Improvements of the Beta-Order Minimum Mean-Square Error (MMSE) Spectral Amplitude Estimator using Chi Priors

Marek B. Trawicki[1], *Michael T. Johnson*[1]

[1]Department of Electrical and Computer Engineering, Marquette University, Milwaukee, WI 53201

`marek.trawicki@marquette.edu, mike.johnson@marquette.edu`

Abstract

In this paper, the authors propose the Beta-Order Minimum Mean-Square Error (MMSE) Spectral Amplitude estimator with Chi statistical models for the speech priors. The new estimator incorporates both a shape parameter on the distribution and cost function parameter. The performance of the MMSE Beta-Order Spectral Amplitude estimator with Chi speech prior is evaluated using the Segmental Signal-to-Noise Ratio (SSNR) and Perceptual Evaluation of Speech Quality (PESQ) objective quality measures. From the experimental results, the new estimator provides gains of 0-3 dB and 0-0.3 in SSNR and PESQ improvements over the corresponding MMSE Beta-Order MMSE Spectral Amplitude estimator with the standard Rayleigh statistical models for the speech prior.

Index Terms: speech enhancement, amplitude estimation, phase estimation, parameter estimation

1. Introduction

Based on the performance of the Minimum Mean-Square Short-Time (MMSE) Spectral Amplitude (STSA) estimator [1], researchers began to modify the squared-error of the spectral amplitude cost function to utilize more subjectively meaningful cost functions. Ephraim and Malah also developed and implemented the MMSE Log-Spectral Amplitude (LSA) [2] estimator that minimizes the squared-error of the log-spectral amplitude, which is a more subjectively meaningful cost function that correlates well with human perception. You, Koh, and Rahardja [3, 4] and then Plourde and Champagne [5] investigated the beta-order cost function, which minimizes the Mean-Square Error (MSE) between the MMSE Beta-Order Spectral Amplitude and estimated MMSE Beta-Order Spectral Amplitude by incorporating a power law parameter on the spectral amplitude cost function. The beta-order cost function yielded a good trade-off between speech distortion and residual noise reduction, particularly for weak spectral components. In each of those corresponding spectral amplitude and log-spectral amplitude estimators, the cost functions employed Rayleigh distributions for the statistical models of the speech priors.

In order to improve performance of the spectral amplitude estimators using objective measures such as Segmental Signal-to-Noise Ratio (SSNR) and Perceptual Evaluation of Speech Quality (PESQ) for generating gains in noise reduction and overall speech quality, researchers began to exploit alternative and more accurate statistical modeling assumptions to the Rayleigh distribution for both the speech prior using the spectral amplitude cost function. Andrianakis and White [6, 7] continued with the MMSE and Maximum A Posteriori (MAP) spectral amplitude estimators with Gamma distribution but introduced Chi distribution for modeling the speech priors. The Chi speech

prior contains a shaping parameter that was varied to determine its effect on the quality of enhanced speech. From the results, the performance of the estimators was dependent on the shaping parameter, which controlled the trade-off between the level of residual noise and musical tones. As a generalization to the MMSE STSA, LSA, and Beta-Order Spectral Amplitude estimators, Breithaupt, Krawczyk, and Martin [8] developed a MMSE STSA estimator that uses both a variable compression function in the error criterion and the Chi distribution as a prior model. The resulting two parameters provide for the reduction of musical noise, speech distortion, and noise distortion. Through the incorporation of Chi distribution statistical models for the speech prior, the squared-error cost functions demonstrated distinct improvement over the Rayleigh statistical models.

Despite the success of the MMSE beta-order cost function with Rayleigh speech prior and MMSE spectral amplitude cost functions with the Rayleigh and Chi speech prior, there has not been any work to merge their positive impacts together into a single estimator. Specifically, the improved statistical models for the speech prior have only been incorporated with the original MMSE STSA estimator, not with the MMSE Beta-Order Spectral Amplitude estimator. Instead of utilizing the Rayleigh distribution for the speech prior, the Chi distribution is employed in this work since it leads to more general, less complicated, and more closed-form estimator solutions. Thus, the focus of this work is to use the MMSE Beta-Order Spectral Amplitude estimator with the Chi spectral speech prior distribution [9] for reducing the background noise and improving overall speech quality.

The remainder of this paper is organized into the following sections: system and statistical models (Section 2), beta-order spectral amplitude estimation (Section 3), experiments and results (Section 4), and conclusion (Section 5).

