

Efficient Alternatives to the Ephraim and Malah Suppression Rule for Audio Signal Enhancement

Patrick J. Wolfe

*Signal Processing Group, Department of Engineering, University of Cambridge, CB2 1PZ Cambridge, UK
Email: pjw47@eng.cam.ac.uk*

Simon J. Godsill

*Signal Processing Group, Department of Engineering, University of Cambridge, CB2 1PZ Cambridge, UK
Email: sjg@eng.cam.ac.uk*

Received 31 May 2002 and in revised form 20 February 2003

Audio signal enhancement often involves the application of a time-varying filter, or suppression rule, to the frequency-domain transform of a corrupted signal. Here we address suppression rules derived under a Gaussian model and interpret them as spectral estimators in a Bayesian statistical framework. With regard to the optimal spectral amplitude estimator of Ephraim and Malah, we show that under the same modelling assumptions, alternative methods of Bayesian estimation lead to much simpler suppression rules exhibiting similarly effective behaviour. We derive three of such rules and demonstrate that, in addition to permitting a more straightforward implementation, they yield a more intuitive interpretation of the Ephraim and Malah solution.

Keywords and phrases: noise reduction, speech enhancement, Bayesian estimation.

1. INTRODUCTION

Herein we address an important issue in audio signal processing for multimedia communications, that of broadband noise reduction for audio signals via statistical modelling of their spectral components. Due to its ubiquity in applications of this nature, we concentrate on short-time spectral attenuation, a popular method of broadband noise reduction in which a time-varying filter, or suppression rule, is applied to the frequency-domain transform of a corrupted signal. We first address existing suppression rules derived under a Gaussian statistical model and interpret them in a Bayesian framework. We then employ the same model and framework to derive three new suppression rules exhibiting similarly effective behaviour, preliminary details of which may also be found in [1]. These derivations lead in turn to a more intuitive means of understanding the behaviour of the well-known Ephraim and Malah suppression rule [2], as well as to an extension of certain others [3, 4].

This paper is organised as follows. In the remainder of Section 1, we introduce the assumed statistical model and estimation framework, and then employ these in an alternate derivation of the minimum mean square error (MMSE) suppression rules due to Wiener [5] and Ephraim and Malah [2]. In Section 2, we derive three alternatives to the MMSE spec-

tral amplitude estimator of [2], all of which may be formulated as suppression rules. Finally, in Section 3, we investigate the behaviour of these solutions and compare their performance to that of the Ephraim and Malah suppression rule. Throughout the ensuing discussion, we consider—for simplicity of notation and without loss of generality—the case of a single, windowed segment of audio data. To facilitate a comparison, our notation follows that of [2], except that complex quantities appear in bold.

1.1. A simple Gaussian model

To date, the most popular methods of broadband noise reduction involve the application of a time-varying filter to the frequency-domain transform of a noisy signal. Let $x_n = x(nT)$ in general represent values from a finite-duration analogue signal sampled at a regular interval T , in which case a corrupted sequence may be represented by the additive observation model

$$y_n = x_n + d_n, \quad (1)$$

where y_n represents the observed signal at time index n , x_n is the original signal, and d_n is additive random noise, uncorrelated with the original signal. The goal of signal enhancement is then to form an estimate \hat{x}_n of the underlying signal x_n based on the observed signal y_n , as shown in Figure 1.

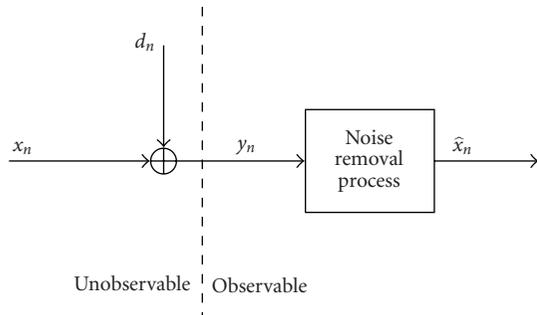


FIGURE 1: Signal enhancement in the case of additive noise.

In many implementations where efficient online performance is required, the set of observations $\{y_n\}$ is filtered using the overlap-add method of short-time Fourier analysis and synthesis, in a manner known as short-time spectral attenuation. Taking the discrete Fourier transform on windowed intervals of length N yields K frequency bins per interval:

$$\mathbf{Y}_k = \mathbf{X}_k + \mathbf{D}_k, \quad (2)$$

where these quantities are denoted in bold to indicate that they are complex. Noise reduction in this manner may be viewed as the application of a suppression rule, or nonnegative real-valued gain H_k , to each bin k of the observed signal spectrum \mathbf{Y}_k , in order to form an estimate $\hat{\mathbf{X}}_k$ of the original signal spectrum:

$$\hat{\mathbf{X}}_k = H_k \cdot \mathbf{Y}_k. \quad (3)$$

As shown in Figure 2, this spectral estimate is then inverse-transformed to obtain the time-domain signal reconstruction.

Within such a framework, a simple Gaussian model often proves effective [6, Chapter 6]. In this case, the elements of $\{\mathbf{X}_k\}$ and $\{\mathbf{D}_k\}$ are modelled as independent, zero-mean, complex Gaussian random variables with variances $\lambda_x(k)$ and $\lambda_d(k)$, respectively:

$$\mathbf{X}_k \sim \mathcal{N}_2(\mathbf{0}, \lambda_x(k)\mathbf{I}), \quad \mathbf{D}_k \sim \mathcal{N}_2(\mathbf{0}, \lambda_d(k)\mathbf{I}). \quad (4)$$

1.2. A Bayesian interpretation of suppression rules

It is instructive to consider an interpretation of suppression rules based on the Gaussian model of (4) in terms of a Bayesian statistical framework. Viewed in this light, the required task is to estimate each component \mathbf{X}_k of the underlying signal spectrum as a function of the corresponding observed spectral component \mathbf{Y}_k . To do so, we may define a nonnegative cost function $C(\mathbf{x}_k, \hat{\mathbf{x}}_k)$ of \mathbf{x}_k (the realisation of \mathbf{X}_k) and its estimate $\hat{\mathbf{x}}_k$, and then minimise the risk $\mathcal{R} \triangleq E[C(\mathbf{x}_k, \hat{\mathbf{x}}_k)|\mathbf{Y}_k]$ in order to obtain the optimal estimator of \mathbf{x}_k .

