A New Speech Enhancement Method based on wavelet thresholding the multitaper spectrum

Yi Hu and Philipos C. Loizou

Abstract— **It is known that the "musical noise" encountered in most frequency domain algorithms is partially due to the large variance estimates of the spectra. To address this issue, we propose in this paper the use of low-variance spectral estimators based on wavelet thresholding the multitaper spectra. We extend the signal subspace approach in the frequency domain and incorporate the wavelet thresholded multitaper spectra into the proposed short-time spectral amplitude estimator. Listening tests showed that the use of multitaper spectrum estimation combined with wavelet thresholding suppressed the musical noise and yielded better quality than the recently proposed subspace approach.**

I. Introduction

Spectral subtraction is probably one of the most popular frequency domain speech enhancement methods used today. Unfortunately it suffers from an annoying artifact known as "musical noise", which is caused by randomly spaced spectral peaks that come and go in each frame. The randomly spaced peaks are partially due to the large-variance estimates of the spectra of the noise and noisy signals, typically computed using the periodogram method.

Several methods have been proposed to reduce the musical noise. Among them, the minimum mean square error (MMSE) short time spectral amplitude (STSA) estimator proposed by Eph-raim and Malah in [1]. Cappé [2] provided an explanation of the mechanism by which the MMSE method eliminates the musical noise. The MMSE estimator applies a spectral gain which is a function of two parameters: the *a priori SNR,* $\gamma_{prio}(\omega_k)$ and the *a posteriori SNR*. Cappé concluded that in the low SNR areas where musical noise frequently dominates, the estimate of the *a priori SNR* proposed in [1] corresponds to a highly smoothed version of the *a posteriori SNR* over successive short-time frames. As a consequence, the variance of $\gamma_{prio}(\omega_k)$ is much smaller. The fundamental mechanism used in [1] for suppressing the musical noise was therefore the smoothness of $\gamma_{prio}(\omega_k)$. Similar conclusions were also reached by Vary [3] who examined the theoretical limits of spectral-magnitude estimation .

Clearly an accurate estimate of the *a priori SNR* is critical for eliminating the musical noise. Rather than smoothing the *a priori SNR* estimate as done in [1], we took a different approach. The variance of the *a priori SNR* estimate is greatly influenced by the variance of the spectral estimate of the noisy speech signal. Hence, we focused in this paper on finding spectrum estimators with low variance. In particular, we considered using the multitaper method proposed by Thomson [4] for power spectrum estimation, which was shown to have good bias and variance properties. To further refine the spectrum estimate, we wavelet thresholded the log multitaper spectra. Unlike others who applied wavelet denoising to the time-domain signal, we applied wavelet denoising to the speech spectrum. It should be pointed out that we do not use wavelet denoising techniques to remove the noise, but rather to get better (lower variance) spectral estimates. Section II provides background information on low-variance spectrum estimators, and section III presents the proposed approach which utilizes these spectral estimators. The implementation details are presented in section IV, the experimental results are given in section V, and the conclusions are given in section VI.

The authors are with Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688. * Email: loizou@utdallas.edu Fax: (972) 883-2710

Research supported in part by Grant No. R01 DC03421 from NIDCD/NIH.

II. Low-variance spectrum estimators: Background

Direct spectrum estimation based on Hamming windowing is the most often used power spectrum estimation method for speech enhancement. Although windowing reduces the bias, it does not reduce the variance of the spectral estimate. The idea behind the multitaper spectrum estimator [4] is to reduce this variance by computing a small number L of direct spectrum estimators each with a different taper (window), and then average the L spectral estimates. The multitaper spectrum estimator is given by

$$\hat{S}^{mt}(\omega) = \frac{1}{L}\sum_{k=0}^{L-1}\hat{S}_k^{mt}(\omega) \tag{1}$$