2. System and statistical models

In the time domain, the single channel additive noise model is given as

$$y(t) = s(t) + d(t), \tag{1}$$

where $s(t)$, $d(t)$, and $y(t)$ represent the clean, noise, and noisy signals. By taking the short-time Fourier Transform, (1) can be written in the frequency domain as

$$Y(\lambda,k) = S(\lambda,k) + D(\lambda,k)$$
$$R(\lambda,k)e^{j\vartheta(\lambda,k)} = X(\lambda,k)e^{j\alpha(\lambda,k)} + N(\lambda,k)e^{j\theta(\lambda,k)}, \tag{2}$$

where λ and k are the particular frame and frequency bin index.

As opposed to using the traditional Rayleigh statistical models for the speech prior and noise likelihood, the speech prior

is modified through the use of Chi speech priors [9]. Specifically, the Chi speech prior is given as

$$p(X,\alpha) = \frac{2}{\theta^a \Gamma(a)} X^{2a-1} \exp\left(-\frac{X^2}{\theta}\right),$$ (3)

where $\sigma_X^2 = \theta a$ with shape parameter a. With $a = 0.5$ and $a = 1$, (3) is equivalent to the Half-Rayleigh and Rayleigh distributions. The noise likelihood is still modeled as a Rayleigh distribution given as

$$p(Y|X,\alpha) = \frac{1}{\pi \sigma_N^2} \exp\left(-\frac{|Y - Xe^{j\alpha}|^2}{\sigma_N^2}\right).$$ (4)

3. Beta-order spectral amplitude estimation

From [3], the beta-order cost function is given as

$$d_\beta(X, \hat{X}) = (X^\beta - \hat{X}^\beta)^2,$$ (5)

where β is the beta-order parameter value with estimator equation

$$\hat{X}_\beta = \left[\frac{\int_0^\infty \int_0^{2\pi} X^\beta p(Y|X,\alpha) p(X,\alpha) d\alpha dX}{\int_0^\infty \int_0^{2\pi} p(Y|X,\alpha) p(X,\alpha) d\alpha dX}\right]^{\frac{1}{\beta}}$$ (6)

for $\beta \neq 0$. With substitution of the statistical models in (3) and (4) and using 8.431.5 and 8.406.1 in [10], the spectral phase is integrated from the numerator integral in (6) as

$$\int_0^\infty \int_0^{2\pi} X^\beta p(Y|X,\alpha) p(X,\alpha) d\alpha dX \propto$$
$$\int_0^\infty X^{2a+\beta-1} \exp\left(-\frac{X^2}{\lambda_a}\right) J_0\left(2iX\sqrt{\frac{v}{\lambda}}\right) dX$$ (7)

and

$$\int_0^\infty \int_0^{2\pi} p(Y|X,\alpha) p(X,\alpha) d\alpha dX \propto$$
$$\int_0^\infty X^{2a-1} \exp\left(-\frac{X^2}{\lambda_a}\right) J_0\left(2iX\sqrt{\frac{v}{\lambda}}\right) dX$$ (8)

where λ is defined in [1] as

$$\frac{1}{\lambda} = \frac{1}{\lambda_X} + \frac{1}{\lambda_D},$$ (9)

$J_0(\square)$ is the 0th-order Bessel function of the first-kind, and

$$\frac{1}{\lambda_a} = \frac{a}{\lambda_X} + \frac{1}{\lambda_D},$$ (10)

which is equivalent to $1/\lambda$ for $a = 1$. By utilizing 6.631.1 and 9.212.1 in [10], (7) and (8) are given as

$$\int_0^\infty \int_0^{2\pi} X^\beta p(Y|X,\alpha) p(X,\alpha) d\alpha dX \propto$$
$$\int_0^\infty X^{2a+\beta-1} \exp\left(-\frac{X^2}{\lambda_a}\right) J_0\left(2iX\sqrt{\frac{v}{\lambda}}\right) dX \propto,$$ (11)
$$\frac{\Gamma\left(a + \frac{\beta}{2}\right)}{2} \lambda_a^{a+\frac{\beta}{2}} {}_1F_1\left(1 - a - \frac{\beta}{2}; 1; -\frac{v/\lambda}{1/\lambda_a}\right)$$

and

$$\int_0^\infty \int_0^{2\pi} p(Y|X,\alpha) p(X,\alpha) d\alpha dX \propto$$
$$\int_0^\infty X^{2a-1} \exp\left(-\frac{X^2}{\lambda_a}\right) J_0\left(2iX\sqrt{\frac{v}{\lambda}}\right) dX \propto,$$ (12)
$$\frac{\Gamma(a)}{2} \lambda_a^a {}_1F_1\left(1 - a; 1; -\frac{v/\lambda}{1/\lambda_a}\right)$$