1.2.1. The Wiener suppression rule

A frequent goal in signal enhancement is to minimise the mean square error of an estimator; within the framework of Bayesian risk theory, this MMSE criterion may be viewed as a

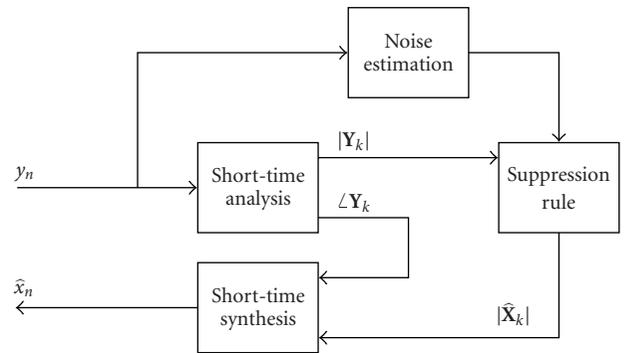


FIGURE 2: Short-time spectral attenuation.

squared-error cost function. Considering the model of (2), it follows from Bayes' rule and the prior distributions defined in (4) that we seek to minimise

$$E[C(\mathbf{x}_k, \hat{\mathbf{x}}_k)|\mathbf{Y}_k] \propto \int_{\mathbf{x}_k} |\hat{\mathbf{x}}_k - \mathbf{x}_k|^2 \exp\left(-\frac{|\mathbf{y}_k - \mathbf{x}_k|^2}{\lambda_d(k)} - \frac{|\mathbf{x}_k|^2}{\lambda_x(k)}\right) d\mathbf{x}_k. \quad (5)$$

The corresponding Bayes estimator is the optimal solution in an MMSE sense, and is given by the mean of the posterior density appearing in (5), which follows directly from its Gaussian form:

$$E[\mathbf{X}_k|\mathbf{Y}_k] = \frac{\lambda_x(k)}{\lambda_x(k) + \lambda_d(k)} \mathbf{Y}_k. \quad (6)$$

The result given by (6) is recognisable as the well-known Wiener filter [5].

In fact, it can be shown (see, e.g., [7, pages 59–63]) that when the posterior density is unimodal and symmetric about its mean, the conditional mean is the resultant Bayes estimator for a large class of nondecreasing, symmetric cost functions. However, we soon move to consider densities that are inherently asymmetric. Thus we will also employ the so-called *uniform* cost function, for which the optimal estimator may be shown to be that which maximises the posterior density—that is, the maximum a posteriori (MAP) estimator.

1.2.2. The Ephraim and Malah suppression rule

While, from a perceptual point of view, the ear is by no means insensitive to phase, the relative importance of spectral amplitude rather than phase in audio signal enhancement [8, 9] has led researchers to recast the spectral estimation problem in terms of the former quantity. In this vein, McAulay and Malpass [4] derive a maximum-likelihood (ML) spectral amplitude estimator under the assumption of Gaussian noise and an original signal characterised by a deterministic waveform of unknown amplitude and phase:

$$H_k = \frac{1}{2} + \frac{1}{2} \sqrt{\frac{\lambda_x(k)}{\lambda_x(k) + \lambda_d(k)}}. \quad (7)$$

As an extension of the model underlying (7), Ephraim and Malah [2] derive an MMSE short-time spectral amplitude estimator based on the model of (4); that is, under the assumption that the Fourier expansion coefficients of the original signal and the noise may be modelled as statistically independent, zero-mean, Gaussian random variables. Thus the observed spectral component in bin k , $\mathbf{Y}_k \triangleq R_k \exp(j\vartheta_k)$, is equal to the sum of the spectral components of the signal, $\mathbf{X}_k \triangleq A_k \exp(j\alpha_k)$, and the noise, \mathbf{D}_k . This model leads to the following marginal, joint, and conditional distributions:

$$p(a_k) = \begin{cases} \frac{2a_k}{\lambda_x(k)} \exp\left(-\frac{a_k^2}{\lambda_x(k)}\right) & \text{if } a_k \in [0, \infty), \\ 0 & \text{otherwise,} \end{cases} \quad (8)$$

$$p(\alpha_k) = \begin{cases} \frac{1}{2\pi} & \text{if } \alpha_k \in [-\pi, \pi), \\ 0 & \text{otherwise,} \end{cases} \quad (9)$$

$$p(a_k, \alpha_k) = \frac{a_k}{\pi\lambda_x(k)} \exp\left(-\frac{a_k^2}{\lambda_x(k)}\right), \quad (10)$$

$$p(\mathbf{Y}_k | a_k, \alpha_k) = \frac{1}{\pi\lambda_d(k)} \exp\left(-\frac{|\mathbf{Y}_k - a_k e^{j\alpha_k}|^2}{\lambda_d(k)}\right), \quad (11)$$

where it is understood that (10) and (11) are defined over the range of a_k and α_k , as given in (8) and (9), respectively; again $\lambda_x(k) \triangleq E[|\mathbf{X}_k|^2]$ and $\lambda_d(k) \triangleq E[|\mathbf{D}_k|^2]$ denote the respective variances of the k th short-time spectral component of the signal and noise. Additionally, define

$$\frac{1}{\lambda(k)} \triangleq \frac{1}{\lambda_x(k)} + \frac{1}{\lambda_d(k)}, \quad (12)$$

$$v_k \triangleq \frac{\xi_k}{1 + \xi_k} \gamma_k; \quad \xi_k \triangleq \frac{\lambda_x(k)}{\lambda_d(k)}, \quad \gamma_k \triangleq \frac{R_k^2}{\lambda_d(k)}, \quad (13)$$

where ξ_k and γ_k are interpreted after [4] as the a priori and a posteriori signal-to-noise ratios (SNRs), respectively.

