3 Statistical Methods for the Enhancement of Noisy Speech

Rainer Martin

Ruhr-Universität Bochum, Institute of Communication Acoustics
Bochum 44780, Germany
E-mail: rainer.martin@rub.de

Abstract. Speech signals are frequently disturbed by statistically independent additive noise signals. When the power fluctuation of the noise signal is significantly slower than that of the speech signal, a single-microphone approach may be successfully used to reduce the level of the disturbing noise. This chapter outlines algorithms for noise reduction which are based on short term spectral representations of speech and on optimal estimation techniques. We present some of the more prominent estimation methods for complex spectral coefficients, for the amplitude and phase of spectral coefficients, and for related parameters such as the *a priori* signal-to-noise ratio. We interpret these algorithms in terms of their input-output characteristics. Some recent developments such as the use of super-Gaussian speech models and the properties of the resulting estimators are highlighted. Furthermore, we discuss the estimation of the background noise power and the application of these techniques in conjunction with a low bit rate speech coder.

3.1 Introduction

Speech communication devices are often used in environments with high levels of ambient noise such as cars and public places. The noise picked up by the microphones of the device can significantly impair the quality of the transmitted speech signal – especially when the speech source is far from the microphones. When the intelligibility of the transmitted speech is also impaired, the device cannot be used in the desired way. It is therefore sensible to include a noise reduction processor in such devices.

Algorithms for noise reduction have been the subject of intensive research over the last two decades [1–7]. The wide-spread use of mobile communication devices and the introduction of digital hearing aids have contributed to the significant interest in this field. While early approaches focused only on speech quality, it is now generally acknowledged that the perceived quality of the residual noise is also of great importance, e.g., random narrowband fluctuations in the processed noise, also known as *musical tones*, are not accepted by the human listener.

Over the last two decades researchers have found ways to improve the performance of noise reduction algorithms such that musical tones can be avoided and the algorithms are more robust with respect to the great vari-

ability of environmental conditions. In this context, statistical models and methods play a prominent role [4,8].

In this chapter, we will outline some well known results as well as some of the recent developments for single-microphone noise reduction algorithms. We will focus on systems which use a short term spectral representation of the speech and noise signals. The noisy signal may be analyzed, for example, by means of a short time discrete Fourier transform (DFT). Most of the results, however, also apply to other non-parametric spectral analysis methods such as filterbanks, subspace algorithms, or wavelet transforms, see e.g., [9,10].

3.2 Spectral Analysis

The advantages of moving into the spectral domain are at least threefold. In the spectral domain we achieve:

- a good separation of speech and noise — especially for voiced speech; thus optimal and/or heuristic approaches can be easily implemented,
- a decorrelation of spectral components; thus frequency bins can be treated independently to some extent and statistical models are simplified, and
- a possibility of integration of psychoacoustic models [11,12].

In most of the relevant applications the noise signal is additive and statistically independent from the source signal. In particular, the noisy speech signal $y(k)$ is generally modeled as the sum of an undisturbed speech signal $s(k)$ and a noise signal $n(k)$. The task of noise reduction is then to recover $s(k)$ "in the best possible way" when only the noisy signal $y(k)$ is given. The estimate of the undisturbed speech signal is denoted by $\widehat{s}(k)$.

Figure 3.1 depicts a typical implementation of a single-channel noise reduction system where the noisy signal is processed in a succession of short signal segments and the spectral coefficients are computed by means of a DFT. The DFT of a segment of M samples of $y(\ell)$, $\ell = k - M + 1, \ldots, k$, is denoted by

$$\boldsymbol{Y}(k) = (Y_0(k), \ldots, Y_\mu(k), \ldots, Y_{M-1}(k))^T, \tag{3.1}$$

with

$$\begin{aligned} Y_\mu(k) &= R_\mu(k) \exp\left(j\theta_\mu(k)\right) \\ &= \sum_{\ell=0}^{M-1} w(\ell) y(k - M + 1 + \ell) \exp\left(-\frac{j2\pi\mu\ell}{M}\right), \end{aligned} \tag{3.2}$$

where a tapered analysis window $w(\ell)$ of length M is applied to the time domain segment before the DFT is computed. k denotes the time index at which the segment of M signal samples is extracted. $\mu = 0, \ldots, M-1$ is the index of the DFT bin which is related to the normalized center frequency Ω_μ

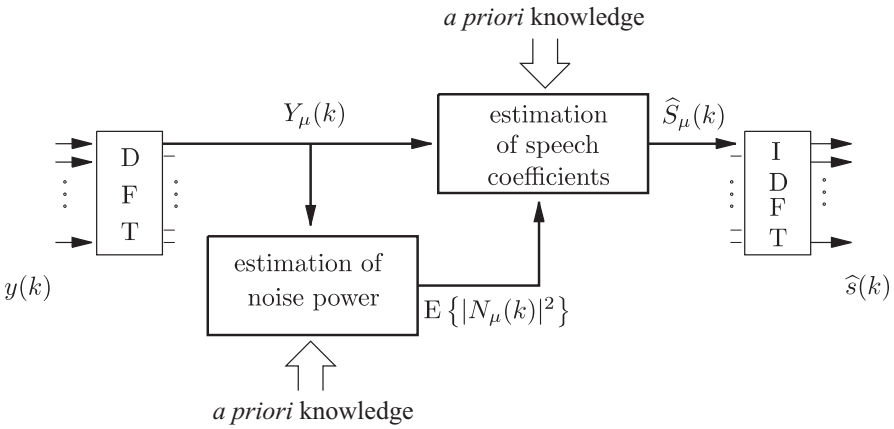

Fig. 3.1. DFT based speech enhancement. k and μ denote the time and the frequency bin index, respectively.

of that bin by $\Omega_\mu = 2\pi\mu/M = 2\pi f_\mu/f_S$ where f_μ and f_S denote the absolute center frequency and the sampling frequency, respectively. An enhanced DFT coefficient is denoted by $\widehat{S}_\mu(k)$. Vectors of the undisturbed speech signal and the enhanced speech signal are defined in the same way. The enhanced signal segments are computed by means of an inverse DFT and a continuous signal is produced by the overlap-add method. For the overlap-add operation the use of a tapered synthesis window is generally beneficial [13,14].

After the short time spectral components are computed by means of a DFT, there are two major tasks which must be addressed:

- estimation of the spectral components $S_\mu(k)$ of the undisturbed speech signal given the noisy spectral components $Y_\mu(k)$,
- estimation of the noise power $\sigma_n^2 = \mathrm{E}\left\{|N_\mu(k)|^2\right\}$ in each frequency bin μ.

Both of these tasks require the application of *a priori* knowledge and will be discussed below.

3.3 The Wiener Filter and its Implementation

Numerous approaches are available for the estimation of the complex coefficients $S_\mu(k) = A_\mu(k)\exp(\alpha_\mu(k))$ of the undisturbed speech signal or functions thereof. Among these are methods based on linear processing models and minimum mean square error (MMSE) estimation such as the Wiener filter. MMSE estimation is suitable for speech processing purposes as large estimation errors are given more weight in the optimization than small estimation errors. The latter might be masked in the human auditory system

and might therefore be inaudible. Under the assumption that all signals are wide-sense stationary, the Wiener filter minimizes

$$\mathrm{E}\left\{(\widehat{s}(k) - s(k))^2\right\}, \tag{3.3}$$

where $\mathrm{E}\{\cdot\}$ denotes the statistical expectation operator and

$$\widehat{s}(k) = \sum_{\ell=-\infty}^{\infty} h(\ell) y(k-\ell) \tag{3.4}$$

is the convolution of an impulse response $h(\ell)$ with the noisy signal $y(k)$. For statistically independent and additive speech and noise signals, the frequency response of the Wiener filter is given by

$$G(\Omega) = \mathrm{DTFT}\{h(\ell)\} = \frac{P_{ss}(\Omega)}{P_{ss}(\Omega) + P_{nn}(\Omega)}, \tag{3.5}$$

where $P_{xx}(\Omega)$ denotes the power spectral density of the signal in the subscript and $\mathrm{DTFT}\{\cdot\}$ is the discrete time Fourier transform. Thus, in the case of stationary signals, the spectrum of the enhanced output signal is computed as

$$\widehat{S}(\Omega) = \frac{P_{ss}(\Omega)}{P_{ss}(\Omega) + P_{nn}(\Omega)} Y(\Omega) = \frac{P_{ss}(\Omega)}{P_{yy}(\Omega)} Y(\Omega) = G(\Omega) Y(\Omega). \tag{3.6}$$

In this context, $G(\Omega)$ is frequently called the *spectral gain* function. For the Wiener filter this function depends on the noisy input $y(k)$ or its Fourier transform $Y(\Omega)$ and on the undisturbed speech signal only via statistical expectations. However, an exact numerical implementation of the Wiener filter is not completely straightforward as this filter has an infinite impulse response and a continuous frequency response.

