CONVEX COMBINATION FRAMEWORK FOR A PRIORI SNR ESTIMATION IN SPEECH ENHANCEMENT

Lara Nahma, Pei Chee Yong, Hai Huyen Dam, Sven Nordholm

Curtin University, Kent Street, Bentley, WA 6102, Australia
l.alibreesm@postgrad.curtin.edu.au { P.Yong, H.Dam, S.Nordholm }@curtin.edu.au

ABSTRACT

The paper proposes a convex combination fusion function based on a sigmoid function for the estimation of the a priori SNR in a speech enhancement framework with critical frequency band processing. The proposed method does not only eliminate the one frame delay generated by the well-known decision directed approach but also increases the adaptation speed during abrupt changes in the SNR estimation. As a result, the advantage of low musical noise has been maintained while more weak speech components have been preserved. Experimental results using instrumental and subjective measures also indicate improvement in speech quality compared to the reference methods.

Index Terms— Single channel speech enhancement, a priori SNR estimation, critical band processing.

1. INTRODUCTION

In many circumstances such as normal voice communications, the application of hearings aids and automatic speech recognition, the speech signals can be severely degraded due to different types of background noises. Therefore, the removal of the noise components from the degraded speech has been the main purpose of research work in the field of acoustic signal processing over the past few decades and it still remains an open problem today. The main task of a single channel speech enhancement is to reduce the background noise without generating musical artifacts while preserving the desired speech components [1, 2].

Most of the speech enhancement techniques are designed in the frequency domain where short time Fourier transform (STFT) is used as a tool to process the input data in overlapping blocks [3, 4, 5, 6]. Processing methods using STFT directly has a constant bandwidth which is different from the natural filtering operation of the human auditory system. Therefore, many recent research papers have employed the human auditory system in the noise reduction process in order to improve the speech quality and intelligibility [7, 8, 9, 10]. The human auditory system works as a banks of bandpass filters known as critical band filters, frequency components in the same critical band perceived equally by human auditory system [8]. In [8] the over-subtraction factor and the floor of the spectral subtraction gain function are adapted in time and frequency based on auditory masking properties along a Bark scale. In [10] an analysis-synthesis filterbank based on Gammatone filters has been employed in single channel noise reduction algorithm, which yields a comparable estimated speech quality as STFT based approach.

The gain function of a STFT based speech enhancement processing is usually a function of the a posteriori signal to noise ratio (SNR) and/ or the a priori SNR [11]. Among the many methods presented in the literature [6, 11, 12, 13, 14], the most common a priori SNR estimator is the decision directed (DD) approach proposed in [12], which consists of a weighted sum of two terms, the a priori SNR estimate from the previous frame and a maximum likelihood (ML) SNR estimate from the current frame. The main advantage of this approach is its ability to eliminate the musical noise artifacts by reducing the variance of the a priori SNR estimate especially during noise frames. The drawback is that it leads to a slow adaptation towards speech onsets and offsets since it uses a constant weighting factor close to unity. As the DD approach depends on the a priori SNR estimation in the previous frame, an extra one frame delay is obtained during speech transient and that leads to a degradation in the speech quality.

A modified decision directed approach (MDD) proposed in [6] overcomes the one frame delay problem by matching the current noisy speech spectrum with the a priori SNR estimate rather than the previous one. However, since the value of the constant weighting factor is close to one, the adaptation speed of the a priori SNR estimate between non-speech and speech frames is still not as fast as the a posteriori SNR's.

In this paper, we propose an improved a priori SNR estimation approach by utilizing a fusion function based on a sigmoidal shape in order to control the adaptation speed of the a priori SNR estimation. The fusion function can be viewed as a convex combination function that selects either the DD term or the ML estimate in the a priori estimate update. We observed that for positive SNR values the a priori estimate and the a posteriori are almost the same. Thus a flexibility to select either of the two terms for SNR values below or above a certain threshold is plausible, which can be achieved by the sigmoidal function. By utilizing a tuned threshold and sigmoid shape, an improved adaptation of the a priori SNR estimate is obtained, which results in better preservation of weak speech components. In conjunction with that, we utilize a critical band mapping from STFT analysis-resynthesis system in the speech enhancement framework for human perceptual processing and lower complexity.

