GENERALIZED GAMMA MODELING OF SPEECH AND ITS ONLINE ESTIMATION FOR SPEECH ENHANCEMENT

Tran Huy Dat, Kazuya Takeda, and Fumitada Itakura
Nagoya University, Japan

ABSTRACT

Generalized gamma modeling and its online parameter estimation method of the speech spectral magnitude are proposed for the MAP based speech enhancement systems. The generalized gamma modeling is shown to be a natural extension of the Gaussian modeling of speech spectral components distribution, therefore, able to fit the prior distribution better than the conventional method. Then an online parameter estimation method for the gamma distribution based on moment matching method is proposed. The effectiveness of the proposed methods are confirmed in both SNR and ASR improvement points of view using AURORA2 standard database, where about 4 dB improvement in SNR and 20% improvement in relative ASR performance are obtained.

1. INTRODUCTION

The reduction of additive noise is an important problem in speech enhancement and speech recognition. Among the one-channel enhancement methods, statistical decomposition methods using the statistical modeling of speech and noise are shown to be more effective for the noise reduction and to produce less speech distortion [1]. Recently we have proposed MAP magnitude estimation based on the Rice conditional distribution and gamma prior modeling of the speech spectral magnitude [2]. This method has shown better performances in terms of both segmental SNR and speech recognition, compared to MMSE and spectral subtraction. In this work, we generalize the gamma modeling on arbitrary power order of spectral magnitude and develop online prior parameter estimations using the moment matching method. The organization of the paper is as follows. Section 2 reviews the MAP magnitude estimation problem. Section 3 applies the generalized gamma modeling on speech spectral magnitude to the MAP estimation. In section 4 we derive the online parameter estimations. Section 5 reports the experimental results and section 6 summarizes the work.

2. MAP SPECTRAL MAGNITUDE ESTIMATION

We consider an additive model of noisy speech,

$$\mathbf{X}(n,k) = \mathbf{S}(n,k) + \mathbf{N}(n,k), \quad (1)$$

where: $\mathbf{X}$, $\mathbf{S}$ and $\mathbf{N}$ are the complex spectrum of noisy speech, the clean speech and the noise signal at each time-frequency index (n,k). Hereafter we use $(.)_R$, $(.)_I$ $(.)$ and $\varphi_{(.)}$ to represent the real and imaginary parts, the magnitude and the phase of a complex spectrum, i.e., $\mathbf{X} = X_R + jX_I = Xe^{j\varphi_X}$ respectively.

Conventionally [1] the zero-mean independent Gaussian distribution is assumed for the real and imaginary parts (spectral components) of the noise spectrum:

$$N_R, N_I \sim Gaussian\left(0, \frac{\sigma_N^2}{2}\right), \quad (2)$$

where σ_N^2 denotes the spectral density (or 'local' power) of the noise signal at the time-frequency index (n,k):

$$\sigma_N^2(n,k) = \left\langle \widehat{|N(n,k)|^2} \right\rangle. \quad (3)$$

Hereafter, we use σ_S^2 and σ_N^2 for the local power of speech and the noise signal, respectively. Since neither σ_S^2 or σ_N^2 can be observed we must estimate them from the noisy speech [3]. The MAP estimate of the clean speech spectral magnitude is defined by $\hat{S} = \arg\max_S \left[\log p(S \mid X)\right]$ and is denoted by

$$\hat{S} = \arg\max_S \left[\log\left(p(X \mid S)\, p(S)\right)\right]. \quad (4)$$

Under assumption (2), the Rice conditional probability can be used [2]:

$$p(X \mid S) = \frac{X}{2\pi\sigma_N^2} \exp\left(-\frac{X^2 + S^2}{\sigma_N^2}\right) I_0\left(\frac{2XS}{\sigma_N^2}\right). \quad (5)$$

Here, $I_0(x)$ is the modified Bessel function of the first kind. The fact that the MAP decomposition can be carried out without making any assumptions on the phase prior distribution should make the estimation more robust. Equation (4) implies the following equation:

$$\frac{\partial}{\partial S}\left[\log\left(p(X \mid S) + \log\left(p(S)\right)\right)\right] = 0. \quad (6)$$

