SPEECH ENHANCEMENT USING GENERALIZED MAXIMUM A POSTERIORI SPECTRAL AMPLITUDE ESTIMATOR

Yu-Cheng Su[1], Yu Tsao[1], Jung-En Wu[2], and Fu-Rong Jean[3]

[1] Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
[2] Biomedical Engineering Department, Georgia Institute of Technology, Atlanta, GA, USA
[3]Department of Electrical Engineering, National Taipei University of Technology, Taipei, Taiwan

ABSTRACT

This paper proposes a generalized maximum a posteriori spectral amplitude (GMAPA) algorithm to spectral restoration for speech enhancement. The proposed GMAPA algorithm dynamically adjusts the scale of prior information to calculate the gain function for spectral restoration. In higher signal-to-noise ratio (SNR) conditions, GMAPA adopts a smaller scale to prevent over-compensations that may result in speech distortions. On the other hand, in lower SNR conditions, GMAPA uses a larger scale to enable the gain function to more effectively remove noise components from noisy speech. We also develop a mapping function to optimally determine the prior information scale according to the SNR of speech utterances. Two standardized speech databases, Aurora-4 and Aurora-2, are used to conduct objective and recognition evaluations, respectively, to test the proposed GMAPA algorithm. For comparison, three conventional spectral restoration algorithms are also evaluated; they are minimum mean-square error spectral estimator (MMSE), maximum likelihood spectral amplitude estimator (MLSA), and maximum a posteriori spectral amplitude estimator (MAPA). The experimental results first confirm that GMAPA provides better objective evaluation scores than MMSE, MLSA, and MAPA in lower SNR conditions, with comparable scores to MLSA in higher SNR conditions. Moreover, our recognition results indicate that GMAPA outperforms the three conventional algorithms consistently over different testing conditions.

Index Terms—Speech enhancement, spectral restoration, MMSE, MAPA, MLSA, Generalized MAPA

1. INTRODUCTION

Speech enhancement aims to reduce background noise from noisy speech signals while preventing possible speech distortions. In speech processing systems, e.g., speech recognition, speech coder, and voice over IP, speech enhancement schemes are often used as a pre-processer to enhance the speech quality. Generally, speech enhancement algorithms can be divided into three categories, namely filtering, spectral restoration, and speech model techniques [1]. First for the filtering techniques, the goal is to design a filter or transformation that attenuates noise components to generate clean speech. Notable filtering techniques include time- and frequency-domain Wiener filters [1, 2, 3], and parametric Wiener filter [1]. For spectral restoration, a gain function is estimated to perform noise reductions in the frequency domain to obtain clean speech spectrums from the noisy speech spectrums. Successful examples include minimum mean square error spectral estimator (MMSE) [2, 4, 5, 6], minimum mean-square error log-spectral amplitude estimator (LSA) [7, 8, 9], maximum a posteriori spectral amplitude estimator (MAPA) [1,10, 11] and maximum likelihood spectral amplitude estimator (MLSA) [1, 12, 13]. Finally, speech model techniques combine human speech production models and speech reduction functions to remove noise components from noisy speech signals. Well-known speech models used for speech enhancement include the harmonic model [1, 14, 15, 16], the linear prediction (LP) model [1, 17, 18], and the hidden Markov model (HMM) [1, 19, 20, 21].

In this study, we focus our discussion on the spectral restoration algorithms for speech enhancement. Although many conventional spectral restoration techniques have shown effectiveness on noise reduction, they may have limited capability to achieve high performance for both high and low signal-to-noise ratio (SNR) conditions. For example, MAPA provides good noise reduction performance in low SNR conditions but possibly generate distortions due to over-compensations in high SNR conditions. On the other hand, MLSA maintains high quality in clean conditions along with limited noise attenuation capability in low SNR conditions. In this study, we propose a generalized maximum a posteriori spectral amplitude (GMAPA) estimator to overcome the limitation of the conventional techniques. GMAPA incorporates an adjustable scale of prior information to calculate the gain function for spectral restoration. In higher SNR conditions, GMAPA uses a smaller scale of prior information to maintain the quality of speech data; on the other hand in lower SNR conditions, GMAPA adopts a larger scale of prior information to enable the gain function to more effectively remove noise components. We also design a mapping function to determine the optimal scale value of the prior information according to the SNR of the speech utterance to be enhanced.