with $\hat{S}_k^{mt}(\omega) = \left|\sum_{m=0}^{N-1} h_k(m)x(m)e^{-j\omega m}\right|^2$, where N is the taper length and $h_k(m)$ is the k-th data taper used for the spectral estimate $\hat{S}_k^{mt}(\cdot)$, also called the k-th eigenspectrum. The tapers $h_k(m)$ are chosen to be orthonormal. A good set of L orthonormal data tapers with good leakage properties are given by the Slepian sequences which are a function of a prescribed mainlobe width [4]. Another orthogonal family of tapers are the sine tapers given by [5]: $h_k(m) = \sqrt{\frac{2}{N+1}}\sin\frac{\pi k m}{N+1}$ $(m = 1, \cdots, N)$. The sine tapers were shown in [5] to produce smaller local bias than the Slepian tapers, with roughly the same spectral concentration. For that reason, we adopted the sine tapers in this paper.

Recent work has shown that wavelet thresholding techniques can be used to further refine the spectral estimate and produce a smooth estimate of the logarithm of the spectrum. Improved periodogram estimates were proposed in [6] and improved multitaper spectrum estimates were proposed in [7]. The underlying idea behind these techniques is to represent the estimated log spectra as a signal plus noise, where the signal is the true log spectrum and noise is the estimation error. If the noise is Gaussian, then standard wavelet denoising techniques can be used to eliminate the noise and obtain better spectral estimates.

It was shown in [7] that if the eigenspectra $\hat{S}_k^{mt}(\omega)$ are uncorrelated, the ratio of the estimated multitaper spectrum $\hat{S}^{mt}(\omega)$ and the true power spectrum $S(\omega)$ conforms to a chi-square distribution with $2L$ degrees of freedom, i.e.,

$$v(\omega) = \frac{\hat{S}^{mt}(\omega)}{S(\omega)} \sim \frac{\chi^2_{2L}}{2L}, \quad 0 < \omega < \pi \tag{2}$$

If L is at least 5, it can be shown that [7] for all ω (except near $\omega = 0$ and π) the random variable $\eta(\omega)$

$$\eta(\omega) = \log v(\omega) - \phi(L) + \log L \tag{3}$$

is approximately Gaussian with zero mean and known variance $\sigma^2_\eta = \phi'(L)$, where $\phi(L)$ and $\phi'(L)$ denote the digamma and trigamma functions [8] respectively. If $Z(\omega)$ is defined as $Z(\omega) = \log \hat{S}^{mt}(\omega) - \phi(L) + \log L$, then

$$Z(\omega) = \log S(\omega) + \eta(\omega) \tag{4}$$

So, the log multitaper power spectrum plus a constant $(\log L - \phi(L))$ can be written as the true log power spectrum plus a nearly Gaussian noise $\eta(\omega)$ with zero mean and known variance σ^2_η [7]. The model in (4) is well suited for wavelet denoising techniques [9] to eliminate the noise $\eta(\omega)$ and obtain a better estimate of the log spectrum.

Figure 1 shows example plots of power spectra estimated using the conventional direct spectrum estimator with Hamming window, the multitaper method (with sine tapers) and the SURE wave-let thresholding method. Clearly, the wavelet thresholding method produced an estimate of the spectrum that was closer to the true spectrum. Also, the resulting spectrum had less variance than the multitaper spectrum.

