ON THE OPTIMIZATION OF SIGMOID FUNCTION FOR SPEECH ENHANCEMENT

Pei Chee Yong⋆, *Sven Nordholm*⋆ *Hai Huyen Dam*† *Siow Yong Low*⋆†

⋆Department of Electrical and Computer Engineering, Curtin University, Kent Street, Bentley, WA 6102, Australia
†Department of Mathematics and Statistics, Curtin University, Kent Street, Bentley, WA 6102, Australia
⋆†Department of Electrical and Computer Engineering, Curtin University, CDT 250, 98009 Miri, Sarawak, Malaysia
⋆peichee.yong@postgrad.curtin.edu.au ⋆S.Nordholm@curtin.edu.au
†H.Dam@exchange.curtin.edu.au ⋆†siowyong@curtin.edu.my

ABSTRACT

This paper develops a methodology to optimize sigmoid function parameters based on a weighted sum of two objective measures, which are the perceptual evaluation of speech quality (PESQ) measure and the log-likelihood ratio (LLR) measure. The sigmoid function has been investigated for speech enhancement as an alternative gain function to the conventional MMSE function and the spectral subtraction function. The benefit of using this function is that it has tunable parameters for both its slope and its mean. It also provides a potential to preserve more speech signal at high SNR level. The SNR estimate and the gain function impact the value of the objective measures such as PESQ and LLR, and provide varying subjective quality. Thus, by studying the relationship between the SNR estimate and the gain function, the performance of a single channel speech enhancement scheme can be optimized. Here, we aim to optimize the parameters of sigmoid function for different types of noise conditions and SNRs. Subjective listening tests demonstrate a significant improvement in the objective measures with proper choice of parameters.

1. INTRODUCTION

Many solutions have been proposed over the years to enhance speech signals under the influence of noise, particularly the spectral subtractive (SS) based algorithms [2, 3, 7] and the Log-MMSE methods [6]. It is well known that these solutions have a “classic trade-off” between SNR and speech distortion [12]. Moreover, SS based algorithms are prone to generating speech artifacts commonly known as “musical tones”, a phenomenon due to errors in noise statistics estimation [12]. The challenge in noise estimation is to control the update so it is not affected by the speech. Consequently, when speech is coming into the noise estimate, it will be biased. One of the solutions for noise estimation is by employing voice activity detection (VAD) based algorithms [5]. However, VAD often miss-detect speech onsets at low SNR and cause the noise estimate to be affected by the speech energy [4]. There are a multitude of methods suggested for the control of noise update [4, 8, 13]. All of them can be employed in this work but we assume an ideal estimation in order to highlight the work in this study.

The use of the sigmoid (SIG) function for speech enhancement has been proposed in [10]. The study showed that SIG function has benefits for hearing impaired people. A more comprehensive description of the use of the SIG function for speech enhancement is found in [1]. Even though both [1, 10] use the apriori SNR estimation in the gain function, they did not provide a clear picture on how the mean and the slope should be estimated. Sigmoid functions naturally maps the SNR estimate into a gain function between zero and one. Thus, in this paper, we propose to investigate the optimization of the parameters in SIG function in order to have a full use of the gain function and its applicability over a wide range of scenarios. More specifically, SIG function is optimized based on the perceptual evaluation of speech quality (PESQ) measure and the log-likelihood ratio (LLR) measure. Both of these measures correlates well with subjective listening evaluations when compared to other objective measures [9]. This has also been verified in subjective evaluations.

A main task for practical usage of speech enhancement techniques is how they map an SNR measure into the gain function which is applied on the input data. The posteriori SNR and the apriori SNR are the common SNR measures used for this task. The apriori SNR is more complex to use since it involves the access to the original speech signal or the need to provide an estimate of the original speech signal. Thus, instead of using the apriori SNR as in [1, 10], we propose to use the posteriori SNR estimate. This estimate provides an efficient way to optimize the mean and the slope of the SIG function as well as the noise floor.

The contributions in this paper include the direct use of the posteriori SNR in the SIG function and the establishment of the relationship between the SNR estimate and the gain functions. This study has direct impact for other speech enhancement techniques and gives a framework for finding new and improved enhancement functions.

