INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

A Novel Risk-Estimation-Theoretic Framework for Speech Enhancement in Nonstationary and Non-Gaussian Noise Conditions

Jishnu Sadasivan[1], *Chandra Sekhar Seelamantula*[2]

[1]Department of Electrical Communication Engineering, [2]Department of Electrical Engineering
Indian Institute of Science, Bangalore 560012, India

jishnus@ece.iisc.ernet.in, chandra.sekhar@ieee.org

Abstract

We address the problem of suppressing background noise from noisy speech within a risk estimation framework, where the clean signal is estimated from the noisy observations by minimizing an unbiased estimate of a chosen risk function. For Gaussian noise, such a risk estimate was derived by Stein, which eventually went on to be called Stein's unbiased risk estimate (SURE). Stein's formalism is restricted to Gaussian noise and exclusive risk estimators have been developed for each noise type. On the other hand, we consider linear denoising functions and derive an unbiased risk estimate without making any assumption about the noise distribution. The proposed unbiased estimate depends only on the second-order statistics of noise and makes the proposed framework applicable to many practical denoising problems where the noise distribution is not known a priori, but one has access only to the samples of noise. We demonstrate the usefulness of the proposed methodology for speech enhancement using subband shrinkage, where the shrinkage parameters are obtained by minimizing the newly developed risk estimator. The proposed methodology is also applicable to nonstationary noise conditions. We show that the proposed denoising algorithm outperforms the state-of-the art algorithms in terms of standard speech-quality evaluation metrics.

Index Terms: Speech denoising, Shrinkage estimator, Stein's unbiased risk estimate (SURE), Mean-squared error

1. Introduction

Noise reduction techniques find applications in mobile communication systems, hearing aids, speech compression, automatic speech recognition, multimedia systems, etc. Speech enhancement techniques can be broadly classified as: (i) Spectral subtraction algorithms, (ii) Wiener filtering techniques, (iii) Subspace methods, (iv) Statistical model-based methods, and those that lie at the intersection of one or more of these classes. In spectral subtraction algorithms, an estimate of the noise power spectrum is subtracted from the noisy signal spectrum, to obtain an estimate of the clean signal spectrum [1]. The key assumption is that the noise is additive and stationary. Improved versions of spectral subtraction algorithms are available in [2] and [3]. In Wiener filter techniques, using an estimate of the clean speech power spectrum and the noise power spectrum obtained from the noisy speech, the Wiener filter is constructed and employed for denoising speech. Hu and Loizou [4, 5] incorporated psychoacoustic constraints in the Wiener filtering paradigm. Chen et al. [6] studied the interdependency between speech distortion and noise reduction after Wiener filtering and also quantified them. In the Wiener filtering framework, one requires an estimate of the power spectrum of the clean signal and noise. Typical power spectrum estimation strategies are presented in [7–9].

The basic principle in the subspace approach is that the noisy subspace is divided into two orthogonal subspaces, signal-plus-noise and noise-only subspace. Enhancement is performed by projecting the noisy signal on to the signal-plus-noise subspace. Signal subspace decompositions can be obtained using eigenvalue decompositions of the data covariance matrix [10, 11] or singular value decomposition of the data matrix [12, 13]. Statistical model-based methods minimize a risk function (that measures the closeness between the clean speech parameter and its estimate) based on the assumed statistical model of clean speech and noise parameters to obtain the estimate of clean speech parameters from the noisy speech. Depending on the cost function used and the statistical assumption on clean signal and noise parameters, different variants of estimators have been proposed in [14–21]. Loizou's book on speech enhancement [22] is a recent and authoritative reference on the topic. Recently non-negative matrix factorization (NMF) based methods have been successfully used for speech enhancement [23]. In this paper, we propose a novel single channel speech enhancement method based on the new risk estimation approach.

