SPEECH MODELLING BASED ON GENERALIZED GAUSSIAN PROBABILITY DENSITY FUNCTIONS

Kostas Kokkinakis and Asoke K. Nandi

Signal Processing and Communications Group,
Department of Electrical Engineering and Electronics,
The University of Liverpool, Brownlow Hill, Liverpool, L69 3GJ, U.K.
{kokkinak,a.nandi}@liv.ac.uk

ABSTRACT

A number of commonly used methods for estimating the exponent parameter of a generalized Gaussian density (GGD) are reviewed, described and compared. More importantly, focusing on the family of entropy matching estimators (EMEs), a novel entropic expression with respect to higher-order moments of the modelled data is proposed. This yields an elegant generalized entropy matching estimator (G-EME). Comparative experimental results illustrate the high accuracy of the proposed estimator, for both light- and heavy-tailed distributions, as well as speech data.

1. INTRODUCTION

Accurately modelling the unknown probability density functions (PDFs) of data encountered in practical applications, can play an important role towards designing more efficient modern signal processing systems. To this end, the parametric family of densities stemming from the generalized Gaussian density (GGD) model, are known to approximate successfully a large number of different signals, encompassing a wide range of statistical distributions. In the context of image coding, GGD source modelling has been used extensively to approximate the distributions associated with discrete cosine transform (DCT) subband coefficients of natural images [1]–[2] and [6], while its use has also been common in video coding systems, as shown in [8] and [10]. Other useful applications of the GGD include, speech recognition and enhancement, speech modelling [5] and more interestingly as suggested in [7], blind source separation (BSS), where such a model has been shown to provide, in most cases, a fairly accurate representation of the unknown source distributions. Confronted with a set of unknown distributions modelled as generalized Gaussian densities, this contribution focuses on the estimation of the exponent parameter of the fitted GGD model, and hence on the complete statistical characterization of the given data. The main objective of our investigation, is the application of such model to speech and ultimately the estimation of first-order (univariate) probability density functions, which can fully approximate such data. In this context, past extensive experimental studies carried out in [3] and later in [9], have provided strong evidence that, in general, speech signals exhibit either Laplacian or generalized Gamma densities, respectively.

This work is supported by the Engineering and Physical Sciences Research Council of the U.K. and the University of Liverpool.

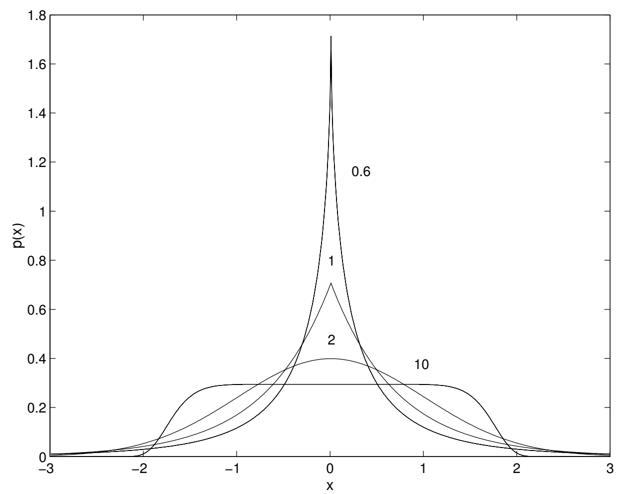

Fig. 1. The probability density of the generalized Gaussian model plotted for different values of the shape parameter $\nu = 0.6, 1, 2$ and 10. All distributions are normalized to unit variance ($\sigma^2 = 1$) and have zero mean.