2. SINGLE CHANNEL SPEECH ENHANCEMENT

Consider a noisy signal in discrete time domain to be expressed as

$$x(n) = s(n) + v(n) \tag{1}$$

where $s(n)$ is the clean speech signal and $v(n)$ is the additive noise. Both speech and noise are assumed to be uncorrelated. By using either a Short Time Fourier Transform (STFT) or a filter bank, the observed signal can be transformed to the frequency domain and can be represented as

$$X(k,m) = S(k,m) + V(k,m) \tag{2}$$

where $k = 1,2,...,K$ denotes the frequency bin index, $m = 1,2,...,M$ denotes the time frame index, K is the number of bands and M is the total number of frames. Here, $X(k,m)$, $S(k,m)$ and $V(k,m)$ denote the short-time spectral components of $x(n)$, $s(n)$ and $v(n)$, respectively.

The profound task for a single channel speech enhancement scheme is to estimate the original speech signal from the noisy speech signal. It involves estimating the speech signal spectrum by applying an adaptive gain function to every frequency bin and time index of the noisy signal spectrum

$$\hat{S}(k,m) = H(k,m)X(k,m) \tag{3}$$

where $\hat{S}(k,m)$ is the estimated speech signal spectrum and $H(k,m)$ is the suppression function. Although many noise suppression gain functions $H(k,m)$ have been proposed in literature, they commonly can be expressed as the function of an SNR estimate, such as the apriori SNR $\xi(k,m)$ and the posteriori SNR $\gamma(k,m)$ given by

$$\xi(k,m) = \frac{E\left\{|S(k,m)|^2\right\}}{E\left\{|V(k,m)|^2\right\}} \tag{4}$$

and

$$\gamma(k,m) = \frac{E\left\{|X(k,m)|^2\right\}}{E\left\{|V(k,m)|^2\right\}} \tag{5}$$

where $E\left\{|S(k,m)|^2\right\}$, $E\left\{|X(k,m)|^2\right\}$ and $E\left\{|V(k,m)|^2\right\}$ are the power spectrum of the clean speech, the noisy speech and the noise, respectively. The noise power spectrum can be estimated during speech pauses. The enhanced speech signal is obtained as

$$s(n) = \text{ISTFT}\{|\hat{S}(k,m)|.\exp(j\phi(X(k,m)))\} \tag{6}$$

where $|\mathbf{.}|$ denotes the magnitude of the estimated speech signal spectrum. The original noisy phase remains the same.

3. SNR ESTIMATION AND GAIN FUNCTION

In a short-time time-frequency interval, the background noise statistics can be assumed to be stationary whilst the speech statistics are non-stationary. By relating the posteriori SNR in Eq. (5) to this assumption, both the background noise spectra and the noisy speech spectra can be estimated by a long term-averaging and a short-term averaging, respectively. This can be done by using the first-order recursive smoothing filter with different time-related smoothing constants. In this work, the amplitude spectrum has been used since the power spectrum estimate tends to increase the musical noise in the enhanced speech. In this case, the posterior SNR can be estimated as

$$\hat{\gamma}(k,m) = \frac{\hat{\lambda}_X(k,m)}{\hat{\lambda}_V(k,m)} \tag{7}$$

where both the speech estimate $\hat{\lambda}_X(k,m)$ and the noise estimate $\hat{\lambda}_V(k,m)$ can be obtained as

$$\hat{\lambda}_X(k,m) = \alpha_X \hat{\lambda}_X(k,m-1) + (1-\alpha_X)|X(k,m)| \tag{8}$$

$$\hat{\lambda}_V(k,m) = \alpha_V \hat{\lambda}_V(k,m-1) + (1-\alpha_V)|V(k,m)| \tag{9}$$

and α_V and α_X are the smoothing constants for noise and speech, respectively. These constants are obtained from

$$\alpha_X = \exp\left(\frac{-2.2R}{t_x f_s}\right) \tag{10}$$

and

$$\alpha_V = \exp\left(\frac{-2.2R}{t_v f_s}\right) \tag{11}$$

where R is the frame rate, f_s is the sampling rate and t_x and t_v denotes the time averaging constant for both speech and noise, respectively. A longer averaging time in these estimates leads to a lower variance in the estimates. The distribution of the SNR estimate is determined by the short-term estimate, which is the speech estimate $\hat{\lambda}_X(k,m)$. In order to get a low variability of the SNR estimate, it is necessary for $\hat{\lambda}_X(k,m)$ to have relatively low variations. However, this is contradictive to maintaining fast variations from the speech signal since too much averaging in SNR estimate will produce muffled speech and give echo.