Under the assumed model, the posterior density $p(a_k | \mathbf{Y}_k)$ (following integration with respect to the phase term α_k) is Rician [10] with parameters (σ_k^2, s_k^2) :

$$p(a_k | \mathbf{Y}_k) = \frac{a_k}{\sigma_k^2} \exp\left(-\frac{a_k^2 + s_k^2}{2\sigma_k^2}\right) I_0\left(\frac{a_k s_k}{\sigma_k^2}\right), \quad (14)$$

$$\sigma_k^2 \triangleq \frac{\lambda(k)}{2}, \quad s_k^2 \triangleq v_k \lambda(k), \quad (15)$$

where $I_i(\cdot)$ denotes the modified Bessel function of order i . The m th moment of a Rician distribution is given by

$$E[X^m] = (2\sigma^2)^{m/2} \Gamma\left(\frac{m+2}{2}\right) \times \Phi\left(\frac{m+2}{2}, 1; \frac{s^2}{2\sigma^2}\right) \exp\left(-\frac{s^2}{2\sigma^2}\right), \quad m \geq 0, \quad (16)$$

where $\Gamma(\cdot)$ is the gamma function [11, equation (8.310.1)]

and $\Phi(\cdot)$ is the confluent hypergeometric function [11, equation (9.210.1)].

The MMSE solution of Ephraim and Malah is simply the first moment of (14); when combined with the optimal phase estimator (found by Ephraim and Malah to be the observed phase ϑ_k [2]), it takes the form of a suppression rule:

$$\hat{A}_k = \lambda(k)^{1/2} \Gamma(1.5) \Phi(1.5, 1; v_k) \exp(-v_k) \\ = \lambda(k)^{1/2} \Gamma(1.5) \Phi(-0.5, 1; -v_k) \quad (17)$$

$$\Rightarrow H_k = \frac{\sqrt{\pi v_k}}{2\gamma_k} \left[(1 + v_k) I_0\left(\frac{v_k}{2}\right) + v_k I_1\left(\frac{v_k}{2}\right) \right] \exp\left(\frac{-v_k}{2}\right). \quad (18)$$

2. THREE ALTERNATIVE SUPPRESSION RULES

The spectral amplitude estimator given by (18), while being optimal in an MMSE sense, requires the computation of exponential and Bessel functions. We now proceed to derive three alternative suppression rules under the same model, each of which admits a more straightforward implementation.

2.1. Joint maximum a posteriori spectral amplitude and phase estimator

As shown earlier, joint estimation of the real and imaginary components of \mathbf{X}_k under either the MAP or MMSE criterion leads to the Wiener estimator (due to symmetry of the Gaussian posterior distribution). However, as we have seen, the problem may be reformulated in terms of spectral amplitude A_k and phase α_k ; it is then possible to obtain a joint MAP estimate by maximising the posterior distribution $p(a_k, \alpha_k | \mathbf{Y}_k)$:

$$p(a_k, \alpha_k | \mathbf{Y}_k) \\ \propto p(\mathbf{Y}_k | a_k, \alpha_k) p(a_k, \alpha_k) \\ \propto \frac{a_k}{\pi^2 \lambda_x(k) \lambda_d(k)} \exp\left(-\frac{|\mathbf{Y}_k - a_k e^{j\alpha_k}|^2}{\lambda_d(k)} - \frac{a_k^2}{\lambda_x(k)}\right). \quad (19)$$

Since $\ln(\cdot)$ is a monotonically increasing function, one may equivalently maximise the natural logarithm of $p(a_k, \alpha_k | \mathbf{Y}_k)$. Define

$$J_1 = -\frac{|\mathbf{Y}_k - a_k e^{j\alpha_k}|^2}{\lambda_d(k)} - \frac{a_k^2}{\lambda_x(k)} + \ln a_k + \text{constant}. \quad (20)$$

Differentiating J_1 with respect to α_k yields

$$\frac{\partial}{\partial \alpha_k} J_1 = -\frac{1}{\lambda_d(k)} [(\mathbf{Y}_k^* - a_k e^{-j\alpha_k})(-j a_k e^{j\alpha_k}) \\ + (\mathbf{Y}_k - a_k e^{j\alpha_k})(j a_k e^{-j\alpha_k})], \quad (21)$$

where \mathbf{Y}_k^* denotes the complex conjugate of \mathbf{Y}_k . Setting to zero and substituting $\mathbf{Y}_k = R_k \exp(j\vartheta_k)$, we obtain

$$0 = j \hat{a}_k R_k e^{j(\vartheta_k - \hat{\alpha}_k)} - j \hat{a}_k R_k e^{-j(\vartheta_k - \hat{\alpha}_k)} \\ = 2j \sin(\vartheta_k - \hat{\alpha}_k) \quad (22)$$

since $\hat{a}_k \neq 0$ if the phase estimate is to be meaningful. Therefore

$$\hat{\alpha}_k = \vartheta_k; \quad (23)$$

that is, the joint MAP phase estimate is simply the noisy phase—just as in the case of the MMSE solution due to Ephraim and Malah [2]. Differentiating J_1 with respect to a_k yields

$$\begin{aligned} \frac{\partial}{\partial a_k} J_1 = & -\frac{1}{\lambda_d(k)} [(\mathbf{Y}_k^* - a_k e^{-j\alpha_k})(-e^{j\alpha_k}) \\ & + (\mathbf{Y}_k - a_k e^{j\alpha_k})(-e^{-j\alpha_k})] \\ & - \frac{2a_k}{\lambda_x(k)} + \frac{1}{a_k}. \end{aligned} \quad (24)$$

