

Speech Enhancement Using a Noncausal *A Priori* SNR Estimator

Israel Cohen, *Senior Member, IEEE*

Abstract—In this letter, we propose a noncausal estimator for the *a priori* signal-to-noise ratio (SNR), and a corresponding noncausal speech enhancement algorithm. In contrast to the decision-directed estimator of Ephraim and Malah, the noncausal estimator is capable of discriminating between speech onsets and noise irregularities. Onsets of speech are better preserved, while a further reduction of musical noise is achieved. Experimental results show that the noncausal estimator yields a higher improvement in the segmental SNR, lower log-spectral distortion, and better Perceptual Evaluation of Speech Quality scores (PESQ, ITU-T P.862).

Index Terms—Parameter estimation, speech enhancement, time-frequency analysis.

I. INTRODUCTION

A N IMPORTANT parameter of numerous speech enhancement algorithms is the *a priori* signal-to-noise ratio (SNR) [1]–[6]. The most practical and computationally efficient approach to determine this parameter is to use the decision-directed estimator of Ephraim and Malah [2]. Unfortunately, the decision-directed estimation approach applies a constraint on the response to speech onsets. The *a priori* SNR estimator cannot respond too fast to an abrupt increase in the instantaneous SNR, since it inevitably yields an increase in the level of musical residual noise. Furthermore, the decision-directed approach is not based on a statistical model, but is heuristically motivated. The *a priori* SNR estimate heavily relies on the strong time-correlation between successive speech spectral magnitudes.

Recently, we introduced a statistical model for speech enhancement, which takes into account the correlation between speech spectral components in the time-frequency domain [7]. We derived a recursive estimator for the *a priori* SNR, based on the statistical model, and showed its close relation to the decision-directed estimator. The recursive *a priori* SNR estimator degenerates, as a special case, to a “decision-directed” estimator with a *time-varying frequency-dependent* weighting factor. The main consequence is that the estimators for the *a priori* SNR and the speech spectral components are both based on a statistical model. However, the performance of the corresponding speech enhancement algorithm, compared with that resulting from the decision-directed approach, is only slightly better.

Manuscript received October 16, 2003; December 14, 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mark Hasegawa-Johnson.

The author is with the Department of Electrical Engineering, The Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: icohen@ee.technion.ac.il).

Digital Object Identifier 10.1109/LSP.2004.833478

In this letter, we extend our recursive estimation approach to noncausal estimation, which is useful in applications that can tolerate a delay of at least 0.1 s in the estimated signal (e.g., digital voice recording and surveillance). A noncausal estimator for the *a priori* SNR is derived in agreement with the model assumptions and the estimation of the speech spectral components. The noncausal estimator employs future spectral measurements to better predict the spectral variances of the clean speech. In contrast with the decision-directed estimator, the noncausal estimator, having a few subsequent spectral measurements at hand, is capable of discriminating between speech onsets and noise irregularities. Experimental results show that compared with the decision-directed estimator, the noncausal estimator yields a higher improvement in the segmental SNR, lower log-spectral distortion, and better Perceptual Evaluation of Speech Quality scores (PESQ, ITU-T P.862). The advantages of the noncausal estimator are particularly perceived during speech onsets and noise only frames. Onsets of speech are better preserved, while a further reduction of musical noise is achieved.

This letter is organized as follows. In Section II, we formulate the speech enhancement problem. In Section III, we present the statistical model. In Section IV, we derive a noncausal estimator for the *a priori* SNR, and introduce the speech enhancement algorithm. In Section V, we demonstrate the improved performance of the noncausal estimation, compared with the decision-directed approach.

II. PROBLEM FORMULATION

Let x and d denote speech and uncorrelated additive noise signals, and let $y = x + d$ represent the observed signal. Applying the short-time Fourier transform (STFT) to the observed signal, we have in the time-frequency domain

$$Y(k, \ell) = X(k, \ell) + D(k, \ell) \quad (1)$$

where k is the frequency-bin index ($k = 0, 1, \dots, K - 1$) and ℓ is the time frame index ($\ell = 0, 1, \dots$). Let $d(X, \hat{X})$ be a given distortion measure between X and \hat{X} , and let $\mathcal{Y}_0^{\ell+L} = \{Y(k, \ell') \mid 0 \leq k \leq K - 1, \ell' \leq \ell + L\}$ represent the set of spectral measurements up to frame $\ell + L$, where $L(L > 0)$ denotes an admissible time delay in frames. The speech enhancement problem is formulated as finding a noncausal estimator $\hat{X}(k, \ell)$, which minimizes the conditional expected value of the distortion measure given $\mathcal{Y}_0^{\ell+L}$

$$\hat{X}(k, \ell) = \arg \min_{\hat{X}} E \left\{ d[X(k, \ell), \hat{X}] \mid \mathcal{Y}_0^{\ell+L} \right\}. \quad (2)$$