The SNR estimate can be mapped by the speech estimation gain function such that noise components will be attenuated and speech components will be maintained. In this case, the parameters of the gain function can be tuned. It is also natural for a gain function to operate between zero and one.

In this paper, two gain functions have been studied. We consider the well-known SS function, $H_{sp}(k,m)$, given by

$$H_{sp}(k,m) = \max\left(\varepsilon, 1-\beta\frac{1}{\hat{\gamma}(k,m)^p}\right) \tag{12}$$

where β and p are the oversubtraction factor and the power factor, respectively. The factor β is used to control the amount of speech spectral distortion, while the power factor $p < 1$ can be used to achieve high noise suppression under low SNR [11]. In addition to that, a lower p value can also allow more variations in the noise estimate. However, for $p < 1$, the gain function at high SNR region will also be attenuated and will not approach unity gain. Thus, the noise floor ε is introduced to control the amount of perceived residual noise and to avoid annoying musical noise [2]. The amount of musical noise is depending on the slope of the gain function and how often the SNR values during noise only periods come above the noise floor. A lower ε threshold can be chosen for a larger β value, which gives higher noise suppression with little musical noise, but at the same time suppresses low energy speech parts.

In order to provide a higher flexibility for speech enhancement and to control the shape of the gain function, the SIG gain function is investigated

$$H_{sig}(k,m) = \max\left(\varepsilon, \frac{1}{1+\exp\left[-a\left(\hat{\gamma}(k,m)-c\right)\right]}\right) \tag{13}$$

or

$$H_{sig}(k,m) = \max\left(\varepsilon, 1-\left[\frac{1-\tanh\left(\frac{a(\hat{\gamma}(k,m)-c)}{2}\right)}{2}\right]\right) \tag{14}$$

where a and c are the slope and the mean respectively. This function allows control of the SIG function's mean and slope. As such it provides a mean to suppress the noise as well as to maintain the unity gain at high SNR region.

The parameters of a gain function are optimized in terms of the level of noise suppression and the amount of musical noise generated. As such, the gain function has high sensitivity to changes in the SNR estimates when speech is active but have a constant value for noise only periods. According to Figure 1, which plots the PDF of SNR estimate for white noise at 938 Hz mapped with several gain functions, the SNR estimate at noise only periods is distributed approximately between 0.5 and 1.5. This means that attenuation

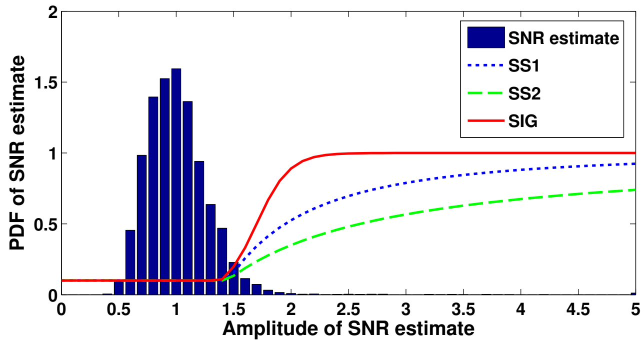

Figure 1: PDF of SNR estimate for white noise at 938 Hz mapped with (i) a spectral subtraction function with power spectrum estimates (SS1), (ii) a spectral subtraction function with amplitude spectrum estimates (SS2), and (iii) a sigmoid function (SIG)

shall only be performed when the SNR estimate falls within that region. For SIG function in Eq. (13) or (14), by mapping the gain function to the SNR estimate, the parameters a and c can be optimized for varying noise types and SNRs. The mean value c can be optimized based on objective evaluation and the SNR estimate, while the slope a is more of a challenge since a larger a indicates more speech distortion while a smaller a indicates lower noise reduction. Furthermore, the optimization problem is exacerbated by the type of noise that corrupts the noisy speech. Thus, it is important to understand how general a selection of parameters is.

