ACOUSTIC MODELING BASED ON A GENERALIZED
LAPLACIAN DISTRIBUTION

Atsushi NAKAMURA and Tomoko MATSUI
ATR Interpreting Telecommunications Research Laboratories
2-2 Hikaridai Seika-Cho Soraku-Gun Kyoto 619-0288, JAPAN
{atsushi,tmatsui}@itl.atr.co.jp

ABSTRACT
In speech recognition, acoustic modeling mismatches for the shapes of distributions are unavoidable when we use fixed-shape base distributions, such as the Gaussian. This paper proposes a new base distribution called the "generalized Laplacian distribution (GLD)", which offers more flexibility in terms of the distribution shape. The effectiveness of the GLD-based modeling is proven through speech recognition experiments.

Keywords: acoustic model, non-Gaussian modeling, shape of distribution, kurtosis

1. INTRODUCTION
In acoustic modeling based on continuous distributions for speech recognition, Gaussians or mixtures of Gaussians are widely used. Although the general advantage of Gaussians to model unseen populations is theoretically obvious, we still do not have a basis showing that the Gaussians are the best solution for concrete problems, to model acoustic features for speech recognition. In particular, mismatches have been unavoidable in modeling the shapes of distributions assuming fixed-shape distributions, such as Gaussian or Laplacian distributions [1], and this has caused inaccurate evaluations of recognition hypotheses. If infinite number of components were available in the Gaussian-mixture type modeling, we could have modeled any populations with arbitrary shapes of distributions. This is, however, infeasible practically. In this paper, we propose a new base distribution called the "generalized Laplacian distribution (GLD)", which offers more flexibility in terms of the shape of the distribution.

This paper is organized as follows. In section 2, a problem caused by mismatches in the distribution shape is discussed. Section 3 presents the formulations for generalizing a subset of an exponential family of distributions and deriving the probability density function (pdf) of the GLD. An algorithm to estimate distribution parameters in GLD is given in section 4. Section 5 shows experimental results that prove the effectiveness of GLD-based acoustic modeling.

2. MISMATCH IN MODELING THE DISTRIBUTION SHAPE
The pdfs and parameters of Laplacian and Gaussian distributions are summarized in Table 1. We suppose a population that completely follows Laplacian distribution. When we model the population using Gaussian distribution, the resultant model involves a mismatch in terms of the shape of the distribution as shown in Figure 1. This mismatch is more obvious on a logarithmic scale.

Table 1. Laplacian and Gaussian distributions.

<table>
<thead>
<tr>
<th></th>
<th>Laplacian</th>
<th>Gaussian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability density</td>
<td>( \frac{1}{2b} \exp \left( \frac{</td>
<td>x-\mu</td>
</tr>
<tr>
<td>Maximum likelihood</td>
<td>( \hat{\mu} = \text{median} { x_i } )</td>
<td>( \hat{\mu} = \frac{\sum x_i}{N} )</td>
</tr>
<tr>
<td>estimator for location parameter</td>
<td>( \hat{\sigma} = \frac{\sum [x_i - \hat{\mu}]^2}{N} )</td>
<td>( \hat{\sigma} = \frac{\sum (x_i - \hat{\mu})^2}{N} )</td>
</tr>
</tbody>
</table>

Linear Logarithmic

Figure 1. Mismatch in modeling the shape of the distribution.
This mismatch is unavoidable when modeling under the assumption of a fixed distribution-shape. This may cause inaccurate evaluations for recognition hypotheses. To cope with this problem, we introduce a new modeling parameter concerning the shape of the distributions and aim to reduce the mismatch in acoustic modeling.

3. FORMULATIONS

By integrating Laplacian and Gaussian distributions, we will produce a distribution with a new parameter for its shape. Table 1 shows two common characteristics of Laplacian and Gaussian distributions.

(Characteristic-1) Each distribution belongs to an exponential family of distributions, and the logarithm of the pdf makes a simple \( N \)-order function (\( N = 1 \) or 2).