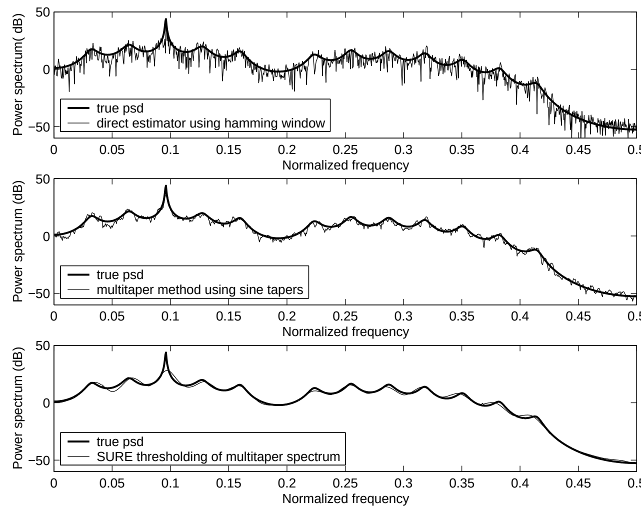

Fig. 1. Comparison of the power spectrum of an AR(24) process [7] estimated by the direct spectrum method using a Hamming window (top panel), the multitaper method using sine tapers (middle panel, with N=2048, L=5) and the SURE wavelet thresholding method (bottom panel, with q_0=5 and 16-tap symlets wavelets).

III. Speech enhancement by wavelet thresholding the multitaper spectrum

In this section we derive the proposed STSA estimator which uses the above mentioned low-variance spectral estimator. The STSA estimator is based on extending the subspace approach proposed in [10] in the frequency domain.

A. Proposed STSA estimator

We assume that the noise signal is additive and uncorrelated with the speech signal, i.e., $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where $\mathbf{y}$, $\mathbf{x}$ and $\mathbf{n}$ are the N-dimensional noisy speech, clean speech and noise vectors respectively. By denoting the N-point discrete Fourier Transform matrix by F, the Fourier transform of the noisy speech vector $\mathbf{y}$ can be written as $\mathbf{Y}(\omega) = F^H \cdot \mathbf{y} = F^H \cdot \mathbf{x} + F^H \cdot \mathbf{n} = \mathbf{X}(\omega) + \mathbf{N}(\omega)$, where $\mathbf{X}(\omega)$ and $\mathbf{N}(\omega)$ are the $N \times 1$ vectors containing the spectral components of the clean speech vector $\mathbf{x}$ and the noise vector $\mathbf{n}$, respectively.

Let $\hat{\mathbf{X}}(\omega) = G \cdot \mathbf{Y}(\omega)$ be the linear estimator of $\mathbf{X}(\omega)$, where G is a $N \times N$ matrix. The error signal obtained in this estimation is given by $\boldsymbol{\varepsilon}(\omega) = \hat{\mathbf{X}}(\omega) - \mathbf{X}(\omega) = \boldsymbol{\varepsilon}_{\mathbf{x}}(\omega) + \boldsymbol{\varepsilon}_{\mathbf{n}}(\omega)$, where $\boldsymbol{\varepsilon}_{\mathbf{x}}(\omega) = (G - I) \cdot \mathbf{X}(\omega)$ represents the speech distortion in the frequency domain and $\boldsymbol{\varepsilon}_{\mathbf{n}}(\omega) = G \cdot \mathbf{N}(\omega)$ represents the residual noise in the frequency domain. After defining the energy of the frequency domain speech distortion as $\varepsilon^2_{\mathbf{x}}(\omega) = E\left(\boldsymbol{\varepsilon}^H_{\mathbf{x}}(\omega) \cdot \boldsymbol{\varepsilon}_{\mathbf{x}}(\omega)\right)$ and the energy of the frequency domain residual noise as $\varepsilon^2_{\mathbf{n}}(\omega) = E\left(\boldsymbol{\varepsilon}^H_{\mathbf{n}}(\omega) \cdot \boldsymbol{\varepsilon}_{\mathbf{n}}(\omega)\right)$, we can obtain the optimal linear estimator by solving the following constrained optimization problem:

$$\min_G \varepsilon^{\mathbf{2}}_{\mathbf{x}}(\omega)$$
$$\textit{subject to}: \frac{1}{N}\varepsilon^{\mathbf{2}}_{\mathbf{n}}(\omega) \leq \mathbf{c} \tag{5}$$

where c is a positive number. It can be shown, using the method of Lagrangian multiplier, that the solution to the above constrained minimization problem is given by:

$$G(F^H \cdot R_{\mathbf{x}} \cdot F + \mu \cdot F^H \cdot R_{\mathbf{n}} \cdot F) = F^H \cdot R_{\mathbf{x}} \cdot F \tag{6}$$