By making an approximation of the Bessel function,

$$I_0(x) \approx \frac{1}{\sqrt{2\pi x}} e^x, x > 0, \quad (7)$$

the first term in (6) can be calculated as follows,

$$\frac{\partial}{\partial S}\left[\log\left(p\left(X \mid S\right)\right)\right] = -\frac{2S}{\sigma_N^2} - \frac{1}{2S} + \frac{2X}{\sigma_N^2}. \quad (8)$$

Note that the relative error given by approximation (7) is less than 1% at $x > 0.2$ or: $10\log_{10}\frac{\sigma_S^2}{\sigma_N^2} > -20$ dB. Since our interest is in the noisy speech with SNR in the range of -5dB to 20dB, the approximation error is negligible.

3. GENERALIZED GAMMA MODELING

3.1 Gaussian model of speech

Naturally, the MAP estimation equation (6) would give a more accurate solution if we could better fit the prior distribution $p\left(S\right)$. The conventional model assumes the zero mean Gaussian distribution of the speech spectral components,

$$S_R, S_I \sim Gaussian\left(0, \frac{\sigma_S^2}{2}\right). \quad (9)$$

Since the prior magnitude distribution $p\left(S\right)$ is scaled by the local speech powers which are estimated separately, we normalize the speech magnitude square to their local powers. From (9), the distribution of the normalized magnitude square is given by a unit exponential distribution namely: $p\left(\frac{S^2}{\sigma_S^2}\right) = \exp\left(-\frac{S^2}{\sigma_S^2}\right)$. (10)

Substituting (10) into (6) yields the MAP spectral magnitude estimation:

$$\hat{S} = \left(\frac{1}{\sigma_N^2} + \frac{1}{\sigma_S^2}\right)^{-1}\left[\frac{X}{2\sigma_N^2} + \sqrt{\left(\frac{X}{2\sigma_N^2}\right)^2 + \left(\frac{1}{\sigma_N^2} + \frac{1}{\sigma_S^2}\right)}\right]. \quad (11)$$

Denoting (11) as a noise suppression filter, $\hat{S} = GX$ where the gain function is given by

$$G = \frac{1}{2\left(1+\frac{1}{\xi}\right)} + \sqrt{\frac{1}{4\left(1+\frac{1}{\xi}\right)^2} + \frac{1}{4\gamma\left(1+\frac{1}{\xi}\right)}}, \quad (12)$$

$\xi \sim \frac{\sigma_S^2}{\sigma_N^2}, \gamma \sim \frac{X^2}{\sigma_N^2}$ are the local prior and posterior SNRs.

3.2 Gamma modeling on the power domain

Distribution (10) is the special case of a gamma distribution with the fixed parameters $a = b = 1$:

$$p\left(\frac{S^2}{\sigma_S^2}\right) = \frac{b^a}{\Gamma\left(a\right)}\left(\frac{S^2}{\sigma_S^2}\right)^{a-1}\exp\left(-b\frac{S^2}{\sigma_S^2}\right). \quad (13)$$

On the basis of the above consideration, the prior distribution of the normalized spectral magnitude square is expected to be modeled more accurately by optimizing the distribution parameters $\left(a, b\right)$ of the gamma distribution (13). Note that the constraint $a = b$ is implied such that (13) has a unit mean, or equivalently:

$$\left\langle S^2 \right\rangle = \sigma_S^2. \quad (14)$$

The gamma distribution (13) yields the generalized Rayleigh distribution on the magnitude domain, which is denoted by

$$p\left(S\right) = \frac{b^a}{\Gamma\left(a\right)\sigma_S}\left(\frac{S}{\sigma_S}\right)^{2a-1}\exp\left(-b\frac{S^2}{\sigma_S^2}\right). \quad (15)$$

Substituting (15) into (6) yields the gain function of the noise suppression filtering by

$$G = \frac{1}{2\left(1+\frac{b}{\xi}\right)} + \sqrt{\frac{1}{4\left(1+\frac{b}{\xi}\right)^2} + \frac{4a-3}{4\gamma\left(1+\frac{b}{\xi}\right)}}. \quad (16)$$