2. GENERALIZED GAUSSIAN PDF ESTIMATION

Normally, a large number of symmetric distributions may be sufficiently approximated by employing the *generalized Gaussian density* (GGD) model. Thus, for a zero mean ($\mu = 0$) signal $x \in \mathbb{R}$, the PDF of a generalized Gaussian distribution with standard deviation σ is defined as:

$$p_x(x|\nu,\sigma) = \left[\frac{\nu \cdot \mathrm{A}(\nu,\sigma)}{2 \cdot \Gamma(1/\nu)}\right] \cdot \exp\left(-\left[\mathrm{A}(\nu,\sigma) \cdot |x|\right]^{\nu}\right) \tag{1}$$

in which

$$\mathrm{A}(\nu,\sigma) = \sigma^{-1}\sqrt{\frac{\Gamma\left(\frac{3}{\nu}\right)}{\Gamma\left(\frac{1}{\nu}\right)}} \tag{2}$$

where $\sigma > 0$ and $\Gamma(\cdot)$ defines the complete Gamma function given by:

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx, \quad z > 0 \tag{3}$$

$\mathrm{A}(\nu,\sigma)$ is a generalized measure of the variance and defines the *dispersion* or *scale* of the distribution, while parameter ν describes the exponential rate of decay and, in general, the *shape* or *skewness* of the distribution $p_x(x|\nu,\sigma)$. Well-known special cases of the GGD function include a Laplacian distribution ($\nu = 1$) and a standard Gaussian or normal distribution ($\nu = 2$). In effect, smaller values of the shape parameter ν correspond to heavier tails and therefore to more peaked distributions. In the limiting cases, when $\nu \to +\infty$, $p_x(x|\nu,\sigma)$ converges to a uniform distribution, whereas for $\nu \to 0+$, (1) approaches an impulse function. The shape of the GGD for different values of the exponent parameter ν is shown in Fig. 1.

2.1. Moment Matching Estimator (MME)

For all values of $\mathrm{A}(\nu,\sigma)$ and ν, the rth-order absolute central moment for a generalized Gaussian distributed signal is given by:

$$E\left[|X|^r\right] = \int_{-\infty}^{+\infty} |x|^r \, p_x(x|\nu,\sigma)\, dx \tag{4}$$

where $E[\,\cdot\,]$ represents the expectation operator. Substituting (1) into (4), we can further show that the rth-order moments are essentially defined as:

$$m_r = E\left[|X|^r\right] = \mathrm{A}^{-r}(\nu,\sigma)\cdot\left[\frac{\Gamma\left(\frac{r+1}{\nu}\right)}{\Gamma\left(\frac{1}{\nu}\right)}\right], \quad \nu > 0 \tag{5}$$

which, for example in the case of $r = 2$, yields the following:

$$m_2 = E\left[|X|^2\right] = \frac{1}{\mathrm{A}^2(\nu,\sigma)}\cdot\left[\frac{\Gamma\left(\frac{3}{\nu}\right)}{\Gamma\left(\frac{1}{\nu}\right)}\right] \tag{6}$$

Mallat first obtained a complete statistical description for GGD modelled distributions, by simply matching the underlying moments of the data set with those of the assumed distribution [8]. Combining the ratio of the square of the first absolute moment m_1^2 to the second-order moment m_2, was shown to produce a consistent *moment matching estimator* (MME) and furthermore an analytic expression for the unknown shape parameter ν. In this paper, as in [10], the inverse of Mallat's ratio, generally known as the *generalized Gaussian ratio function*, is used instead, which in fact is also a monotonically increasing function of ν:

$$\mathcal{M}(\nu) = \frac{m_2}{m_1^2} = \frac{\Gamma\left(\frac{3}{\nu}\right)\cdot\Gamma\left(\frac{1}{\nu}\right)}{\Gamma^2\left(\frac{2}{\nu}\right)} \tag{7}$$

further yielding the shape parameter estimate $\hat{\nu}$:

$$\hat{\nu}_{\mathcal{M}} \triangleq \mathcal{M}^{-1}\left(\frac{m_2}{m_1^2}\right) \tag{8}$$

A numerical solution for $\hat{\nu}$ and hence the MME defined in (7) is obtained by resorting to basic linear interpolation. We first sample $\mathcal{M}(\nu)$ with an accurate enough step (evidence shows that 10^{-3} ensures enough precision) and then invert with the aid of a look-up table whose entries are the corresponding moment ratios for all possible values of $\hat{\nu}$.