Setting the above to zero implies

$$\begin{aligned} 2\hat{a}_k^2 = & \lambda_x(k) - \frac{\lambda_x(k)}{\lambda_d(k)} \hat{a}_k [2\hat{a}_k - R_k e^{-j(\vartheta_k - \hat{\alpha}_k)} - R_k e^{j(\vartheta_k - \hat{\alpha}_k)}] \\ = & \lambda_x(k) - \xi_k \hat{a}_k [2\hat{a}_k - 2R_k \cos(\vartheta_k - \hat{\alpha}_k)]. \end{aligned} \quad (25)$$

From (23), we have $\cos(\vartheta_k - \hat{\alpha}_k) = 1$; therefore

$$0 = 2(1 + \xi_k) \hat{a}_k^2 - 2R_k \xi_k \hat{a}_k - \lambda_x(k), \quad (26)$$

where ξ_k is as defined in (13). Solving the above quadratic equation and substituting

$$\lambda_x(k) = \frac{\xi_k}{\gamma_k} R_k^2, \quad (27)$$

which follows from the definitions of ξ_k and γ_k in (13), we have

$$\hat{A}_k = \frac{\xi_k + \sqrt{\xi_k^2 + 2(1 + \xi_k)(\xi_k/\gamma_k)}}{2(1 + \xi_k)} R_k. \quad (28)$$

Equations (23) and (28) together define the following suppression rule:

$$H_k = \frac{\xi_k + \sqrt{\xi_k^2 + 2(1 + \xi_k)(\xi_k/\gamma_k)}}{2(1 + \xi_k)}. \quad (29)$$

2.2. Maximum a posteriori spectral amplitude estimator

Recall that the posterior density $p(a_k | \mathbf{Y}_k)$ of (14), arising from integration over the phase term α_k , is Rician with parameters (σ_k^2, s_k^2) . Following McAulay and Malpass [4], we may for large arguments of $I_0(\cdot)$ (i.e., when, for $\lambda_x(k) = A_k^2$, $\xi_k R_k \sqrt{1/[(1 + \xi_k)\lambda(k)]} \geq 3$) substitute the approximation

$$I_0(|x|) \approx \frac{1}{\sqrt{2\pi|x|}} \exp(|x|) \quad (30)$$

into (14), yielding

$$p(a_k | \mathbf{Y}_k) \approx \frac{1}{\sqrt{2\pi\sigma_k^2}} \left(\frac{a_k}{s_k}\right)^{1/2} \exp\left(-\frac{1}{2} \left[\frac{a_k - s_k}{\sigma_k}\right]^2\right), \quad (31)$$

which we note is “almost” Gaussian. Considering (31), and again taking the natural logarithm and maximising with respect to a_k , we obtain

$$J_2 = -\frac{1}{2} \left[\frac{a_k - s_k}{\sigma_k}\right]^2 + \frac{1}{2} \ln a_k + \text{constant}, \quad (32)$$

in which case

$$\frac{d}{da_k} J_2 = \frac{s_k - a_k}{\sigma_k^2} + \frac{1}{2a_k} \quad (33)$$

$$\Rightarrow 0 = \hat{a}_k^2 - s_k \hat{a}_k - \frac{\sigma_k^2}{2}. \quad (34)$$

Substituting (15) and (27) into (34) and solving, we arrive at the following equation, which represents an approximate closed-form MAP solution corresponding to the maximisation of (14) with respect to a_k :

$$\hat{A}_k = \frac{\xi_k + \sqrt{\xi_k^2 + (1 + \xi_k)(\xi_k/\gamma_k)}}{2(1 + \xi_k)} R_k. \quad (35)$$

Note that this estimator differs from that of the joint MAP solution only by a factor of two under the square root (owing to the factor $\sqrt{a_k}$ in (31), replacement with a_k would yield the spectral estimator of (28)).

Combining (35) with the Ephraim and Malah phase estimator (i.e., the observed phase ϑ_k) yields the following suppression rule:

$$H_k = \frac{\xi_k + \sqrt{\xi_k^2 + (1 + \xi_k)(\xi_k/\gamma_k)}}{2(1 + \xi_k)}. \quad (36)$$

In fact, this solution extends that of McAulay and Malpass [4], who use the same approximation of $I_0(\cdot)$ to enable the derivation of the ML estimator given by (7). In this sense, the suppression rule of (36) represents a generalisation of the (approximate) ML spectral amplitude estimator proposed in [4].

2.3. Minimum mean square error spectral power estimator

Recall that Ephraim and Malah formulated the first moment of a Rician posterior distribution, $E[A_k | \mathbf{Y}_k]$, as a suppression rule. The second moment of that distribution, $E[A_k^2 | \mathbf{Y}_k]$, reduces to a much simpler expression

$$E[A_k^2 | \mathbf{Y}_k] = 2\sigma_k^2 + s_k^2, \quad (37)$$

where σ_k^2 and s_k^2 are as defined in (15). Letting $B_k = A_k^2$ and substituting for σ_k^2 and s_k^2 in (37) yields

$$\hat{B}_k = \frac{\xi_k}{1 + \xi_k} \left(\frac{1 + v_k}{\gamma_k}\right) R_k^2, \quad (38)$$

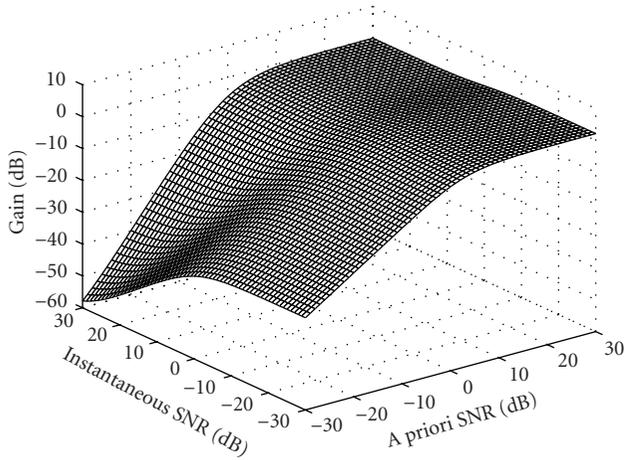


FIGURE 3: Ephraim and Malah MMSE suppression rule.