Denoising is essentially an estimation problem where a risk function is minimized to obtain an estimate of the clean signal parameters. Direct minimization of the risk function results in an estimator that is a function of the parameter to be estimated or its statistics, which is often difficult to get in practice. An alternative approach is the risk estimation framework, in which, instead of minimizing the original risk, an unbiased estimate of risk that is a function of the observations is minimized to obtain the unknown clean signal parameter. Assuming the observations to be independent and identically Gaussian distributed, Stein proposed an unbiased estimate of the mean-squared error (MSE) [24], which is now called Stein's unbiased risk estimate (SURE). It is widely used for signal denoising applications [25–33]. The original formulation of SURE based on the independent and identically distributed (i.i.d.) Gaussian assumption is further extended to non-i.i.d. multivariate exponential family of distribution in [33]. SURE specifically depends on the observation distribution, which limits its applicability to the practical problem where the observation distribution is not known a priori. In this paper, we derive an unbiased estimate of the MSE assuming the denoising function to be linear, without making any assumption on the observation distribution. The solution is seen to be dependent on the second-order statistics. Hence, the proposed unbiased estimate of the MSE, which depends only on the second-order statistics, gives the flexibility of

http://dx.doi.org/10.21437/Interspeech.2016-151

using risk estimation framework in conditions where the noise distribution is not known. We also demonstrate the usefulness of the proposed risk estimate in speech denoising applications. A subband-shrinkage denoiser is developed where the shrinkage parameter is obtained by minimizing the proposed unbiased estimate of MSE. We also show performance comparisons of the proposed denoising algorithm with three state-of-the art algorithms in three different noise conditions.

2. An Unbiased Estimate of the MSE

Consider the additive noise model $\mathbf{x} = \theta + \mathbf{w}$, where $\theta \in \mathbf{R}^n$ (non-random) and the noise vector $\mathbf{w}$ is assumed to be independent and identically distributed (i.i.d.) entries with zero mean and variance σ^2. Our goal is to obtain an estimate $\hat{\theta}(\mathbf{x})$, of the parameter $\theta \in \mathbf{R}^n$ starting from the observation $\mathbf{x}$ that minimizes the MSE:

$$\begin{aligned}\mathcal{R} &= \mathcal{E}\left\{\|\theta - \hat{\theta}\|^2\right\}, \\ &= \mathcal{E}\left\{\theta^\mathsf{T}\theta\right\} - 2\mathcal{E}\left\{\theta^\mathsf{T}\hat{\theta}\right\} + \mathcal{E}\left\{\hat{\theta}^\mathsf{T}\hat{\theta}\right\}. \qquad (1)\end{aligned}$$

In the above cost function, the first term does not affect the minimization and so it need not be considered, but the second term is a function of θ, which is the parameter to be estimated. This makes the minimization infeasible. An alternative is produced by the risk-estimation framework. In this approach, instead of minimizing $\mathcal{R}$, one minimizes an unbiased estimate $\hat{\mathcal{R}}$, of $\mathcal{R}$, to obtain $\hat{\theta}$. To find an unbiased estimate of $\mathcal{R}$, one computes an unbiased estimate of $\mathcal{E}\left\{\theta^\mathsf{T}\hat{\theta}\right\}$. Consider the estimator, $\hat{\theta}(\mathbf{x}) = \mathbf{H}\mathbf{x}$, where $\mathbf{H} \in \mathbf{R}^{n\times n}$, results $\mathcal{E}\left\{\theta^\mathsf{T}\hat{\theta}\right\} = \mathcal{E}\left\{\theta^\mathsf{T}\mathbf{H}\mathbf{x}\right\}$. We have that

$$\begin{aligned}\mathcal{E}\left\{\theta^\mathsf{T}\mathbf{H}\mathbf{x}\right\} &= \mathcal{E}\left\{\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x}\right\} - \mathcal{E}\left\{\mathbf{w}^\mathsf{T}\mathbf{H}\mathbf{x}\right\}, \\ &= \mathcal{E}\left\{\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x}\right\} - \mathcal{E}\left\{\mathbf{w}^\mathsf{T}\mathbf{H}\mathbf{w}\right\}, \\ &= \mathcal{E}\left\{\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x}\right\} - \sum_{l=1}^{n}\sum_{k=1}^{n} h_{l,k}\mathcal{E}\left\{w_l w_k\right\}, \\ &= \mathcal{E}\left\{\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x}\right\} - \sum_{l=1}^{n} h_{l,l}\sigma^2, \qquad (2)\end{aligned}$$

