are used for comparison in the experiment.

4.1 Subjective assessment

MOS score is a widely used subjective assessment method. Ten testers with normal hearing are invited to hear, and grade according to the MOS grading standard. Under White noise, Pink noise, SpeechBabble noise and HFchannel noise, the MOS score results of three noise reduction methods are showed in Fig. 1.

It is can be infer from Fig. 1 that the MOS score of WF is the lowest of three methods, because there is still residual music noise after WF noise reduction. There is almost no music noise to be heard after PF noise reduction, the MOS score of PF is higher than WF method. Because the MAN phenomenon of WPF is reduced based on PF method, the speech quality is improved. The proposed WPF method is more effective in noise reduction, and the speech hearing is better.

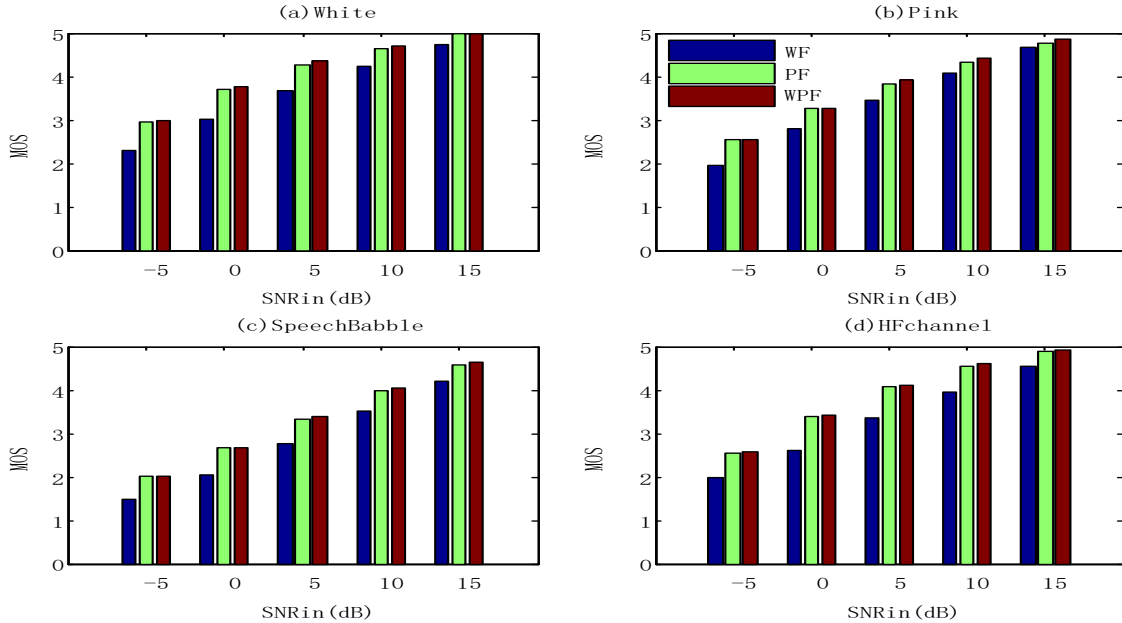


Figure 1: MOS comparisons

4.2 Objective assessment

(1) segSNR

Under White noise, Pink noise, SpeechBabble noise and HFchannel noise, the Segmental SNR of three noise reduction methods are shown in Fig. 2:

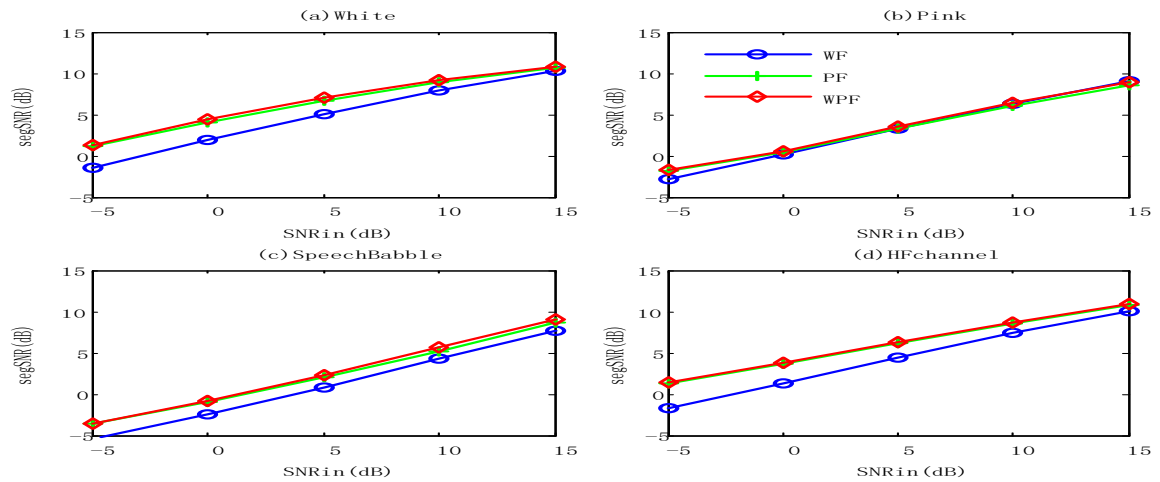


Figure 2: Segmental SNR comparisons

(2) WSS

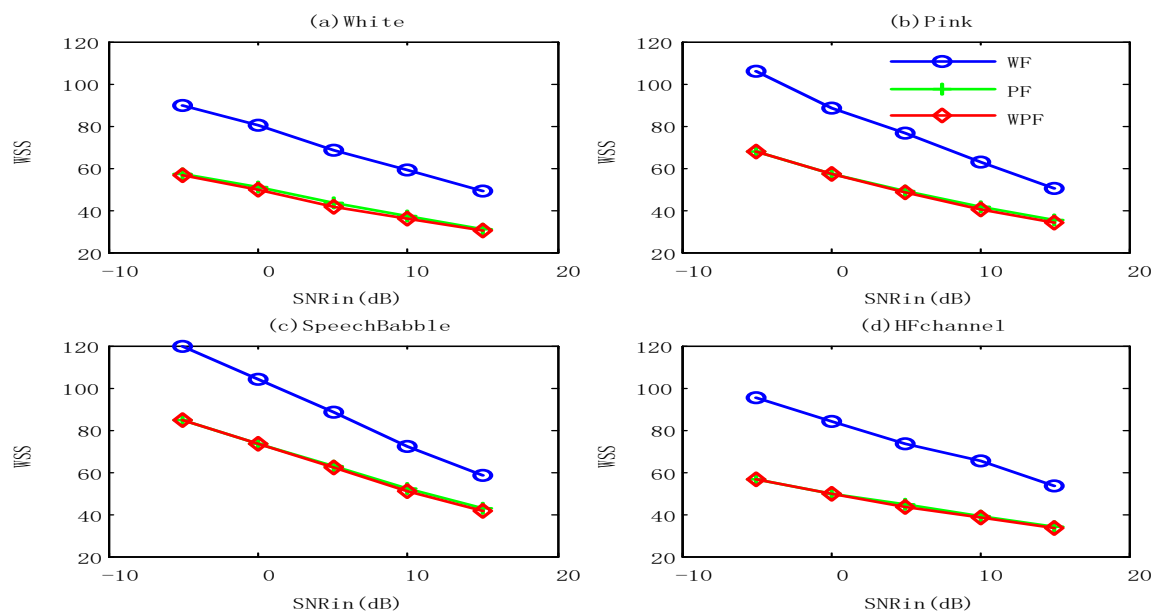


Figure 3: WSS comparisons

Under White noise, Pink noise, SpeechBabble noise and HFchannel noise, WSS of three methods are shown in Fig. 3.

From the objective assessments, the WPF noise reduction method is superior to WF method and PF method, the speech after WPF noise reduction is closer to pure speech, the distortion is less, and the speech quality is better. A better performance in back ground noise eliminate, residual noise suppression and speech distortion reduction is achieved.

5. Conclusions

A weighting perceptual filter algorithm imitated hearing masking is proposed. Compared to the traditional speech noise reduction algorithm, real time weighting factor is introduced to improve the speech noise reduction performance, speech quality and naturalness, and the speech recognition rate of hearing-impaired

persons, which has a high reference value and guidance for subsequent digital hearing aids designing and application.

Acknowledgments

The work was supported by the National Natural Science Foundation of China under Grant No. 61301219, No. 61375028 and No. 61301295, the Natural Science Foundation of Jiangsu Province under Grant No. BK20130241. The authors would like to thank the reviewers for their valuable suggestions and comments.

References

- Boll S., 1979. Suppression of acoustic noise in speech using spectral subtraction [J]. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 27(2): 113-120. DOI: 10.1109/TASSP.1979.1163209
- Cappe O., 1994. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor [J]. *IEEE Transactions on Speech and Audio Processing*, 2(2): 345-349. DOI: 10.1109/89.279283
- Chen B., Loizou P.C., 2007. A Laplacian-based MMSE estimator for speech enhancement [J]. *Speech communication*, 49(2): 134-143. DOI: 10.1016/j.specom.2006.12.005
- Cohen I., Berdugo B., 2001. Speech enhancement for non-stationary noise environments [J]. *Signal processing*, 81(11): 2403-2418. DOI: 10.1016/S0165-1684(01)00128-1
- Ephraim Y., Malah D., 1984. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator [J]. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 32(6): 1109-1121. DOI: 10.1109/TASSP.1984.1164453
- Ephraim Y., Malah D., 1985. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator [J]. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 33(2): 443-445. DOI: 10.1109/TASSP.1985.1164550
- Gustafsson H., Nordholm S.E., Claesson I., 2001. Spectral subtraction using reduced delay convolution and adaptive averaging [J]. *Speech and Audio Processing, IEEE Transactions on*, 9(8): 799-807. DOI: 10.1109/89.966083
- Lim J S, Oppenheim A.V., 1979, Enhancement and bandwidth compression of noisy speech [J]. *Proceedings of the IEEE*, 67(12): 1586-1604. DOI: 10.1109/PROC.1979.11540
- Sim B L, Tong Y.C., Chang J.S., et al. 1998, A parametric formulation of the generalized spectral subtraction method [J]. *Speech and Audio Processing, IEEE Transactions on*, 6(4): 328-337. DOI: 10.1109/89.701361
- Udrea R.M., Vizireanu N.D., Ciochina S., 2008. An improved spectral subtraction method for speech enhancement using a perceptual weighting filter [J]. *Digital Signal Processing*, 18(4): 581-587. DOI: 10.1016/j.dsp.2007.08.002
- Virag N., 1999. Single channel speech enhancement based on masking properties of the human auditory system [J]. *Speech and Audio Processing, IEEE Transactions on*, 7(2): 126-137. DOI: 10.1109/89.748118
- Wiener N. 1949. *Extrapolation, interpolation, and smoothing of stationary time series* [M]. Cambridge, MA: MIT press. [3]