For a numerical implementation in conjunction with the above spectral analysis-synthesis system, the gain function is evaluated at the center frequencies of the spectral bins. Furthermore, as speech and noise signals are not stationary, short-term approximations to the power spectra must be used. However, for the segment-by-segment processing approach outlined above, we prefer an alternative derivation. In analogy to the Wiener filter in (3.6), the output of the filter for the signal segment at time k, $\widehat{\boldsymbol{S}}(k) = (\widehat{S}_0(k), \ldots, \widehat{S}_\mu(k), \ldots, \widehat{S}_{M-1}(k))^T$, is computed by an elementwise multiplication

$$\widehat{\boldsymbol{S}}(k) = \boldsymbol{G}(k) \otimes \boldsymbol{Y}(k) \tag{3.7}$$

of the DFT vector $\boldsymbol{Y}(k)$ and a gain vector

$$\boldsymbol{G}(k) = (G_0(k), G_1(k), \ldots, G_{M-1}(k))^T. \tag{3.8}$$

For independent additive speech and noise signals the minimization of $\mathrm{E}\left\{\left(\widehat{S}_\mu(k) - S_\mu(k)\right)^2\right\}$ with respect to $G_\mu(k)$ leads to

$$G_\mu(k) = \frac{\mathrm{E}\left\{|S_\mu(k)|^2\right\}}{\mathrm{E}\left\{|S_\mu(k)|^2\right\} + \mathrm{E}\left\{|N_\mu(k)|^2\right\}} = \frac{\eta_\mu(k)}{1+\eta_\mu(k)}, \tag{3.9}$$

where the right hand side of (3.9) makes use of the *a priori* SNR

$$\eta_\mu(k) = \frac{\mathrm{E}\left\{|S_\mu(k)|^2\right\}}{\mathrm{E}\left\{|N_\mu(k)|^2\right\}}. \tag{3.10}$$

$\mathrm{E}\left\{|S_\mu(k)|^2\right\} = \sigma^2_{s,\mu}(k)$ and $\mathrm{E}\left\{|N_\mu(k)|^2\right\} = \sigma^2_{n,\mu}(k)$ are the power of the undisturbed speech signal and the noise signal in frequency bin μ, respectively.

In a linear systems framework, the multiplication of the two DFT vectors and the subsequent inverse DFT of the result corresponds to a cyclic convolution in the time domain. Therefore, to implement this Wiener-like filter as a segmentwise linear system the signal and the gain vectors must be zero-padded to the appropriate length.

It is, however, instructive to consider the above estimation task in the framework of *non-linear* estimation, i.e., to derive the best estimator in the MMSE sense for the *short term* spectral coefficients of the undisturbed speech signal given the short term coefficients of the noisy signal. Contrary to the Wiener-like filter (3.9) which relies on second order statistics only, the non-linear solution generally requires knowledge of the probability density functions (pdf) of the speech and noise spectral coefficients. Under the assumption that all frequency bins are mutually independent, the MMSE solution can be stated as the conditional expectation

$$\begin{aligned}\widehat{S}_\mu(k) &= \mathrm{E}\left\{S_\mu(k) \mid Y_\mu(k)\right\} \\ &= \int\int S_\mu(k) p_{S|Y}(S_\mu(k) \mid Y_\mu(k)) dS_\mu(k) \\ &= \frac{1}{p(Y_\mu(k))} \int\int S_\mu(k) p_{Y|S}(Y_\mu(k) \mid S_\mu(k)) p(S_\mu(k)) dS_\mu(k),\end{aligned} \tag{3.11}$$

where $p_{S|Y}(S_\mu(k) \mid Y_\mu(k))$ is the pdf of an undisturbed speech coefficient given the coefficient of the noisy signal and $p(S_\mu(k))$ is the density of the undisturbed speech coefficients. Note that $S_\mu(k)$ is a complex quantity and therefore a double integration over the real and imaginary parts or over the magnitude and phase is required.

For additive noise which is statistically independent of the speech signal we have $p_{Y|S}(Y_\mu(k) \mid S_\mu(k)) = p_N(Y_\mu(k) - S_\mu(k))$. Therefore, the application of Bayes theorem in (3.11) leads to a nice decomposition of the density $p_{S|Y}(S_\mu(k) \mid Y_\mu(k))$ in terms of the probability density functions of the noise and the density of the undisturbed speech spectral coefficients. To model

the probability density function of the real and the imaginary part of these coefficients, $S_\mu^{<R>}$ and $S_\mu^{<I>}$ respectively, the Gaussian density

$$p(S_\mu^{<R>}) = \frac{1}{\sqrt{\pi}\sigma_s} \exp\left(-\frac{(S_\mu^{<R>})^2}{\sigma_s^2}\right),$$
$$p(S_\mu^{<I>}) = \frac{1}{\sqrt{\pi}\sigma_s} \exp\left(-\frac{(S_\mu^{<I>})^2}{\sigma_s^2}\right), \tag{3.12}$$

is frequently used. These probability densities depend on the speech power σ_s^2 which is, in general, time-variant. When the noise coefficients are also Gaussian distributed it is straightforward to show that for statistically independent and additive speech and noise coefficients, (3.7) with (3.9) is the solution to the estimation problem. For Gaussian signals the non-linear optimal estimator yields a linear function of the observations. However, this does not necessarily hold for practical implementations of these filters.

To illustrate the non-linearity of practical implementations we consider the estimation of the *a priori* SNR $\eta_\mu(k)$ which is required for the computation of the Wiener-like filter in (3.9). $\eta_\mu(k)$ is frequently estimated using the *decision-directed* approach [4]. This scheme assumes that an estimate $\widehat{A}_\mu(k-r)$ for the undisturbed speech amplitudes $A_\mu(k-r) = |S_\mu(k-r)|$ from a previous signal segment at time $k-r$ is available and sufficiently close to the undisturbed speech amplitudes of the current segment. The decision-directed approach then feeds back the estimate of the previous segment and combines it with an instantaneous estimate of the SNR,

$$\gamma_\mu(k) - 1 = \frac{|Y_\mu(k)|^2}{\mathrm{E}\left\{|N_\mu(k)|^2\right\}} - 1 = \frac{R_\mu^2(k)}{\sigma_{n,\mu}^2(k)} - 1, \tag{3.13}$$

such that the estimated SNR $\widehat{\eta}_\mu(k)$ is obtained as

$$\widehat{\eta}_\mu(k) = \alpha_\eta \frac{|\widehat{S_\mu(k-r)}|^2}{\mathrm{E}\left\{|N_\mu(k)|^2\right\}} + (1-\alpha_\eta)\max\left(0, \gamma_\mu(k) - 1\right), \tag{3.14}$$

where the latter contribution is forced to be non-negative and α_η is a smoothing parameter. The term

$$\gamma_\mu(k) = \frac{|Y_\mu(k)|^2}{\mathrm{E}\left\{|N_\mu(k)|^2\right\}} = \frac{R_\mu^2(k)}{\sigma_{n,\mu}^2(k)} \tag{3.15}$$

is the *a posteriori* SNR. For low SNR conditions, this estimator is clearly biased. The bias can be reduced if the maximum operation is applied to the sum of the two contributions:

$$\widehat{\eta}_\mu(k) = \max\left(0, \alpha_\eta \frac{|\widehat{S_\mu(k-r)}|^2}{\mathrm{E}\left\{|N_\mu(k)|^2\right\}} + (1-\alpha_\eta)(\gamma_\mu(k) - 1)\right). \tag{3.16}$$