The remainder of this paper is organized as follows. In section 2, a single channel speech enhancement framework with critical band processing is presented. Section 3 shows the decision directed based a priori SNR estimators. Section 4 presents the proposed a priori SNR estimation approach. Section 5 demonstrates the results of the experimental evaluation and section 6 concludes the paper.

2. CRITICAL BAND SPEECH ENHANCEMENT

Let $s(n)$ and $v(n)$ denote clean and additive noise signals respectively, and $y(n) = s(n) + v(n)$ is the noisy signal where the clean speech and noise signals are assumed to be uncorrelated. The time-frequency domain of the noisy signal can be obtained by applying the short time Fourier transform (STFT) as follows

$$Y(k,m) = S(k,m) + V(k,m) \tag{1}$$

where k is the frequency index and m is the time frame index. An analytical expression to describe the relationship between frequency f (in Hz) and critical band z (in bark scale) can be approximately formulated [15] by

$$f = 650 \sinh\left(\frac{z}{7}\right). \tag{2}$$

The number of critical bands I depends on the sampling frequency f_s and frequency band limits (minimum frequency=0 Hz and maximum frequency=$f_s/2$). The noisy spectrum is then expressed in terms of the critical band numbers i and frame index m by combining the FFT frequency bins into I critical bands as follows

$$Y_{cb}(i,m) = \sum_{k=0}^{K/2+1} M(i,k)\,|Y(k,m)| \tag{3}$$

where $i = [1, 2, \cdots, I]$, and $M(i,k)$ is the critical bandpass filter coefficients which can be defined [16] by

$$M(i,k) = \begin{cases} 10^{(z(k)-z_c(i)+0.5)} & z(k) < z_c(i) - 0.5 \\ 1 & z_c(i) - 0.5 < z(k) < z_c(i) + 0.5 \\ 10^{-2.5(z(k)-z_c(i)-0.5)} & z(k) > z_c(i) + 0.5 \end{cases} \tag{4}$$

where $z_c(i)$ denotes to the center frequency of the i^{th} critical band.

An estimate of the clean speech signal can be obtained by applying a spectral gain $G(i,m)$ to each time-frequency component of noisy spectrum

$$\hat{S}(i,m) = G(i,m) Y_{cb}(i,m) \tag{5}$$

where $0 < G(i,m) < 1$, which is a function of the a priori SNR $\xi(i,m)$ and/or the a posteriori SNR $\gamma(i,m)$. In this work, we have chosen Wiener gain as given by

$$G(i,m) = \frac{\xi(i,m)}{1 + \xi(i,m)} \tag{6}$$

which is a function of the a priori SNR. The a priori SNR $\xi(i,m)$ and the a posteriori SNR $\gamma(i,m)$ can be defined by

$$\xi(i,m) = \frac{\lambda_s(i,m)}{\lambda_v(i,m)} \tag{7}$$

and

$$\gamma(i,m) = \frac{|Y_{cb}(i,m)|^2}{\lambda_v(i,m)} \tag{8}$$

where $\lambda_v(i,m) = E\left[|V_{cb}(i,m)|^2\right]$, $\lambda_s(i,m) = E\left[|S_{cb}(i,m)|^2\right]$ are the power spectral density of noise and clean speech, respectively. Since only the noisy signal is given and both PSD of noise and speech are unknown, the a priori SNR and the a posteriori SNR have to be estimated.

The enhanced speech is then reconstructed by first transforming back into STFT form by

$$\hat{S}_s(k,m) = W(k,m) Y(k,m) \tag{9}$$

where $W(k,m)$ is obtained from $\mathbf{W} = \mathbf{A}\mathbf{g}$. Here, $\mathbf{g}$ is a vector given by $\mathbf{g} = [G(1,m), G(2,m), ..., G(I,m)]^T$ and $\mathbf{A}$ is the $K \times I$ rescaling matrix and can be defined by least square approximation as $\mathbf{A} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T$, with $\mathbf{M}$ denotes the matrix form of $M(i,k)$. From empirical findings, a better result can be obtained by simplifying the reconstruction matrix as

$$\mathbf{A} = \mathrm{diag}\left(\frac{1}{\mathbf{1M}}\right)\mathbf{M}^T$$

where $\mathbf{1}$ is $1 \times I$ row vector, then taking the inverse STFT of the enhanced speech spectrum by using the phase of the noisy observation and overlap-add method

$$\hat{s}(n) = \mathrm{IFFT}\left(\left|\hat{S}_s(k,m)\right| e^{j \arg(Y(k,m))}\right). \tag{10}$$

