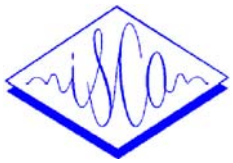
ISCA Archive
http://www.isca-speech.org/archive



INTERSPEECH 2004 - ICSLP
8th International Conference on Spoken Language Processing
ICC Jeju, Jeju Island, Korea
October 4-8, 2004

Speech Probability Distribution based on Generalized Gamma Distribution

Jong Won Shin, Joon-Hyuk Chang and Nam Soo Kim

School of Electrical Engineering and INMC
Seoul National University, Seoul, Korea
{jwshin, changjh}@hi.snu.ac.kr, nkim@snu.ac.kr

10.21437/Interspeech.2004-402

Abstract

In this paper, we propose a new speech probability distribution, two-sided generalized gamma distribution (GΓD) for an efficient parametric characterization of speech spectra. GΓD forms a generalized class of parametric distributions including the Gaussian, Laplacian and Gamma probability density functions (pdf's) as special cases. All the parameters associated with the GΓD are estimated by the on-line tracking procedure according to the maximum likelihood principle. Likelihoods, coefficients of variation (CV's), and Kolmogorov-Smirnov (KS) tests show that GΓD can model the distribution of the real speech signal more accurately than the conventional Gaussian, Laplacian, Gamma pdf or generalized Gaussian distribution (GGD).

1. Introduction

In many areas of speech signal processing such as speech coding, recognition and classification, adopting statistical models has been proven to be very attractive [1]-[3]. Accurate estimation of the probability density functions (pdf's) of clean speech, noisy speech and noise is an essential part of the reliable and effective speech signal processing techniques that are based on statistical modeling. In most of the speech processing algorithms adopting statistical models, the distribution of the speech spectra is assumed to be Gaussian. Recently, Martin [4] and Gazor et. al. [5] showed that the distribution of the speech signal can be approximated more closely by the Laplacian or Gamma pdf than by the traditional Gaussian pdf in various domains of speech representation. Moreover, the generalized Gaussian distribution (GGD) was proposed as a parametric pdf of speech signal, which includes the Gaussian and Laplacian pdf's as special cases [5].

In this paper, we propose the two-sided generalized gamma distribution (GΓD), which includes not only the Gaussian and Laplacian pdf's but also the conventional Gamma pdf as special cases, for a more efficient parametric modeling of speech distribution. The generalized gamma distribution, first proposed by Stacy [6], has been applied in various areas due mainly to its highly flexible form. Principal parameters associated with GΓD are estimated according to the maximum likelihood (ML) principle by means of a computationally inexpensive on-line algorithm. A variety of measures, e.g., likelihoods, coefficients of variation (CV's), and Kolmogorov-Smirnov (KS) statistics, are evaluated to compare the performance of each parametric model. From the results, it is shown that GΓD provides a more precise approximation of the speech spectra distribution than the Gaussian, Laplacian, Gamma pdf or GGD.

2. Two-sided generalized gamma distribution (GΓD)

GΓD is defined as follows:

$$f_{\mathbf{x}}(x) = \frac{\gamma\beta^{\eta}}{2\Gamma(\eta)}|x|^{\eta\gamma-1}\exp(-\beta|x|^{\gamma}) \qquad (1)$$

where $\Gamma(z)$ denotes the gamma function, and η, β and γ are positive real valued parameters. This covers a fairly flexible family of distributions which includes most of the commonly used speech signal distributions. Observe that for $\gamma = 1$ or $\eta\gamma = 1$ GΓD defines a (general) gamma pdf or GGD respectively. Furthermore, if $\gamma = 2$ and $\eta = 0.5$, it becomes the Gaussian pdf, and if $\gamma = 1$ and $\eta = 1$, it represents the Laplacian pdf. The pdf commonly referred to as just the 'Gamma pdf' is a special case of the gamma pdf with $\gamma = 1$ and $\eta = 0.5$. As in the conventional Gamma pdf case [4], GΓD diverges when the argument approaches zero if $\eta\gamma < 1$. In fact, GΓD can not be specified when the argument equals exactly to zero. For that reason, zero inputs are ignored during the parameter estimation procedure and also in the calculation of measures of fit.