3.3 Gamma modeling on the magnitude domain

The gamma modeling can also be applied to the normalized spectral magnitude, and is denoted as follow

$$\frac{S}{\sigma_S} \sim gamma\left(a, b\right). \quad (17)$$

Here, the normalization condition implies the constraint:

$$\left\langle S^2 \right\rangle = \sigma_S^2 \Leftrightarrow b = \sqrt{a\left(a+1\right)}. \quad (18)$$

In this case, the noise suppression rule is given by

$$G = \left(\frac{1}{2} - \frac{b}{4\sqrt{\xi\gamma}}\right) + \sqrt{\left(\frac{1}{2} - \frac{b}{4\sqrt{\xi\gamma}}\right)^2 + \frac{a-1.5}{2\xi}}. \quad (19)$$

3.4 Gamma modeling on the arbitrary magnitude power domain

The gamma modeling in sections 3.2 and 3.3 can be generalized to an arbitrary magnitude power domain:

$$\left(\frac{S}{\sigma_S}\right)^L \sim Gamma\left(a, b\right). \quad (20)$$

Now we call the magnitude distribution a generalized gamma distribution and it is denoted by:

$$p\left(S\right) = \frac{b^a}{\Gamma\left(a\right)\sigma_S}\left(\frac{S}{\sigma_S}\right)^{L(a-1)+L-1}\exp\left[-b\left(\frac{S}{\sigma_S}\right)^L\right]. \quad (21)$$

The generalized gamma distribution is a wide class of distributions including the gamma distribution, the generalized Rayleigh distribution and the positive generalized Gaussian distribution. Substituting (21) into (6) yields the MAP estimation equation as follows:

$$G - 1 - \frac{\left(La-1.5\right)}{2\gamma G} + G^{L-1}\frac{cb}{2\gamma}\left(\frac{\gamma}{\xi}\right)^{L/2} = 0. \quad (22)$$

Since equation (22) does not always have a closed form solution we implement a numerical solution using the Newton-Raphson method. Figure 1 gives a summary of the gamma modeling of spectral magnitude.

$S_R, S_I \sim Gaussian$

c $(equivalent)$

$S^2 \sim exponential \Leftrightarrow S \sim Rayleigh$

$\Downarrow (generalization)$ $\Uparrow (special\ case)$

$S^2 \sim gamma \Leftrightarrow S \sim generalized\ Rayleigh$

$\Downarrow (generalization)$ $\Uparrow (special\ case)$

$S^L \sim gamma \Leftrightarrow S \sim generalized\ gamma$

Figure 1 Spectral magnitude modeling of speech

4. ONLINE GAMMA PARAMETER ESTIMATION

In this section, we develop a parameter estimation based on the moment matching method. In order to use the generalized gamma distribution we must fix the parameters (a,b). Note that the unit second moment imposes a relationship between a and b and therefore there is only one free parameter.

4.1 Gamma prior estimation on the magnitude square domain

The fourth order moment of the spectral magnitude of observed noisy speech can be denoted by

$$\left\langle X^4 \right\rangle = \left\langle \left(S^2 + N^2 + 2SN\cos\left(\varphi_S - \varphi_N\right)\right)^2 \right\rangle . \quad (23)$$

According to the above discussion, the speech and noise spectral magnitude square should have gamma (13) and unit exponential distributions. Approximating the distribution of the phase difference by a uniform distribution, the fourth order moment is given by

$$\left\langle X^4 \right\rangle = \left\langle S^4 \right\rangle + \left\langle N^4 \right\rangle + 4\left\langle S^2 \right\rangle \left\langle N^2 \right\rangle . \quad (24)$$

Denoting (24) in terms of moments of exponential and gamma distributions of the normalized speech and noise magnitude squares, and substituting the moments of Gamma distributions and averaging σ_S and σ_N in each frequency bin, yields: $\left\langle S^4 \right\rangle = \overline{\sigma_S^4}\,\frac{a(a+1)}{a^2}$, $\left\langle N^4 \right\rangle = 2\overline{\sigma_N^4}$.