3. CONVENTIONAL A PRIORI SNR ESTIMATION

3.1. Decision Directed Approach (DD)

The most commonly used method to estimate the a priori SNR from noisy speech is the decision directed (DD) approach [12], which is updated based on the amplitude estimate from previous frames. Specifically, the estimate consists of two terms, where the first one indicates the amplitude estimator of the previous frame, and the second term represents Maximum Likelihood (ML) estimate of the a priori SNR as a function of the a posteriori SNR. This approach is defined by

$$\hat{\xi}_{\mathrm{DD}}(i,m) = \beta \frac{|\hat{S}(i,m-1)|^2}{\hat{\lambda}_{\mathrm{v}}(i,m-1)} + (1-\beta) P\left[\hat{\gamma}(i,m) - 1\right] \tag{11}$$

where $\hat{\lambda}_{\mathrm{v}}(i,m-1)$ is the noise PSD estimate at previous frames, P is the half wave rectification, and β denotes a weighting factor that controls the trade-off between the a priori SNR from previous frames and the a posteriori SNR estimate $\hat{\gamma}(i,m)$ at current frames.

The advantage of the DD approach is its capability to significantly reduce background noise while avoiding musical noise phenomenon, given that the weighting factor is a value very close to 1 ($\beta = 0.98$) [17]. In particular, during speech onsets when the a posteriori SNR is $\gg$1, the first term of the DD approach will correspond to the a posteriori SNR estimate from the preceding frames, such that

$$\begin{aligned} \hat{\xi}^{\uparrow\uparrow}_{\mathrm{DD}}(i,m) &= \beta \frac{G^2(i,m-1)|Y_{\mathrm{cb}}(i,m-1)|^2}{\hat{\lambda}_{\mathrm{v}}(i,m)} + (1-\beta) P[\hat{\gamma}(i,m) - 1] \\ &\approx \beta G^2(i,m-1)\hat{\gamma}(i,m-1) + (1-\beta) P[\hat{\gamma}(i,m) - 1]. \end{aligned}$$

Since β is close to 1, the second term of the above equation would not have a significant impact on the estimation process and could be assumed negligible. That means the a priori SNR estimate will follow the a posteriori SNR with a one frame delay during abrupt changes in SNR. When the a posteriori SNR is ≤ 1 (noise frame), the second term of the DD approach is equal to zero and the a priori SNR estimate corresponds to a scaled version of the a posteriori SNR. A priori SNR estimate can be defined by

$$\hat{\xi}^{\downarrow}_{\mathrm{DD}}(i,m) = \beta G^2(i,m-1)\hat{\gamma}(i,m-1).$$

From the above discussion it can be noticed the following: 1) As observed from the first term of both scenarios, the DD approach utilizes the estimate of the clean speech in the preceding frames instead of the current ones, which leads to a one frame delay. 2) During abrupt changes in SNR, given that the parameter β is chosen very close to 1, the second term gives little influence in the update of the a priori SNR estimate. In this case, the a priori SNR estimate becomes a weighted version of the a posteriori SNR estimate at previous frame.