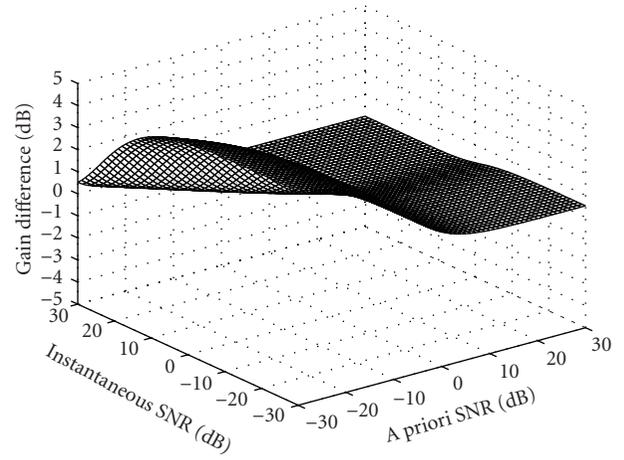


FIGURE 5: MAP approximation suppression rule gain difference.

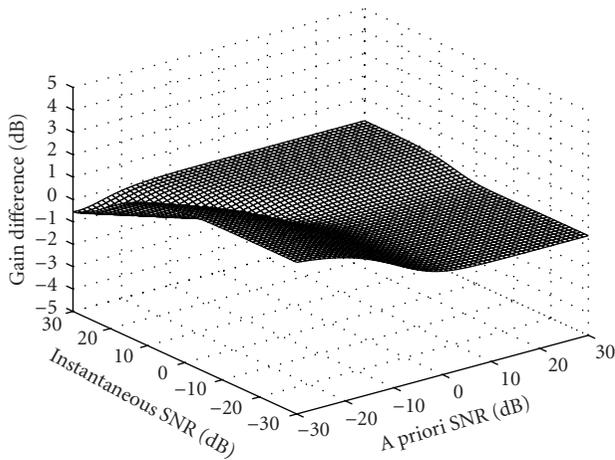


FIGURE 4: Joint MAP suppression rule gain difference.

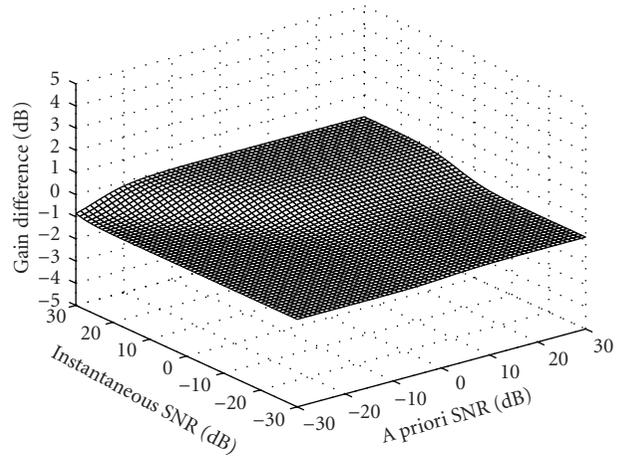


FIGURE 6: MMSE power suppression rule gain difference.

where \hat{B}_k is the optimal spectral power estimator in an MMSE sense, as it is also the first moment of a new posterior distribution $p(b_k | \mathbf{Y}_k)$ having a noncentral chi-square probability density function with two degrees of freedom and parameters (σ_k^2, s_k^2) .

When combined with the optimal phase estimator of Ephraim and Malah (i.e., the observed phase ϑ_k), this estimator also takes the form of a suppression rule

$$H_k = \sqrt{\frac{\xi_k}{1 + \xi_k} \left(\frac{1 + v_k}{\gamma_k} \right)}. \quad (39)$$

3. ANALYSIS OF ESTIMATOR BEHAVIOUR

Figure 3 shows the Ephraim and Malah suppression rule as a function of instantaneous SNR (defined in [2] as $\gamma_k - 1$)

and a priori SNR ξ_k .¹ Figures 4, 5, and 6 show the gain difference (in decibels) between it and each of the three derived suppression rules, given by (29), (36), and (39), respectively (note the difference in scale). A comparison of the magnitude of these gain differences is shown in Table 1.

From these figures, it is apparent that the MMSE spectral power suppression rule of (39) follows the Ephraim and Malah solution most closely and consistently, with only slightly less suppression in regions of low a priori SNR. Table 1 also indicates that the approximate MAP suppression rule of (36) is still within 5 dB of the Ephraim and Malah rule value over a wide SNR range, despite the approximation

¹Recall that the a priori SNR is the “true but unobserved” SNR, whereas the instantaneous SNR is the “spectral subtraction estimate” thereof.

TABLE 1: Magnitude of deviation from MMSE suppression rule gain.

Suppression rule	$(\gamma_k - 1, \xi_k) \in [-30, 30]$ dB			$(\gamma_k - 1, \xi_k) \in [-100, 100]$ dB		
	Mean	Maximum	Range	Mean	Maximum	Range
MMSE power	0.68473	-1.0491	1.0469	0.63092	-1.0491	1.0491
Joint MAP	0.52192	+1.7713	2.3352	0.74507	+1.9611	2.5250
Approximate MAP	1.2612	+4.7012	4.7012	1.7423	+4.9714	4.9714

of (30).² While the sign of the deviation of both the MMSE spectral power and approximate MAP rules is constant, that of the joint MAP suppression rule of (29) depends on the instantaneous and a priori SNRs.