We conducted objective and recognition evaluations on Aurora-4 [22] and Aurora-2 [23, 24] speech databases, respectively, to test the proposed GMAPA algorithm. For objective evaluations, we tested speech distortion index (SDI) values [1, 25] and perceptual estimation of speech quality (PESQ) [26, 27, 28] using the speech data from Aurora-4. For recognition evaluations, we trained acoustic models and tested recognition on the Aurora-2 task. Our experimental results first indicate that GMAPA gives better objective evaluation results comparing to MMSE, MLSA, and MAPA in lower SNR conditions (under 15 dB), with comparable to MLSA in higher SNR conditions (over 20 dB). Moreover, the recognition evaluation results show that GMAPA outperforms the three algorithms consistently over different testing conditions.

2. SPECTRAL RESTORATION TECHNIQUES

In this section, we review the overall spectral restoration process and two notable algorithms, namely MLSA and MAPA.

2.1 Spectral Restoration Process

In the time domain, we consider a noisy speech signal, $y[n]$, as a sum of a clean speech, $s[n]$, and a noise signal, $v[n]$, as

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

where n denotes the time index. In the frequency domain, the noisy speech spectrum of the m-th frame, $Y[m,l]$, can be expressed as

$$Y[m,l] = S[m,l] + V[m,l], 0 \le l \le L-1\,, \tag{2}$$

where l is the frequency bin correspond to the frequency ω_l, where $\omega_l = \frac{2\pi l}{L}$, $l = 0, 1, \dots, L-1$; $m = 1, \dots, M$; $S[m,l]$ and $V[m,l]$ are speech and noise spectrums, respectively.

Fig. 1 shows the overall spectral restoration process, which can be divided into noise tracking and gain estimation stages. The noise tracking stage computes noise power from the noisy speech, $Y[m,l]$, to obtain a priori and a posteriori SNR statistics [29, 30]. Then the gain estimation stage calculates a gain function, $G[m,l]$, based on the computed a priori and a posteriori SNR statistics, to obtain enhanced speech, $\hat{S}[m,l]$, by filtering $Y[m,l]$ through $G[m,l]$. In the following discussion, we denote $Y[m,l]$, $S[m,l]$, $V[m,l]$, and $G[m,l]$, respectively, as Y, S, V, and G, for simplicity.

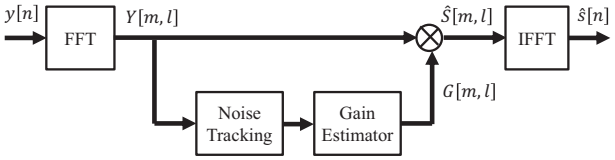


Fig. 1. Block diagram of a speech enhancement system.

By decomposing noisy and clean speech spectrums, Y and S in (2), into amplitude and phase parts, we have

$$Y = Y_k exp\left(j\theta_{Y_k}\right), \tag{3}$$

$$S = S_k exp\left(j\theta_{S_k}\right), \tag{4}$$

where $Y_k = |Y|$, $S_k = |S|$, $\theta_{Y_k} = \angle Y$, and $\theta_{S_k} = \angle S$. To restore S from Y, we first estimate the phase of clean speech spectrum by

$$exp\left(j\hat{\theta}_{S_k}\right) = \arg \min_{exp\left(j\hat{\theta}_{S_k}\right)} E\left[\left|exp\left(j\theta_{S_k}\right) - exp\left(j\hat{\theta}_{S_k}\right)\right|^2\right]. \tag{5}$$

Then, we have

$$exp\left(j\hat{\theta}_{S_k}\right) = exp\left(j\theta_{Y_k}\right). \tag{6}$$

Accordingly, the clean speech spectrum is estimated as

$$\hat{S} = \hat{S}_k exp\left(j\theta_{Y_k}\right). \tag{7}$$

More details about the phase estimation can be found in [1, 4].