3.2. Modified Decision Directed Approach (MDD)

The modified decision directed (MDD) approach is proposed in [18], which has resolved the problem of extra one frame delay by matching the a priori SNR and clean speech estimation with the current noisy frame instead of the previous one. So this approach reduces the speech transient distortion resulting from an extra frame delay. This approach can be defined by

$$\hat{\xi}_{\mathrm{MDD}}(i,m) = \beta \frac{G^2(i,m-1)\,|Y_{\mathrm{cb}}(i,m)|^2}{\hat{\lambda}_{\mathrm{v}}(i,m)} + (1-\beta)P[\hat{\gamma}(i,m)-1]\,. \quad (12)$$

This method is not a first order recursive averaging as it employs the current noisy frame instead of the previous one which can increase the variance of the a priori SNR estimate and increase the amount of musical noise. So in order to reduce the sensitivity of the estimated a priori SNR and reduce the musical noise, the magnitude square of noisy signal has been smoothed by using first order recursive smoothing procedure as given by [6]

$$\lambda_y(i,m) = \alpha_y \lambda_y(i,m-1) + (1-\alpha_y)\,|Y_{cb}(i,m)|^2 \quad (13)$$

where α_y is the smoothing constant. Then, λ_y is used to smooth the a posteriori SNR from (8). Similar to the DD approach, the drawback lies in the large smoothing constant which reduces the influence from the second term towards the a priori SNR update, causing it to become a scaled down estimate when compared to the true a priori SNR.

4. PROPOSED A PRIORI SNR ESTIMATION

In this work, we view the modified a priori SNR estimation not as a recursion but as a convex combination filter, where the weighting factor β is a fusion function that combines the weighting of the first and second term in Eq. ((12)) based on a sigmoid function. The sigmoid function is a function that varies between 0 to 1. It consists of two parameters, a to control the transition speed and c to determine the threshold for active speech signal versus only noise. As we know that the a priori SNR and a posterior SNR will be the same for high SNR, this can be utilized in the parameter selection for the sigmoid function. With this in mind, a fusion function $\hat{\beta}(i,m)$ is proposed based on the instantaneous SNR as given by

$$\hat{\beta}(i,m) = \left| \frac{1}{1+\exp\left[-a\left(\hat{\kappa}(i,m)-c\right)\right]} - \varsigma \right| \quad (14)$$

where $0 < \hat{\beta}(i,m) < 1$ and $\hat{\kappa}(i,m) = \hat{\gamma}(i,m) - 1$ is the instantaneous SNR, and ς denotes the parameter controlling the upper and lower limits of the fusion function values. The modified a priori SNR estimation approach is then defined by

$$\hat{\xi}_{\mathrm{prop}}(i,m) = \hat{\beta}(i,m)\frac{|G(i,m-1)Y_{\mathrm{cb}}(i,m)|^2}{\hat{\lambda}_v(i,m)} + (1-\hat{\beta}(i,m))P\left[\tilde{\gamma}(i,m)-1\right] \quad (15)$$

where $\tilde{\gamma}(i,m)$ is the a posteriori SNR estimate employing the smoothed estimate of the noisy speech from Eq. ((13)).

Figure 1 shows the variation of the proposed adaptive fusion function with the instantaneous SNR. For $\hat{\kappa}(i,m) \leq$ 0dB (noise frame), the second term is zero, fusion function takes the maximum value (close to 1), which means that the proposed method will have the same behavior as MDD approach during noise frames, and as a result, the a priori SNR estimate will corresponds to a scaled version of the a posteriori SNR estimate as given by

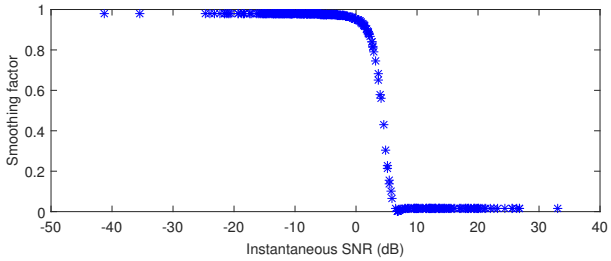

Fig. 1. A fusion function as a function of instantaneous SNR with $a = -2$, $c = 2.7$ and $\varsigma = 0.015$.

$$\hat{\xi}^{\downarrow}_{\mathrm{prop}}(i,m) = \hat{\beta}(i,m)G^2(i,m-1)\hat{\gamma}(i,m).$$