The parameters η, β, and γ should be estimated to test the goodness-of-fit or to take advantage of the assumed pdf for various applications such as voice activity detection (VAD) and speech enhancement . Moment and ML estimators are the standard estimators for the GΓD. Moment estimators are simple to derive, but their sampling errors sometimes turn out to be unacceptably large. ML estimators are more efficient even though they are

somewhat more complicated to derive and to calculate from sample data [7]. Here, we apply the ML criterion to estimate the parameters of GΓD. Given N data $\mathbf{x} = \{x_1, x_2, \cdots, x_N\}$, with the assumption that the data are mutually independent, the log-likelihood function is given as follows:

$$\log f_{\mathbf{x}}(\mathbf{x}; \eta, \beta, \gamma) = \; N \log \frac{\gamma \beta^{\eta}}{2\Gamma(\eta)} + (\eta\gamma - 1) \sum_{i=1}^{N} \log |x_i| \; - \beta \sum_{i=1}^{N} |x_i|^{\gamma}. \quad (2)$$

By differentiating the log-likelihood function with respect to η, β and γ, and setting them to zero, we obtain the following three equations [7]:

$$\psi_0(\eta) \; = \; \log \beta + \frac{1}{N} \sum_{i=1}^{N} \log |x_i|^{\gamma} \quad (3)$$

$$\beta \; = \; \eta \frac{1}{\frac{1}{N} \sum_{i=1}^{N} |x_i|^{\gamma}} \quad (4)$$

$$\frac{1}{\eta} + \psi_0(\eta) - \log \beta - \frac{\beta}{\eta} \frac{1}{N} \sum_{i=1}^{N} |x_i|^{\gamma} \log |x_i|^{\gamma} \; = \; 0 \quad (5)$$

where $\psi_0(z)$ is the digamma function, which indicates the first-order derivative of $\log \Gamma(z)$. After some mathematical manipulation, the ML estimate of γ is obtained by the root of the single nonlinear equation

$$\psi_0\Big(\frac{\frac{1}{N} \sum_{i=1}^{N} |x_i|^{\gamma}}{\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |x_i|^{\gamma} \log \frac{|x_i|^{\gamma}}{|x_j|^{\gamma}}} \Big) - \frac{1}{N} \sum_{i=1}^{N} \log |x_i|^{\gamma} + \log \Big(\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |x_i|^{\gamma} \log \frac{|x_i|^{\gamma}}{|x_j|^{\gamma}} \Big) = 0. \quad (6)$$

Given the estimate of γ, it is straightforward to derive the estimates for η and β.

Since, however, it is difficult to solve (6) analytically, we employ the gradient ascent algorithm to obtain the estimate of γ, and determine the estimates of η and β based on the obtained value of γ. From now on, let us denote the estimates of γ, η and β by $\hat{\gamma}$, $\hat{\eta}$ and $\hat{\beta}$, respectively. It was reported that due to the persistent divergence of iterative numerical methods, the solution to the aforementioned nonlinear equation can not be obtained unless the sample size is quite large ($n > 400$, say) [8]. Fortunately, in the case of speech signal processing, the condition, $n > 400$ on the sample size is usually satisfied, and moreover, for the initial estimates for the parameters, we can start the algorithm with the values that specify the Laplacian or Gamma pdf, which is a pretty good approximation to the distribution of speech signal in various domains [3]-[5].Experimental results have shown that the large sample size and the reasonable initial estimates yield a satisfactory convergence property of the gradient ascent algorithm.

3. On-line ML algorithm for parameter estimation

In this section, we propose an on-line algorithm based on a forgetting scheme which emphasizes the data incoming most recently. On-line algorithm is fairly effective to deal with the input whose statistical property evolves as time flows. Although we model only the distribution of the clean speech signal in this paper, the on-line nature will make this algorithm even more powerful for characterizing the noisy speech signals, in which the statistical property of the added noise or the signal-to-noise ratio changes in a non-stationary manner.