Ephraim and Malah [2] show that at high SNRs, their derived suppression rule converges to the Wiener suppression rule detailed in Section 1.2.1, formulated as a function of a priori SNR ξ_k :

$$H_k = \frac{\xi_k}{1 + \xi_k}. \quad (40)$$

This relationship is easily seen from the MMSE spectral power suppression rule given by (39), expanded slightly to the following equation:

$$H_k = \sqrt{\frac{\xi_k}{1 + \xi_k} \left(\frac{1}{\gamma_k} + \frac{\xi_k}{1 + \xi_k} \right)}. \quad (41)$$

As the instantaneous SNR becomes large, (41) may be seen to approach the Wiener suppression rule of (40). As it becomes small, the $1/\gamma_k$ term in (41) lessens the severity of the attenuation. Cappé [12] makes the same observation concerning the behaviour of the Ephraim and Malah suppression rule, although the simpler form of the MMSE spectral power estimator shows the influence of the a priori and a posteriori SNRs more explicitly.

We also note that the success of the Ephraim and Malah suppression rule is largely due to the authors' decision-directed approach for estimating the a priori SNR ξ_k [12]. For a given short-time block n , the decision-directed a priori SNR estimate $\hat{\xi}_k$ is given by a geometric weighting of the SNRs in the previous and current blocks:

$$\hat{\xi}_k = \alpha \frac{|\hat{\mathbf{X}}_k(n-1)|^2}{\lambda_d(n-1, k)} + (1 - \alpha) \max[0, \gamma_k(n) - 1], \quad \alpha \in [0, 1]. \quad (42)$$

It is instructive to consider the case in which $\xi_k = \gamma_k - 1$, that is, $\alpha = 0$ in (42) so that the estimate of the a priori SNR is based only on the spectral subtraction estimate of the

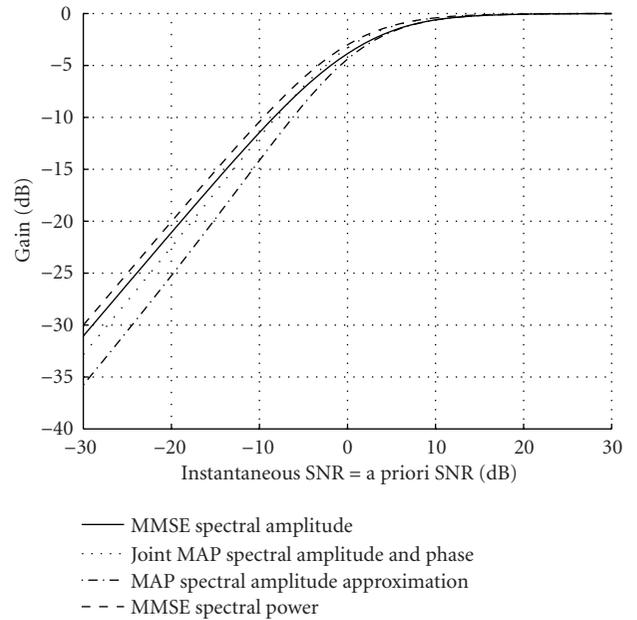


FIGURE 7: Optimal and derived suppression rules.

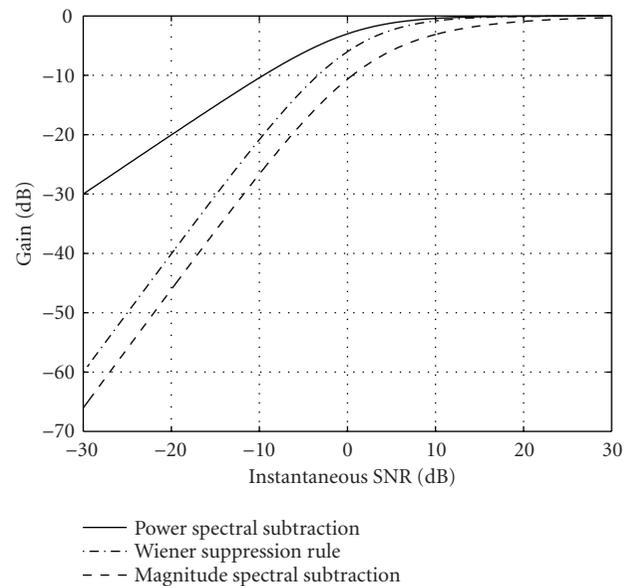


FIGURE 8: Standard suppression rules.

²For a fixed spectral magnitude observation R_k , and with $\lambda_x(k) = A_k^2$, the approximation of (30) is dominated by the a priori SNR ξ_k . Hence we see that when ξ_k is large, the resultant suppression rule gain exhibits less deviation from that of the other rules.