Distortion measures that are of particular interest in speech enhancement applications are the mean squared-error (MSE) [9], the spectral magnitude MSE [2], the log-spectral magnitude MSE [7], and the spectral power MSE [10]. The last three distortion measures are insensitive to the spectral phase estimation error, but can be combined with an estimator for the complex exponential of the phase, constrained to not affecting the spectral magnitude estimate [2].

III. SPEECH SPECTRAL MODEL

A statistical model, which takes into account the time-frequency correlation of speech signals, was recently proposed in [8]. Accordingly,

- 1) The noise spectral components $\{D(k, \ell)\}$ are statistically independent zero-mean complex Gaussian random variables. The real and imaginary parts of $D(k, \ell)$ are independent and identically distributed (IID).
- 2) The speech spectral phases $\{\angle X(k, \ell)\}$ are IID uniform random variables on $[-\pi, \pi]$.
- 3) For fixed k and ℓ , a speech spectral component $X(k, \ell)$ is *conditionally* a zero-mean complex Gaussian random variable given its variance $\lambda_X(k, \ell) \triangleq E\{|X(k, \ell)|^2\}$.
- 4) The sequence of speech spectral variances $\{\lambda_X(k, \ell) | \ell = 0, 1, \dots\}$, where $\lambda_X(k, \ell) \triangleq E\{|X(k, \ell)|^2\}$, is a random process, generally correlated with the speech spectral magnitudes $A(k', \ell') \triangleq |X(k', \ell')|$. However, given $\lambda_X(k, \ell)$, $A(k, \ell)$ is statistically independent of $A(k', \ell')$ for all $(k', \ell') \neq (k, \ell)$.

This statistical model enables a derivation of efficient estimators for the speech spectral components and the *a priori* SNR, while keeping the resulting algorithms simple [7]

IV. NONCAUSAL SPECTRAL ENHANCEMENT

In this section, we derive noncausal estimators for $X(k, \ell)$ and the *a priori* SNR, and introduce a noncausal speech enhancement algorithm. We assume knowledge of the noise spectrum, which in practice can be estimated by using the *Minima Controlled Recursive Averaging* approach [11]. For notational simplicity, we often omit the arguments k and ℓ when there is no confusion.

Let $\lambda_{X| \ell+L} \triangleq E\{|X|^2 | \mathcal{Y}_0^{\ell+L}\}$ denote the conditional variance of X given the noisy measurements $\mathcal{Y}_0^{\ell+L}$. Let $p(X|Y, \lambda_{X| \ell+L})$ denote the conditional pdf of X given Y and $\lambda_{X| \ell+L}$. Then, we consider an estimator for X which minimizes the expected distortion given $\mathcal{Y}_0^{\ell+L}$ and $\lambda_{X| \ell+L}$. The proposed statistical model implies

$$p(X | \mathcal{Y}_0^{\ell+L}, \lambda_X) = p(X|Y, \lambda_X). \quad (3)$$

Therefore, we obtain the estimator \hat{X} from

$$\begin{aligned} \min_{\hat{X}} \int d(X, \hat{X}) p(X | \mathcal{Y}_0^{\ell+L}, \lambda_{X| \ell+L}) dX \\ \approx \min_{\hat{X}} \int d(X, \hat{X}) p(X|Y, \lambda_{X| \ell+L}) dX \end{aligned} \quad (4)$$

That is, given the set of noisy measurements $\mathcal{Y}_0^{\ell+L}$, we first derive an estimate for the clean speech spectral variance $\lambda_{X| \ell+L}$. Subsequently, the estimation problem for the speech spectral component X reduces to that of estimating X from Y alone, assuming knowledge of the variance of X . The latter problem, when the *a priori* SNR is defined appropriately, is in fact the classical spectral enhancement problem as formulated by Ephraim and Malah [2], [9]. As a result, an estimate for X is obtained by applying a spectral gain function to the corresponding noisy spectral component

$$\hat{X} = G(\xi, \gamma)Y \quad (5)$$

where the *a priori* and *a posteriori* SNRs are defined, respectively, by

$$\xi \triangleq \frac{\lambda_{X| \ell+L}}{\lambda_D}, \quad \gamma \triangleq \frac{|Y|^2}{\lambda_D} \quad (6)$$

and where $\lambda_D(k, \ell) \triangleq E\{|D(k, \ell)|^2\}$ denotes the noise spectral variance.