4. OPTIMIZATION

The single channel speech enhancement is a system which contains a set of parameters that cannot be estimated ad hoc. This paper aims to optimize the parameters in the gain function based on a proposed multi-objective optimization algorithm, which can be formulated as

$$\max_{a} \quad w\mathrm{PESQ} - (1-w)\mathrm{LLR} \tag{15}$$

where w denotes the trade off between the two objective measures, $0 \leq w \leq 1$. Here, we choose $w = 0.5$. PESQ and LLR measures were used as the criteria of the optimization problem. PESQ measure has been proposed in ITU-T Recommendation P.862 and has recently been suggested to be more reliable than other traditional objective measures for speech quality [9]. It was implemented based on the steps in [12], which consists of pre-processing and filtering, time alignment, auditory transformation, computation of the difference between loudness spectra and time averaging of both reference and test signals. A higher PESQ score yields a better perceived speech quality [12]. Whilst LLR measure, which was reported in [9] as a reliable objective measure for speech distortion, is one of the speech quality objective measures that evaluate the dissimilarity of the all-pole models between the clean and the processed speech signals [12],

$$d_{LLR}\left(\vec{l}_{\hat{s}}, \vec{l}_s\right) = \frac{\vec{l}_{\hat{s}} \mathrm{R}_s \vec{l}_{\hat{s}}^T}{\vec{l}_s \mathrm{R}_s \vec{l}_s^T} \tag{16}$$

where $\vec{l}_s$ and $\vec{l}_{\hat{s}}$ are the linear predictive coding (LPC) coefficients of the clean speech signal and the processed speech signal respectively, and R_s is the autocorrelation matrix of the clean speech signal. A lower LLR score indicates a better speech quality.

We begin with the evaluation of (i) SS function with power spectrum estimates (SS1); (ii) SS function with amplitude spectrum estimates (SS2) and (iii) SIG function. By mapping the gain functions to the PDF of SNR estimate as shown in Figure 1, parameters β, a and c were chosen such that the gain functions would stay constant during noise only periods to advoid musical noise. Since the SNR estimate is distributed mainly between 0.5 and 1.5 during noise only periods, the gain functions should be constant up to SNR= 1.5 in order to minimize the amount of musical noise. With this in mind, the parameters for SS1 and SS2 are optimized based on the mapping of the gain functions to the fitted distribution in Figure 1, the objective measures (PESQ and LLR) and the informal listening tests. Hence, the optimal parameters are $\beta = 1.9$ for SS1 with $p = 2$, and $\beta = 1.3$ for SS2 with $p = 1$. This is consistent with the findings in [12] that the oversubtraction factor should range from 1.3 to 2.0 for low SNR conditions. The noise floor for the gain functions was set as a constant value $\varepsilon = -20$ dB.

For SIG function, from an exhaustive study based on the similar procedure used for SS function, we have obtained the optimized mean value as $c = 1.7$ from the distribution of SNR estimate. The slope can be set to $a = 7$ to achieve the same amount of noise suppression when compared to SS1 and SS2 as shown in Figure 1. However, in order to find the optimal performance of SIG function in different noise conditions and SNRs, we optimize a based on the objective function defined in Eq. (15).

5. EXPERIMENTAL RESULTS

The evaluation of the speech enhancement gain functions was done by using a database of noisy speech corpus named NOIZEUS [12]. The database contains 30 IEEE sentences produced by 3 male and 3 female speakers and corrupted by 8 different types of noise at global SNR levels of 0 dB, 5 dB, 10 dB and 15 dB. In this work, white noise, pink noise and factory noise were used for evaluation.