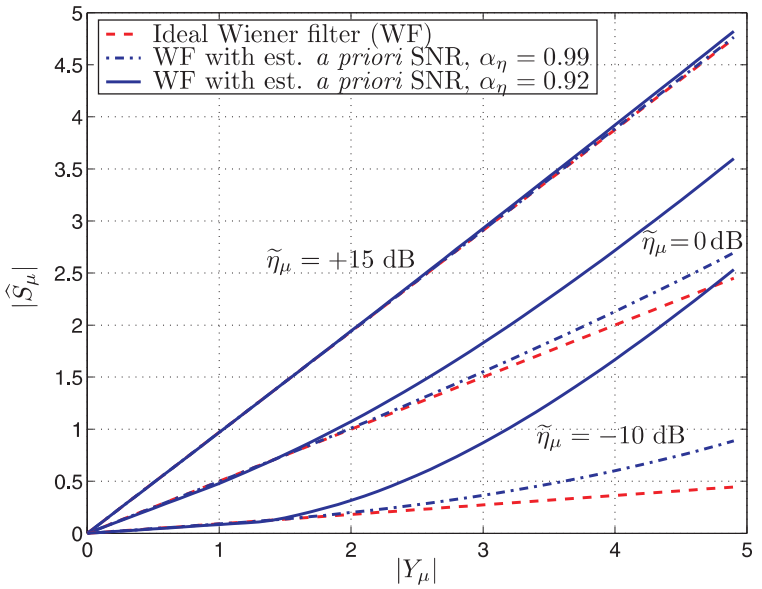

Fig. 3.2. Estimator characteristics for the ideal Wiener-like filter (dashed), the Wiener-like filter with $\alpha_\eta = 0.99$ (dash-dotted) and with $\alpha_\eta = 0.92$ (solid) for three different *a priori* SNR $\widetilde{\eta}_\mu(k-r)$. The decision-directed SNR estimator (3.14) was used and $\sigma_y^2 = \sigma_s^2 + \sigma_n^2 = 2$.

Using (3.14) in (3.9) we find that the spectral components $Y_\mu(k)$ of the current signal segment now have a direct influence on the gain function. Therefore, the combination of the Wiener-like filter and the decision-directed SNR estimation leads to a non-linear system. This non-linear dependency on the observation is clearly visible in Fig. 3.2 which plots the magnitude of the estimated spectral coefficient as a function of the magnitude of the noisy coefficient for $\sigma_y^2 = \sigma_s^2 + \sigma_n^2 = 2$. Three different values for the *a priori* SNR $\widetilde{\eta}_\mu(k-r) = \widehat{A}_\mu^2(k-r)/\mathrm{E}\left\{|N_\mu(k)|^2\right\}$ related to the previous frame are selected. Compared to the ideal Wiener-like filter which is also shown, the non-linear behaviour is visible, especially for low *a priori* SNR conditions. For comparison purposes, the same graphs are shown for the less biased *a priori* SNR estimation (3.16) in Fig. 3.3. For low *a priori* SNR values and small input coefficients, more attenuation is achieved than with the decision-directed approach in (3.14).

For the ideal Wiener-like filter the slope of the filter characteristic does not depend on the noisy input coefficient. On the other hand, the practical implementation using the decision-directed approach provides a larger gain than the Wiener filter when the observed coefficient is larger than its standard deviation. In this case, it is likely that speech is contained in the current segment of the input signal and thus speech distortions are reduced. When the

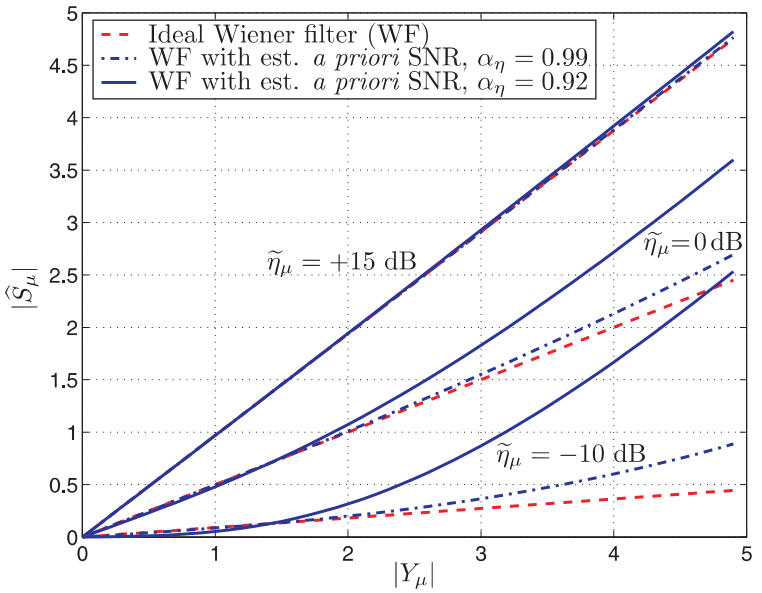

Fig. 3.3. Estimator characteristics for the ideal Wiener-like filter (dashed), the Wiener-like filter with $\alpha_\eta = 0.99$ (dash-dotted) and with $\alpha_\eta = 0.92$ (solid) for three different *a priori* SNR $\widetilde{\eta}_\mu(k-r)$. The decision-directed SNR estimator (3.16) was used and $\sigma_y^2 = \sigma_s^2 + \sigma_n^2 = 2$.

noisy coefficient is relatively small the input coefficient contains mostly noise. In this case it is important to avoid large fluctuations of the output coefficients as these translate into musical tones. With the noise reduction scheme discussed here this can be achieved by choosing the smoothing parameter α_η close to unity and thus smoothing the estimated *a priori* SNR. However, a large amount of smoothing will reduce the non-linearity of the estimation scheme for large amplitudes and thus lead to less transparent speech reproduction. The combination of the Wiener-like filter and the decision-directed estimator therefore requires a balance between these conflicting objectives [8,15]. Nevertheless, the decision-directed estimation procedure is advantageously combined with many noise reduction algorithms where the *a priori* SNR plays a role [15]. Furthermore, there are other ways to exploit the idea of recursive estimation, e.g., [16,17] which in general lead to less musical noise than the standard methods. As an accurate *a priori* SNR estimate is a key factor in the performance of these algorithms, improved *a priori* SNR estimators have also been developed [18,19].

To conclude this discussion we note that noise reduction schemes are frequently non-linear. In general, it is therefore not appropriate to cast spectral estimation procedures into the form of a multiplication of the noisy spectral coefficients with a spectral gain function as in (3.7). Moreover, there

are immediate consequences for the synthesis of the enhanced signal. In the framework of non-linear estimation in the spectral domain we strive for the optimal estimate of the spectral coefficients of a short signal segment. The enhanced segments will then be synthesized using an inverse spectral transform and concatenated to produce a continuous signal. By virtue of this approach zero-padding is not necessarily required. For example, the (non-realizable) gain vector $G_\mu(k) = S_\mu(k)/Y_\mu(k)$ will result in a perfect reconstruction of the spectral coefficients and hence of the undisturbed speech signal without zero-padding and any cyclic effects. On the other hand, an MMSE-optimal estimate in the spectral domain does not deliver MMSE-optimal time domain segments. Also, simplifying assumptions such as the independence of adjacent frequency bins lead to estimation errors. Thus, there are no strict guidelines for the implementation of the spectral analysis-synthesis system. To suppress estimation errors in the synthesized signal it is, however, advisable to use a tapered analysis and a tapered synthesis window [14].

3.4 Estimation of Spectral Amplitudes

In the context of single-microphone speech enhancement, the short term spectral amplitudes are much more important than the short term spectral phases [20]. It is therefore sensible to estimate the spectral amplitudes $A_\mu(k)$ of the undisturbed speech signal jointly with the phase $\alpha_\mu(k)$ or directly by using the marginal distribution of the spectral amplitudes. We briefly present *minimum mean square error* (MMSE) and *maximum a posteriori* (MAP) solutions to this problem. These estimators require explicit knowledge of the probability density functions of the spectral coefficients of speech and noise.