where μ is the Lagrangian multiplier. The above equation can be simplified if we assume that each frequency component is modified individually by a gain, that is, if we assume that G is a diagonal matrix. The matrices $F^H \cdot R_{\mathbf{x}} \cdot F$ and $F^H \cdot R_{\mathbf{n}} \cdot F$ are asymptotically diagonal (assuming that $R_{\mathbf{x}}$ and $R_{\mathbf{n}}$ are Toeplitz) and the diagonal elements of $F^H \cdot R_{\mathbf{x}} \cdot F$ and $F^H \cdot R_{\mathbf{n}} \cdot F$ are the power spectrum components $S_{\mathbf{x}}(\omega)$ and $S_{\mathbf{n}}(\omega)$ of the clean speech vector $\mathbf{x}$ and noise vector $\mathbf{n}$, respectively. Denoting the kth diagonal element of G by $g(k)$, (6) can be rewritten as: $g(k) \cdot (S_{\mathbf{x}}(k) + \mu \cdot S_{\mathbf{n}}(k)) = S_{\mathbf{x}}(k)$. The gain function $g(k)$ for the kth frequency component is therefore given by:

$$g(k) = \frac{S_{\mathbf{x}}(k)}{S_{\mathbf{x}}(k) + \mu \cdot S_{\mathbf{n}}(k)} = \frac{\gamma_{prio}(k)}{\gamma_{prio}(k) + \mu} \tag{7}$$

where $\gamma_{prio}(k) = \frac{S_{\mathbf{x}}(k)}{S_{\mathbf{n}}(k)}$ is the *a priori SNR* at frequency ω_k. Note that the above equation is the generalized Wiener filter equation.

The power spectrum $S_{\mathbf{x}}(\omega)$ in (7) of the clean speech signal is not available, but in practice can be estimated as: $\widehat{S}_{\mathbf{x}}(\omega) = S_{\mathbf{y}}(\omega) - \widehat{S}_{\mathbf{n}}(\omega)$, where $\widehat{S}_{\mathbf{n}}(\omega)$ denotes the estimate of the noise spectrum obtained during speech-absent frames. As discussed in the Introduction, the estimate of the $\gamma_{prio}(k)$ is crucial for eliminating musical noise. In this paper, we considered two different methods for obtaining a good estimate of $\gamma_{prio}(k)$.

In the first method, we form the ratio of the multitaper spectra $\widehat{S}^{mt}_{\mathbf{x}}(\omega)/\widehat{S}^{mt}_{\mathbf{n}}(\omega)$, and wavelet threshold the log of the ratio of the two spectra to get an estimate of $\gamma_{prio}(k)$. It can be proven (see Appendix) that the log *a priori SNR* estimate, based on multitaper spectra [denoted as $\gamma^{mt}_{prio}(k)$], can be modeled as the true log *a priori SNR* plus a Gaussian distributed noise $\xi(k)$, i.e.,

$$\log \gamma^{mt}_{prio}(k) = \log \gamma_{prio}(k) + \xi(k) \tag{8}$$

where $\xi(k)$ is approximately Gaussian distributed with zero mean and known variance $2\sigma^2_{\eta}$. Because of the nature of $\xi(k)$, wavelet denoising techniques can be used to eliminate $\xi(k)$.

The second method is based on the assumption that a good estimate of the *a priori SNR*, can be obtained using a good low-variance spectral estimate of $\widehat{S}_{\mathbf{x}}(\omega)$ and $\widehat{S}_{\mathbf{n}}(\omega)$. We considered first obtaining the multitaper spectral estimates of $S_{\mathbf{y}}(\omega)$ and $\widehat{S}_{\mathbf{n}}(\omega)$ and then wavelet thresholding the log of those estimates individually to obtain $\widehat{S}_{\mathbf{x}}(\omega)$. The refined spectrum of $\widehat{S}_{\mathbf{x}}(\omega)$, along with the wavelet thresholded estimate of $\widehat{S}_{\mathbf{n}}(\omega)$ are used to obtain a better estimate of the *a priori SNR.*