The recursive averaging constant were chosen as $\alpha_V = 0.9912$ with 1 second averaging time and $\alpha_X = 0.8636$ with a 60 millisecond averaging time. The frame size were chosen to be $K = 256$ with frame rate $R = 64$ corresponding to 75% overlap. A sampling frequency of $fs = 8000$ Hz and a 256 points Hamming Window were applied.

By using the cost function as defined in Eq. (15), the optimal points for SIG function (SIGopt) at different noise conditions and SNRs were searched. Tables 1, 2, and 3 show the mean value for 30 NOIZEUS sentences corrupted by white, pink and factory noise, respectively. In these three tables, SIGopt at different SNRs and noise conditions are compared to the corresponding results obtained from the the noisy signal, SS1 and SS2. The optimum points for SIG function with the corresponding slope value a, which were obtained from the optimization cost function, can be identified from Figure 2. As observed, with the flexibility of the parameters a and c, the optimal values of SIGopt are much higher than the results of both the SS functions. However, for pink noise at SNR = 0 dB, the performance of SIG is slightly lower than SS1 and SS2. Despite that, the tables show that there are still a significant improvement between SIGopt at 0 dB SNR

Table 1: Optimal value of objective function, white noise

SNR	Noisy	SS1	SS2	SIGopt
0	-0.0378	0.1575	0.1303	0.2726
5	0.1235	0.2569	0.2352	0.4082
10	0.3346	0.3930	0.3925	0.5670
15	0.5739	0.5307	0.5577	0.7264

Table 2: Optimal value of objective function, pink noise

SNR	Noisy	SS1	SS2	SIGopt
0	0.0855	0.4423	0.4005	0.2910
5	0.3123	0.5502	0.7288	0.7736
10	0.5579	0.6359	0.7903	0.9265
15	0.8035	0.7399	0.8814	1.0580

Table 3: Optimal value of objective function, factory noise

SNR	Noisy	SS1	SS2	SIGopt
0	0.4446	0.5559	0.6405	0.7713
5	0.6846	0.6764	0.7444	0.9111
10	0.9192	0.7960	0.8592	1.056
15	1.1379	0.9396	1.0018	1.217

pink noise and the noisy signal. The possible solution to increase the objective scores for SIG function at 0 dB SNR pink noise is to increase its mean value, c. Besides that, we can also observe that the optimal a becomes smaller with the increasing of SNR. This is because the cost function finds the optimal points of the gain function that minimize speech distortion. Although a larger a leads to a higher noise reduction, it will increase the amount of speech distortion in the enhanced speech signals. This indicates that the gain function can be less aggressive at higher SNR for lower speech distortion.

Tables 4 - 6 show the results from LLR measure whilst Tables 7, 8, and 9 show the PESQ scores. Similarly, for both individual objective measures, the performance of SIGopt is slightly better than the performance of SS1 and SS2 except for 0 dB SNR pink noise that acts as an outliner in these results. From the tables, it can be observed that both measures show similar behaviour in defining the quality of a speech signal in terms of speech distortion.

In order to validate the performance of objective measures, informal subjective subjective listening tests have been performed in factory noise at both 0 dB and 10 dB SNRs. The listening tests have been conducted with ten listeners. According to the amount of perceived noise and speech distortion, each listener was required to rate each signal from a scale between one and five: 5 = Excellent, 4 = Good, 3 = Fair and 1 = Bad. All listener results were averaged to represent the mean opinion score (MOS) as described in [12]. The test were conducted with closed-headphone. For each listeners, the applied procedures are: (1) clean speech and noisy speech were played and repeated upon request; (2) test signals were randomly played. The used paramters are: $b = 1.9$ for SS1, $b = 1.3$ for SS2, $a = 1.1, c = 1.7$ for SIG at 0 dB factory noise, and $a = 1.7, c = 1.7$ for SIG at 10 dB factory noise.