To estimate the required parameters, only three statistics should be computed over the given data, $\frac{1}{N} \sum_{i=1}^{N} |x_i|^{\hat{\gamma}}$, $\frac{1}{N} \sum_{i=1}^{N} \log |x_i|^{\hat{\gamma}}$, and $\frac{1}{N} \sum_{i=1}^{N} |x_i|^{\hat{\gamma}} \log |x_i|^{\hat{\gamma}}$. For the implementation of an on-line algorithm, these statistics are modified to incorporate a forgetting factor λ:

$$\begin{aligned} S_1(n) \; &= \; (1-\lambda) S_1(n-1) + \lambda |x_n|^{\hat{\gamma}(n)} \\ S_2(n) \; &= \; (1-\lambda) S_2(n-1) + \lambda \log |x_n|^{\hat{\gamma}(n)} \\ S_3(n) \; &= \; (1-\lambda) S_3(n-1) + \lambda |x_n|^{\hat{\gamma}(n)} \log |x_n|^{\hat{\gamma}(n)}. \end{aligned} \quad (7)$$

In our experiments, the initial value for $\hat{\gamma}$ is set to 1, which specifies the Laplacian or Gamma pdf. Once $\hat{\gamma}$ is given, we can obtain $\hat{\eta}$ and $\hat{\beta}$ from (3) and (4) such that

$$\psi_0(\hat{\eta}(n)) - \log \hat{\eta}(n) \; = \; S_2(n) - \log S_1(n) \quad (8)$$

$$\hat{\beta}(n) \; = \; \frac{\hat{\eta}(n)}{S_1(n)} \quad (9)$$

by taking the forgetting scheme into consideration. Since $\psi_0(z) - \log z$ is a monotonically increasing function of z, the value of $\hat{\eta}$ can be uniquely determined if the solution exists. The value of $\hat{\gamma}$ is updated at each time based on the gradient ascent approach given as follows:

$$\hat{\gamma}(n+1) = \hat{\gamma}(n) + \mu \phi(\hat{\gamma}(n), \hat{\eta}(n), \mathbf{x}) \quad (10)$$

where μ is a learning rate and $\phi(\hat{\gamma}(n), \hat{\eta}(n), \mathbf{x})$ is an on-line version of the gradient of the 'average' log-likelihood function with respect to γ. The 'average' log-likelihood function is given as (2) divided by N, and its gradient with respect to γ equals to the left-hand side of (5). Using (3), (4), (5) and (7), the on-line version of the gradient is defined by

$$\phi(\hat{\gamma}(n), \hat{\eta}(n), \mathbf{x}) = \frac{1}{\hat{\eta}(n)} + S_2(n) - \frac{S_3(n)}{S_1(n)}. \quad (11)$$

As we can see, the estimation procedure is not computationally expensive if we store the values of the function $\psi_0(z) - \log z$ or the inverse of it on a table.

Note that for one new input, parameter updating can be performed many times, say, M times. Increasing the number of iterations can provide better estimation, but also increases computations needed for estimation.

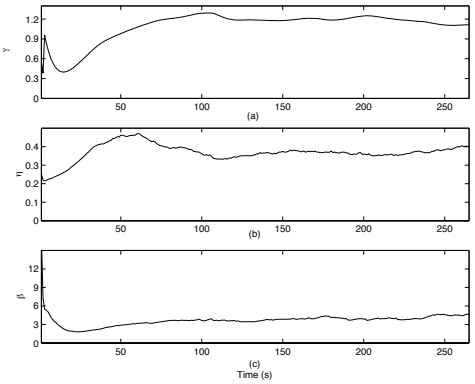


Figure 1: *Evolution of the parameters of GΓD for the real part of the 6th DFT coefficient*

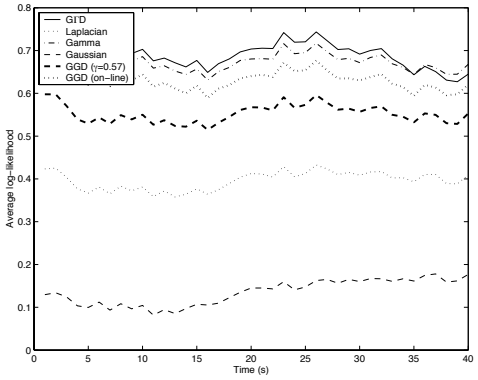


Figure 2: *The 'average' log-likelihoods evaluated for the four parametric pdf's*