2.2 MLSA and MAPA Algorithms

This section introduces two well-known gain estimators—MLSA and MAPA. The calculations of noise power and gain function are derived based on two assumptions: a) speech and noise signals are independent, and noise signal is additive; b) both speech and noise signals are random processes. Two statistics for spectral restoration, namely a priori SNR (ξ_k) and a posteriori SNR (γ_k), are defined as $\xi_k = \sigma_s^2/\sigma_v^2$ and $\gamma_k = Y_k^2/\sigma_v^2$, where $\sigma_s^2 = E[|S[m,l]|^2]$ and $\sigma_v^2 = E[|V[m,l]|^2]$. We denote ξ_k and γ_k, respectively, as ξ and γ.

2.2.1 MLSA Algorithm

For MLSA, the spectral amplitude, $\hat{S}_k$, is calculated by [1, 12, 13]

$$\hat{S}_k = \arg \max_{S_k} J_{MLSA}(S_k)\,, \tag{8}$$

where $J_{MLSA}(S_k)$ is the MLSA cost function and is defined as

$$J_{MLSA}(S_k) = \ln\{p[Y|S_k]\}. \tag{9}$$

By differentiating the MLSA cost function in Eq. (9) with respect to S_k and equating the result to zero, we can obtain

$$\hat{S}_k = \frac{Y_k + \sqrt{Y_k^2 - \sigma_v^2}}{2}. \tag{10}$$

Thus, the MLSA-based gain function, G_{MLSA}, is

$$G_{MLSA} = \frac{1 + \sqrt{(Y_k^2 - \sigma_v^2)/Y_k^2}}{2}. \tag{11}$$

2.2.2 MAPA Algorithm

MAPA estimates the spectral amplitude, $\hat{S}_k$, based on [1, 10, 11]

$$\hat{S}_k = \arg \max_{S_k} J_{MAPA}(S_k). \tag{12}$$

$J_{MAPA}(S_k)$ is the MAPA cost function and can be expressed as

$$J_{MAPA}(S_k) = \ln\{p[Y|S_k]p[S_k]\}. \tag{13}$$

By differentiating the MAPA cost function in Eq. (13) with respect to S_k and equating the result to zero, we can obtain

$$\hat{S}_k = \frac{\xi + \sqrt{\xi^2 + (1+\xi)\xi/\gamma}}{2(1+\xi)} Y_k. \tag{14}$$

Thus, the MAPA-based gain function, G_{MAPA}, can be expressed as

$$G_{MAPA} = \frac{\xi + \sqrt{\xi^2 + (1+\xi)\xi/\gamma}}{2(1+\xi)}. \tag{15}$$

3. GMAPA ESTIMATOR

In this section, we introduce the proposed GMAPA algorithm and the mapping function to determine the scale of prior information.

3.1 GMAPA Algorithm

For GMAPA, the spectral amplitude, $\hat{S}_k$, is calculated by

$$\hat{S}_k = \arg \max_{S_k} J_{GMAPA}(S_k). \tag{16}$$

$J_{GMAPA}(S_k)$ is the GMAPA cost function and can be expressed as

$$J_{GMAPA}(S_k) = \ln\{p[Y|S_k]\,(p[S_k])^{\alpha}\}. \tag{17}$$

By differentiating the GMAPA cost function in Eq. (17) with respect to S_k and equating the result to zero, we can obtain

$$\hat{S}_k = \frac{\xi + \sqrt{\xi^2 + (2\alpha-1)(\alpha+\xi)\xi/\gamma}}{2(\alpha+\xi)} Y_k, \tag{18}$$

where $\hat{S}_k$ is the enhanced speech. Thus, GMAPA gain function is

$$G_{GMAPA} = \frac{\xi + \sqrt{\xi^2 + (2\alpha - 1)(\alpha + \xi)\xi/\gamma}}{2(\alpha + \xi)}. \quad (19)$$

Please note that when setting α=1 in Eq. (17), $J_{GMAPA}(S_k)$ becomes $J_{MAPA}(S_k)$ in Eq. (13). When setting α=0 in Eq. (17), $J_{GMAPA}(S_k)$ becomes $J_{MLSA}(S_k)$ in Eq. (9). Table I summarizes the gain functions of MLSA, MAPA, and GMAPA. The gain function of the well-known MMSE [1] is also listed in the first row of Table I for comparison. For MMSE in Table I, $\delta = [\xi/(1+\xi)]\gamma$; $\Gamma(.)$ is the Gamma function; $I_0(.)$ and $I_1(.)$ are the modified Bessel function of the zero-order and first-order, respectively.