where $h_{l,l}$ is the $(l,l)^{\text{th}}$ element of $\mathbf{H}$. Substituting (2) in (1) and using the fact that $\mathcal{E}\left\{\theta^\mathsf{T}\theta\right\} = \mathcal{E}\left\{\mathbf{x}^\mathsf{T}\mathbf{x}\right\} - n\sigma^2$, we obtain $\mathcal{R} = \mathcal{E}\left\{\mathbf{x}^\mathsf{T}\mathbf{x}\right\} - n\sigma^2 - 2\mathcal{E}\left\{\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x}\right\} + 2\sum_{l=1}^{n} h_{l,l}\sigma^2 + \mathcal{E}\left\{\mathbf{x}^\mathsf{T}\mathbf{H}^\mathsf{T}\mathbf{H}\mathbf{x}\right\}$. An unbiased estimate of $\mathcal{R}$ is $\hat{\mathcal{R}} = \mathbf{x}^\mathsf{T}\mathbf{x} - n\sigma^2 - 2\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + 2\sum_{l=1}^{n} h_{l,l}\sigma^2 + \mathbf{x}^\mathsf{T}\mathbf{H}^\mathsf{T}\mathbf{H}\mathbf{x}$. Instead of minimizing $\mathcal{R}$, we minimize $\hat{\mathcal{R}}$ to obtain optimal $\mathbf{H}$. During the derivation of $\hat{\mathcal{R}}$, no specific distribution assumption on the observation has been made. Hence, the proposed unbiased estimate of MSE is useful in scenarios where the observation distribution is unknown, and only the knowledge of first- and second-order statistics is available. In contrast, the SURE formulation requires the observation distribution to be Gaussian. For other noise distributions, specific risk estimators have to be developed.

3. Speech Denoising

We develop a subband-shrinkage estimator for speech denoising based on the proposed unbiased estimate of MSE. Subband-shrinkage estimator is defined as $\hat{\theta} = \sum_{j=1}^{N}\alpha_j\mathbf{S}_j\mathbf{A}_j\mathbf{x}$, where $\mathbf{A}_j$ and $\mathbf{S}_j$ correspond to the j^{th} analysis and synthesis filters, respectively, N is the number of subbands, and α_j corresponds to the shrinkage parameter in the j^{th} subband. The subband-shrinkage estimator is given as

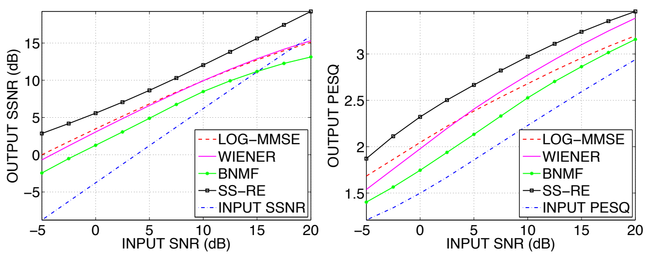
$$\begin{aligned}\hat{\theta} = \sum_{j=1}^{N}\alpha_j\mathbf{S}_j\mathbf{A}_j\mathbf{x} &= \sum_{j=1}^{N}\alpha_j\mathbf{H}^{(j)}\mathbf{x} = \sum_{j=1}^{N}\alpha_j\phi^{(j)} \\ &= \mathbf{H}\mathbf{x} = \mathbf{\Phi}\alpha,\end{aligned}$$

where $\mathbf{H} = \sum_{j=1}^{N}\alpha_j\mathbf{H}^{(j)}$, $\mathbf{\Phi} = \left[\phi^{(1)}\phi^{(2)}\cdots\phi^{(N)}\right]$, and $\phi^{(j)} = \mathbf{H}^{(j)}\mathbf{x}$. The goal is to obtain the optimum shrinkage parameter that minimizes $\mathcal{R} = \mathcal{E}\left\{\|\theta - \mathbf{\Phi}\alpha\|^2\right\}$. Since direct minimization results in an unrealizable estimator, instead of minimizing $\mathcal{R}$, we minimize the unbiased estimate $\hat{\mathcal{R}}$ to obtain the optimum shrinkage parameter, i.e., $\alpha^* = \arg\min_{\alpha} \hat{\mathcal{R}}\left(\mathbf{\Phi}\alpha\right)$. For subband-shrinkage estimator, $\hat{\mathcal{R}} = \mathbf{x}^\mathsf{T}\mathbf{x} - n\sigma^2 - 2\mathbf{x}^\mathsf{T}\mathbf{\Phi}\alpha + 2\mathbf{q}^\mathsf{T}\alpha + \|\mathbf{\Phi}\alpha\|^2$ where q_j the j^{th} component of $\mathbf{q}$ equals, $\sum_{l=1}^{n} h_{l,l}^{(j)}\sigma^2$. To obtain the optimum α, setting $\frac{\partial\hat{\mathcal{R}}}{\partial\alpha} = 0$ results in $\alpha^* = \left(\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi}\right)^{-1}\left(\mathbf{\Phi}^\mathsf{T}\mathbf{x} - \mathbf{q}\right)$. To avoid the sign change of subband waveform by negative shrinkage value we use final shrinkage parameter as $\max\left(\alpha_j^*, 0\right)$.