4. Experimental results

In this section, we apply the proposed model to characterize the distribution of speech spectra. Since a single frequency component of a spectrum consists of both the real and imaginary parts, GΓD needs to be extended to the complex version to incorporate the two separate parts simultaneously. As in most of the former techniques, we assume that the real and imaginary parts are statistically independent with each other. Each part is assumed to follow GΓD and the parameters of GΓD's for the real and imaginary parts are also treated independently.

In our experiments, speech data spoken by 4 male and 4 female speakers were sampled at 8000 Hz and used to investigate each parametric model. Active speech frames were extracted and then analyzed for the ML estimation and goodness-of-fit tests. Discrete Fourier transform (DFT) of 128-points block size was applied at 10 ms frame rate after windowing the speech signal with a trapezoidal window. Due to the space limitation, we show here only the test results for the real part of the 6*th* DFT coefficient, which corresponds to about 375 Hz. However, we have observed that a different choice of the real/imaginary part or different frequency components yielded similar results. The parameters used in the experiments were $\lambda = 0.0004$, $\mu = 0.0001$, and $M = 11$.

Fig. 1 illustrates the evolution of the parameter estimates of GΓD. The result shows a good convergence property of the proposed algorithm which can track the real parameter values smoothly. Because of the on-line nature of the algorithm, the curve does not converge to a single fixed value but evolved slowly even after getting out of the initial fluctuation.

Three tests were performed to measure the goodness-of-fit of several parametric models including the Laplacian, Gamma, Gaussian pdf's and GGD as well as GΓD. For GGD, which is the special case of GΓD with $\eta = 1/\gamma$, the on-line version which was obtained by similar approach with GΓD and the fixed version with $\gamma = 0.57$, which was selected to fit the empirical CV as in the moment estimator case, are used for the comparison. First, using (2) and (7) the on-line version of the 'average' log-likelihood function, which is given by

$$J(\mathbf{x};\eta,\beta,\gamma) = \log\frac{\gamma\beta^{\eta}}{2\Gamma(\eta)} + (\eta - \frac{1}{\gamma})S_2 - \beta S_1, \qquad (12)$$

was evaluated for each distribution. Fig. 2 shows the traces of the 'average' log-likelihood values for the six pdf's. Since the criterion according to which the parameters were estimated was the ML principle, GΓD showed the uniformly best performance in terms of the log-likelihood value.

Secondly, the CV [7], which is defined as $CV(\mathbf{x}) = \sqrt{E[\mathbf{x}^2]}/E[|\mathbf{x}|]$, was calculated for each of the six parametric pdf's and compared with that computed from the empirical input data. The CV of GΓD can be obtained in the following closed form:

$$CV(\mathbf{x})_{\mathbf{G\Gamma D}} = \frac{\sqrt{\Gamma(\eta)\Gamma(\eta+\frac{2}{\gamma})}}{\Gamma(\eta+\frac{1}{\gamma})}. \qquad (13)$$

CV's of the other five parametric distributions can be obtained by substituting appropriate η and γ values in (13), since all the five distributions are special cases of GΓD. The empirical CV's were calculated by using the same forgetting factor as in (7). The results are plotted in Fig. 3 where we can see not only that GΓD outperformed the other distributions but also that GΓD could track the statistical property of the empirical distribution quite closely.

Lastly, the KS test [3] was carried out for each 2 s time frame where the overlap between successive frames was 1 s. The smaller the KS statistic is, the better the hypothesized model fits to the empirical distribution. The KS test statistic for each distribution is shown in Fig. 4 where we can once again discover the superiority of GΓD.