Table I. Gain functions of different algorithms

MMSE	$\Gamma\left(\frac{3}{2}\right)\frac{\sqrt{\delta}}{\gamma} exp\left(-\frac{\delta}{2}\right)\left[(1+\delta)I_0\left(\frac{\delta}{2}\right) + \delta I_1\left(\frac{\delta}{2}\right)\right]$
MLSA	$\frac{1+\sqrt{(Y_k^2 - \sigma_v^2)/Y_k^2}}{2}$
MAPA	$\frac{\xi + \sqrt{\xi^2 + (1+\xi)\xi/\gamma}}{2(1+\xi)}$
GMAPA	$\frac{\xi + \sqrt{\xi^2 + (2\alpha - 1)(\alpha + \xi)\xi/\gamma}}{2(\alpha + \xi)}$

3.2 Determining the Scale of Prior Information

We designed a sigmoid function [31, 32] to optimally determine the scale α for G_{GMAPA} in Eq. (17) for each utterance by

$$\alpha = \frac{\alpha_{max}}{1 + exp[-b(\bar{\gamma} - c)]}, \quad (20)$$

where α_{max} is the maximum value for α; b and c are coefficients of the sigmoid function; $\bar{\gamma}$ is the mean of a posteriori SNR for a given utterance, where $\bar{\gamma} = mean[\gamma^{(m)}, m = 1,2 \ldots M]$. Fig. 2 shows the designed function that determines the scale factor α based on $\bar{\gamma}$. Fig. 2 shows that the mapping function gives a larger α in lower SNR conditions and a smaller α in higher SNR conditions.

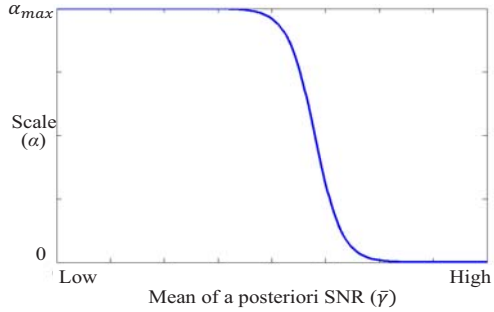


Fig. 2. The mapping function to determine α for each SNR.

4. EXPERIMENT

In this section, we discuss experimental setup and results. Since this study focuses on comparing different gain function estimators, we adopted a same noise tracking method, minimum statistics (MS) [33, 34], throughout the experiments for a fair comparison.

4.1. Experimental Setup

In this study, we performed objective evaluations and recognition tests for MMSE, MLSA, MAPA, and the proposed GMAPA.

4.1.1. Database

We conducted objective and recognition evaluations on the Aurora-4 [22] and Aurora-2 [23, 24] databases, respectively. For the objective evaluation, we selected 36 utterances from the Aurora-4 clean training set. Six speakers (three males and three females) pronounced these utterances. We intentionally selected the utterances in various lengths. With these clean utterances, we artificially simulated noisy speech utterances, using three noise types—white Gaussian noise (WGN), Babble, and Train, at seven SNRs—0 dB, 5 dB, 10 dB, 15 dB, 20 dB, 30 dB, and 40 dB. Accordingly, we prepared 21 different conditions, each including 36 utterances. For the recognition test, we used the clean condition training set in Aurora-2 to prepare a set of acoustic models. This training set contains 8440 utterances recorded in a clean condition. We tested recognitions using speech data in 60 conditions (10 noise types, at 0- to 20-dB SNR levels with the clean condition) in the Aurora-2 test set. Here, we report the average performance of each SNR level. In addition, we report an average result (denoted as Avg) for the average performance over 0-20 dB conditions. For Aurora-2 experiments, Avg is often used to present the overall performance.

For both objective and recognition evaluations, we decided α_{max}, b and c, in Eq. (20) using an additional development set. This set consisted of noisy data with given SNR levels. For each utterance in the development set, we calculated its mean of a posteriori SNR, $\bar{\gamma}$, and found the optimal α to this particular utterance. With the collection of $\bar{\gamma}$ and α pairs from all the utterances in the development set, we can determine α_{max}, b and c in Eq. (20). In the online, we first calculated $\bar{\gamma}$ for each testing utterance and then estimated the corresponding α by Eq. (20) to perform GMAPA.