Table 10 shows the subjective MOS results of the noisy speech signal and the enhanced signals. As shown in the table, for factory noise at both 0 dB and 10 dB, a human listener prefers the SIGopt approach when compared to SS1 and SS2 methods. These results match the performance of

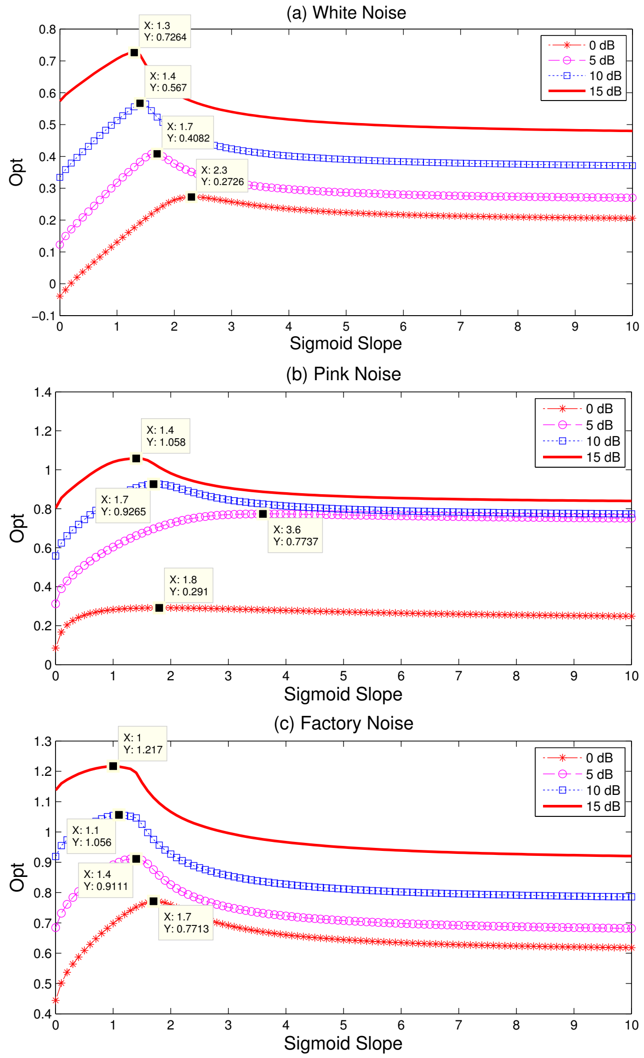

Figure 2: Cost function for optimization for different slope a with mean $c = 1.7$ at different SNRs: (a) white noise; (b) pink noise; and (c) factory noise

LLR and PESQ measures in Tables 6 and 9, respectively.

6. CONCLUSION AND FUTURE WORK

This paper presents a methodology to optimize the mean and the slope of the sigmoid function based on a proposed objective function. It was shown that the SNR estimate and the gain function impact the objective measures and provide varying subjective quality. The gain function parameters were designed such that during the noise only periods it provides a constant suppression thus avoiding annoying nonlinear artefacts (musical noise). This was done by mapping the function to the distribution of the SNR estimate. Optimization of the sigmoid function was done based on two widely used objective measures: PESQ and LLR. Experimental result shows that with proper choice of parameters, the sigmoid function can be optimized, which can enhance the quality of the noisy speech while maintaining more energy of the speech components when compared to the spectral subtraction function.

Table 4: LLR, white noise

SNR	Noisy	SS1	SS2	SIGopt
0	1.5978	1.4843	1.4949	1.4340
5	1.4978	1.4383	1.4352	1.3480
10	1.3708	1.3609	1.3471	1.2570
15	1.2245	1.2601	1.2386	1.1530

Table 5: LLR, pink noise

SNR	Noisy	SS1	SS2	SIGopt
0	1.4225	1.3221	1.2659	1.3100
5	1.2867	1.2429	1.1057	1.0750
10	1.1383	1.1794	1.0915	0.9941
15	0.9920	1.1021	1.0441	0.9189

Table 6: LLR, factory noise

SNR	Noisy	SS1	SS2	SIGopt
0	1.1480	1.2071	1.1066	1.0220
5	1.0047	1.0873	1.0339	0.9331
10	0.8680	0.9897	0.9498	0.8395
15	0.7519	0.9002	0.8682	0.7448

Future work will extend the optimization work based on the level of speech distortion and noise suppression in test signals. Evaluation can be done by using a combination of more objective measures as the cost function of the optimization problem. Measurement and reduction of musical noise will be included. More formal subjective listening tests will be conducted and compared with the results from objective measures. Investigation for more noise scenarios such as babble noise and music will be performed.