2.2. Entropy Matching Estimator (EME)

Another class of estimators, associated with entropy, has been recently seen as a promising alternative for estimating the shape factor of a generalized Gaussian [1]. The *entropy matching estimator* (EME) relies on matching the entropy of the GGD modelled distribution with that of a set of empirical data. Given a continuous distribution $p_x(x|\nu,\sigma)$ obeying the GGD model, its differential entropy is defined as [1]–[2]:

$$\begin{aligned} H_{\mathrm{GGD}}(\nu,\sigma) &= -\int_{-\infty}^{+\infty} p_x(x|\nu,\sigma)\cdot\log_2\left(p_x(x|\nu,\sigma)\right)\, dx \\ &= -\log_2\left[\frac{\nu\cdot\mathrm{A}(\nu,\sigma)}{2\cdot\Gamma(1/\nu)}\right] + \frac{1}{\nu\ln 2} \end{aligned} \tag{9}$$

Associating the above with the entropy $H(X)$ of the data obtained at the output of an optimum entropy constrained uniform threshold quantizer (UTQ) with step size Δ, yields $H(X) = H_{\mathrm{GGD}} - \log_2\Delta$ which after replacing H_{GGD} with (9) and substituting $\mathrm{A}(\nu,\sigma)$ for (2) leads to:

$$H(X) - \frac{1}{2}\log_2\frac{m_2}{\Delta^2} = \log_2\left[\frac{2\cdot\Gamma(1/\nu)^{3/2}}{\nu\cdot\Gamma(3/\nu)^{1/2}}\right] + \frac{1}{\nu\ln 2} \tag{10}$$

with the right hand side depending solely on the shape factor ν. A possible estimator for the assumed distribution having entropy $H(X)$ is hence realized by the monotonically increasing function:

$$\mathcal{H}(\nu) = H(X) - \frac{1}{2}\log_2\frac{m_2}{\Delta^2} \tag{11}$$

Employing higher-order moments of the data, we can formulate a rather more general expression for the entropy matching estimator (EME) defined above. Based on the unique relation between the second- and the rth-order absolute moments we may write:

$$m_2 = E\left[|X|^r\right]^{\frac{2}{r}}\cdot\left[\frac{\Gamma\left(\frac{3}{\nu}\right)\cdot\Gamma\left(\frac{1}{\nu}\right)^{\frac{(2-r)}{r}}}{\Gamma\left(\frac{r+1}{\nu}\right)^{\frac{2}{r}}}\right] \tag{12}$$

which when substituted in (10) yields a generalized entropic function with respect to every rth-order moment m_r:

$$H(X) - \frac{1}{r}\log_2\frac{m_r}{\Delta^r} = \log_2\left[\frac{2\cdot\Gamma\left(\frac{1}{\nu}\right)^{\frac{r+1}{r}}}{\nu\cdot\Gamma\left(\frac{r+1}{\nu}\right)^{\frac{1}{r}}}\right] + \frac{1}{\nu\ln 2} \tag{13}$$

from which it is now straightforward to derive the *generalized entropy matching estimator* (G-EME):

$$\mathcal{H}_r(\nu) = H(X) - \frac{1}{r}\log_2\frac{m_r}{\Delta^r} \tag{14}$$

which also includes (11) as a special case for $r = 2$ and further provides the following shape parameter estimate:

$$\hat{\nu}_{\mathcal{H}_r} \triangleq \mathcal{H}_r^{-1}\left(H(\hat{X}) - \frac{1}{r}\log_2\frac{\hat{m}_r}{\Delta^r}\right) \tag{15}$$

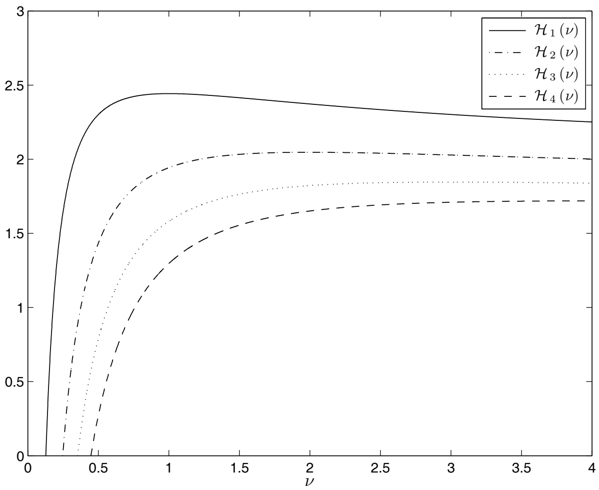

Fig. 2. Generalized rth-order entropic functions $\mathcal{H}_1(\nu)$, $\mathcal{H}_2(\nu)$, $\mathcal{H}_3(\nu)$ and $\mathcal{H}_4(\nu)$ plotted against the GGD shape parameter $\nu \in [0, 4]$.