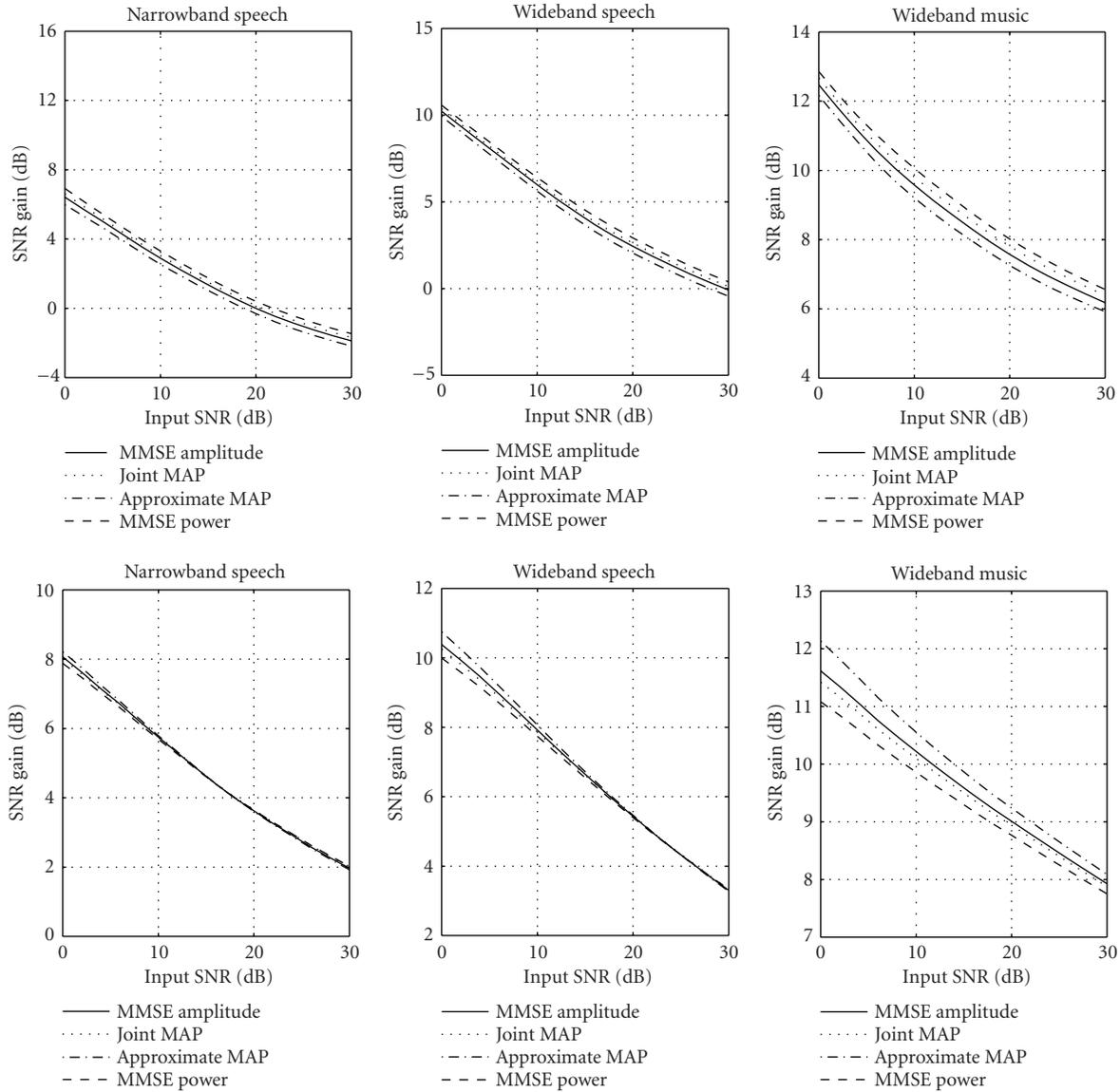


FIGURE 9: A performance comparison of the derived suppression rules. The top row of figures corresponds to a priori SNR estimation using the decision-directed approach of (42), with $\alpha = 0.98$ as recommended in [2]. The bottom row corresponds to $\alpha = 0$, in which case the gain surfaces of Figures 3, 4, 5, and 6 reduce to the gain curves of Figure 7.

current block. In this case, the MMSE spectral power suppression rule given by (41) reduces to the method of power spectral subtraction (see, e.g., [3]). Figure 7 shows a comparison of the derived suppression rules under this constraint; by way of comparison, Figure 8 shows some standard suppression rules, including power spectral subtraction and the Wiener filter, as a function of instantaneous SNR (note the difference in ordinate scale).

Lastly, we mention the results of informal listening tests conducted across a range of audio material. These tests indicate that, especially when coupled with the decision-directed approach for estimating ξ_k , each of the derived estimators yields an enhancement similar in quality to that obtained us-

ing the Ephraim and Malah suppression rule. To this end, Figure 9 shows a comparison of SNR gain over a range of input SNRs for three typical 16-bit audio examples, artificially degraded with additive white Gaussian noise, and processed using the overlap-add method with a 50% window overlap: narrowband speech (sampled at 16 kHz and analysed using a 256-sample hanning window), wideband speech (sampled at 44.1 kHz and analysed using a 512-sample hanning window), and wideband music (solo piano, sampled at 44.1 kHz and analysed using a 2048-sample Hanning window).³

³Segmental SNR gain measurements yield a similar pattern of results.

As we intend these results to be illustrative rather than exhaustive, we limit our direct comparison here to the Ephraim and Malah suppression rule. Comparisons have been made both with and without smoothing in the a priori SNR calculation, as described in the caption of Figure 9. It may be seen from Figure 9 that in the case of smoothing (upper row), the spectral power estimator appears to provide a small increase in SNR gain. In terms of sound quality, a small decrease in residual musical noise results from the approximate MAP solution, albeit at the expense of slightly more signal distortion. The joint MAP suppression rule lies in between these two extremes. Without smoothing, the methods produce a residual with approximately the same amount of musical noise as power spectral subtraction (as is expected in light of the comparison of these curves given by Figure 7). In comparison to Wiener filtering and magnitude spectral subtraction, the derived methods yield a slightly greater level of musical noise (as is to be expected according to Figure 8).

Audio examples illustrating these features, along with a Matlab toolbox allowing for the reproduction of results presented here, as well as further experimentation and comparison with other suppression rules, are available online at <http://www-sigproc.eng.cam.ac.uk/~pjw47>.

4. DISCUSSION

In the first part of this paper, we have provided a common interpretation of existing suppression rules based on a simple Gaussian statistical model. Within the framework of Bayesian estimation, we have seen how two MMSE suppression rules due to Wiener [5] and Ephraim and Malah [2] may be derived. While the Ephraim and Malah MMSE spectral amplitude estimator is well known and widely used, its implementation requires the evaluation of computationally expensive exponential and Bessel functions. Moreover, an intuitive interpretation of its behaviour is obscured by these same functions. With this motivation, we have presented in the second part of this paper a derivation and comparison of three alternatives to the Ephraim and Malah MMSE spectral amplitude estimator.