The specific expression for the spectral gain function $G(\xi, \gamma)$ depends on the particular choice of a distortion measure $d(X, \hat{X})$. For squared-error distortion, the gain function is a Wiener filter given by [9]

$$G_{SE}(\xi) = \frac{\xi}{1 + \xi}. \quad (7)$$

In case of combining the spectral amplitude, the log-spectral amplitude, or the spectral power distortion measures with the constrained estimator for the complex exponential of the phase, the gain functions can respectively be written as [2], [7], [10]

$$G_{SA}(\xi, \gamma) = \frac{\sqrt{\pi v}}{2\gamma} \left[(1+v)I_0\left(\frac{v}{2}\right) + vI_1\left(\frac{v}{2}\right) \right] \exp\left(-\frac{v}{2}\right) \quad (8)$$

$$G_{LSA}(\xi, \gamma) = \frac{\xi}{1 + \xi} \exp\left(\frac{1}{2} \int_v^\infty \frac{e^{-t}}{t} dt\right) \quad (9)$$

$$G_{SP}(\xi, \gamma) = \sqrt{\frac{\xi}{1 + \xi} \left(\frac{1}{\gamma} + \frac{\xi}{1 + \xi} \right)} \quad (10)$$

where $I_0(\cdot)$ and $I_1(\cdot)$ denote the modified Bessel functions of zero and first order, respectively, and v is defined by $v \triangleq \xi\gamma/(1 + \xi)$.

Let $\lambda'_{X| \ell+L}(k, \ell) \triangleq E\{|X(k, \ell)|^2 | \mathcal{Y}_0^{\ell+L} \setminus \{Y(k, \ell)\}\}$ denote the conditional variance of X given $\mathcal{Y}_0^{\ell+L}$ excluding the noisy measurement Y . Let $\lambda'_{X| [\ell, \ell+L]}(k, \ell) \triangleq E\{|X(k, \ell)|^2 | \mathcal{Y}_\ell^{\ell+L} \setminus \{Y(k, \ell)\}\}$ denote the conditional variance of X given the noisy measurements $\mathcal{Y}_\ell^{\ell+L} \setminus \{Y\}$. Then, an estimate for $\lambda_{X| \ell+L}$ can be obtained by computing the conditional variance of X given Y and $\lambda'_{X| \ell+L}$

$$\hat{\lambda}_{X| \ell+L} = E\{|X|^2 | \lambda'_{X| \ell+L}, Y\}. \quad (11)$$

This is obtained by applying the gain function $G_{\text{SP}}(\xi, \gamma)$ to Y , with $\xi \equiv \hat{\lambda}'_{X|\ell+L}/\lambda_D$, and computing the squared absolute value of the result¹

$$\begin{aligned}\hat{\lambda}_{X|\ell+L} &= G_{\text{SP}}^2\left(\frac{\hat{\lambda}'_{X|\ell+L}}{\lambda_D}, \gamma\right) |Y|^2 \\ &= \frac{\hat{\lambda}'_{X|\ell+L}}{\lambda_D + \hat{\lambda}'_{X|\ell+L}} \left(\lambda_D + \frac{\hat{\lambda}'_{X|\ell+L} |Y|^2}{\lambda_D + \hat{\lambda}'_{X|\ell+L}} \right).\end{aligned}\quad (12)$$