When 0 dB $< \hat{\kappa}(i,m) <$ 7 dB, the value of $\hat{\kappa}(i,m)$ decreases with the increment in the a posteriori SNR. As a consequence, the a priori SNR estimate in (12) will be given by

$$\hat{\xi}^{\uparrow}_{\mathrm{prop}}(i,m) = \hat{\beta}(i,m)G^2(i,m-1)\hat{\gamma}(i,m) + (1-\hat{\beta}(i,m))P[\tilde{\gamma}(i,m)-1]$$

where the second term in the above equation feeds the a priori SNR estimation in order to track the abrupt SNR changes. In this case, the a priori SNR estimate corresponds to a combination of a scaled amplitude estimate and the smoothed instantaneous SNR estimate. A higher influence of that second term results in the capability of the estimator to pick up more weak speech components, which is shown in the next section. For $\hat{\kappa}(i,m) >$ 7 dB, the fusion function takes the lowest value (close to 0) and as a result, the a priori SNR estimate corresponds to the smoothed a posteriori SNR as given by

$$\hat{\xi}^{\uparrow\uparrow}_{\mathrm{prop}}(i,m) = (1-\hat{\beta}(i,m))P[\tilde{\gamma}(i,m)-1].$$

5. EXPERIMENTAL EVALUATION

In order to evaluate the proposed a priori SNR estimation method, we performed experiments using speech sequences and noise from NOISEUS and NOISEX database respectively. The speech sequences are corrupted by pink noise at input SNRs of 0 dB and 10 dB. A sampling frequency of $f_s = 8000$ Hz with $K = 512$ frequency bins were used, and square root hanning window with 50% overlapping was applied for STFT analysis. The noisy spectrum was then processed by critical band processing and grouped into $I = 17$ critical bands as shown in Eq. (3). The overlap-add method were used to reconstruct the estimated signal. The value of the smoothing constant in Eq. (13) was chosen as $\alpha_y = 0.2$. The values of the parameters in Eq. (14) are chosen, respectively, as $a = -2$, $c = 2.7$ and $\varsigma = 0.015$. The fixed weighting constant for DD and MDD approaches was chosen as value of $\beta = 0.98$. In this work, noise PSD was estimated based on the probability of speech presence [19] for all the a priori SNR estimators.

Figure 2 shows an example of the behaviour of the a priori SNR estimated by DD, MDD and the proposed method. The a priori SNR estimates are displayed at 9^{th} critical band. During noise frames, it can be seen that the behaviour of the a priori SNR estimation of the evaluated methods are almost identical since at low SNR the proposed fusion function takes the maximum value (close to 1). As a result, the estimated a priori SNR estimation is a highly scaled version of the a posteriori SNR. However, during speech transition, $\hat{\xi}_{\mathrm{DD}}$

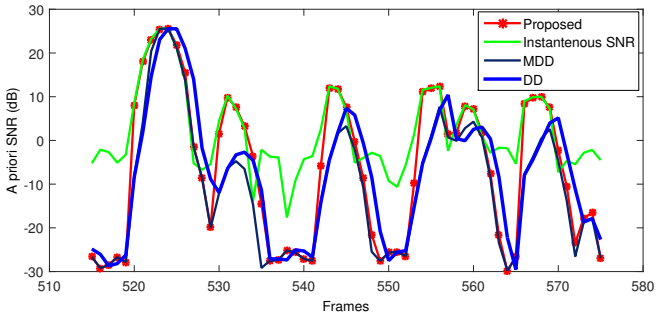

Fig. 2. Comparison of the a priori SNR estimation over a short time period between $\hat{\kappa}$ (green solid line) , $\hat{\xi}_{\mathrm{DD}}$ (blue solid line), $\hat{\xi}_{\mathrm{MDD}}$ (black solid line), and $\hat{\xi}_{\mathrm{prop}}$ (red solid line), at 9^{th} critical band, and 10 dB SNR white noise.

and $\hat{\xi}_{\mathrm{MDD}}$ follow the instantaneous SNR with delays, which result in degradation in speech quality. Meanwhile, $\hat{\xi}_{\mathrm{prop}}$ eliminates the delay and improve the adaptation speed to follow any abrupt changes in the a priori SNR estimation. Figure 3 depicts five spectrograms represent the clean signal, noisy speech corrupted by pink noise at 10 dB, enhanced speech estimated by DD, MDD and proposed method, respectively. It can be observed that the proposed method preserved more weak speech components when compared to DD and MDD approaches.