where, $H(\hat{X})$ and $\hat{m}_r = E\left[|\hat{X}|^r\right]$, correspond to the entropy and the rth-order absolute moment respectively, both evaluated from actual (source) data. Fig. 2 depicts the standard generalized rth-order entropic functions $\mathcal{H}_r(\nu)$ plotted for several values of ν in the interval $[0, 4]$. For $0 < \nu \leq 1$, $\mathcal{H}_1(\nu)$ is increasing monotonically arriving at its theoretical maximum when $\nu = 1$. On the other hand, second- and third-order entropic functions $\mathcal{H}_2(\nu)$ and $\mathcal{H}_3(\nu)$ attain their maximum values for $\nu = 2$ and $\nu = 3$ instead, where $\mathcal{H}_2(2) = 2.0471$[1] and $\mathcal{H}_3(3) = 1.8459$, respectively. Worth noting is also $\mathcal{H}_4(\nu)$ with a maximum at $\mathcal{H}_4(4) = 1.7189$.

2.3. Maximum Likelihood Estimator (MLE)

In general, for a sequence of data $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ of sample size N, taken from a GGD with density $p_{x_i}(x_i|\nu, \sigma)$, the *maximum likelihood* (ML) estimate is uniquely defined by the values of parameters ν and σ that maximize the likelihood function:

$$L(\mathbf{x}|\nu, \sigma) = \log \prod_{i=1}^{N} p_{x_i}(x_i|\nu, \sigma) \tag{16}$$

Assuming now that mean and variance are known, the ML solution for the unknown shape parameter is found by setting the partial derivative of (16) with respect to ν equal to zero yielding the likelihood equation as in [6]:

$$1 + \frac{\psi(1/\nu)}{\nu} + \frac{\log\left[\frac{\nu}{N} \cdot \sum_{i=1}^{N} |x_i|^{\nu}\right]}{\nu} - \frac{\sum_{i=1}^{N} |x_i|^{\nu} \, \log|x_i|}{\sum_{i=1}^{N} |x_i|^{\nu}} = 0 \tag{17}$$

which only holds true for $\nu > 0$ and where $\psi(\cdot) = \Gamma'(\cdot)/\Gamma(\cdot)$ is the digamma function defined as:

$$\psi(z) = -\gamma + \int_0^1 (1 - x^{z-1})(1 - x)^{-1} dx, \quad z > 0 \tag{18}$$

[1] As one would expect $\mathcal{H}_2(\nu)$ attains its maximum at $\nu = 2$, corresponding to a Gaussian PDF, since from theory this should yield the largest entropy among all distributions of the same variance.

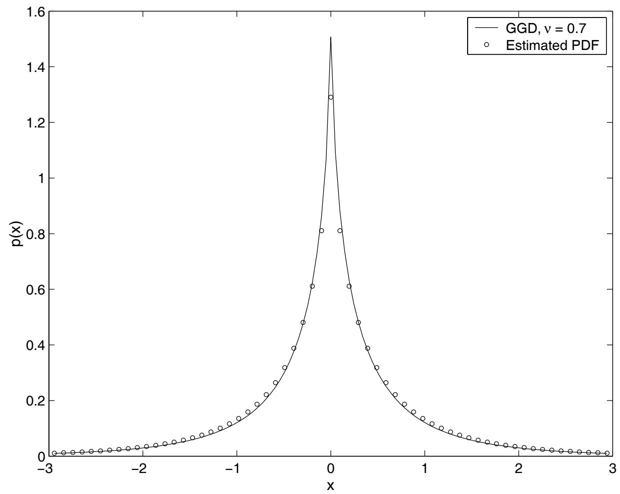

Fig. 3. The probability density function of the generalized Gaussian model ($\nu = 0.7$) fitted against the actual distribution of the given data.

where γ denotes the Euler constant. Solving (17) yields the ML estimate of the shape parameter $\hat{\nu}_L$, while differentiating (16) with respect to the standard deviation σ for known ν can also provide an ML estimate for the variance of the distribution. In our implementation, the starting value for the ML estimate $\hat{\nu}_L$ is calculated using the method of bisection instead, which despite being slow, is known to always converge to a unique solution for the root of (17).