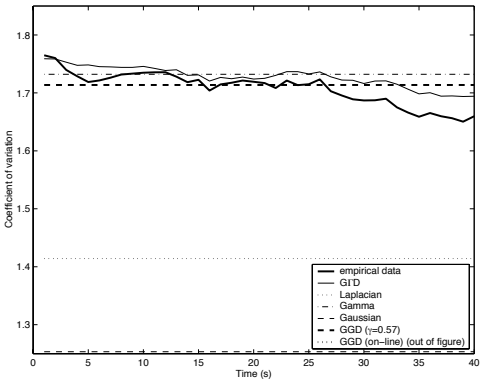


Figure 3: *The coefficient of variations (CV) evaluated for the four parametric pdf's*

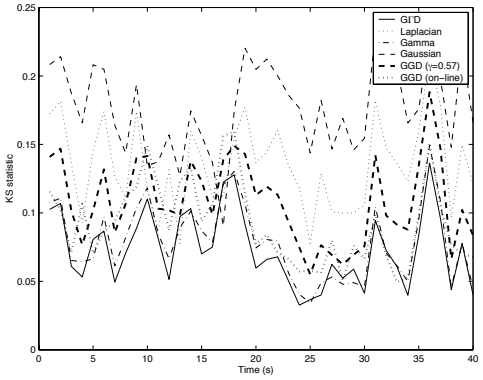


Figure 4: *The Kolmogorov-Smirnov test statistics evaluated for the four parametric pdf's*

The average results are summarized in Table 1. As seen from the results, GΓD outperformed the other parametric distributions in all the three measures of fit, the likelihood, CV, and KS test. We can consider that the gradient ascent algorithm behaves well despite the small number of iteration, and the approximation of the statistics in (7) is acceptable. The results also imply that the parameters of GΓD estimated according to the ML criterion not only produce a high likelihood, but also show a better performance in terms of other evaluation measures such as the CV, and KS test.

5. Conclusions

We have proposed a new statistical model, the two-sided generalized gamma distribution for the speech signal. An on-line algorithm for learning the parameters of GΓD according to the ML principle has been presented. The performance of GΓD has been found superior to other widely used pdf's through a number of goodness-of-fit tests. The computationally inexpensive nature of this algorithm may provide a good applicability to many speech processing areas such as VAD and speech enhancement.

Table 1: *The averages of the goodness-of-fit test results*

	Average of the 'average' log -likelihood	Root mean square error of the CV	Average KS statistics
GΓD	0.737988	0.036701	0.084848
Laplacian	0.441006	0.295524	0.143529
Gamma	0.713358	0.055355	0.086801
Gaussian	0.180821	0.454999	0.185474
GGD ($\gamma = 0.57$)	0.596978	0.049330	0.115173
GGD (online)	0.672689	0.661389	0.098040

6. References

[1] Paez, M. D. and Glisson, T.H., "Minimum-mean squared-error quantization in speech PCM and DPCM," IEEE Trans. Commun., vol. COM-20, pp. 225-230, Apr. 1972.

[2] Huang, J. and Zhao, Y., "A DCT-based fast signal subspace technique for robust speech recognition," IEEE Trans. Speech Audio Processing, vol. 8, pp. 747-751, Nov. 2000.

[3] Chang, J. -H., Shin, J. W. and Kim, N. S., "Likelihood ratio test with complex Laplacian model for voice activity detection," Proc. Eurospeech, Geneva, Switzerland, pp. 1065-1068, Aug. 2003.

[4] Martin, R., "Speech enhancement using short time spectral estimation with Gamma distributed priors," Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, Orlando, FL, USA, vol. 1, pp. I-253 - I-256, May 2002.

[5] Gazor, S. and Zhang, W., "Speech probability distribution," IEEE Signal Processing Letters, vol. 10, no. 7, pp. 204-207, Jul. 2003.

[6] Stacy, E. W., "A generalization of the gamma distribution", Annals Mathematical Statistics, vol. 33, pp. 1187-1192, 1962.

[7] Cohen, A. C. and Whitten, B. J., Parameter Estimation in Reliability and Life Span Models, 1988; New York: Marcel Dekker. 1988.

[8] Wingo, D. R., "Computing maximum-likelihood parameter estimates of the generalized gamma distribution by numerical root isolation," IEEE Trans. Reliability, vol. R-36, no. 5, pp. 586-590, Dec. 1987.