4.1.2. Objective Evaluations

In this paper, we used speech distortion index (SDI) [1, 25] and perceptual estimation of speech quality (PESQ) [26, 27, 28] as the objective evaluations. SDI evaluates the distortion of the enhanced speech signal with respect to the original clean speech signal by

$$\text{SDI} = \frac{E[(s[n] - \hat{s}[n])^2]}{E[s^2[n]]}, \quad (21)$$

where $s[n]$ and $\hat{s}[n]$ are the clean and enhanced speech signals.

The PESQ evaluation was proposed by the International Telecommunication Union (ITU) to evaluate the mean opinion score (MOS) [26, 27, 28]. The PESQ value indicates the quality difference between the enhanced and clean speech signals. The score range of PESQ is from 0.5 to 4.5. A higher score implies that the enhanced speech signal is closer to the clean speech signal.

4.1.3. Recognition Tests

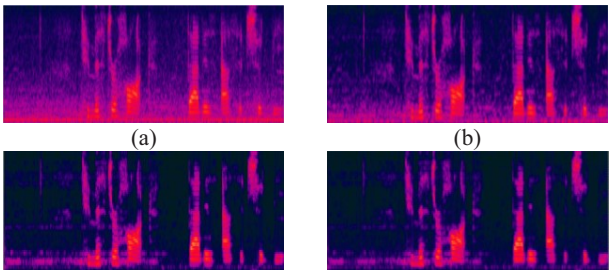
The complex back-end HMM topology was adopted to prepare the acoustic models [23]. Each digit model was characterized by 16 states, with 20 Gaussian mixtures per state. Silence and short pause models included three and one states, respectively, both with 36 Gaussian mixtures per state. Mel-frequency cepstral coefficients (MFCC) were used as speech features. Every feature vector comprised 13 static plus their first- and second-order time derivatives. Word error rate (WER) was used to evaluate the recognition performance. A lower WER indicated a better recognition result.

4.2. Experiment Results

This section presents our experimental results, including a spectrogram comparison, SDI and PESQ objective evaluations on Aurora-4, and the speech recognition results on Aurora-2.

4.2.1. Spectrogram Analysis

A spectrogram shows the spectral representations of a time-varying signal and is often used to analyze frequency and level properties of speech signals [35, 36]. Fig. 3 illustrates four spectrograms: (a) noisy speech at 5 dB SNR; (b), (c), and (d) enhanced speech using MLSA, MAPA, and GMAPA, respectively. The spectrograms are from a female voice saying "*To Mr. Hawke that is as it should be*".



(a) (b) (c) (d)

Fig. 3. Spectrograms: (a) noisy speech at 5 dB SNR; (b), (c), and (d) enhanced speech using MLSA, MAPA, and GMAPA, respectively.

From (a), (b), and (c) in Fig. 3, both MLSA and MAPA effectively removed noise components, while MAPA provided better noise reduction performance than MLSA. From (c) and (d), GMAPA showed even better noise reduction capability than MAPA.

4.2.2. Objective Evaluation

Tables II and III show the results of SDI and PESQ, respectively, of MLSA, MAPA, and GMAPA. The results of MMSE are also listed for comparison. Each value in the tables is an average of three noise types—babble, train, and WGN, under a specific SNR.

Table II. SDI values for four estimators in different SNRs.

SNR	MMSE	MLSA	MAPA	GMAPA
0 dB	0.4766	0.7915	0.4617	**0.4050**
5 dB	0.1500	0.2478	0.1455	**0.1317**
10 dB	0.0485	0.0777	0.0471	**0.0447**
15 dB	0.0162	0.0243	0.0159	**0.0158**
20 dB	0.0058	0.0076	**0.0057**	**0.0057**
30 dB	0.0011	**0.0007**	0.0011	0.0008
40 dB	0.0005	**0.0001**	0.0005	0.0002

Table III. PESQ values for four estimators in different SNRs.