3. EXPERIMENTAL RESULTS

3.1. GGD Modelling

This section attempts to measure the efficiency of the generalized entropic estimator $\mathcal{H}_r(\nu)$, by fitting a GGD model-based distribution with the PDF of a set of given data. We first generate 100 different zero mean and unit variance generalized Gaussian[2] i.i.d. sequences for a single value of the shape parameter ($\nu = 0.7$) and a relative small number of samples ($N = 10^3$). Based on the generalized entropic estimator $\mathcal{H}_3(\nu)$, obtained from (14) for $r = 3$, we then calculate the underlying shape parameter of the distribution for all 100 sequences. The final value obtained after averaging over all sequences, yields around $\hat{\nu}_{\mathcal{H}_3} = 0.681$, which clearly demonstrates the extremely high accuracy of the entropy matching method. The estimated PDF from the actual data and the fitted GGD model are shown in Fig. 3. Note also that in this case, inversion of the estimating function is carried out using a look-up table of about 4 000 different values of ν in the range $[0, 4]$ with a step size of 10^{-3}, while the differential entropy associated with $\mathcal{H}_3(\nu)$, is estimated according to a histogram-based method.

3.2. Speech Modelling

The source material used here, is taken from the TIMIT speech corpus [4] and consists of a total of ten speech signals, where the

[2] To simulate a GGD with parameters ν, σ, we first generate Gamma distributed random variables $Z_i \sim \mathrm{G}(\mathrm{A}(\nu, \sigma)^{-\nu}, \nu^{-1})$ and then obtain the generalized Gaussian distributed variables $X_i \sim \mathrm{GG}(\nu, \sigma)$ using the transformation $X_i = Z^{1/\nu}$.

Frame $\rightarrow$	50 ms			100 ms			200 ms		
Speaker $\downarrow$	$\hat{\nu}_{\mathcal{M}}$	$\hat{\nu}_{\mathcal{H}_3}$	$\hat{\nu}_L$	$\hat{\nu}_{\mathcal{M}}$	$\hat{\nu}_{\mathcal{H}_3}$	$\hat{\nu}_L$	$\hat{\nu}_{\mathcal{M}}$	$\hat{\nu}_{\mathcal{H}_3}$	$\hat{\nu}_L$
Male 1	1.489	1.233	1.404	1.197	1.051	1.118	0.870	0.815	0.789
Male 2	1.487	1.128	1.566	1.188	1.010	1.142	0.898	0.850	0.810
Male 3	1.314	1.041	1.246	1.048	0.928	0.974	0.804	0.767	0.730
Male 4	1.292	1.014	1.186	1.049	0.891	0.949	0.779	0.735	0.696
Male 5	1.348	1.120	1.322	1.052	0.965	0.976	0.820	0.803	0.726
Male 6	1.653	1.204	1.604	1.299	1.086	1.224	0.974	0.911	0.888
Male 7	1.337	1.079	1.255	1.091	0.960	1.020	0.834	0.793	0.777
Male 8	1.413	1.097	1.362	1.114	0.983	1.020	0.906	0.840	0.776
Female 1	1.418	1.059	1.475	1.021	0.894	0.915	0.758	0.728	0.616
Female 2	1.733	1.165	1.598	1.374	1.078	1.334	1.031	0.927	0.879

Table 1. GGD exponent parameter (ν) estimated using MME, EME and MLE for ten different speech signals. Values obtained by averaging across different frames of speech data.