3.4.1 MMSE Estimation

For Gaussian speech and noise coefficients the MMSE *short term spectral amplitude* estimator (MMSE-STSA) was derived by Ephraim and Malah [4],

$$\widehat{A}_{\mathrm{STSA},\mu} = \mathrm{E}\{A_\mu \mid Y_\mu\} = \sigma_n \sqrt{\frac{\eta_\mu}{1+\eta_\mu}}\, \Gamma(1.5)\;\; F_1(-0.5; 1, -v_\mu), \qquad (3.17)$$

where we have now dropped the time index k for improved readability. $F_1(\cdot;\cdot,\cdot)$ is a confluent hypergeometric function [21] and v_μ is defined as

$$v_\mu = \frac{\eta_\mu}{1+\eta_\mu}\gamma_\mu. \qquad (3.18)$$

The confluent hypergeometric function can be expanded in terms of Bessel functions and may be tabulated for efficient numerical implementations. Besides the MMSE-STSA estimator, the estimate of the logarithm of the spec-

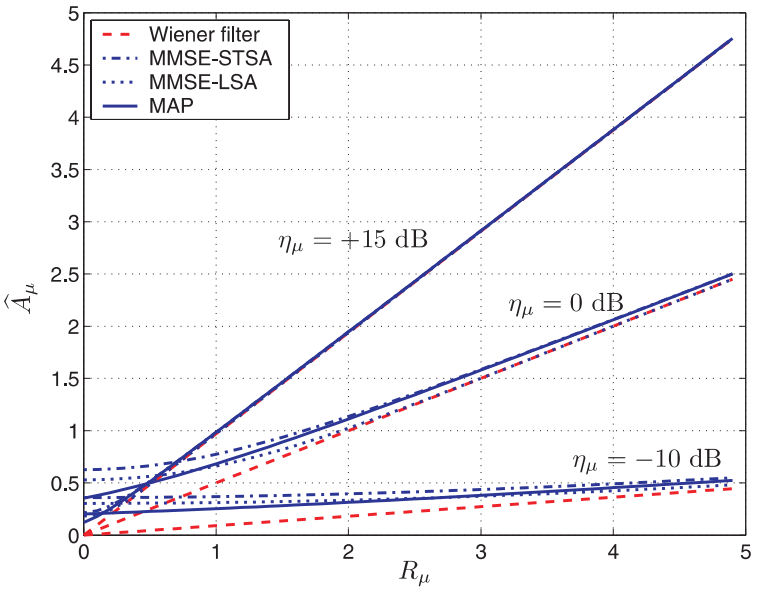

Fig. 3.4. Estimator characteristics for the Wiener filter (dashed), the MMSE-STSA [25] (dash-dotted), the MMSE-LSA [25] (dotted), and the MAP estimator [23] (solid) for three different *a priori* SNR values. $\sigma_y^2 = \sigma_s^2 + \sigma_n^2 = 2$.

tral amplitudes is also widely used. This MMSE *log spectral amplitude* estimator (MMSE-LSA) may be written as

$$\begin{aligned} \widehat{A}_{\mathrm{LSA},\mu} &= \exp\left(\mathrm{E}\left\{\log(A_\mu) \mid Y_\mu\right\}\right) \\ &= \frac{\eta_\mu}{1+\eta_\mu} \exp\left(\frac{1}{2}\int_{v_\mu}^{\infty} \frac{\exp\{-t\}}{t} dt\right) R_\mu, \end{aligned} \tag{3.19}$$

where R_μ denotes the amplitude of the noisy spectral coefficients. For large *a posteriori* SNR values both estimators approach the Wiener filter. For small, noisy amplitudes the estimators deliver an almost constant output value which depends to a greater extent on the *a priori* SNR than on the instantaneous input amplitude. This behaviour contributes significantly to the perceived quality of the residual noise since for small input values the fluctuations of the noisy amplitudes result in much smaller fluctuations in the enhanced output. For a single frequency bin and for $\sigma_s^2 + \sigma_n^2 = 2$ the resulting input-output characteristics are shown in Fig. 3.4 for the *a priori* SNR estimation (3.14) with $\alpha_\eta = 1$ and in Fig. 3.5 for $\alpha_\eta = 0.92$. To compute the enhanced complex spectral coefficient, the estimated spectral amplitude is combined with the short term phase of the noisy input. The observed phase represents the optimal phase estimate in the MMSE sense [4].

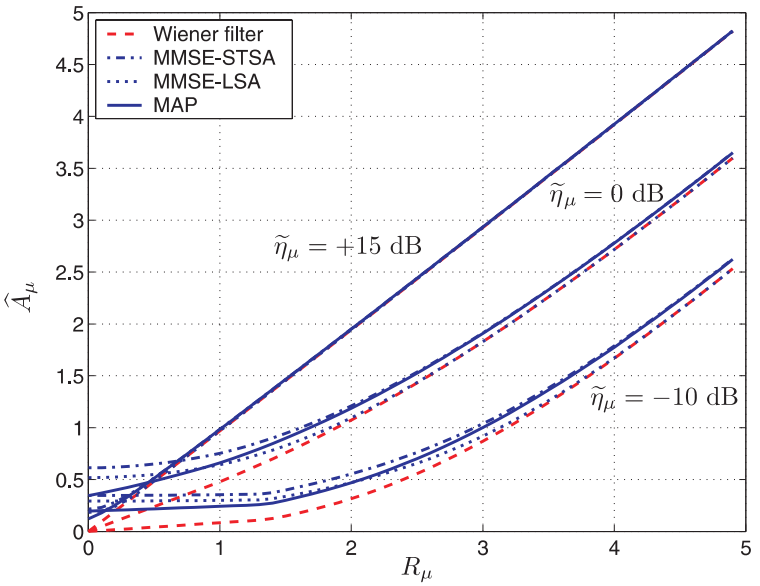

Fig. 3.5. Estimator characteristics for the Wiener filter (dashed), the MMSE-STSA [25] (dash-dotted), the MMSE-LSA [25] (dotted), and the MAP estimator [23] for three different *a priori* SNR and $\sigma_y^2 = \sigma_s^2 + \sigma_n^2 = 2$. The decision-directed SNR estimator (3.14) was used with $\alpha = 0.92$.

3.4.2 Maximum Likelihood and MAP Estimation

The maximum likelihood (ML) and the maximum *a posteriori* (MAP) estimation techniques avoid hard-to-compute integrals and lead to relatively simple solutions. In the case of complex Gaussian distributed spectral coefficients, the ML and the MAP estimators yield the well known Wiener-like solution. An ML estimate for deterministic spectral amplitudes in Gaussian noise was derived in [22],

$$\widehat{A}_{\mathrm{ML},\mu} = \left(0.5R_\mu + 0.5\sqrt{R_\mu^2 - \sigma_{n,\mu}^2}\right). \tag{3.20}$$

This estimator provides only a modest amount of noise reduction and is therefore not often used. Joint MAP estimation of the spectral amplitude and the spectral phase was proposed by Wolfe and Godsill [23]. Also in this case, the optimal estimate of the phase of the undisturbed spectral coefficients is the phase of the noisy input. The estimate of the amplitude is given by

$$\widehat{A}_{\mathrm{JMAP},\mu} = \frac{\eta_\mu + \sqrt{\eta_\mu^2 + 2\,(1+\eta_\mu)\,\frac{\eta_\mu}{\gamma_\mu}}}{2(1+\eta_\mu)} R_\mu. \tag{3.21}$$

Estimation of the spectral amplitude using the marginal density is also feasible but for closed form analytic solutions approximations to the Rician

density are required. Using such approximations, the MAP estimation of the spectral amplitudes leads to a solution which, like (3.21), is close in performance to the MMSE methods [23],

$$\widehat{A}_{\mathrm{MAP},\mu} = \frac{\eta_\mu + \sqrt{\eta_\mu^2 + (1+\eta_\mu)\,\frac{\eta_\mu}{\gamma_\mu}}}{2(1+\eta_\mu)} R_\mu. \tag{3.22}$$

The attenuation characteristics of this latter estimator in conjunction with (3.14) is shown in Fig. 3.4 for $\alpha = 1$ and in Fig. 3.5 for $\alpha = 0.92$. A MAP amplitude estimator using super-Gaussian speech models is derived in [24] and discussed in Chapter 4.