To obtain an estimate for $\lambda'_{X|\ell+L}(k, \ell)$, we employ the estimates $\hat{A}(k, \ell - 1)$ and $\{\hat{\lambda}_{X|\ell+L-1}(k, \ell - 1)\}_{k=0}^{K-1}$ from the previous frame, and derive an estimate for $\lambda'_{X|[\ell, \ell+L]}(k, \ell)$ from the measurements $\mathcal{Y}_\ell^{\ell+L} \setminus \{Y\}$. Suppose an estimate $\hat{\lambda}'_{X|[\ell, \ell+L]}(k, \ell)$ is given, and let b denote a normalized window function of length $2w + 1$, i.e., $\sum_{i=-w}^w b(i) = 1$. We propose to propagate the estimates from frame $\ell - 1$ to frame ℓ by

$$\begin{aligned}\hat{\lambda}'_{X|\ell+L}(k, \ell) &= \max \left\{ \mu \hat{A}^2(k, \ell - 1) + (1 - \mu) \right. \\ &\quad \times \left[\mu' \sum_{i=-w}^w b(i) \hat{\lambda}_{X|\ell+L-1}(k - i, \ell - 1) \right. \\ &\quad \left. \left. + (1 - \mu') \hat{\lambda}'_{X|[\ell, \ell+L]}(k, \ell) \right], \lambda_{\min} \right\}\end{aligned}\quad (13)$$

where $\mu(0 \leq \mu \leq 1)$ is related to the degree of nonstationarity of the random process $\{\lambda_X(k, \ell) | \ell = 0, 1, \dots\}$, b is related to the correlation between frequency bins of λ_X , $\mu'(0 \leq \mu' \leq 1)$ is associated with the reliability of the estimate $\hat{\lambda}'_{X|[\ell, \ell+L]}$ in comparison with that of $\hat{\lambda}_{X|\ell+L-1}$, and λ_{\min} is a lower bound on the variance of X . An estimate for $\lambda'_{X|[\ell, \ell+L]}(k, \ell)$ given the measurements $\mathcal{Y}_\ell^{\ell+L} \setminus \{Y\}$ is obtained by local averaging. Specifically,

$$\begin{aligned}\hat{\lambda}'_{X|[\ell, \ell+L]}(k, \ell) &= \max \left\{ \frac{\sum_{(n,i) \in \Gamma} b(i) |Y(k - i, \ell + n)|^2}{\sum_{(n,i) \in \Gamma} b(i)} - \beta \lambda_D, 0 \right\}\end{aligned}\quad (14)$$

where $\Gamma \triangleq \{(n, i) | 0 \leq n \leq L, -w \leq i \leq w, (n, i) \neq (0, 0)\}$ designates the time-frequency indices of the measurements, and $\beta(\beta \geq 1)$ is an over-subtraction factor to compensate for a sudden increase in the noise level. The steps of the noncausal spectral enhancement algorithm are summarized in Table I.

For comparison, the decision-directed *a priori* SNR estimate is given by

$$\hat{\xi}^{\text{DD}}(k, \ell) = \max \left\{ \alpha \frac{\hat{A}^2(k, \ell - 1)}{\lambda_D} + (1 - \alpha)[\gamma(k, \ell) - 1], \xi_{\min} \right\}\quad (15)$$

¹Recall that G_{SP} minimizes the expected spectral power distortion, yielding the square root of the conditional expected spectral power. That is, $G_{\text{SP}}(\xi, \gamma) |Y| = [E\{|X|^2 | \xi, Y\}]^{1/2}$.

TABLE I
SUMMARY OF THE NONCAUSAL SPEECH ENHANCEMENT ALGORITHM

Initialization at the first frame for all frequency bins k :
$\hat{A}(k, -1) = 0, \hat{\lambda}_{X L-1}(k, -1) = \lambda_{\min}.$
For all short-time frames $\ell = 0, 1, \dots$
For all frequency bins $k = 0, \dots, K - 1$
Compute $\hat{\lambda}'_{X [\ell, \ell+L]}(k, \ell)$ by using (13).
Compute $\hat{\lambda}'_{X \ell+L}(k, \ell)$ by using the “propagation” step (12).
Compute $\hat{\lambda}_{X \ell+L}(k, \ell)$ by using the “update” step (11).
Compute $\hat{X}(k, \ell)$ by using (4) and (5).