B. Wavelet thresholding techniques

Critical to the performance of wavelet denoising techniques is the choice of threshold levels. Several methods have been proposed in the literature for thresholding the wavelet coefficients. In this paper we focus on evaluating two popular thresholding schemes, namely the universal thresholding method and the SURE method [9]. One of the main differences between the two thresholding schemes is that the universal thresholding technique is not dependent on the data, while the SURE technique is. In this paper, the SURE method was implemented as per [9]. Next, we give a brief description of the universal wavelet thresholding method used in this paper.

As pointed out in [9], if the noise term $\eta(\omega)$ is stationary and colored (as it is in our case), the variance of the noise wavelet coefficients, $n_{j,k}$ (j denotes the scale or level of wavelet decomposition and k denotes the k-th coefficient), will be different for each scale in the wavelet decomposition. Consequently, scale-dependent thresholding can be used to account for the different variances of the noise wavelet coefficients in each scale. Walden *et al.* [7] extended that idea and derived the variances of $n_{j,k}$ of the nearly Gaussian noise $\eta(\omega)$ in (4). The level-dependent variances of the noise wavelet coefficients, $n_{j,k}$, were estimated according to [7]:

$$\mathrm{var}(n_{j,k}) = \sigma^2_j \equiv \frac{1}{N} \sum_{k=0}^{N-1} S(k)|H_j(k)|^2 \tag{9}$$

where $H_j(k)$ is the frequency response of the length N periodized wavelet filter of level j, and $S(k)$ is the Fourier transform of the autocorrelation function $r_{\eta\eta}$ of the noise $\eta(\omega)$ in (4) which is given by $r_{\eta\eta}(i) = \sigma^2_{\eta}(1 - |i|/(L+1))$ for $|i| \leq L+1$. The universal threshold T_j for scale j is selected as $T_j = \sigma_j\sqrt{2\log N}$.

IV. Implementation details

The proposed method can be implemented in four steps. For each speech frame:

Step 1: Compute the multitaper power spectrum $S^{mt}_{\mathbf{y}}$ of the noisy speech $\mathbf{y}$ using (1), and estimate the multitaper power spectrum $S^{mt}_{\mathbf{x}}$ of the clean speech signal by: $S^{mt}_{\mathbf{x}}(\omega) = S^{mt}_{\mathbf{y}}(\omega) - S^{mt}_{\mathbf{n}}(\omega)$, where $S^{mt}_{\mathbf{n}}(\omega)$ is the multitaper power spectrum of the noise. $S^{mt}_{\mathbf{n}}(\omega)$ can be obtained using noise samples collected during speech-absent frames. Here L is set to 5. Any negative elements of $S^{mt}_{\mathbf{x}}(\omega)$ are floored to $0.002 \cdot S^{mt}_{\mathbf{n}}(\omega)$.

Step 2: Estimate the *a priori SNR* using one of the two methods described in section III-A. For the second method, for instance, first compute $Z(\omega) = \log S^{mt}_{\mathbf{y}}(\omega) - \phi(L) + \log L$ and then apply the Discrete Wavelet Transform (DWT) of $Z(\omega)$ out to level q_0 to obtain the empirical DWT coefficients $z_{j,k}$ for each level j, where q_0 is set to 5. Sixteen-tap Daubechie's least asymmetry wavelets were used in this paper. Threshold the wavelet coefficients $z_{j,k}$ using one of the two thresholding techniques described in Section III-B, and apply the inverse DWT to the thresholded wavelet coefficients to obtain the refined log spectrum, $\log S^{wmt}_{\mathbf{y}}(\omega)$, of the noisy signal. Repeat above procedure to obtain the refined log spectrum, $\log S^{wmt}_{\mathbf{n}}(\omega)$, of the noise signal. The estimated power spectrum $S^{wmt}_{\mathbf{x}}$ of the clean speech signal can be estimated by: $S^{wmt}_{\mathbf{x}}(\omega) = S^{wmt}_{\mathbf{y}}(\omega) - S^{wmt}_{\mathbf{n}}(\omega)$. The *a priori SNR* $\gamma_{prio}(k)$ for frequency ω_k can be estimated as $S^{wmt}_{\mathbf{x}}(\omega_k)/S^{wmt}_{\mathbf{n}}(\omega_k)$.