To conclude this section, we firstly note that all of these estimators and the underlying statistical models, e.g., (3.12), are conditioned on the signal power. The power of the undisturbed speech signal as well as of the noise signal are random processes by themselves and must be estimated, e.g., using the decision-directed approach. Secondly, all of the above approaches assume that speech is actually present in the frequency bin under consideration. This is, of course, not always the case as there are speech pauses and possibly also a concentration of speech power onto a few dozen harmonics during voiced speech. Frequently, these estimators are used in conjunction with a statistical two-state speech presence/absence model which leads to a soft-decision gain modification procedure. The resulting soft-decision gain functions are dependent on the signal model and are discussed in detail in [22,4,18,30].

3.5 MMSE Estimation Using Super-Gaussian Speech Models

In the time domain, the probability density function of speech samples may be modelled by Laplacian (bilateral exponential) or Gamma, i.e. *super-Gaussian*, densities rather than Gaussian densities [26, page 235]. It has been suggested [27,28] that also in the short time Discrete Fourier domain (frame size $<$ 100 ms), the Laplace and Gamma densities are much better models for the probability density function of the real and imaginary parts of the speech coefficients than the commonly used Gaussian density. In fact, the Gaussian assumption is based on the central limit theorem [29]. However, when the DFT length is shorter than the span of correlation of the signal, the asymptotic arguments do not hold. While for many applications the spectral coefficients of the noise can be modeled by a complex Gaussian random variable, the span of correlation of voiced speech is certainly larger than the typical segment size used in voice communications. Note again that all of these probability functions are conditioned on the signal power which is, in general, time-variant. Therefore, in an experimental verification of the density model great care must be exercised to generate quasi-stationary conditions [30].

Only recently, analytic solutions to the estimation problem under super-Gaussian model assumptions have been found [28,31,32,24]. In this section, we will present an example based on a Laplacian speech pdf and a Gaussian noise model [32]. Estimators for complex spectral coefficients based on Gamma densities as well as soft-decision gain functions for various combinations of speech and noise densities are discussed, e.g., in [30].

When the spectral coefficients of the speech and noise signals are mutually independent with respect to frequency bins and time segments, the optimal instantaneous estimate can be written as a conditional expectation

$$\widehat{S}_\mu(k) = \mathrm{E}\left\{S_\mu(k) \mid Y_\mu(k)\right\} = \mathrm{E}\left\{S \mid Y\right\}. \tag{3.23}$$

On the right hand side we now drop time and frequency bin indices to simplify our notation. For statistically independent real and imaginary parts, we may decompose the optimal estimate into an estimate of its real and its imaginary part

$$\mathrm{E}\left\{S \mid Y\right\} = \mathrm{E}\left\{S^{<R>} \mid Y^{<R>}\right\} + j\mathrm{E}\left\{S^{<I>} \mid Y^{<I>}\right\}, \tag{3.24}$$

where $<R>$ and $<I>$ in the superscript indicate the real and the imaginary parts, respectively. When $\diamondsuit$ denotes either the real or the imaginary part, the MMSE estimate of one of these is given by

$$\mathrm{E}\left\{S^{\diamondsuit} \mid Y^{\diamondsuit}\right\} = \int_{-\infty}^{\infty} S^{\diamondsuit} p(S^{\diamondsuit} \mid Y^{\diamondsuit}) dS^{\diamondsuit}. \tag{3.25}$$

With Bayes theorem we obtain

$$\mathrm{E}\left\{S^{\diamondsuit} \mid Y^{\diamondsuit}\right\} = \frac{1}{p(Y^{\diamondsuit})} \int_{-\infty}^{\infty} S^{\diamondsuit} p(Y^{\diamondsuit} \mid S^{\diamondsuit}) p(S^{\diamondsuit}) dS^{\diamondsuit}. \tag{3.26}$$

Good candidates for the pdf of the real and the imaginary parts of DFT coefficients of speech signals are the Laplacian pdf,

$$p(S^{\diamondsuit}) = \frac{1}{\sigma_s} \exp\left(-\frac{2|S^{\diamondsuit}|}{\sigma_s}\right), \tag{3.27}$$

and the Gamma pdf,

$$p(S^{\diamondsuit}) = \frac{\sqrt[4]{3}}{2\sqrt{\pi\sigma_s}\sqrt[4]{2}} |S^{\diamondsuit}|^{-\frac{1}{2}} \exp\left(-\frac{\sqrt{3}|S^{\diamondsuit}|}{\sqrt{2}\sigma_s}\right). \tag{3.28}$$

These two densities are better models than the Gaussian pdf, not only for small amplitudes, but also for large amplitudes where a *heavy-tailed* density leads to a better fit for the observed data [30]. The complexity of the analytic solutions depends upon the density models and the optimization criterion. A relatively simple analytical MMSE solution is based on the Gaussian noise and the Laplacian speech models.

To facilitate the development we introduce the shorthand notations

$$L^{\diamond +} = \frac{\sigma_n}{\sigma_s} + \frac{Y^\diamond}{\sigma_n} = \frac{1}{\sqrt{\eta}} + \frac{Y^\diamond}{\sigma_n},$$
$$L^{\diamond -} = \frac{\sigma_n}{\sigma_s} - \frac{Y^\diamond}{\sigma_n} = \frac{1}{\sqrt{\eta}} - \frac{Y^\diamond}{\sigma_n}, \tag{3.29}$$

where $\eta = \sigma_s^2/\sigma_n^2$ denotes the *a priori* SNR as before.

For the Laplacian speech pdf we obtain the optimal MMSE estimator of either the real part or the imaginary part [21, Theorem 3.462,1] as:

$$\begin{aligned}
&\mathrm{E}\{S^\diamond \mid Y^\diamond\} = \\
&\frac{1}{\sqrt{\pi}\sigma_n\sigma_s p(Y^\diamond)} \int_{-\infty}^{\infty} S^\diamond \exp\left(-\frac{(Y^\diamond - S^\diamond)^2}{\sigma_n^2}\right) \exp\left(-\frac{2|S^\diamond|}{\sigma_s}\right) dS^\diamond \\
&= \frac{\sigma_n \exp(\sigma_n^2/\sigma_s^2)}{2\sigma_s p(Y^\diamond)} \Bigg\{ L^{\diamond +} \exp\left(2\frac{Y^\diamond}{\sigma_s}\right) \mathrm{erfc}(L^{\diamond +}) \\
&\qquad\qquad -L^{\diamond -} \exp\left(-2\frac{Y^\diamond}{\sigma_s}\right) \mathrm{erfc}(L^{\diamond -}) \Bigg\},
\end{aligned} \tag{3.30}$$

with [21, Theorem 3.322,2]

$$\begin{aligned}
&p\ (Y^\diamond) = \\
&\frac{1}{\sqrt{\pi}\sigma_n\sigma_s} \int_{-\infty}^{\infty} \exp\left(-\frac{(Y^\diamond - S^\diamond)^2}{\sigma_n^2}\right) \exp\left(-\frac{2|S^\diamond|}{\sigma_s}\right) dS^\diamond \\
&= \frac{\exp\left(\sigma_n^2/\sigma_s^2\right)}{2\sigma_s} \left\{ \exp\left(2\frac{Y^\diamond}{\sigma_s}\right) \mathrm{erfc}(L^{\diamond +}) + \exp\left(-2\frac{Y^\diamond}{\sigma_s}\right) \mathrm{erfc}(L^{\diamond -}) \right\},
\end{aligned} \tag{3.31}$$

where $\mathrm{erfc}(z)$ denotes the complementary error function [21, Theorem 8.250]. The optimal estimator for the undisturbed complex speech coefficient is therefore given by $E\{S \mid Y\} = E\{S^{<R>} \mid Y^{<R>}\} + jE\{S^{<I>} \mid Y^{<I>}\}$ with

$$\begin{aligned}
&\mathrm{E}\{S^\diamond \mid Y^\diamond\} \\
&= \frac{\sigma_n \left[L^{\diamond +} \exp(2Y^\diamond/\sigma_s)\mathrm{erfc}(L^{\diamond +}) - L^{\diamond -} \exp(-2Y^\diamond/\sigma_s)\mathrm{erfc}(L^{\diamond -})\right]}{\exp(2Y^\diamond/\sigma_s)\mathrm{erfc}(L^{\diamond +}) + \exp(-2Y^\diamond/\sigma_s)\mathrm{erfc}(L^{\diamond -})}.
\end{aligned} \tag{3.32}$$

