SIMPLE ALTERNATIVES TO THE EPHRAIM AND MALAH SUPPRESSION RULE FOR SPEECH ENHANCEMENT

Patrick J. Wolfe and Simon J. Godsill*

Signal Processing Group, University of Cambridge
Department of Engineering, Trumpington Street
CB2 1PZ, Cambridge, UK
{pjw47, sjg}@eng.cam.ac.uk
http://www-sigproc.eng.cam.ac.uk

ABSTRACT

Short-time spectral attenuation is a common form of audio signal enhancement in which a time-varying filter, or suppression rule, is applied to the frequency-domain transform of a corrupted signal. The Ephraim and Malah suppression rule for speech enhancement is both optimal in the minimum mean-square error sense and well-known for its associated colourless residual noise; however, it requires the computation of exponential and Bessel functions. In this paper we show that, under the same modelling assumptions, alternative Bayesian approaches lead to suppression rules exhibiting almost identical behaviour. We derive three such rules and show that they are efficient to implement and yield a more intuitive interpretation.

1. INTRODUCTION

1.1. Short-Time Spectral Attenuation

Short-time spectral attenuation is a popular method of broadband noise reduction in which a time-varying filter is applied to the frequency-domain transform of a corrupted audio signal. Often such a signal is modelled as follows: let $\{x_n\} \triangleq \{x(nT)\}$ in general represent a set of values from a finite-duration analogue signal sampled regularly at intervals of T, so that at time n one has the additive observation model $y_n = x_n + d_n$, where y_n is the observed signal, x_n is the original signal, and d_n is random noise.

In many implementations the set of observations $\{y_n\}$ is analysed using the discrete Fourier transform (DFT), via the overlap-add method of short-time Fourier analysis and synthesis. 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 $\widehat{\mathbf{X}}_k$ of the original signal spectrum.

In the ensuing discussion of such suppression rules we consider, for simplicity of notation and without loss of generality, the case of a single (windowed) short-time block. To facilitate a comparison our notation follows that of Ephraim and Malah [1], except that complex quantities appear in bold throughout.

*Material by the first author is based upon work supported under a U.S. National Science Foundation Graduate Fellowship. The authors also wish to acknowledge the contribution of Shyue Ping Ong to this paper.

1.2. The Ephraim and Malah Suppression Rule

Ephraim and Malah [1] derive a minimum mean-square error (MMSE) short-time spectral amplitude estimator for speech enhancement under the assumption that the Fourier expansion coefficients of the original signal x_n and the noise d_n may be modelled as independent, zero-mean, Gaussian random variables. Thus the observed spectral component in DFT 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:

$$p(a_k) = \begin{cases} \frac{2a_k}{\lambda_x(k)} \exp\left(-\frac{a_k^2}{\lambda_x(k)}\right) & \text{if } a_k \in [0, \infty)\,, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

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

$$p(a_k, \alpha_k) = \frac{a_k}{\pi \lambda_x(k)} \exp\left(-\frac{a_k^2}{\lambda_x(k)}\right) \tag{3}$$

$$p(\mathbf{Y}_k | a_k, \alpha_k) = \frac{1}{\pi \lambda_d(k)} \exp\left(-\frac{\left|\mathbf{Y}_k - a_k e^{j\alpha_k}\right|^2}{\lambda_d(k)}\right) \tag{4}$$

where it is understood that (3) and (4) are defined over the range of a_k in (1) and α_k in (2); $\lambda_x(k) \triangleq E[|\mathbf{X}_k|^2]$ and $\lambda_d(k) \triangleq E[|\mathbf{D}_k|^2]$ denote the respective variances of the kth short-time spectral component of the signal and noise. The MMSE spectral amplitude estimator derived by Ephraim and Malah, when combined with their derived optimal phase estimator (the observed phase ϑ_k [1]), takes the form of a suppression rule:

$$H_k = \frac{\sqrt{\pi \upsilon_k}}{2\gamma_k} \left[(1 + \upsilon_k) I_0\left(\frac{\upsilon_k}{2}\right) + \upsilon_k I_1\left(\frac{\upsilon_k}{2}\right)\right] \exp\left(\frac{-\upsilon_k}{2}\right), \tag{5}$$

where $I_0(\cdot)$ and $I_1(\cdot)$ denote the modified Bessel functions of order zero and one, respectively. Additionally,

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

and

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

where ξ_k and γ_k are interpreted after [2] as the *a priori* and *a posteriori* SNR, respectively.