Step 3: Compute the μ value in (7) according to the segmental SNR:

$$\mu = \begin{cases} \mu_0 - (SNR_{dB})/s & -5 < SNR_{dB} < 20 \\ 1 & SNR_{dB} \geq 20 \\ \mu_{\max} & SNR_{dB} \leq -5 \end{cases} \tag{10}$$

where $\mu_{\max}$ is the maximum allowable value of μ, which we set to 10, $\mu_0 = (1 + 4\mu_{\max})/5$, $s = 25/(\mu_{\max} - 1)$, $SNR_{dB} = 10\log_{10} SNR$ and SNR is computed as: $SNR = \frac{\sum_{i=0}^{N-1} S^{wmt}_{\mathbf{x}}(i)}{\sum_{i=0}^{N-1} S^{wmt}_{\mathbf{n}}(i)}$.

Step 4: Estimate the gain function $g(k)$ for component ω_k using (7). Obtain the enhanced frequency spectrum $\hat{X}(\omega_k)$ by $\hat{X}(\omega_k) = g(k) \cdot Y(\omega_k)$. Apply the inverse FFT of $\hat{X}(\omega)$ to obtain the enhanced speech signal.

The estimator was applied to 32-ms duration frames of the noisy signal with 50% overlap between frames. The enhanced speech signal was combined using the overlap and add approach.

V. Experimental results

Subjective listening tests were conducted to evaluate the quality of the proposed STSA estimator. Ten sentences from the HINT database [11] produced by a male speaker and ten sentences from the TIMIT database produced by female speakers were used. Spee-ch shaped noise and Volvo car interior noise was added to the clean speech files at 5 dB and 0 dB SNR respectively. Each speech file was enhanced by four methods: the proposed estimator in (7) using the multitaper spectrum (MT), the proposed estimator in (7) using wavelet-thresholded multitaper spectra with SURE thresholding (MT_SURE), the proposed estimator in (7) using wavelet-thresholded multitaper spectra with universal level-dependent soft thresholding (MT_UNIV), and an improved version of the signal subspace approach (SigSub) proposed in [12] which utilizes the multiwindow covariance matrix estimator as in [13]. The second method, discussed in Section 3.1, was used to obtain $\gamma_{prio}(k)$ in the MT_SURE and MT_UNIV methods. No significant differences in quality were noted between the two methods proposed for the estimation of $\gamma_{prio}(k)$.

Ten native speakers of English participated in the subjective test and were asked to compare the speech quality of pairs of sentences processed with the above methods. Table I summarizes the subjective evaluation tests for the 20 sentences in terms of preference percentage. The numbers in Table I indicate the percentage of time that the listeners preferred the speech quality of the MT_SURE method over the other methods.

Noise type	Comparison with MT	Comparison with MT_UNIV	Comparison with SigSub
Speech shaped	75%	46%	91%
Car noise	53%	56%	98%

TABLE I

Results of listening tests in terms of percentage of time that listeners preferred the MT_SURE method over the other methods.

From Table I we can see that listeners preferred the quality of the MT_SURE method to the quality of the MT method when speech was corrupted with speech-shaped noise. Speech enhanced with the MT method had some musical noise. In contrast, speech enhanced with the MT_SURE or MT_UNIV methods had *no* musical noise. The wavelet thresholding techniques applied to the multitaper spectra eliminated the musical noise. The quality of the MT_SURE method was found to be superior to the quality of the signal subspace method for both noise types. As indicated by the results in Table I, there was a small but subtle difference in quality between the two wavelet thresholding techniques. The SURE technique had a small advantage in the car noise environment.