where $\Gamma(\bullet)$ and ${}_1F_1(\bullet;\bullet;\bullet)$ are the gamma and confluent hypergeometric functions. With the combination and simplification of the integrals in (11) and (12), the final form of the new MMSE Beta-Order Spectral Amplitude estimator with Chi speech prior in (3) is given as

$$\hat{X}_{\beta,CHI} =$$
$$\left[\frac{\Gamma\left(a + \frac{\beta}{2}\right)}{\Gamma(a)}\right]^{\frac{1}{\beta}} \frac{\sqrt{v_a}}{\gamma} \left[\frac{{}_1F_1\left(1 - a - \frac{\beta}{2}; 1; -v\zeta_a\right)}{{}_1F_1(1 - a; 1; -v\zeta_a)}\right]^{\frac{1}{\beta}} R,$$ (13)

where v is defined in [1] as

$$v = \frac{\xi}{1 + \xi} \gamma$$ (14)

and

$$v_a = \frac{\xi}{a + \xi} \gamma$$ (15)

and

$$\zeta_a = \frac{1 + \xi}{a + \xi}.$$ (16)

with $1/\lambda_a > 0$, $a > 0$, and $2a + \beta > 0$. For $a = 1$, (13) is exactly equivalent to the MMSE Beta-Order Spectral Amplitude estimator with Rayleigh speech prior [3].

4. Experiments and results

The proposed optimal MMSE Beta-Order Spectral Amplitude estimator with Chi speech prior given in (13) was evaluated using the objective measure of SSNR and PESQ. Clean and noisy speech were taken from the noisy speech corpus (NOIZEUS) [11], which contains 30 IEEE sentences [12] (produced by three male and three female speakers) corrupted by eight different real-world noises at different SNRs ranging from 0 dB to 15 dB at increments of 5 dB, where the noises were taken from the AURORA database [13], which includes train, babble, car, exhibition hall, restaurant, street, airport, and station noises. The analysis conditions consisted of frames of 256 samples (25.6 ms) with 50% overlap using Hanning windows.
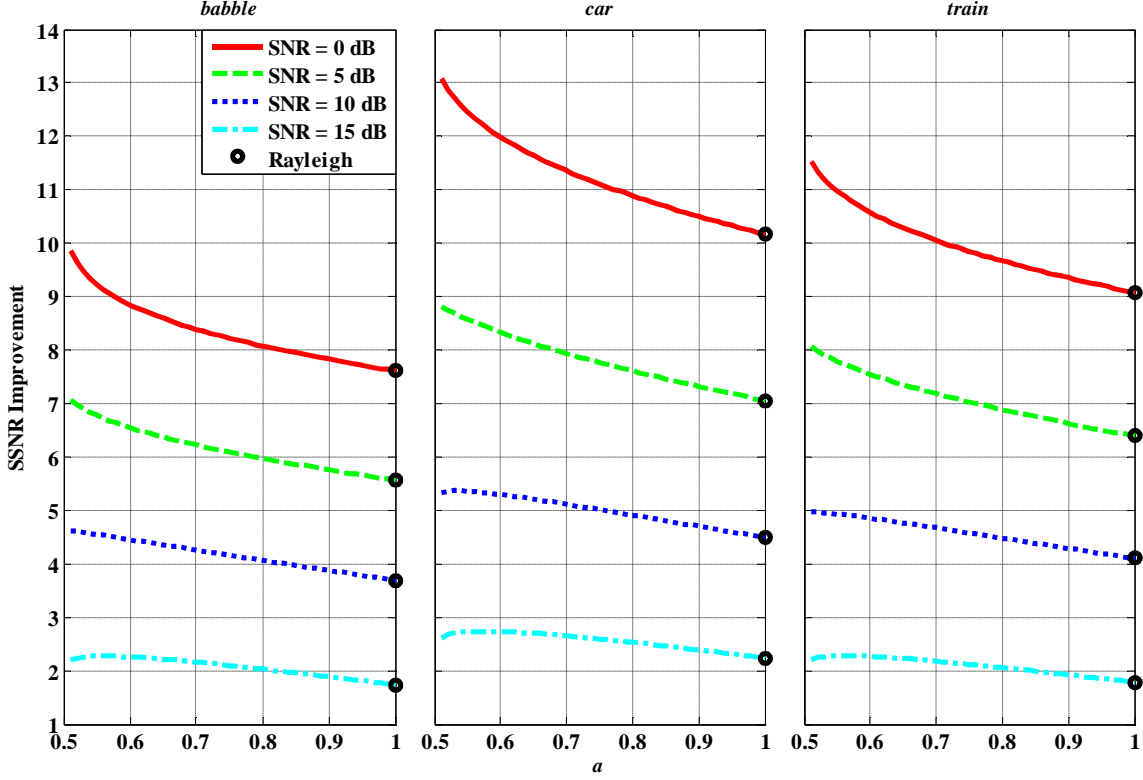
Figure 1: *SSNR Improvement for MMSE Beta-Order Spectral Amplitude Estimator with Chi Prior*