The a priori SNR estimators were also evaluated in the critical band based speech enhancement system in terms the musical noise measurement (KurtR), perceptual speech quality (PESQ), and segmental signal to noise ratio ($\mathrm{SNR}_{\mathrm{seg}}$) [6]. Lower values of KurtR with a larger PESQ and $\mathrm{SNR}_{\mathrm{seg}}$ indicate an improved performance.

Table 1 shows the mean performance of KurtR, PESQ, and $\mathrm{SNR}_{\mathrm{seg}}$ measurement for different a priori SNR estimation methods for NOIZEUS databse corrupted by pink noise. It can be clearly seen that the proposed method has an improvement in terms of PESQ and segmental SNR with higher scores comparing to DD and MDD methods. For musical noise measure KurtR, the proposed method maintain the advantage of DD and MDD approaches in eliminating the musical noise in low SNR case. Meanwhile at high SNR, the proposed method generates slightly higher musical noise comparing to the MDD approach because of the sensitivity of the a priori SNR estimation to abrupt changes.

	DD		MDD		Proposed	
SNR	0	10	0	10	0	10
KurtR	1.0484	1.5930	1.0046	1.1714	1.0301	1.2590
PESQ	1.8720	2.5864	1.8571	2.6246	2.0561	2.7439
$\mathrm{SNR}_{\mathrm{seg}}$	-0.2356	4.3209	-0.1361	4.5123	1.2753	5.7417

Table 1: Objective measurement mean performance for pink noise.

	0dB			10dB		
	DD	MDD	Proposed	DD	MDD	Proposed
Speech	2.50	2.63	3.13	2.63	3.50	3.88
Noise	1.88	2.25	2.88	2.50	3.13	3.50
Musical Noise	3.38	3.13	3.38	3.50	3.75	4.13

Table 2: Informal listening test results for pink noise.

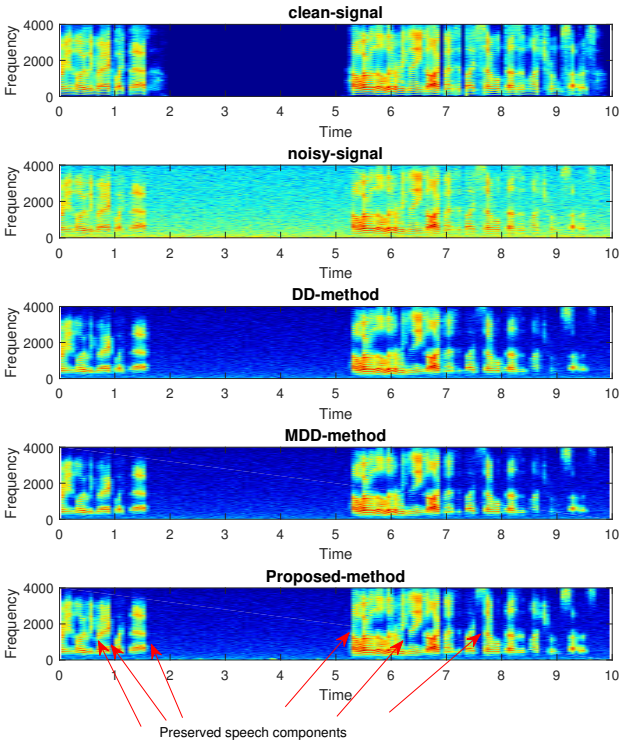

Fig. 3. Spectrogram comparison between different a priori SNR estimators at 10 dB input SNR with pink noise.

5.1. Listening test

A subjective evaluation was performed by organising an informal listening test with eight participants to validate the results from objective measurement [20]. The listeners were asked to rate the enhanced signals in three criteria: the audibility of speech, the amount of suppressed background noise and musical noise, using a 5 point scale. The results are tabulated in Table 2, where it can be clearly noticed that the listeners preferred the signals estimated by the proposed method more than the other aforementioned approaches.