We note that both $\mathrm{E}\left\{S^{<R>} \mid Y^{<R>}\right\}$ and $\mathrm{E}\left\{S^{<I>} \mid Y^{<I>}\right\}$ are odd symmetric functions of $Y^{<R>}$ and $Y^{<I>}$, respectively. Figure 3.6 plots the input-output characteristics of this estimator and the Wiener-like filter for $0 \le Y^\diamond \le 5$, $\sigma_y^2 = \sigma_s^2 + \sigma_n^2 = 2$, and three different *a priori* SNR values. Again, the decision-directed SNR estimator is used with two different values of α_η. For high *a priori* SNR values the estimate is almost identical to the estimate delivered by the Wiener filter. Clearly, for a fixed *a priori* SNR, the Wiener filter is a linear estimator, characterized by its constant slope. The estimator based on super-Gaussian densities leads to an increased attenuation of the input when the instantaneous input value is smaller than its standard deviation and a significantly larger output value when the input is larger than

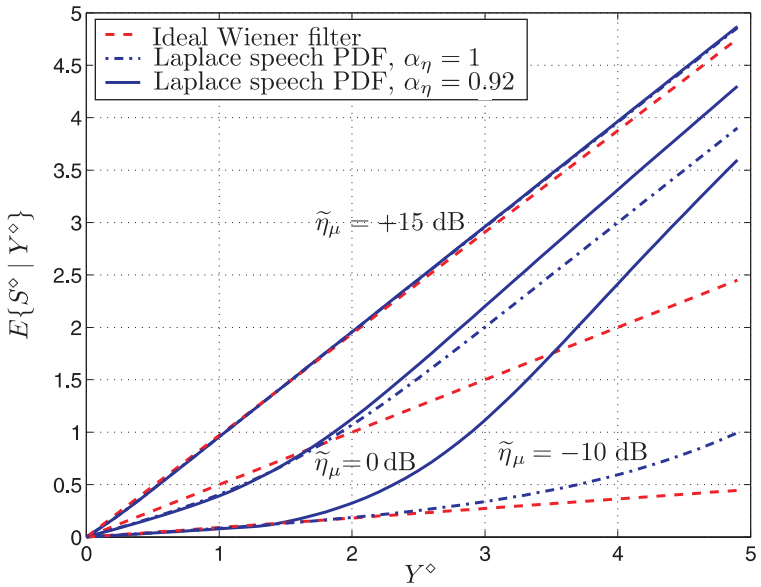

Fig. 3.6. Estimator characteristics $E\{S^\diamond \mid Y^\diamond\}$ for the ideal Wiener filter (dashed) and for the Laplacian speech pdf and the Gaussian noise pdf for $\alpha_\eta = 1$ (dash-dotted) and for $\alpha_\eta = 0.92$ (solid) and for three *a priori* SNR values $\widetilde{\eta}_\mu = 15, 0, -10$ dB. The decision-directed SNR estimator (3.14) was used and $\sigma_y^2 = \sigma_s^2 + \sigma_n^2 = 2$.

the standard deviation. Due to the heavy-tailed speech density, it is highly likely that speech is present in this latter case. Both of these characteristics contribute to the improved SNR of the output coefficients with respect to the linear estimator.

Figure 3.6 also plots the characteristics using the decision-directed SNR estimation technique (3.14) with $\alpha_\eta = 0.92$. The *a priori* SNR $\widetilde{\eta}_\mu(k-r)$ of the preceding signal segment is fixed. The SNR estimate of the present segment is then a function of the instantaneous, magnitude-squared input value which leads to an additional non-linear effect. Compared to the estimators based on Gaussian densities we find that in conjunction with the decision-directed estimator more smoothing can be applied to the SNR estimate without sacrificing the transparency of the enhanced speech components. Furthermore, we note that the proposed estimators may be applied to the magnitude of the spectral coefficients as well if we assume a fixed (hypothetical) phase angle. These procedures are outlined in [30].

3.6 Background Noise Power Estimation

The second estimation task which arises in the processing model of Fig. 3.1 is the estimation of the background noise power in the spectral bins. Most of the proposals in the literature are based on either

- voice activity detection and recursive averaging [22,33],
- soft-decision methods [34,35],
- bias compensated tracking of spectral minima ("minimum statistics") [36,37],

or a combination of these, as, e.g., developed by Cohen [38]. In general, these methods rely on the assumptions that

- speech and noise are statistically independent,
- speech is not always present, and
- noise is more stationary than speech.

For single-microphone systems it is in general difficult to track non-stationary noise mostly because a sudden increase in noise power in one or several frequency bins cannot easily distinguished from a speech onset. Only after a few hundred milliseconds can speech and noise components be reliably discriminated. Therefore, it is difficult to identify and to suppress short noise bursts or competing speakers. Current developments strive to improve the performance of noise estimation under non-stationary conditions [37,38].

In what follows, we briefly outline the minimum statistics approach. The power of this approach relies on the intrinsically non-linear minimum extraction and the subsequent bias compensation. It has been shown that this method can contribute significantly to the intelligibility and the listening ease of the enhanced signal especially in conjunction with a low bit rate speech coder.

3.6.1 Minimum Statistics Noise Power Estimation

Since speech and noise are additive and statistically independent we have

$$\mathrm{E}\left\{|Y_\mu(k)|^2\right\} = \mathrm{E}\left\{|S_\mu(k)|^2\right\} + \mathrm{E}\left\{|N_\mu(k)|^2\right\}. \tag{3.33}$$

Recursive smoothing of the magnitude-squared spectral coefficients leads to

$$P_\mu(k) = \beta_\mu(k)\, P_\mu(k-r) + (1-\beta_\mu(k))\, |Y_\mu(k)|^2, \tag{3.34}$$

where $\beta_\mu(k)$ is a time and frequency dependent smoothing parameter. We now search for the minimum from D samples of the smoothed power $P_\mu(k-\lambda r)$, $\lambda = 0, 1, \ldots, D-1$. Then, we might use this minimum as a first coarse estimate of the noise floor since

$$\begin{aligned} &\min\left(P_\mu(k), \ldots, P_\mu(k-(D-1)r)\right) \\ &\approx \min\left(P_{N,\mu}(k), \ldots, P_{N,\mu}(k-(D-1)r)\right). \end{aligned} \tag{3.35}$$

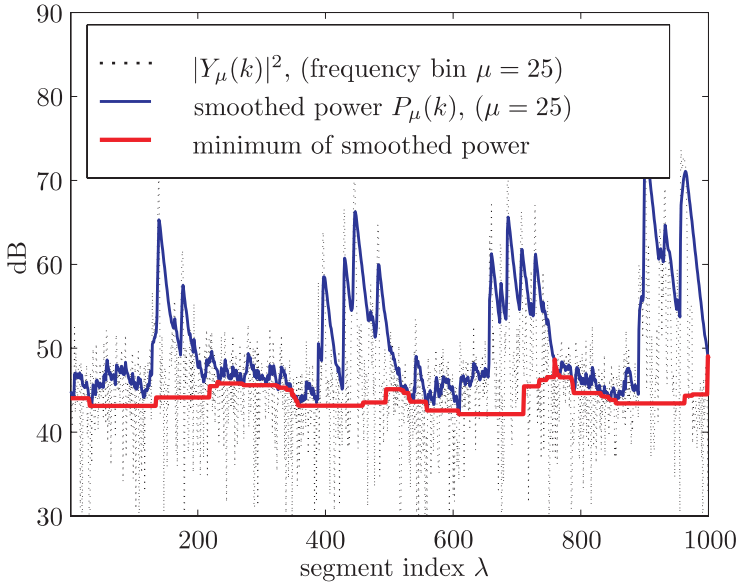

Fig. 3.7. Magnitude-squared DFT coefficient (dotted), smoothed power, and noise floor for a noisy speech signal (6 dB SNR).