The derivations also yield an extension of two existing suppression rules: the ML spectral estimator due to McAulay and Malpass [4], and the estimator defined by power spectral subtraction. Specifically, the ML suppression rule has been generalised to an approximate MAP solution in the case of an independent Gaussian prior for each spectral component. It has also been shown that the well-known method of power spectral subtraction, previously developed in a non-Bayesian context, arises as a special case of the MMSE spectral power estimator derived herein.

In addition to providing the aforementioned theoretical insights, these solutions may be of use themselves in situations where a straightforward implementation involving simpler functional forms is required; alternative approaches along a similar line of motivation are developed in [13, 14]. Additionally, for the purposes of speech enhancement, each may be coupled with hypotheses concerning uncertainty of

speech presence, as in [2, 4, 13, 14]. Moreover, the form of the MMSE spectral power suppression rule given by (41) provides a clearer insight into the behaviour of the Ephraim and Malah solution. Finally, we note that just as Ephraim and Malah argued that log-spectral amplitude estimation may be more appropriate for speech perception [15], so in other cases may be MMSE spectral power estimation—for example, when calculating auditory masked thresholds for use in perceptually motivated noise reduction [16].

ACKNOWLEDGMENTS

Material by the first author is based upon work supported under a US National Science Foundation Graduate Fellowship. The authors also gratefully acknowledge the contribution of Shyue Ping Ong to this paper, as well as the helpful comments of the anonymous reviewers.

REFERENCES

- [1] P. J. Wolfe and S. J. Godsill, "Simple alternatives to the Ephraim and Malah suppression rule for speech enhancement," in *Proc. 11th IEEE Workshop on Statistical Signal Processing*, pp. 496–499, Orchid Country Club, Singapore, August 2001.
- [2] 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, no. 6, pp. 1109–1121, 1984.
- [3] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 208–211, Washington, DC, USA, April 1979.
- [4] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 28, no. 2, pp. 137–145, 1980.
- [5] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*, Principles of Electrical Engineering Series, MIT Press, Cambridge, Mass, USA, 1949.
- [6] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration: A Statistical Model Based Approach*, Springer-Verlag, Berlin, Germany, 1998.
- [7] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Part 1, Detection, Estimation and Linear Modulation Theory*, John Wiley & Sons, New York, NY, USA, 1968.
- [8] D. L. Wang and J. S. Lim, "The unimportance of phase in speech enhancement," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 30, no. 4, pp. 679–681, 1982.
- [9] P. Vary, "Noise suppression by spectral magnitude estimation—Mechanism and theoretical limits," *Signal Processing*, vol. 8, no. 4, pp. 387–400, 1985.
- [10] S. O. Rice, "Statistical properties of a sine wave plus random noise," *Bell System Technical Journal*, vol. 27, pp. 109–157, 1948.
- [11] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, San Diego, Calif, USA, 5th edition, 1994.
- [12] O. Cappé, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. Speech, and Audio Processing*, vol. 2, no. 2, pp. 345–349, 1994.
- [13] A. Akbari Azirani, R. le Bouquin Jeannès, and G. Faucon, "Optimizing speech enhancement by exploiting masking

- properties of the human ear,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 1, pp. 800–803, Detroit, Mich, USA, May 1995.
- [14] A. Akbari Azirani, R. le Bouquin Jeannès, and G. Faucon, “Speech enhancement using a Wiener filtering under signal presence uncertainty,” in *Signal Processing VIII: Theories and Applications*, G. Ramponi, G. L. Sicuranza, S. Carrato, and S. Marsi, Eds., vol. 2 of *Proceedings of the European Signal Processing Conference*, pp. 971–974, Trieste, Italy, September 1996.
- [15] 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, no. 2, pp. 443–445, 1985.
- [16] P. J. Wolfe and S. J. Godsill, “Towards a perceptually optimal spectral amplitude estimator for audio signal enhancement,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 2, pp. 821–824, Istanbul, Turkey, June 2000.

Patrick J. Wolfe attended the University of Illinois at Urbana-Champaign (UIUC) from 1993–1998, where he completed a self-designed programme leading to undergraduate degrees in electrical engineering and music. After working at the UIUC Experimental Music Studios in his final year and later at Studer Professional Audio AG, he joined the Signal Processing Group at the University of Cambridge. There he held a US National Science Foundation Graduate Research Fellowship at Churchill College, working towards his Ph.D. with Dr. Simon Godsill on the application of perceptual criteria to statistical audio signal processing, prior to his appointment in 2001 as a Fellow and College Lecturer in engineering and computer science at New Hall, University of Cambridge, Cambridge. His research interests lie in the intersection of statistical signal processing and time-frequency analysis, and include general applications as well as those related specifically to audio and auditory perception.



Simon J. Godsill is a Reader in statistical signal processing in the Engineering Department of Cambridge University. In 1988, following graduation in electrical and information sciences from Cambridge University, he led the technical development team at the audio enhancement company, CEDAR Audio, Ltd., researching and developing DSP algorithms for restoration of audio signals. Following this, he completed a Ph.D. with Professor Peter Rayner at Cambridge University and went on to be a Research Fellow of Corpus Christi College, Cambridge. He has research interests in Bayesian and statistical methods for signal processing, Monte Carlo algorithms for Bayesian problems, modelling and enhancement of audio signals, nonlinear and non-Gaussian signal processing, image sequence analysis, and genomic signal processing. He has published over 70 papers in refereed journals, conference proceedings, and edited books. He has authored a research text on sound processing, *Digital Audio Restoration*, with Peter Rayner, published by Springer-Verlag.