Noise estimation was performed on an initial silence of 5 frames. The decision-directed (DD) [1] smoothing approach was utilized to estimate ξ with $\alpha_{SNR} = 0.98$ using thresholds of $\xi_{min} = 10^{-25/10}$ and $\gamma_{min} = 40$. The enhanced signals were reconstructed using the overlap-add technique.

The shape parameter a in the Chi speech prior was varied for a specific β parameter value to determine its impact on noise reduction and speech quality with results averaged over 30 utterances using babble, car, and train noises. As advised by Plourde and Champagn [5], $\beta = -1$ was utilized as the β parameter value to balance the gains in noise reduction and speech quality. Figure 1 and Figure 2 demonstrate the SSNR and PESQ improvements using the baseline MMSE Beta-Order Spectral Amplitude estimator with Rayleigh speech prior [3-5], where SSNR/PESQ improvement was computed as SSNR/PESQ output (enhanced signal) minus SSNR/PESQ input (noisy signal). For the SSNR improvements, the MMSE Beta-Order Spectral Amplitude estimator with Chi speech prior delivered gains over the baseline MMSE Beta-Order Spectral Amplitude estimator with Rayleigh speech prior of approximately 0-3 dB across the different noises with the best results at the lower SNRs (i.e., 0 dB and 5 dB) for the car noise following by the train and babble noises. By proceeding to decrease the β parameter value, which would result in the increase of the a shape parameter value through the relationship $2a + \beta > 0$ from (13), the SSNR improvements would continue to increase for all

the input SNRs, especially at 5 dB and 10 dB. In regards to the PESQ improvements, the MMSE Beta-Order Spectral Amplitude estimator with Chi speech prior generated returns of 0-0.3 over the baseline MMSE Beta-Order Spectral Amplitude estimator with Rayleigh speech prior. At input SNRs of 0 dB and 5 dB, the PESQ improvement reached around 0.55 and 0.50 (car noise), 0.48 and 0.44 (train noise), and 0.29 and 0.29 (babble noise), which did not decrease below the baseline MMSE Beta-Order Spectral Amplitude estimator with Rayleigh speech prior until the a shape parameter extended to $a = 0.7 - 0.8$. In the end, the best performance occurred for the car noise at input SNR of 5-10 dB using the SSNR and PESQ objective metrics.