Implementation details: We use clean speech from TIMIT database [34] and the noise samples from Noisex-92 database [35]. We consider sampling frequency of 16 kHz and 32 ms speech frames with 75% overlap for processing. For performing multiband decomposition and reconstruction of noisy speech signal, we use a 32-channel cosine-modulated perfect reconstruction filterbank [36]. We consider frame-by-frame processing of the speech signal, where the shrinkage parameters for subbands for a particular frame are obtained by minimizing $\hat{\mathcal{R}}$. A statistical model based voice activity detector (VAD) [37] is employed to update the noise variance in speech absent frames [22, pp. 543–544].

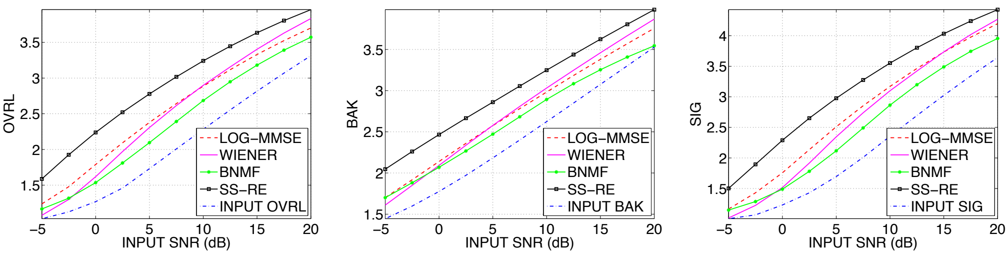
4. Results and Discussion

We compare the performance of the proposed subband-shrinkage denoiser with three benchmark algorithms: (i) Wiener filter algorithm with an a priori signal-to-noise ratio (SNR) estimation method (WIENER), proposed in [7]; (ii) Statistical model-based short-time spectral amplitude (STSA) estimator of speech signal that minimizes the mean-squared error between the log-magnitude of the original STSA and its



(a) (b)

Figure 1: (Color online) Denoising performance comparisons: (a) output SSNR and (b) output PESQ of different algorithms for different input SNRs, where noise considered is white Gaussian noise. The results have been averaged for 10 different speech files and 10 independent noise realizations.



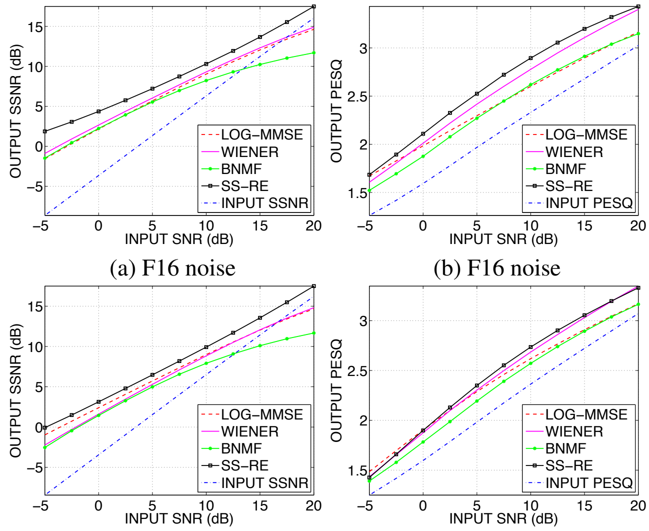
(a) (b) (c)

Figure 2: (Color online) Denoising performance comparisons in terms of composite scores. (a) OVRL, (b) BAK, and (c) OVRL, for different input SNRs. The results have been averaged over 10 different speech files and 10 independent noise realizations.