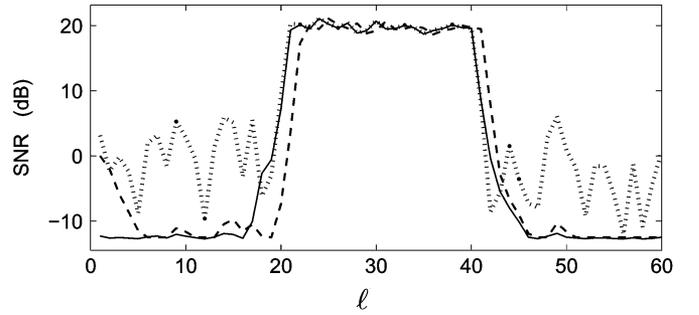


Fig. 1. SNRs in successive short-time frames: *A posteriori* SNR γ (dotted line), decision-directed *a priori* SNR estimate $\hat{\xi}^{\text{DD}}$ (dashed line), and noncausal *a priori* SNR estimate $\hat{\xi}$ (solid line).

where $\alpha(0 \leq \alpha \leq 1)$ is a weighting factor that controls the tradeoff between noise reduction and transient distortion introduced into the signal, and ξ_{\min} is a lower bound on the *a priori* SNR [2], [3]. It can be shown that a special case of the proposed *a priori* SNR estimator, with $L = 0$ (minimum delay in the estimated signal), $b(i) = \delta(i)$ (frequency bins are assumed uncorrelated) and $\mu = 1$ (maximum degree of nonstationarity), degenerates to a “decision-directed” estimator with a time-varying frequency-dependent weighting factor $\alpha(k, \ell)$. In the next section we present experimental results that show the improved performance of the noncausal *a priori* SNR estimator, compared with the decision-directed estimator.

V. EXPERIMENTAL RESULTS

Fig. 1 demonstrates the different behaviors of the noncausal and the decision-directed *a priori* SNR estimators. The analyzed signal is sampled at 16 kHz, and transformed into the STFT domain using half overlapping Hamming windows of 512 samples length (32 ms). It contains only white Gaussian noise (WGN) during the first and last 20 frames, and in between it contains an additional sinusoidal component at the displayed frequency with 0 dB SNR. The *a priori* SNR estimates, $\hat{\xi}$, and $\hat{\xi}^{\text{DD}}$, are obtained by using the parameters $\mu = \mu' = 0.8, b = [0.25 \ 0.5 \ 0.25], L = 3, \beta = 2, \lambda_{\min} = \xi_{\min} \lambda_D, \xi_{\min} = -25$ dB, and $\alpha = 0.98$ (this value of α was determined in [2], [9] by simulations and informal listening tests). The spectral gain used in our evaluation is G_{LSA} (9), since it proved very efficient in reducing musical residual noise phenomena [9]. Fig. 1 shows that when the *a posteriori* SNR γ is sufficiently low, the noncausal *a priori* SNR

estimate is smoother than the decision-directed estimate, which helps reducing the level of musical noise. When γ increases, the noncausal estimator, having a few subsequent spectral measurements at hand, is capable of discriminating between speech onsets and irregularities in γ corresponding to noise only. Whereas the decision-directed estimator cannot respond too fast to an abrupt increase in γ , since it necessarily implies an increase in the level of musical noise. When γ is sufficiently high, $\hat{\xi}^{\text{DD}}$ follows the *a posteriori* SNR with a delay of 1 frame, whereas $\hat{\xi}$ follows the *a posteriori* SNR instantaneously. When γ decreases, the response of $\hat{\xi}$ is immediate, while that of $\hat{\xi}^{\text{DD}}$ is delayed by 1 frame. Consequently, in comparison with the decision-directed estimator, the noncausal *a priori* SNR estimator produces lower levels of musical noise and signal distortion.

The performance of the noncausal *a priori* SNR estimator is evaluated under various noise conditions by enhancing noisy speech signals and computing the average improvement in the segmental SNR, the log-spectral distortion (LSD) [12], and the PESQ scores. The performance results are averaged out using 20 different utterances of 20 different speakers, drawn from the TIMIT database. Half of the utterances are from male speakers, and half are from female speakers. The noise signals include WGN, car interior noise, and F16 cockpit noise, taken from the Noisex92 database. The speech signals are sampled at 16 kHz and degraded by the various noise types with segmental SNRs in the range $[-5, 10]$ dB. The noncausal speech enhancement algorithm (Table I) is applied to the noisy speech signals, with the same parameters as in the example of Fig. 1. Alternatively, the *a priori* SNR is estimated by the decision-directed method (15).

Table II presents the results of the segmental SNR improvement, the LSD and the PESQ scores achieved by using the noncausal and the decision-directed *a priori* SNR estimators. The noncausal estimator yields a higher improvement in the segmental SNR, lower LSD, and higher PESQ scores than the decision-directed estimator under all tested environmental conditions. Informal listening tests confirm that the advantages of the noncausal estimator are particularly perceived during onsets of speech and noise only frames. Onsets of speech are better preserved, while a further reduction of musical noise is achieved.