7. ACKNOWLEDGEMENT

This research was supported in part by Sensear Pty Ltd and Linkage Grant LP100100433 by Australian Research Council (ARC).

Table 7: PESQ scores of the speech signals, white noise

SNR	Noisy	SS1	SS2	SIGopt
0	1.5221	1.7994	1.7556	1.9800
5	1.7448	1.9521	1.9056	2.1680
10	2.0401	2.1470	2.1320	2.3960
15	2.3724	2.3214	2.3541	2.6170

Table 8: PESQ scores of the speech signals, pink noise

SNR	Noisy	SS1	SS2	SIGopt
0	1.5935	2.2067	2.0669	1.8920
5	1.9114	2.3432	2.5633	2.6260
10	2.2541	2.4512	2.6721	2.8530
15	2.5991	2.5820	2.8068	3.0580

Table 9: PESQ scores of the speech signals, factory noise

SNR	Noisy	SS1	SS2	SIGopt
0	2.0371	2.3188	2.3876	2.5680
5	2.3740	2.4401	2.5226	2.7630
10	2.7065	2.5818	2.6683	2.9700
15	3.0277	2.7794	2.8718	3.224

Table 10: MOS of the speech signals in factory noise

SNR	Noisy	SS1	SS2	SIGopt
0 dB	1.83	2.08	2.33	2.75
10 dB	3.33	3.33	2.83	3.50

REFERENCES

[1] M. Alam, D. O'Shaughnessy, and S. Selouani. Speech enhancement employing a sigmoid-type gain function with a modified a priori signal-to-noise ratio (SNR) estimator. *IEEE Canadian Conference on Electrical and Computer Engineering*, pages 000631–000636, 2008.

[2] M. Berouti, R. Schwartz, and J. Makhoul. Enhancement of speech corrupted by acoustic noise. *IEEE Int. Conf. Acoust., Speech, and Signal Process.*, 4:208–211, Apr. 1979.

[3] S. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. on Acoust., Speech, and Signal Process.*, 27(2):113–120, Apr. 1979.

[4] I. Cohen. Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging. *IEEE Trans. on Speech and Audio Process.*, 11(5):466–475, Sept. 2003.

[5] A. Davis, S. Nordholm, and R. Togneri. Statistical voice activity detection using low-variance spectrum estimation and an adaptive threshold. *IEEE. Trans. on Audio, Speech and Language Process.*, 14(2):412–424, March 2006.

[6] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Trans. on Acoust., Speech, Signal Process.*, 33:443–445, Apr. 1985.

[7] H. Gustafsson, S. Nordholm, and I. Claesson. Spectral subtraction using reduced delay convolution and adaptive averaging. *IEEE Trans. on Speech and Audio Process.*, 9(8):799–807, November 2001.

[8] S. Gustafsson, R. Martin, P. Jax, and P. Vary. A psychoacoustic approach to combined acoustic echo cancellation and noise reduction. *IEEE Trans. on Speech and Audio Process.*, 10(5):245–256, Jul. 2002.

[9] Y. Hu and P. Loizou. Evaluation of objective quality measures for speech enhancement. *IEEE Trans. on Audio, Speech, and Language Process.*, 16(1):229–238, Jan. 2008.

[10] Y. Hu, P. Loizou, N. Li, and K. Kasturi. Use of a sigmoidal-shaped function for noise attenuation in cochlear implants. *J. Acoust. Soc. Am.*, 122(4):128–134, Oct. 2007.

[11] J. Li, S. Sakamoto, S. Hongo, M. Akagi, and Y. Suzuki. Adaptive [beta]-order generalized spectral subtraction for speech enhancement. *Signal Processing*, 88(11):2764–2776, 2008.

[12] P. C. Loizou. *Speech Enhancement Theory and Practice*. CRC Press, 2007.

[13] R. Martin. Spectral subtraction based on minimum statistics. *European Signal Process. Conf.*, pages 1182–1185, Sept. 1994.