estimate (LOG-MMSE) [15]; and (iii) Bayesian non-negative matrix factorization method for speech enhancement (BNMF) [23]. We use the same noise variance estimation algorithm for the proposed subband-shrinkage denoiser (SS-RE), WIENER and LOG-MMSE. BNMF is a supervised technique that require training, whereas SS-RE does not require training data. For performance evaluation, we use the objective measures, segmental SNR (SSNR) [38], perceptual evaluation of speech quality (PESQ) [39], and composite objective scores [40]. We use three noise conditions for evaluation, white Gaussian noise and the two real-world noise types from Noisex-92 database (F16 noise and factory noise).

We first consider denoising of speech corrupted with white Gaussian noise. Figure 1(a) shows the output SSNR of the different denoising algorithms for different input SNRs. We observe that the proposed subband-shrinkage denoising algorithm based on risk estimation (SS-RE) shows better denoising in terms of the output SSNR compared with the other algorithms. SS-RE shows an SSNR improvement of $2.5 - 3$ dB compared with the competing techniques over the input SNR range -5 to 20 dB. Figure 1(b) shows the output PESQ of different algorithms for different input SNRs. It is observed that proposed SS-RE consistently dominates over all the other algorithms in terms of the output PESQ. Next, we compare the performance in terms of the composite objective scores (on a five-point scale). It includes a composite measure for signal distortion (SIG), a composite measure for noise distortion (BAK), and a composite measure for overall speech quality (OVRL). Composite scores are obtained by combining the basic objective scores [40]. We use the MATLAB implementation of the composite scores available with [22] for the evaluation. Figure 2 shows the performance comparison of different algorithms for different input SNRs in terms of the composite measures. It is observed that, for input SNR ranges from -5 to 20 dB, the proposed SS-RE consistently outperforms the other denoising algorithms in terms of OVRL, BAK, and SIG scores. The experimental results suggest that SS-RE yields a higher noise attenuation while maintaining low speech distortion, and maintains a high overall quality of the denoised speech compared with the other algorithms.

We next consider denoising performance evaluation of the algorithms when speech is corrupted with real-world noise types. For experiments, we consider F16 engine noise and factory noise from the Noisex-92 database. The F16 noise is relatively stationary whereas factory noise is non-stationary. Since the real-world noises are not white, we estimate noise variance in each subband, and obtain optimum shrinkage parameter separately for each subband by minimizing the corresponding unbiased estimate of the MSE. Figure 3 shows the denoising performance of different algorithms for different input SNRs in terms of SSNR and PESQ, where the noise types considered are F16 noise and factory noise. The results are averaged over 10 different speech files and 10 different noise realizations (noise taken at randomly selected locations from the long noise sequence added to the clean speech to create noisy speech). Figures 3(a) and 3(b) show the denoising performance of different algorithms in terms of SSNR and PESQ, where noise considered is F16 noise. It is observed that SS-RE consistently outperforms the three algorithms in terms of SSNR and PESQ for all input SNRs. Figures 3(c) and 3(d) shows the denoising performance of different algorithms in terms of SSNR and PESQ in factory noise which is highly nonstationary. We observe from the Figure 3(c) that for SS-RE, the output SSNR gain is high compared with the other denoising algorithms. However, the PESQ gain of SS-RE and WIENER are matching closely. The PESQ gain is relatively low compared with the F16 noise case. PESQ gain is high in F16 noise compared with factory noise because F16 noise is relatively stationary and hence noise variance estimate will be more accurate. Figure 4 shows a performance comparison in terms of the composite scores. It is observed that in the case of F16 noise, the SS-RE scores of OVRL, BAK and



(a) F16 noise (b) F16 noise
(c) Factory noise (d) Factory noise

Figure 3: (Color online) Denoising performance comparisons of different algorithms in terms of SSNR and PESQ. The results have been averaged over 10 different speech files and 10 different noise realisations.