$P_{N,\mu}(k)$ denotes a noise power estimate which is smoothed just like $P_\mu(k)$ in (3.34).

An example is shown in Figure 3.7 for a single frequency bin. Obviously, this estimate is biased towards lower values. However, the bias can be computed and compensated. It turns out that the bias depends on the variance of the smoothed power $P_\mu(k)$ which, in turn, is a function of the smoothing parameter $\beta_\mu(k)$ and of the variance of the signal under consideration. For recursively smoothed power estimates and a unity noise power, Fig. 3.8 plots the factor by which the minimum is smaller than the mean as a function of D and $Q_{eq} = 2\mathrm{E}\left\{|N_\mu(k)|^2\right\}^2 / \mathrm{var}\{P_\mu(k)\}$. Q_{eq} is the inverse normalized variance of the smoothed power. When much smoothing is applied $\mathrm{var}\{P_\mu(k)\}$ is relatively small and therefore Q_{eq} is large. Then, the minimum of subsequent values of $P_\mu(k)$ is close to the mean of these values. On the other hand, no smoothing ($Q_{eq} = 2$) requires a large bias compensation.

While earlier versions of the Minimum Statistics algorithm used a fixed smoothing parameter β and hence a fixed bias compensation we note that the full potential is only developed when a time and frequency dependent smoothing method is used. This in turn requires a time and frequency dependent bias compensation [37]. The result when using the adaptive smoothing and bias compensation is shown in Figure 3.9 for the same signal as in Figure 3.7.

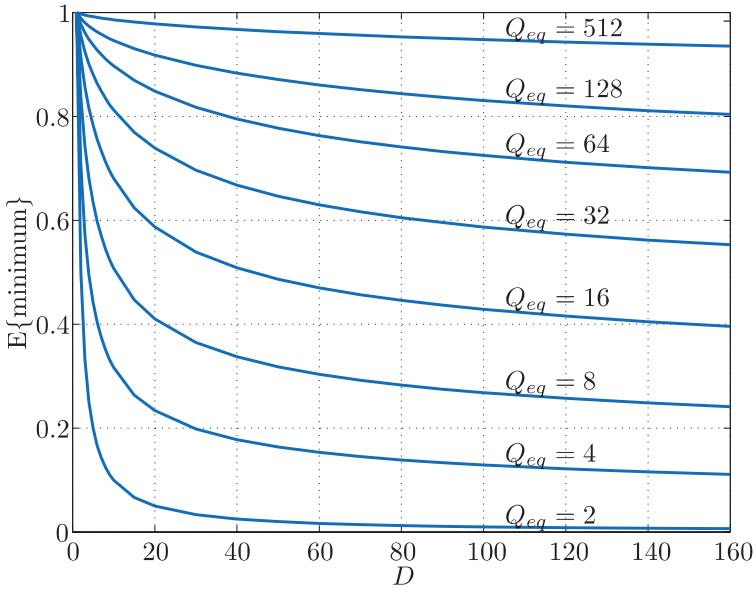

Fig. 3.8. Mean of the minimum of D correlated short term noise power estimates for $\sigma_n^2 = 1$.

3.7 The MELPe Speech Coder

As an application of the above techniques, we consider a speech enhancement algorithm which was developed for a low bit rate speech coder. Low bit rate speech coders are especially susceptible to environmental noise as they use a parametric model to code the input signal. One such example is the *mixed excitation linear prediction* (MELP) coder which operates at bit rates of 1.2 and 2.4 kbps [39]. It is used for secure governmental communications and is expected to succeed the well-known FS 1015 (LPC-10e) and FS 1016 (CELP) speech coding standards. This coder also includes an optional noise reduction preprocessor. The combined system of the preprocessor and the MELP coder is termed *MELPe* [39].

The noise reduction preprocessor [40] of the MELPe coder is based on

- the MMSE log spectral amplitude estimator [25];
- multiplicative soft-decision gain modification [35];
- adaptive gain limiting [14];
- estimation of the *a priori* SNR [35];
- *minimum statistics* noise power estimation [37].

This noise reduction preprocessor turns out to be very robust in a variety of noise environments and SNR conditions. Table 3.1 summarizes the results of a *diagnostic acceptability measure* (DAM) test for undisturbed and noisy

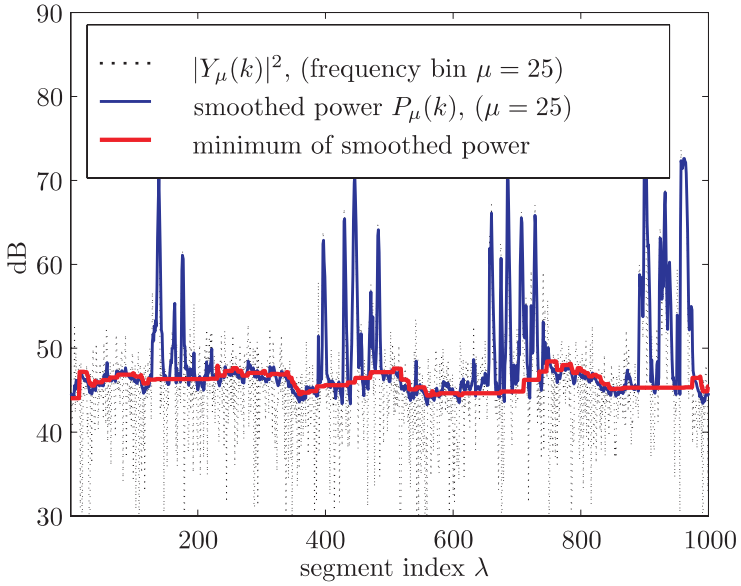

Fig. 3.9. Magnitude-squared DFT coefficient (dotted), smoothed power, and bias corrected noise floor for the same noisy speech signal as in Figure 3.7.

Table 3.1 DAM scores and standard error without noise and with vehicular noise (average SNR ≈ 6 dB).

condition	coder	DAM	standard error
no noise	MELPe	68.6	0.90
noisy	unprocessed	45.0	1.2
noisy	MELP	38.9	1.1
noisy	MELPe	50.3	0.80

conditions. As stated before, the MELP coder is highly sensitive to environmental noise. The noise reduction preprocessor helps to reduce these effects. Table 3.2 shows results of a *diagnostic rhyme test* (DRT) intelligibility evaluation for the same conditions as in the DAM test. We note, that the noisy but unprocessed signal has the highest intelligibility of the noisy conditions in Table 3.2. In conjunction with the MELP coder, the enhancement preprocessor leads to a significant improvement in terms of intelligibility. Thus, for a low bit rate speech coder, single-channel noise reduction systems can improve the quality as well as the intelligibility of the coded speech.

Table 3.2 DRT scores and standard error without noise and with vehicular noise (average SNR $\approx$ 6 dB).

condition	coder	DRT	standard error
no noise	MELPe	93.9	0.53
noisy	unprocessed	91.1	0.37
noisy	MELP	67.3	0.8
noisy	MELPe	72.5	0.58

3.8 Conclusions

Noise reduction technology is still an area of active research. While in the past decade most of these activities were triggered by new developments in mobile communications we now find increasing interest in automatic speech recognition and digital hearing aids applications.

Much of the research in this field is directed towards a better understanding and a better exploitation of the statistical properties of speech signals. As a result, several papers have been published which improve the estimation of critical (yet unknown) quantities such as the *a priori* SNR or the background noise power. Other approaches use optimal time domain estimators like Kalman filters which provide for an easy integration of autoregressive models. The question, however, of how the parameters of such models can be estimated in a robust fashion will require further research.

Further improvements are possible if we can employ more than one microphone and thus sample the sound field at more than one spatial location. There are a number of different ways to exploit multiple microphone signals. The most common are

- to use the spatial directivity of the microphone array [41,42],
- to adapt a single-channel *post-filter* based on the statistics of the microphone signals [43–46],

and combinations thereof. Some of these approaches are discussed, e.g., in [42]. Also, MAP and MMSE estimation of spectral amplitudes has been developed for the multi-microphone case, e.g., [47,48].