The estimation equation (24) is then given by

$$\left\langle X^4 \right\rangle = \overline{\sigma_S^4}\,\frac{a(a+1)}{a^2} + 2\overline{\sigma_N^4} + 4\overline{\sigma_S^2}\,\overline{\sigma_N^2} . \quad (25)$$

Here $\overline{(.)}$ denotes the operation of averaging over time. Finally, we obtain a closed form of the prior parameters:

$$a = b = q\left(\frac{\overline{X^4} - 2\overline{\sigma_N^4} - 4\overline{\sigma_S^2}\,\overline{\sigma_N^2}}{\overline{\sigma_S^4}} - 1 \right)^{-1} . \quad (26)$$

4.2 Gamma prior estimation on the magnitude domain

Analogously the distribution parameters of the gamma prior of speech can be estimated on the magnitude domain. Here, the fourth moment is given by

$$\left\langle X^4 \right\rangle = \overline{\sigma_S^4}\,\frac{(a+2)(a+3)}{a(a+1)} + 2\overline{\sigma_N^4} + 4\overline{\sigma_S^2}\,\overline{\sigma_N^2} , \quad (27)$$

and therefore the closed form solution is given by:

$$a = \frac{4q-1}{2} + \sqrt{\left(\frac{4q-1}{2}\right)^2 + 5q} , \quad (28)$$

where q is given by (26).

4.3 Gamma prior estimation on the arbitrary magnitude power domain

Analogously, the normalization condition and the estimation equation (24) can be denoted by

$$\left\langle S^2 \right\rangle = \sigma_S^2 \Leftrightarrow b = \left(\Gamma\left(a + \frac{2}{L}\right) / \Gamma(a) \right)^{\frac{L}{2}} , \quad (29)$$

$$\left\langle X^4 \right\rangle = \overline{\sigma_S^4}\,\Gamma\left(a + \frac{4}{L}\right)\Gamma(a)\,\Gamma\left(a + \frac{2}{L}\right)^{-2} + 2\overline{\sigma_N^4} + 4\overline{\sigma_S^2}\,\overline{\sigma_N^2} . \quad (30)$$

Since equation (30) has no closed form solution, the numerical method is used. Figures2-4 show examples of parameter estimation for the L=1, L=2 and L=1.45.

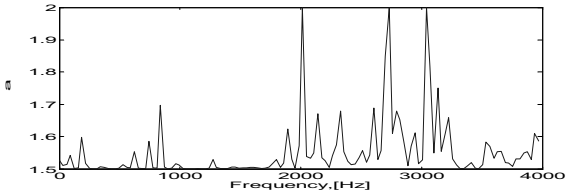

Figure2. Gamma prior estimation for $L = 1$

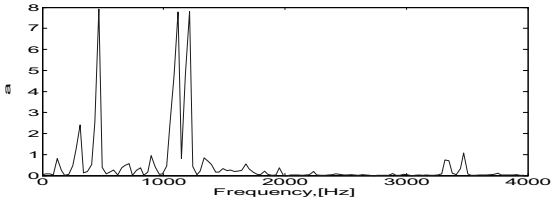

Figure 3. Gamma prior estimation for $L = 2$

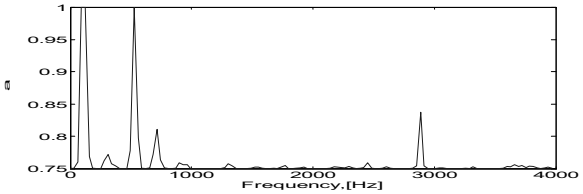

Figure 4. Gamma prior estimation for $L = 1.45$

5. EXPERIMENTS

The MAP based speech enhancements using the proposed gamma modeling on magnitude (L=1 named RGL1) and magnitude square (L=2 named RGL2) domains are implemented, with both online and offline parameter estimation. The gain factors for the filtering are given by (16) and (19), respectively. For the offline versions, we set a=1.5 for the magnitude distribution whereas a=0.5 is used for the magnitude square domain. Those numbers are found by preliminary experiments. The generalized order modeling given by (18) is also implemented with *L=1.45* (RGL1.45), where the online parameter estimation is performed by Newton-Raphson method. For the comparison, we implemented three conventional methods based on the Gaussian model of speech, using MAP (MAP-G), Wiener filtering (WF) and MMSE [1]. Note that the WF is MMSE estimation on power domain for the Gaussian model, the MMSE derived in [1] is done on the magnitude domain and the MAP-G gain function is given by (12). The local noise power is estimated using the minimum statistic method [2]. The local speech power is estimated by a subtraction:

$$\sigma_S^2 = \max\left(|X|^2 - \sigma_N^2, \beta|X|^2\right), \text{ where } \beta = 0.01 \quad (31)$$