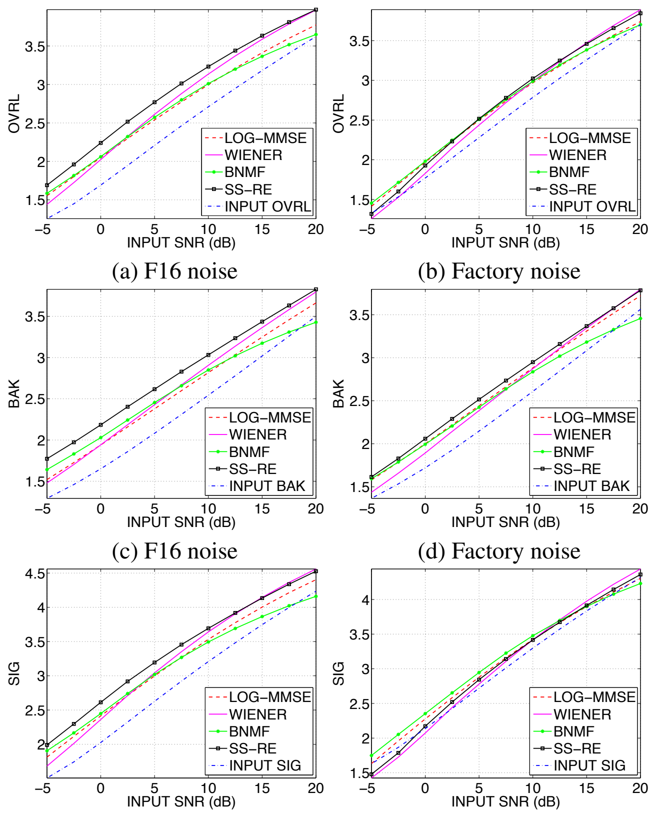


Figure 4: (Color online) Denoising performance comparison of different algorithms for different input SNRs in terms of composite scores.

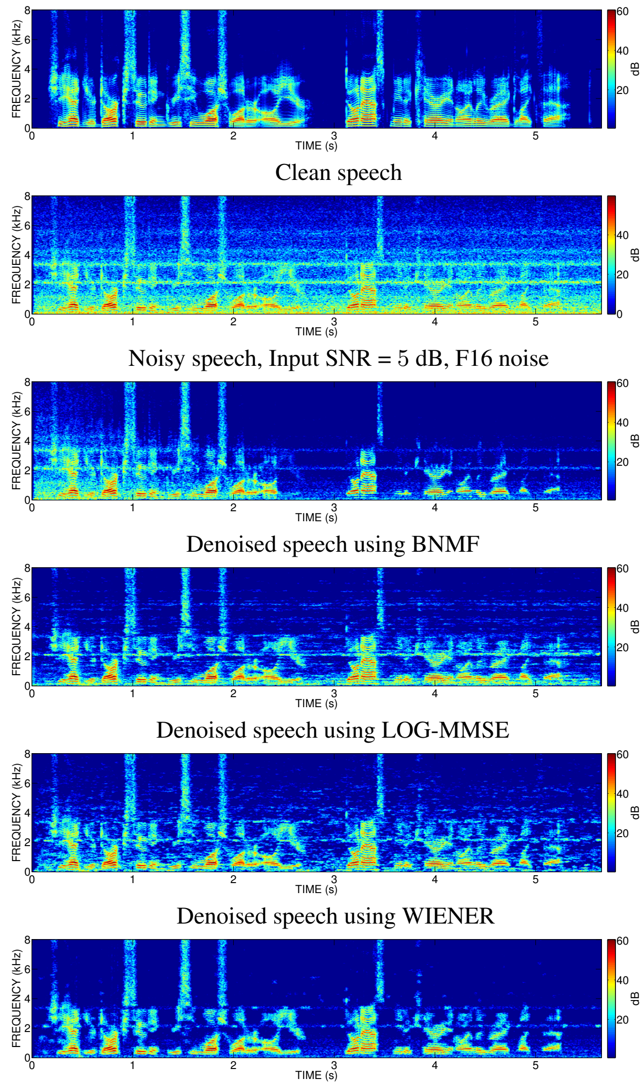


Figure 5: (Color online) Spectrograms of the denoised speech obtained using different algorithms

SIG are high compared with the other algorithms. In the case of factory noise, it is observed that SS-RE shows a high OVRL gain (cf. Figure 4(b)) in the range 0 dB to 20 dB but below 0 dB, BNMF shows high output OVRL. From Figure 4(d), we observe that the BAK score is high for the SS-RE algorithm. For low input SNR values, SIG value is lower than the other algorithms (cf. Figure 4(f)). From the results relating to the real-world noise cases, we infer that, if the noise variance estimate is accurate the performance of the SS-RE algorithm will be significantly better than the other algorithms. In the case of non-stationary noise, where the noise variance estimate may not be accurate, the denoised speech overall quality is comparable to the best performing algorithm under comparison. Moreover, the attenuation of background residual noise is high in SS-RE for all the observed cases as indicated by the BAK scores.