2. DERIVATION OF EFFICIENT APPROXIMATIONS

2.1. Joint Maximum *A Posteriori* Spectral Amplitude and Phase Estimator

Joint estimation of the real and imaginary components of $\mathbf{X}_k$ under either the maximum *a posteriori* (MAP) or MMSE criterion leads to the Wiener estimator (due to symmetry of the resultant posterior distribution, which is Gaussian). However, one may reformulate the problem in terms of spectral amplitude A_k and phase α_k, and then obtain a joint MAP estimate by maximising the posterior distribution $p(a_k, \alpha_k|\mathbf{Y}_k)$:

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

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

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

Differentiating J_1 w.r.t. α_k yields

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

Setting to zero and substituting $\mathbf{Y}_k = R_k \exp(j\vartheta_k)$, we get

$$\begin{aligned}0 &= j\hat{a}_k R_k e^{j(\vartheta_k - \hat{\alpha}_k)} - j\hat{a}_k R_k e^{-j(\vartheta_k - \hat{\alpha}_k)}\\ &= 2j\sin(\vartheta_k - \hat{\alpha}_k),\end{aligned}$$

and therefore

$$\hat{\alpha}_k = \vartheta_k, \tag{6}$$

i.e., the joint MAP phase estimate is simply the noise phase. Differentiating J_1 w.r.t. a_k yields

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

Setting the above to zero implies

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

From (6), we have $\cos(\vartheta_k - \hat{\alpha}_k) = 1$; therefore

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

Solving the above quadratic equation, and substituting

$$\lambda_x(k) = \frac{\xi_k}{\gamma_k}{R_k}^2, \tag{7}$$

we have

$$\widehat{A}_k = \frac{\xi_k + \sqrt{{\xi_k}^2 + 2(1+\xi_k)\frac{\xi_k}{\gamma_k}}}{2(1+\xi_k)}R_k. \tag{8}$$

Together (8) and (6) define the following suppression rule:

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

2.2. Maximum *A Posteriori* Spectral Amplitude Estimator

First we note that the posterior density $p(a_k|\mathbf{Y}_k)$ arising from integration over the phase term α_k is Rician with parameters (σ_k^2, s_k^2):

$$p(a_k|\mathbf{Y}_k) = \frac{a_k}{\sigma_k^2}\exp\left(-\frac{a_k^2 + s_k^2}{2\sigma_k^2}\right)I_0\left(\frac{a_k s_k}{\sigma_k^2}\right) \tag{9}$$

$$\sigma_k^2 \triangleq \frac{\lambda(k)}{2}, \quad s_k^2 \triangleq v_k\lambda(k). \tag{10}$$

for large arguments of $I_0(\cdot)$ we may substitute the approximation

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

into (9), yielding

$$p(a_k|\mathbf{Y}_k) \approx \frac{1}{\sqrt{2\pi\sigma_k^2}}\left(\frac{a_k}{s_k}\right)^{\frac{1}{2}}\exp\left(-\frac{1}{2}\left[\frac{a_k - s_k}{\sigma_k}\right]^2\right), \tag{11}$$

which is almost Gaussian. Considering (11), and maximising its natural logarithm w.r.t. a_k, we obtain

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

$$\begin{aligned}\frac{d}{da_k}J_2 &= \frac{s_k - a_k}{\sigma_k^2} + \frac{1}{2a_k}\\ 0 &= \hat{a}_k^2 - s_k\hat{a}_k - \frac{\sigma_k^2}{2}.\end{aligned} \tag{12}$$

Substituting (10) and (7) into (12) and solving, we arrive at an estimator differing 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 (11); replacement with a_k would yield the joint MAP spectral amplitude solution):

$$\widehat{A}_k = \frac{\xi_k + \sqrt{{\xi_k}^2 + (1+\xi_k)\frac{\xi_k}{\gamma_k}}}{2(1+\xi_k)}R_k. \tag{13}$$

Combining (13) with the Ephraim and Malah optimal phase estimator (i.e., the observed phase ϑ_k; *cf.* (6) also) yields the following suppression rule:

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

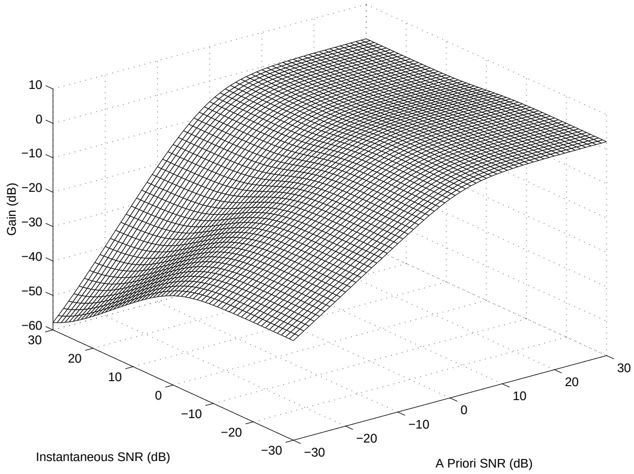

Fig. 1. Ephraim and Malah MMSE suppression rule

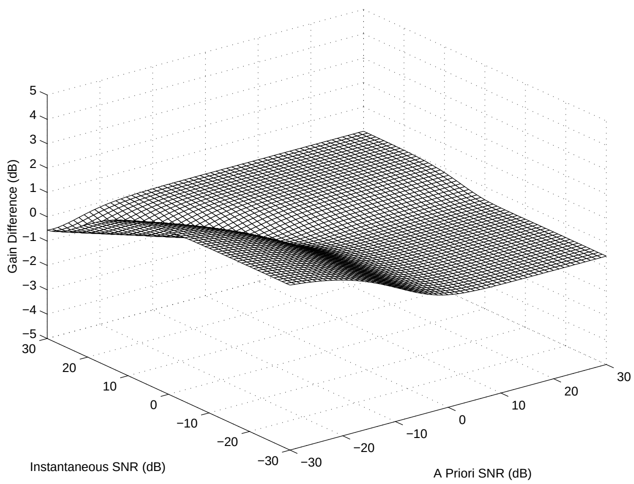

Fig. 2. Joint MAP suppression rule gain difference

2.3. Minimum Mean-Square Error Spectral Power Estimator

Recall that Ephraim and Malah formulate the first moment of a Rician posterior distribution, $E[A_k|\mathbf{Y}_k]$, as a suppression rule. The second moment $E[A_k^2|\mathbf{Y}_k]$ of that distribution is given by a much simpler formula (see, e.g., [3]):

$$E[A_k^2|\mathbf{Y}_k] = 2\sigma_k^2 + s_k^2, \tag{14}$$

where σ_k^2 and s_k^2 are as defined previously in (10). Letting $B_k = A_k^2$ and substituting for σ_k^2 and s_k^2 in (14) yields

$$\widehat{B}_k = \frac{\xi_k}{1+\xi_k}\left(\frac{1+\upsilon_k}{\gamma_k}\right)R_k^2,$$

where $\widehat{B}_k$ is the optimal spectral power estimator in the 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).

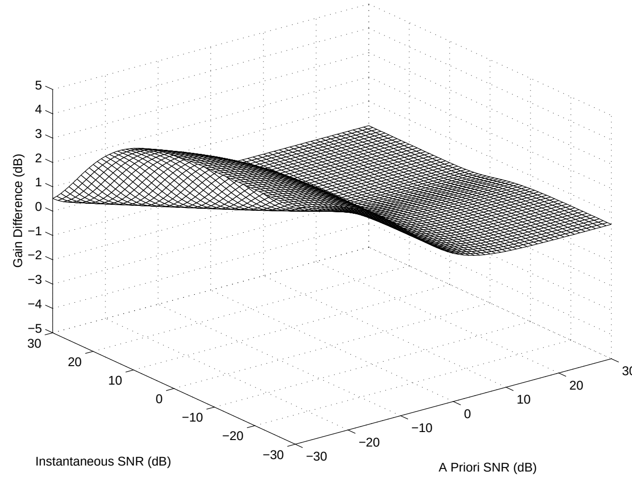

Fig. 3. MAP approximation suppression rule gain difference

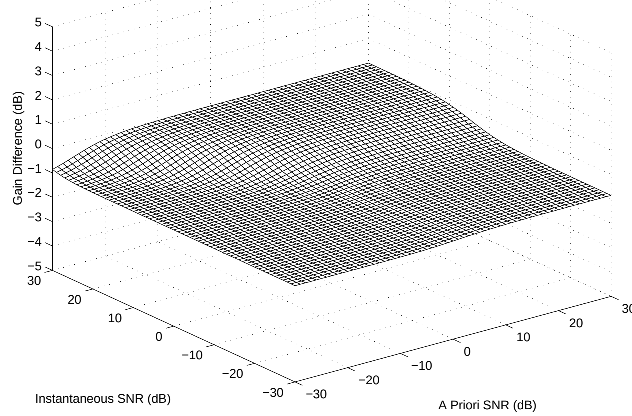

Fig. 4. MMSE power suppression rule gain difference

When combined with the observed phase ϑ_k, this estimator also takes the form of a suppression rule:

$$H_k = \sqrt{\frac{\xi_k}{1+\xi_k}\left(\frac{1+\upsilon_k}{\gamma_k}\right)}. \tag{15}$$

3. COMPARISON OF APPROXIMATIONS

Figure 1 shows the Ephraim and Malah suppression rule as a function of instantaneous SNR (defined in [1] as $\gamma_k - 1$) and *a priori* SNR ξ_k. Figures 2, 3, and 4 show the gain difference (in dB) between it and each of the three derived suppression rules (note the difference in scale). Table 1 on the following page shows a comparison of the magnitude of gain differences for the three approximations. The MMSE spectral power suppression rule provides the best and most consistent approximation to the Ephraim and Malah