same sentence is uttered by 8 male and 2 female speakers. All speech data are sampled at 16 KHz and have a duration ranging from 2-4 s, since no special provision has been taken to re-adjust the incurring silences for each speaker. In order to capture the non-stationary characteristics of speech, the data from each speaker are processed as a series of independent short-time frames. A number of different frame lengths is chosen, ranging from 50-200 ms, while the overlapping between successive frames is set to 50% in all cases. The aforementioned estimators for the exponent of the GGD, are applied to each of the extracted frames of speech for all different speakers. Next, a single value for the exponent parameter uniquely characterizing the processed speech signal, is calculated by averaging across all individual values of $\hat{\nu}$ estimated at each frame. All values obtained are summarized in Table 1. Focusing on short time frames, i.e., on frames of 50 and 100 ms length, it is obvious that the estimated values for all speakers, predominantly point to the conclusion that, in fact, speech follows a Laplacian distribution ($\nu = 1$). This appears to be true for all estimators, yet it becomes more evident in the case of the G-EME (when $r = 3$) and especially for time frames of 100 ms. On the other hand, when the chosen frame length is increased, the estimated exponent parameter $\hat{\nu}$ rapidly decreases, clearly indicating that the PDF of speech is significantly more heavy-tailed ($\nu \leq 1$) and thus more likely to exhibit a near Gamma-like shape. The validity of these findings is strongly corroborated in [5], where an extensive analysis has, among others, revealed that the majority of short-data segments of speech ($\leq$ 100 ms) may, in general, be described most accurately by a Laplacian distribution, whereas longer ones ($\geq$ 200 ms), usually appear to exhibit significantly heavier-tailed densities.

4. CONCLUSIONS

We have derived an elegant generalized entropy matching estimator (G-EME), which we have applied on the estimation of the exponent parameter of a generalized Gaussian density (GGD). A number of experimental results, conducted with both generalized Gaussian sequences and speech data, have illustrated the high efficiency of the proposed entropy-based estimator even for small number of samples. Some interesting conclusions, have also been drawn for the distribution of speech signals, for which we have shown, that in most cases, the GGD model appears to be an accurate enough fit. In this context, however, our findings have also revealed that the optimal statistical approximation for the PDF of speech is only feasible under the interval at which the signal of interest remains widely stationary, thus still remaining a largely open issue.

5. REFERENCES

[1] B. Aiazzi, L. Alparone and S. Baronti, "Estimation Based on Entropy Matching for Generalized Gaussian PDF Modeling," *IEEE Signal Proc. Letters*, Vol. 6, No. 6, June 1999, pp. 138–140.

[2] K. A. Birney and T. R. Fisher, "On the Modeling of DCT and Subband Image Data for Compression," *IEEE Trans. Image Process.*, Vol. 4, No. 2, February 1995, pp. 186–193.

[3] W. B. Davenport, "An Experimental Study of Speech-Wave Probability Distributions," *J. Acoust. Soc. America*, Vol. 24, No. 4, July 1952, pp. 390–399.

[4] W. Fisher, G. Doddington and K. Goudie-Marshall, "The DARPA Speech Recognition Research Database: Specifications and Status," In *Proc. of the DARPA Speech Recog. Workshop*, 1986, pp. 93–99.

[5] S. Gazor and W. Zhang, "Speech Probability Distribution," *IEEE Signal Proc. Letters*, Vol. 10, No. 7, July 2003, pp. 204–207.

[6] R. L. Joshi and T. R. Fisher, "Comparison of Generalized Gaussian and Laplacian Modeling in DCT Image Coding," *IEEE Signal Proc. Letters*, Vol. 2, No. 5, May 1995, pp. 81–82.

[7] K. Kokkinakis and A. K. Nandi, "Optimal Blind Separation of Convolutive Audio Mixtures without Temporal Constraints," In *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Montréal, Canada, May 17–21, 2004, Vol. I, pp. 217–220.

[8] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. Pattern Anal. and Machine Intell.*, Vol. 11, No. 7, July 1989, pp. 674–693.

[9] D. L. Richards, "Statistical Properties of Speech Signals," *Proc. of the IEE*, Vol. 111, No. 5, May 1964, pp. 941–949.

[10] K. Sharifi and A. Leon-Garcia, "Estimation of Shape Parameter for Generalized Gaussian Distributions in Subband Decompositions of Video," *IEEE Trans. Circuits and Syst. for Video Technol.*, Vol. 5, No. 1, February 1995, pp. 52–56.