Figure 5 shows the spectrogram plots of the different denoising algorithms, where the noisy signal is generated by adding F16 noise at input SNR of 5 dB. It is observed that the SS-RE denoising algorithm yields a higher noise attenuation and maintains a low speech distortion compared with the other algorithms. In the case of WIENER and LOG-MMSE, the speech distortion is low but the residual noise is high. In the case of BNMF, noise attenuation is lesser in the initial two seconds. This is because the algorithm learns the noise model during this time. However, BNMF shows a higher noise attenuation after 2 seconds but at the cost of high speech distortion. The denoised speech files are available online at http://spectrumee.wix.com/ssre.

5. Conclusions

We derived an unbiased estimate of the MSE considering a multiplicative denoising function. Unlike conventional SURE, no distribution assumption is made for deriving the unbiased MSE estimate. The proposed unbiased estimate of MSE, which depends only on the second-order statistics allows the flexibility of using the risk estimation frame-work in many real-world noise scenarios where the distribution of noise is unknown. We developed a subband-shrinkage denoiser based on the proposed unbiased estimate of the MSE where the shrinkage parameters are obtained by minimizing unbiased estimate of the MSE. It is observed that the proposed subband-shrinkage denoiser outperform the three benchmark algorithms in terms of SSNR, PESQ, and composite scores.

6. Acknowledgements

We would like to thank to Mohammadiha et al. for providing Matlab implementation of their denoising technique [23] online.

7. References

[1] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust. Speech, Signal Process.*, ASSP-27(2), pp. 113–120, April 1979.

[2] P. Lockwood and J. Boudy, "Experiments with a non-linear spectral subtractor (NSS), hidden Markov models and the projections, for robust recognition in cars," *Speech Comm.*, vol. 11, issue 2, pp. 215–228, Jun. 1992.

[3] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, pp. 4164–4167, May 2002.

[4] Y. Hu and P. Loizou, "A perceptually motivated approach for speech enhancement," *IEEE Trans. Speech, Audio Process.*, vol. 11, no. 5, pp. 457–465, Sep. 2003.

[5] Y. Hu and P. Loizou, "Incorporating psycho-acoustical model in frequency domain speech enhancement," *IEEE Signal Process. Lett.*, vol. 11, no. 1, pp. 270–273, Feb. 2004.

[6] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insight into noise reduction Wiener filter," *IEEE Trans. Speech, Audio Process.* , vol. 14, no. 4, pp. 1218–1234, Jul. 2006.

[7] P. Scalart, and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* , vol. 2, pp. 629–632, May. 1996.

[8] S. Srinivasan, J. Samuelsson, and W. Kleijn, "Codebook driven short-term predictor parameter estimation for speech enhancement," *IEEE Trans. Speech, Audio Process.*, vol. 14, no.1, pp. 163–176, Jan. 2006.

[9] T. Rosenkranz and H. Puder, "Improving robustness of codebook-based noise estimation approaches with delta codebooks," *IEEE Trans. Speech, Audio Process.*, vol. 20, no.4, pp. 1177–1188, May. 2012.

[10] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech, Audio Process.*, vol. 3, no. 4, pp. 251–266, Jul. 1995.

[11] A. Rezayee and S. Gazor, "An adaptive KLT approach for speech enhancement," *IEEE Trans. Speech, Audio Process.*, vol. 9, no. 2, pp. 87–95, Feb. 2001.

[12] P. S. K. Hansen, P. C. Hansen, S. D. Hansen, and J. A. Sorensen, "Experimental comparison of signal subspace based noise reduction methods," *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 1, pp. 101–104, 1999.

[13] P. S. K. Hansen, P. C. Hansen, S. D. Hansen, and J. A. Sorensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Trans. Speech, Audio Process.*, vol. 3, no. 6, pp. 439–448, Nov. 1995.