5. Conclusion

In this paper, the authors derived MMSE Beta-Order Spectral Amplitude estimator using Chi speech priors. By comparison to the baseline MMSE Beta-Order Spectral Amplitude estimators using the Rayleigh speech prior, the Chi speech prior produced gains of 0-3 dB and 0-0.3 in SSNR and PESQ improvements. As measured by SSNR and PESQ metrics, the estimators showed solid performance for noise reduction and overall quality with the Chi speech priors in babble, train, and car noises. For future work, the authors propose developing and implementing an optimal combination of the shape parameter a and β parameter for the Chi distribution and beta-order cost function to provide further improvements in noise reduction and overall speech quality.
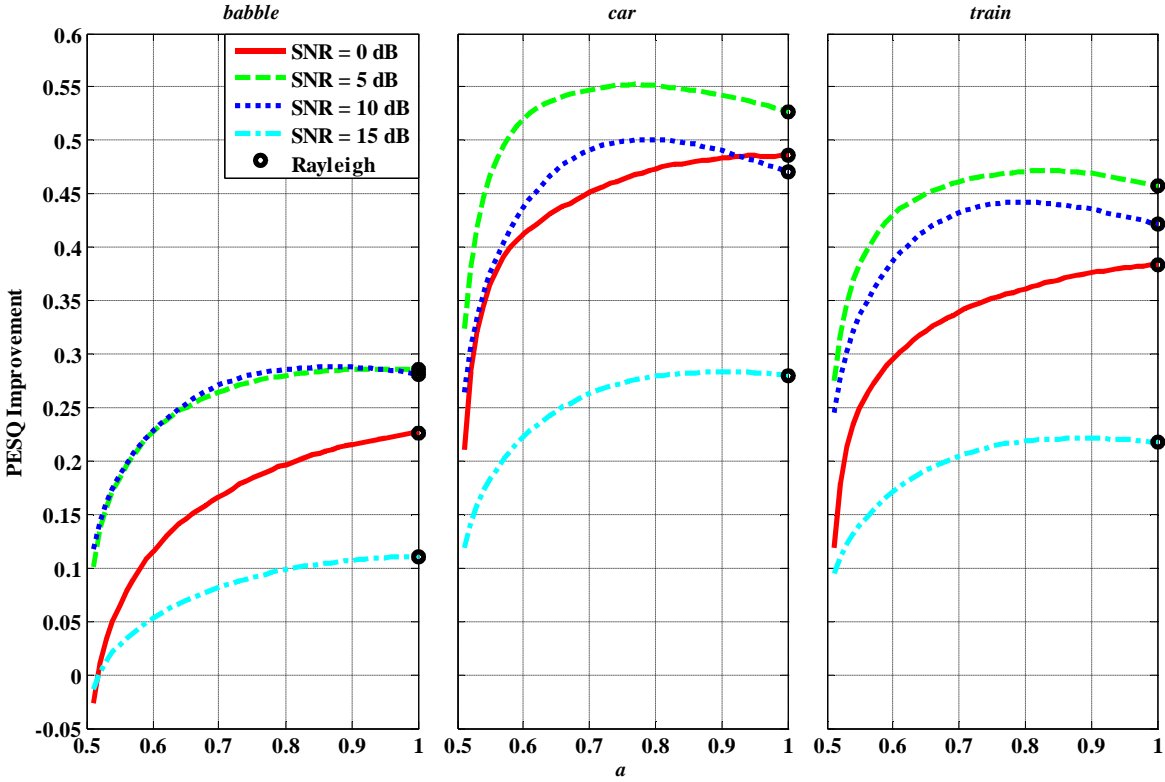
Figure 2: PESQ Output *for MMSE Beta-Order Spectral Amplitude Estimator with Chi Prior*

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, pp. 1109-1121, 1984.

[2] Y. Ephraim and D. Malah, "Speech Enhancement using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, pp. 443-445, 1985.

[3] C. H. You, S. N. Koh, and S. Rahardja, "Beta-Order MMSE Spectral Amplitude Estimation for Speech Enhancement," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 13, pp. 475-486, 2005.

[4] C. H. You, S. N. Koh, and S. Rahardja, "Masking-based beta-order MMSE speech enhancement," *Speech Communication*, vol. 48, pp. 57-70, 2006.

[5] E. Plourde and B. Champagne, "Further Analysis of the Beta-Order MMSE STSA Estimator for Speech Enhancement," presented at International Conference on Acoustics, Speech, and Signal Processing, 2007.

[6] I. Andrianakis and P. R. White, "Speech Spectral Amplitude Estimators using Optimally-Shaped Gamma and Chi Priors," *Speech Communication*, pp. 1-14, 2009.

[7] I. Andrianakis and P. R. White, "MMSE Speech Spectral Amplitude Estimators with Chi and Gamma Speech Priors," presented at International Conference on Acoustics, Speech, and Signal Processing, 2006.

[8] C. Breithaupt, M. Krawczyk, and R. Martin, "Parameterized MMSE Spectral Magnitude Estimation for the Enhancement of Noisy Speech,"
presented at International Conference on Acoustics, Speech, and Signal Processing, 2008.

[9] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 1, 2nd ed: John Wiley and Sons, 1994.

[10] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products*: Academic Press, 2007.

[11] Y. Hu and P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech Communication*, vol. 49, pp. 588-601, 2007.

[12] I. Subcommittee, "IEEE Recommended Practice for Speech Quality Measurements," *IEEE Transactions on Audio and Electroacoustics*, vol. AU-17, pp. 225-246, 1969.

[13] D. Pearce and H.-G. Hirsch, "Performance Evaluation of Speech Recognition Systems under Noisy Conditions," presented at 6th International Conference on Spoken Language Processing (ICSLP), Beijing, China, 2000.