VI. Summary and conclusions

A new speech enhancement method was proposed in this paper which, unlike most spectral domain methods, uses low-variance spectrum estimators. The low-variance spectrum estimators were based on wavelet-thresholding the multitaper spectrum. Listening tests revealed that the enhanced speech had no musical noise and we attribute that to the use of low-variance spectrum estimators. Furthermore, listening tests showed that the proposed method had a superior speech quality to signal subspace methods.

APPENDIX

In this appendix we show how the *a priori SNR* estimate obtained using the log multitaper spectra can be modeled as the true log *a priori SNR* plus a Gaussian distributed noise. Denoting the *a priori SNR* estimated by the multitaper spectra as γ_{prio}^{mt}, it is clear that

$$\frac{\gamma_{prio}^{mt}(k)}{\gamma_{prio}(k)} = \left(\frac{S_{\mathbf{x}}^{mt}(k)}{S_{\mathbf{x}}(k)}\right) / \left(\frac{S_{\mathbf{n}}^{mt}(k)}{S_{\mathbf{n}}(k)}\right) \tag{11}$$

Using the relationship in (2) and taking the log of both sides, we get:

$$\log \gamma_{prio}^{mt}(k) = \log \gamma_{prio}(k) + \log\chi_x^2(k) - \log\chi_n^2(k) \tag{12}$$

where $\chi_x^2(k)$ and $\chi_n^2(k)$ denote the chi-square distributions (with $2L$ degrees of freedom) of $S_{\mathbf{x}}^{mt}(k)/S_{\mathbf{x}}(k)$ and $S_{\mathbf{n}}^{mt}(k)/S_{\mathbf{n}}(k)$ respectively. If $L \geq 5$, the log chi-square distributions are nearly Gaussian with mean $\phi(L) - \log L$ and variance $\phi'(L)$ [8]. Hence, the above equation can be simplified to:

$$\log \gamma_{prio}^{mt}(k) = \log \gamma_{prio}(k) + \xi(k) \tag{13}$$

where $\xi(k)$ is nearly Gaussian with zero mean and variance $2\phi'(L)$.

References

[1] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-32, pp. 1109–1121, 1984.

[2] O. Cappe, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. Speech Audio Processing*, vol. 2, pp. 345–349, 1994.

[3] P. Vary, "Noise suppression by spectral magnitude estimation -mechanism and theoretical limits," *Signal Processing*, vol. 8, pp. 387–400, 1985.

[4] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, Sept. 1982.

[5] K. S. Riedel and A. Sidorenko, "Minimum bias multiple taper spectral estimation," *IEEE Trans. Signal Processing*, vol. 43, pp. 188–195, 1995.

[6] P. Moulin, "Wavelet thresholding techniques for power spectrum estimation," *IEEE Trans. Signal Processing*, vol. 42, pp. 3126–3136, Dec. 1994.

[7] A. T. Walden, D. B. Percival, and E. J. McCoy, "Spectrum estimation by wavelet thresholding of multitaper estimators," *IEEE Trans. Signal Processing*, vol. 46, pp. 3153–3165, 1998.

[8] M. S. Bartlett and D. G. Kendall, "The statistical analysis of variance heterogeneity and the logarithmic transformation," *Suppl. J. R. Statist. Soc.*, vol. 8, pp. 128–138, 1946.

[9] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *J. R. Statist. Soc. B*, vol. 59, pp. 319–351, 1997.

[10] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 251–266, 1995.

[11] M. Nilsson, S. Soli, and J. Sullivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.*, vol. 95, pp. 1085–1099, 1994.

[12] Y. Hu and P. C. Loizou, "A subspace approach for enhancing speech corrupted by colored noise," *IEEE Signal Processing Letters*, pp. 204–206, July 2002.

[13] Y. Hu and P. C. Loizou, "A perceptually motivated subspace approach for speech enhancement," in *International Conference on Spoken Language Processing*, 2002, pp. 1797–1800, Denver.