[14] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-squared error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[15] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-squared error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 443–445, Apr. 1985.

[16] P. C. Loizou, "Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum," *IEEE Trans. Speech, Audio Process.*, vol. 13, no.5, pp. 857–869, Sep. 2005.

[17] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum mean-square error estimation of discrete fourier coefficients with generalized gamma priors," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 15, no. 6, pp. 1741−1752, Aug. 2007.

[18] M. McCallum and B. Guillemin, "Stochastic-deterministic MMSE STFT speech enhancement with general a priori information," *IEEE Trans. Speech, Audio Process.*, vol. 21, no. 7, pp. 1445–1457, Jul. 2013.

[19] Y. Ephraim, "A Bayesian estimation approach for speech enhancement using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 725–735, Apr. 1992.

[20] T. Lotter and P. Vary, "Speech enhancement by maximum a posteriori estimation using super-Gaussian speech model," *EURASIP J. Appl. Signal Process.*, vol. 7, pp. 1110–1126, 2005.

[21] P. Mowlaee and R. Saeidi, "Iterative closed-loop phase-aware single-channel speech enhancement," *IEEE Signal Process. Lett.*, vol. 20, no. 12, pp. 1235−1239, Dec. 2013.

[22] P. Loizou, *Speech Enhancement — Theory and Practice*. CRC Press, 2007.

[23] N. Mohammadiha, P. Smaragdis, and A. Leijon, "Supervised and unsupervised speech enhancement using nonnegative matrix factorization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no.10, pp. 2140–2151, Oct. 2013.

[24] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Stat.*, vol. 9, no. 6, pp. 1135−1151, Nov. 1981.

[25] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Trans. Image Process.*, vol 16, no. 3, pp. 593−606, Mar. 2007.

[26] J. C. Pesquet, A. Benazza-Benyahia, and C. Chaux, "A SURE approach for digital signal/image deconvolution problems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4616−4632, Dec. 2009.

[27] D. V. De Ville and M. Kocher, "Nonlocal means with dimensionality reduction and SURE-based parameter selection," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2683−2690, Sep. 2011.

[28] X. Zhang and M. D. Desai, "Adapting denoising based on SURE risk," *IEEE Signal Process. Lett.*, vol. 5, no. 10, pp. 265−267, Oct. 1998.

[29] F. Luisier, T. Blu, and M. Unser, "SURE-LET for orthonormal wavelet-domain video denoising," *IEEE Trans. on Circ. and Syst. for Video Tech.*, vol. 20, no. 6, pp. 913−919, Jun. 2010.

[30] A. M. Atto, D. Pastor, and G. Mercier, "Optimal SURE parameter for sigmoidal wavelet shrinkage," *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Glasgow, Scotland, pp. 40−43, Aug. 2009.

[31] N. R. Muraka and C. S. Seelamantula, "A risk-estimation-based comparison of mean-square error and Itakura-Saito distortion measures for speech enhancement," in *Proc. Interspeech*, Florence, Italy, pp. 349−352, Aug. 2011.

[32] N. Zheng, X. Li, T. Blu, and T. Lee, "SURE-MSE speech enhancement for robust speech recognition," in *Proc.* 7^{th} *Int. Symp. Chinese Spoken Lang. Process.*, pp. 271−274, Nov. 2010.

[33] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471−481, Feb. 2009.

[34] J. S. Garofolo et al., "TIMIT acoustic-phonetic continous speech corpus," *National Institute of standards and Technology*, 1990.

[35] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition ii: Noisex-92: a database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, vol. 12, no. 3, pp. 247−251, 1993.

[36] Y. Lin and P. P. Vaidyanathan, "A Kaiser window approach for the design of prototype filters of cosine modulated filterbanks," *IEEE Signal Process. Lett.*, vol. 5, no. 6, pp. 132−134, Jun. 1998.

[37] J. Sohn and N. S. Kim, "A statistical model-based voice activity detection," *IEEE Signal Process. Lett.*, vol. 6, no. 1, pp. 1−3, Jan. 1999.

[38] J. Hansen and B. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," in *Proc. Int. Conf. Spoken Lang. Process.*, 1998, vol. 7, pp. 2819−2822.

[39] ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ)-An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," International Telecommunication Union, Feb. 2001.

[40] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no.1, pp. 229–238, Jan. 2008.