Despite these developments and many more which are not discussed here, there are still open questions which need to be addressed in the future:

- What are meaningful optimization criteria for speech enhancement and how can they be mathematically formulated?
- Which method of signal analysis is the most suitable?
- How can we improve the perceived quality of the enhanced signal without compromising intelligibility and vice versa?
- How can we combine signal theoretic and perceptual approaches?

- What kind of processing approach will be optimal for signals perceived by normal or hearing impaired persons, or, for signals processed by speech coders or speech recognition systems, and how are these approaches interrelated?
- What processing takes place in the higher stages of the auditory system and how can we model it?

Given all these questions it is clear that there will not be a single answer. We must, however, pay more attention to how humans process auditory information.

Acknowledgment

The author would like to thank Nilesh Madhu and Dirk Mauler for proofreading this manuscript and for valuable comments.

References

1. S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 27, pp. 113–120, 1979.
2. M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE ICASSP*, 1979, pp. 208–211.
3. J. Lim, ed., *Speech Enhancement*. Prentice-Hall, 1983.
4. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 32, pp. 1109–1121, Dec. 1984.
5. D. Van Compernolle, "DSP techniques for speech enhancement," in *Proc. Speech Processing in Adverse Conditions*, 1992, pp. 21–30.
6. R. Martin, "Statistical methods for the enhancement of noisy speech," in *Proc. IWAENC*, 2003, pp. 1–6.
7. Y. Ephraim and I. Cohen, "Recent advancements in speech enhancement," book chapter, CRC Press, 2004.
8. O. Cappé, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. Speech and Audio Processing*, vol. 2, pp. 345–349, Apr. 1994.
9. Y. Ephraim and H. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 4, pp. 251–266, 1995.
10. T. Gülzow and A. Engelsberg, "Comparison of a discrete wavelet transformation and a nonuniform polyphase filterbank applied to spectral subtraction speech enhancement," *Signal Processing, Elsevier*, vol. 64, no. 1, pp. 5–19, 1998.
11. S. Gustafsson, P. Jax, and P. Vary, "A novel psychoacoustically motivated audio enhancement algorithm preserving background noise characteristics," in *Proc. IEEE ICASSP*, 1998, pp. 397–400.

12. S. Gustafsson, R. Martin, P. Jax, and P. Vary, "A psychoacoustic approach to combined acoustic echo cancellation and noise reduction," *IEEE Trans. Speech and Audio Processing*, vol. 10, no. 5, pp. 245–256, 2002.
13. D. Griffin and J. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 32, pp. 236–243, Apr. 1984.
14. R. Martin and R. Cox, "New speech enhancement techniques for Low bit rate speech coding," in *Proc. IEEE Workshop on Speech Coding*, 1999, pp. 165–167.
15. P. Scalart and J. Vieira Filho, "Speech enhancement based on a priori signal to noise estimation," in *Proc. IEEE ICASSP*, 1996, pp. 629–632.
16. K. Linhard and T. Haulick, "Noise subtraction with parametric recursive gain curves," in *Proc. EUROSPEECH*, vol. 6, 1999, pp. 2611–2614.
17. C. Beaugeant and P. Scalart, "Speech enhancement using a minimum least-squares amplitude estimator," in *Proc. IWAENC*, 2001, pp. 191–194.
18. I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Processing, Elsevier*, vol. 81, pp. 2403–2418, 2001.
19. I. Cohen, "Speech enhancement using a noncausal *a priori* SNR estimator," *IEEE Signal Processing Letters*, vol. 11, pp. 725–728, 2004.
20. D. Wang and J. Lim, "The unimportance of phase in speech enhancement," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 30, no. 4, pp. 679–681, 1982.
21. I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*. Academic Press, 5th ed., 1994.
22. R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 28, pp. 137–145, Dec. 1980.
23. P. Wolfe and S. Godsill, "Simple alternatives to the Ephraim and Malah suppression rule for speech enhancement," in *IEEE Workshop on Statistical Signal Processing*, 2001, pp. 496–499.
24. T. Lotter and P. Vary, "Noise reduction by maximum a posteriori spectral amplitude estimation with supergaussian speech modeling," in *Proc. IWAENC*, 2003, pp. 83–86.
25. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 33, pp. 443–445, Apr. 1985.
26. D. O'Shaughnessy, *Speech Communications*. IEEE Press, 2 ed., 2000.
27. J. Porter and S. Boll, "Optimal estimators for spectral restoration of noisy speech," in *Proc. IEEE ICASSP*, 1984, pp. 18A.2.1–18A.2.4.
28. R. Martin, "Speech enhancement using MMSE short time spectral estimation with Gamma distributed speech priors," in *Proc. IEEE ICASSP*, vol. I, 2002, pp. 253–256.
29. D. Brillinger, *Time Series: Data Analysis and Theory*. Holden-Day, 1981.
30. R. Martin, "Speech enhancement based on minimum mean square error estimation and supergaussian priors," *IEEE Trans. Speech and Audio Processing*,to appear, 2005.
31. C. Breithaupt and R. Martin, "MMSE estimation of magnitude-squared DFT coefficients with supergaussian priors," in *Proc. IEEE ICASSP*, vol. I, 2003, pp. 848–851.
32. R. Martin and C. Breithaupt, "Speech enhancement in the DFT domain using Laplacian speech priors," in *Proc. IWAENC*, 2003, pp. 87–90.

33. D. Van Compernolle, "Noise adaptation in a hidden Markov model speech recognition system," *Computer Speech and Language*, vol. 3, pp. 151–167, 1989.
34. J. Sohn and W. Sung, "A Voice activity detector employing soft decision based noise spectrum adaptation," in *Proc. IEEE ICASSP*, vol. 1, 1998, pp. 365–368.
35. D. Malah, R. Cox, and A. Accardi, "Tracking speech-presence uncertainty to improve speech enhancement in non-stationary noise environments," in *Proc. IEEE ICASSP*, 1999, pp. 789–792.
36. R. Martin, "Spectral subtraction based on minimum statistics," in *Proc. EUSIPCO*, 1994, pp. 1182–1185.
37. R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech and Audio Processing*, vol. 9, pp. 504–512, July 2001.
38. I. Cohen, "Noise estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Trans. Speech and Audio Processing*, vol. 11, pp. 466–475, Sept. 2003.
39. T. Wang, K. Koishida, V. Cuperman, A. Gersho, and J. Collura, "A 1200/2400 BPS coding suite based on MELP," in *IEEE Workshop on Speech Coding*, 2002, pp. 90–92.
40. R. Martin, D. Malah, R. Cox, and A. Accardi, "A noise reduction preprocessor for Mobile voice communication," *EURASIP Journal on Applied Signal Processing*, vol. 2004, pp. 1046–1058, Aug. 2004.
41. G. Elko, "Microphone array systems for hands-free telecommunication," in *Proc. IWAENC*, 1995, pp. 31–38.
42. M. Brandstein and D. B. Ward, eds., *Microphone Arrays*. Springer-Verlag, Berlin, 2001.
43. R. Zelinski, "A Microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *Proc. IEEE ICASSP*, 1988, pp. 2578–2581.
44. C. Marro, Y. Mahieux, and K. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. Speech and Audio Processing*, vol. 6, no. 3, pp. 240–259, 1998.
45. J. Bitzer, K. Simmer, and K.-D. Kammeyer, "Multi-microphone noise reduction by post-filter and superdirective beamformer," in *Proc. IWAENC*, 1999, pp. 100–103.
46. R. Martin, "Small microphone arrays with postfilters for noise and acoustic echo reduction," in *Microphone Arrays* (M. Brandstein and D. B. Ward, eds.), Springer-Verlag, Berlin, 2001.
47. R. Balan and J. Rosca, "Microphone array speech enhancement by Bayesian estimation of spectral amplitude and phase," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2002.
48. T. Lotter, C. Benien, and P. Vary, "Multichannel speech enhancement using Bayesian spectral amplitude estimation," in *Proc. IEEE ICASSP*, 2003.