The Aurora2 speech data set is used for evaluation. The data has moderate SNR conditions from -5dB to 20dB. In Figure 5, the segmental SNR improvements are shown for each of 8 methods. From the figure, it can be seen that the proposed gamma distribution models perform better than the conventional methods. Among proposed methods, the gamma magnitude modeling with online parameter estimation performed best with about 4dB improvement from simple WF. In Figure 6 and 7, the relative improvements of recognition accuracies are also plotted. The effectiveness of using gamma modeling can be also confirmed in the recognition experiments, however, for the recognition experiments, generalized order modeling (L=1.45) performed best.

6. CONCLUSIONS

We proposed the generalized gamma modeling and their online parameter estimation method of speech spectral magnitude for the MAP based speech enhancement systems. We showed that the generalized gamma modeling is a natural extension of the Gaussian modeling of speech amplitude distribution, therefore, can fit the distribution better than the conventional method. Then we developed an online parameter estimation method for the gamma distribution based on moment matching method. The effectiveness of the proposed methods are confirmed in both SNR and ASR improvements using AURORA2 standard database, where about 4 dB improvement in SNR and 20 % improvement in relative ASR rate are obtained.

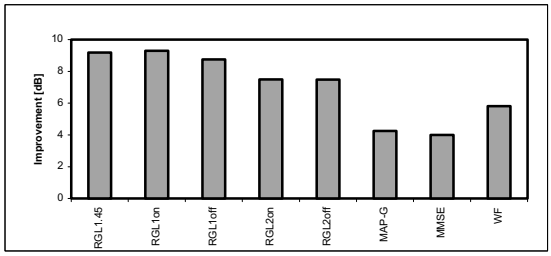

Figure5. Improvements in segmental SNR

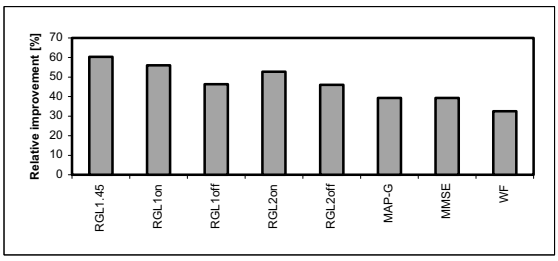

Figure 6. Relative improvement in ASR in clean training

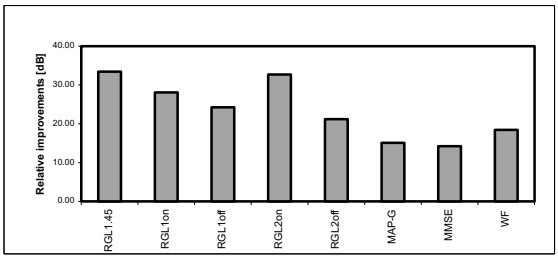

Figure7. Relative improvement in ASR in multi-condition training

Further studies are needed for improving the stabilities as well as the latency of the parameter estimation method for real applications especially for the gamma models of rational power of the speech spectral magnitude.

7. ACKNOWLEGDGMENTS

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for CC Society, Course Management System under Ubiquitous Computing Environment, 2004.

8. REFERENCES

[1] Y. Ephraim, and D. Malah, "Speech enhancement using a MMSE log-spectral amplitude estimation,'' IEEE Trans. ASSP, Vol. 33, No. 2, pp.443-445, 1985

[2] T. H. Dat, W. Lee, K. Takeda, and F.Itakura, "Speech enhancement based on MAP magnitude estimation and gamma prior," in Proc. ICSLP, 2004

[3] R. Martin, ''Noise power spectral estimation based on optimal smoothing and minimum statistics, ''IEEE Trans. ASSP, Vol. 9, No.5, pp.504-512, 2001.

[4] V.I. Tikhonov, Statistical radio technique, Moscow, Soviet Radio, 1983.