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 Spectral Power	0.68473	−1.0491	1.0469	0.63092	−1.0491	1.0491
Joint MAP Spectral Amplitude and Phase	0.52192	+1.7713	2.3352	0.74507	+1.9611	2.5250
MAP Spectral Amplitude Approximation	1.2612	+4.7012	4.7012	1.7423	+4.9714	4.9714

Table 1. Magnitude of deviation from Ephraim and Malah MMSE suppression rule gain

rule, with only slightly less suppression. The MAP spectral amplitude approximation, although still within 5 dB of the optimal value over a wide range of SNR, is the poorest. While the sign of the deviation of each of these two approximations is constant, that of the joint MAP suppression rule depends on the instantaneous and *a priori* SNR.

4. DISCUSSION

Ephraim and Malah [1] show that at high SNR, their derived suppression rule approaches the Wiener suppression rule:

$$H_k = \frac{\xi_k}{1+\xi_k}. \tag{16}$$

Although not immediately obvious upon inspection of (5), this relationship is easily seen in the MMSE spectral power suppression rule given by (15), expanded slightly to the following:

$$H_k = \sqrt{\frac{\xi_k}{1+\xi_k}\left(\frac{1}{\gamma_k} + \frac{\xi_k}{1+\xi_k}\right)}. \tag{17}$$

As the instantaneous SNR γ_k becomes large, (17) may be seen to approach the Wiener suppression rule given by (16). As it becomes small, the $1/\gamma_k$ term in (17) lessens the severity of the attenuation. Cappé [4] makes the same qualitative 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* SNR more explicitly.

Lastly, we note that the success of the Ephraim and Malah suppression rule is largely due to the so-called ‘decision-directed approach’ for estimating the *a priori* SNR ξ_k [4]. For a given short-time block l, the decision-directed *a priori* SNR estimate $\widehat{\xi}_k$ is given by a geometric weighting of the SNR in the previous and current blocks:

$$\widehat{\xi}_k = \alpha \frac{|\widehat{\mathbf{X}}_k(l-1)|^2}{\lambda_d(l-1,\, k)} + (1-\alpha)\max\left[0, \gamma_k(l) - 1\right],\ \alpha \in [0,1). \tag{18}$$

It is instructive to consider the case in which $\xi_k = \gamma_k - 1$; i.e., $\alpha = 0$ in (18) so that the estimate of the *a priori* SNR is based only on the current block. In this case the MMSE spectral power suppression rule given by (17) reduces to the method of power spectral subtraction (see, e.g., [2]). Figure 5 shows a comparison of the derived suppression rules under this constraint.

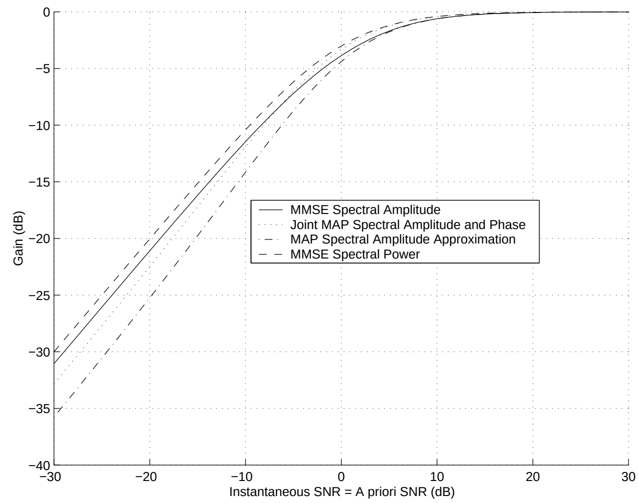

Fig. 5. Optimal and derived suppression rules

5. CONCLUSION

Herein we have presented a derivation and comparison of three simple alternatives to the Ephraim and Malah MMSE spectral amplitude estimator. These may be implemented where increased efficiency is desired, and each may be coupled with hypotheses concerning uncertainty of speech presence, as in [1, 2]. Moreover, the form of the MMSE spectral power suppression rule given by (17) provides a clear insight into the behaviour of the Ephraim and Malah solution, and in particular its connection to simpler suppression rules.

6. REFERENCES

[1] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[2] R. J. McAulay and M. L. Malpass, “Speech enhancement using a soft-decision noise suppression filter,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, no. 2, pp. 137–145, Apr. 1980.

[3] S. O. Rice, “Statistical properties of a sine wave plus random noise,” *Bell System Technical Journal*, vol. 27, pp. 109–157, Jan. 1948.

[4] O. Cappé, “Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor,” *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 345–349, Apr. 1994.