SNR	MMSE	MLSA	MAPA	GMAPA
0 dB	2.0472	2.0162	2.0429	**2.0575**
5 dB	2.3669	2.3161	2.3673	**2.3814**
10 dB	2.6936	2.6366	2.6942	**2.7094**
15 dB	3.0429	2.9747	3.0430	**3.0556**
20 dB	3.4098	3.3406	3.4081	**3.4131**
30 dB	3.9914	**3.9954**	3.9837	3.9846
40 dB	4.2704	**4.3290**	4.2595	4.2919

From Tables II and III, the proposed GMAPA algorithm outperformed MMSE, MLSA, and MAPA in lower SNR conditions (0-10 dB conditions). The results indicated that by optimally determining α of the sigmoid function, GMAPA removed noise from the noisy speech signal more effectively and thus provided better performance for the objective evaluations. On the other hand, in higher SNR conditions (20-40 dB conditions), GMAPA provided better performance than MMSE and MAPA in most cases. This set of results confirmed that by using a smaller α, GMAPA overcame the limitations of MMSE and MAPA that over-compensated and thus generated distorted enhanced speech. However, the results of GMAPA in 30 dB and 40 dB conditions were slightly worse than MLSA. Since α used in GMAPA was determined based on estimated a posteriori SNR, $\bar{\gamma}$, and may not be always zero, it was reasonable that MLSA (with α=0) gave better results than GMAPA in high SNR conditions (SNR=30 dB and 40 dB).

4.2.3. Recognition Evaluation

Table IV lists the recognition results of MMSE, MLSA, MAPA, and GMAPA. The table also lists our baseline result for comparison. The baseline result is conducted by using the original MFCC for testing with no speech enhancement performed.

Table IV. WER values for different algorithms in different SNRs.

SNR	Baseline	MMSE	MLSA	MAPA	GMAPA
0dB	85.15	75.38	82.19	75.79	**72.21**
5dB	63.49	47.28	56.05	47.75	**43.13**
10dB	34.86	22.7	29.13	22.89	**19.62**
15dB	14.41	9.62	12.42	9.56	**7.72**
20dB	4.91	3.62	4.63	3.55	**3.03**
Clean	0.36	0.39	0.34	0.36	**0.33**
Avg	40.56	31.72	36.88	31.91	**29.14**

From Table IV, we first observed that all the four spectral restoration algorithms achieved lower WERs than the baseline in most SNR conditions. Since we applied a same algorithm for both training and testing speech data, this set of results confirmed that all the four algorithms can effectively reduce the mismatch between training and testing conditions for speech recognition. Next when comparing the four algorithms, GMAPA gave the lowest WERs over all SNR conditions. Especially, it was noted that from Tables II and III, MLSA achieved better SDI and PESQ performances than GMAPA in higher SNR conditions (30 dB and 40 dB); the proposed GMAPA still outperformed MLSA in the recognition results for high SNR conditions (20 dB and clean) in Table IV. The results suggest that GMAPA has better capability to handle the mismatch issue and can be more suitable for speech recognition systems than the other three algorithms. Finally when comparing to the baseline, GMAPA provided a clear 28.16% average WER reductions (from 40.56 % to 29.14 %) over 0-20 dB SNR conditions.

5. CONCLUSION

In this paper, we proposed the GMAPA algorithm to spectral restoration for speech enhancement. The GMAPA algorithm used a scale factor, α, to determine the prior information for calculating the gain function. A mapping function was also designed to optimally determine α according to the estimated SNR level of the noisy speech. We conducted both objective and recognition evaluations to test the proposed GMAPA algorithm. The objective evaluation results from both SDI and PESQ confirmed that GMAPA outperformed MMSE, MLSA, and MAPA under lower SNR conditions, with achieving similar scores to MLSA under higher SNR conditions. Meanwhile, the recognition results showed that the proposed GMAPA algorithm provided better performance than the other three algorithms consistently over different SNR conditions.

6. ACKNOWLEDGEMENT

This work was supported by the National Science Council of Taiwan under contracts NSC101-2221-E-001-020-MY3.

7. REFERENCE

[1] J. Chen, *Fundamentals of Noise Reduction in Spring Handbook of Speech Processing*, Springer, 2008.

[2] P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," *in Proc. ICASSP*, pp. 629-632, 1996.

[3] E. Hänsler and G. Schmidt, *Topic in Acoustic Echo and Noise Control*, Chapter 9, Springer, 2006.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 32, pp. 1109-1121, 1984.