VI. CONCLUSION

In some important applications, e.g., digital voice recording, surveillance, speech recognition, and speaker identification, a delay of a few short-term frames between the enhanced speech and the noisy observation is tolerable. In such cases, the noncausal estimation approach produces less signal distortion and less musical residual noise than the decision-directed approach. A further improvement of the speech enhancement results can be achieved by utilizing the uncertainty of speech presence in the noisy measurements [1], [2], [5].

It should be noted that improved segmental SNR, lower log-spectral distortion, and higher PESQ scores do not necessarily imply improved speech recognition or speaker identification performances. The usefulness of the proposed algorithm for such applications needs to be verified.

TABLE II
SEGMENTAL SNR IMPROVEMENT, LOG-SPECTRAL DISTORTION AND PESQ SCORES FOR VARIOUS NOISE TYPES AND LEVELS, OBTAINED BY USING THE DECISION-DIRECTED (DD) AND NONCAUSAL (NC) A PRIORI SNR ESTIMATORS

		Noise type	Method	Input SegSNR [dB]			
				-5	0	5	10
Segmental SNR Improvement	Stationary	DD	8.36	6.54	4.82	3.31	
	WGN	NC	9.00	7.37	5.83	4.47	
	Car interior	DD	8.62	7.16	5.74	4.38	
	noise	NC	9.21	7.81	6.33	4.97	
	F16 cockpit	DD	6.83	5.03	3.42	2.11	
	noise	NC	7.51	5.81	4.33	3.12	
Log Spectral Distance	Stationary	DD	4.96	3.80	2.99	2.23	
	WGN	NC	4.76	3.49	2.56	1.74	
	Car interior	DD	3.18	2.38	1.73	1.20	
	noise	NC	2.85	2.10	1.52	1.04	
	F16 cockpit	DD	4.58	3.59	2.72	1.95	
	noise	NC	4.22	3.19	2.32	1.59	
PESQ scores	Stationary	DD	2.22	2.60	2.93	3.21	
	WGN	NC	2.35	2.69	2.97	3.25	
	Car interior	DD	2.49	2.82	3.14	3.43	
	noise	NC	2.56	2.88	3.17	3.44	
	F16 cockpit	DD	2.14	2.51	2.83	3.13	
	noise	NC	2.25	2.58	2.88	3.17	

REFERENCES

- [1] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 137–145, Apr. 1980.
- [2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 1109–1121, Dec. 1984.
- [3] O. Cappé, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 2, pp. 345–349, Apr. 1994.
- [4] P. Scalart and J. Vieira-Filho, "Speech enhancement based on a priori signal to noise estimation," in *Proc. 21st IEEE Int. Conf. Acoust. Speech Signal Processing*, Atlanta, GA, May 1996, pp. 629–632.
- [5] I. Cohen and B. Berdugo, "Speech enhancement for nonstationary noise environments," *Signal Process.*, vol. 81, no. 11, pp. 2403–2418, Nov. 2001.
- [6] Y. Ephraim and I. Cohen, "Recent advancements in speech enhancement," in *The Electrical Engineering Handbook*, 3rd ed. Boca Raton, FL: CRC, [Online], Available: <http://ece.gmu.edu/~yephraim/ephraim.html>, to be published.
- [7] I. Cohen, *Relaxed Statistical Model for Speech Enhancement and a priori SNR Estimation*. Haifa, Israel: Technion-IIT, Oct. 2003, Tech. Rep. EE 1384.
- [8] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, pp. 1586–1604, Dec. 1979.
- [9] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 443–445, Apr. 1985.
- [10] P. J. Wolfe and S. J. Godsill, "Simple alternatives to the Ephraim and Malah suppression rule for speech enhancement," in *Proc. 11th IEEE Workshop Statist. Signal Processing*, Singapore, Aug. 6–8, 2001, pp. 496–499.
- [11] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Processing*, vol. 11, pp. 466–475, Sept. 2003.
- [12] S. R. Quackenbush, T. P. Barnwell, and M. A. Clements, *Objective Measures of Speech Quality*. Englewood Cliffs, NJ: Prentice-Hall, 1988.