6. CONCLUSION AND RELATION TO PRIOR WORK

We presented a novel method for the a priori SNR estimation based on a convex combination sigmoidal fusion function. Apart from combining the benefits of the conventional decision directed estimation (DD) in [12] and the modified decision directed (MDD) estimation in [6], where a fixed weighting factor has been used, the fusion function in this approach provides a much faster adaptation when there is a speech input. This improved tracking capability of the abrupt changes in SNR improves the preservation of weak speech components which is important for speech quality and intelligibility. The objective comparison and listening test both indicate that the proposed method is the preferred approach over DD and MDD methods.

In addition, we employed a critical band speech enhancement framework with a different speech reconstruction approach compared to the algorithms in literature that utilize a normal STFT structure [5, 7, 8] or an auditory filterbank [9, 10]. The study has been performed for Wiener filter gain function but can also be applied to other types of spectral gain functions.

7. REFERENCES

[1] Jacob Benesty, Shoji Makino, and Jingdong Chen, *Speech enhancement*, Springer Science & Business Media, 2005.

[2] Robert J McAulay and Marilyn L Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 2, pp. 137–145, 1980.

[3] Harald Gustafsson, Sven E Nordholm, and Ingvar Claesson, "Spectral subtraction using reduced delay convolution and adaptive averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 799–807, 2001.

[4] Israel Cohen and Baruch Berdugo, "Speech enhancement for non-stationary noise environments," *Signal processing*, vol. 81, no. 11, pp. 2403–2418, 2001.

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

[6] Pei Chee Yong, Sven Nordholm, and Hai Huyen Dam, "Optimization and evaluation of sigmoid function with a priori snr estimate for real-time speech enhancement," *Speech Communication*, vol. 55, no. 2, pp. 358–376, 2013.

[7] Dionysis E Tsoukalas, John N Mourjopoulos, and George Kokkinakis, "Speech enhancement based on audible noise suppression," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 6, pp. 497–514, 1997.

[8] Nathalie Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Transactions on speech and audio processing*, vol. 7, no. 2, pp. 126–137, 1999.

[9] Toshio Irino and Roy D Patterson, "A dynamic compressive gammachirp auditory filterbank," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2222–2232, 2006.

[10] Steffen Kortlang, Stephan D Ewert, and Timo Gerkmann, "Single channel noise reduction based on an auditory filterbank," in *14th International Workshop on Acoustic Signal Enhancement (IWAENC'14)*. IEEE, 2014, pp. 283–287.

[11] Cyril Plapous, Claude Marro, Laurent Mauuary, and Pascal Scalart, "A two-step noise reduction technique," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)*. IEEE, 2004, vol. 1, pp. I–289.

[12] Yariv Ephraim and David Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 32, no. 6, pp. 1109–1121, 1984.

[13] Rainer Martin, "An efficient algorithm to estimate the instantaneous snr of speech signals.," in *Eurospeech*, 1993, vol. 93, pp. 1093–1096.

[14] Israel Cohen, "Speech enhancement using a noncausal a priori snr estimator," *IEEE Signal Processing Letters*, vol. 11, no. 9, pp. 725–728, 2004.

[15] Manfred R Schroeder, Bishnu S Atal, and JL Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *the Journal of the Acoustical Society of America*, vol. 66, no. 6, pp. 1647–1652, 1979.

[16] Hynek Hermansky, "Perceptual linear predictive (plp) analysis of speech," *the Journal of the Acoustical Society of America*, vol. 87, no. 4, pp. 1738–1752, 1990.

[17] Olivier Cappé, "Elimination of the musical noise phenomenon with the ephraim and malah noise suppressor," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 345–349, 1994.

[18] Pei Chee Yong, Sven Nordholm, Hai Huyen Dam, and Siow Yong Low, "On the optimization of sigmoid function for speech enhancement," in *19th European Signal Processing Conference (EUSIPCO'11)*. IEEE, 2011, pp. 211–215.

[19] Timo Gerkmann and Richard C Hendriks, "Noise power estimation based on the probability of speech presence," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'11)*. IEEE, 2011, pp. 145–148.

[20] Yi Hu and Philipos C Loizou, "Evaluation of objective measures for speech enhancement," in *Interspeech*, 2006.