[5] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 845-856, 2005.

[6] J. H. L. Hansen, V. Radhakrishnan, and K. H. Arehart, "Speech enhancement based on generalized minimum mean square error estimators and masking properties of the auditory system," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 2049-2063. 2006.

[7] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, pp. 443-445, 1984.

[8] D. Malah, R. V. Cox, and A. J. Accardi, "Tracking speech-presence uncertainty to improve speech enhancement non-stationary noise environments," *in Proc. ICASSP,* pp. 789-792, 1999.

[9] A. Das and J. H. L. Hansen "Constrained iterative speech enhancement using phonetic classes," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, pp. 1869-1883, 2012.

[10] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," *EURASIP Journal on Applied Signal Processing*, pp. 1110-1126, 2005.

[11] S. Suhadi, C. Last, and T. Fingscheidt, "A data-driven approach to a priori SNR estimation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, pp. 186-195, 2011.

[12] R. J. McAulay and M.L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 28, pp. 137-145, 1980.

[13] U. Kjems and J. Jensen, "Maximum likelihood based noise covariance matrix estimation for multi-microphone speech enhancement," *in Proc. EUSIPCO*, pp. 295-299, 2012.

[14] R. H. Frazier, S. Samsam, L. D. Braida, and A. V. Oppenheim, "Enhancement of speech by adaptive filtering," *in Proc. ICASSP*, pp.251-253, 1976.

[15] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, pp. 744–754, 1986.

[16] T. F. Quatieri and R. J. McAulay. "Shape-invariant time- scale and pitch modifications of speech," *IEEE Transactions on Signal Processing*, vol. 40, pp. 497–510, 1992.

[17] J. Makhoul, "Linear prediction: A tutorial review," *in Proc. IEEE*, vol. 63, pp. 561-580, 1975.

[18] B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals and subjective error criteria," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-27, pp. 247–254, 1979.

[19] L. R. Rabiner, B. H. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, pp. 4-16, 1986.

[20] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *in Proc. IEEE*, vol. 77, pp. 257-286, 1989.

[21] Y. Ephraim, "Statistical-model-based speech enhancement systems," *Proc. IEEE*, vol. 80, pp. 1526-1555, 1992.

[22] N. Parihar and J. Picone, "Aurora working group: DSR front end LVCSR evaluation AU/384/02," *Institute for Signal & Information Processing Report*, 2002.

[23] D. Pearce and H. G. Hirsch, "The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions," *ICSA ITRW ASR2000*, 1999.

[24] D. Macho, L. Mauuary, B. Noe, Y. M. Cheng, D. Ealey, D. Jouver, H. Kelleher, D. Pearce, and F. Saadoun, "Evaluation of a noise-robust DSR front-end on Aurora databases," *in Proc. ICSLP*, pp. 17–20, 2002.

[25] J. Chen, J. Benesty, Y. Huang, S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 1218–1234, 2006.

[26] ITU-T Recommendation P.862, *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, 2001.

[27] A. W. Rix, J. G. Beerends, M. P. Hollier and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ) – a new method for speech quality assessment of telephone networks and codecs," *in Proc. ICASSP*, pp. 749-752, 2001.

[28] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, pp. 229-238, 2008.

[29] I. Cohen, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Processing Letters*, vol. 9, pp. 12-15, 2002.

[30] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 11, pp. 466-475, 2003.

[31] C. Alippi and G. Storti-Gajani, "Simple approximation of sigmoidal functions: realistic design of digital neural networks capable of learning," *in Proc. ISCAS*, pp. 1505-1508, 1991.

[32] M. Zhang, S. Vassiliadis and J. G. Delgado-Frias, "Sigmoid generators for neural computing using piecewise approximations," *IEEE Transactions on Computers*, vol. 45, pp. 1045-1049, 1996.

[33] R. Martin, "Spectral subtraction based on minimum statistics," *in Proc. EUSIPCO*, pp. 1182-1185, 1994.

[34] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, pp. 504-512, 2001.

[35] J. L. Flanagan, *Speech Analysis, Synthesis and Perception*, Springer- Verlag, 1972.

[36] S. Haykin, *Advances in Spectrum Analysis and Array Processing*, Prentice-Hall, 1991.