(Characteristic-2) The ML estimator for the scale parameter of each distribution is given by the \( N \)-powered mean over absolute errors between location parameters and samples. (\( N = 1 \) or 2).

These characteristics should be inherited, as much as possible, to the new distribution for the simple calculation of the likelihood, and for the easy implementation of the parameter estimation algorithm.

We suppose the following function that formally includes the Laplacian and Gaussian density functions.

\[
f(x) = \alpha \exp\left(\frac{-(|x-\mu|)^\rho}{\rho \beta^\rho}\right),
\]

(1)

Here, \( \rho \) is a positive real number. As the logarithm of \( f(x) \) makes a \( \rho \)-order function, the distribution represented by the pdf \( f(x) \) follows a characteristics given by expanding Characteristic-1. In order to meet a requirement for \( f(x) \) to be a pdf, both sides of the infinite integral of \( f(x) \) must equal 1:

\[
\int_{-\infty}^{\infty} \alpha \exp\left(\frac{-(|x-\mu|)^\rho}{\rho \beta^\rho}\right) dx = 1
\]

(2)

\[
\alpha = \frac{1}{\int_{-\infty}^{\infty} \exp\left(\frac{-(|x-\mu|)^\rho}{\rho \beta^\rho}\right) dx}
\]

(3)

The denominator in (3) can be calculated as,

\[
\int_{-\infty}^{\infty} \exp\left(\frac{-(|x-\mu|)^\rho}{\rho \beta^\rho}\right) dx = \frac{2\beta}{\rho^{\frac{1}{\rho}}} \int_{0}^{\infty} t^{\frac{1}{\rho}} e^{-t} dt
\]

\[
= \frac{2\beta \Gamma(1/\rho)}{\rho^{\frac{1}{\rho}}}.
\]

(4)

Here, \( \Gamma(*) \) is the Euler’s Gamma function. Accordingly, the pdf for the GLD is derived as follows:

\[
f(x) = \frac{\rho^{(\rho-1)/\rho}}{2\beta \Gamma(1/\rho)} \exp\left(\frac{-(|x-\mu|)^\rho}{\rho \beta^\rho}\right),
\]

(5)

where

\( \mu \): location parameter,
\( \beta \): scale parameter, and
\( \rho \): distribution-shape parameter.

We introduced a new parameter \( \rho \), concerned with the shape of the distribution, so that the GLD can properly model populations in a wider variety of distribution-shapes, including Gaussian (\( \rho = 2.0 \)), Laplacian (\( \rho = 1.0 \)), and other populations (Figure 2).

![Figure 2. Shapes of distributions with varying \( \rho \).](image)

We can easily expand this to a mixture density function for a multidimensional variable with a diagonal covariance matrix that has been widely used in conventional Gaussian-based modeling.

4. PARAMETER ESTIMATION ALGORITHMS

4.1 The Location And Scale Parameters

For the parameter estimation in acoustic modeling, a type of the Expectation-Maximization (EM) algorithm based on the maximum likelihood (ML) estimation has been widely used. However, it is not easy to give the exact ML estimators for the GLD parameters, \( \mu \), \( \beta \) and \( \rho \), with a conventional way, e.g., to solve the likelihood equations, because the equation for the location parameter \( \mu \) cannot feasibly be solved. To solve the \( \mu \) estimation, we can employ a type of gradient descend al-
gorithm and numerically solve the equation. However, this also has problems in the computational cost and stability in the convergence. In this paper, we assume that we can use the sample mean as a pseudo ML estimator for $m$. Although the estimation for $m$ does not give the exact ML unless $r$ equals 2, we can still expect a high gain in likelihood for several $r$ through the estimation.

When we have the ML estimator for the $\mu$, we can build a likelihood equation concerning the scale parameter $\beta$:

$$\frac{\partial}{\partial \beta} \log \prod \left( \frac{\rho^{\rho-1/r}}{2\Gamma(1/r)} \exp \left( \frac{|x_i - \mu|^\rho}{\rho \beta^r} \right) \right) = 0 \tag{6}$$

If we solve this equation, the ML estimator for the scale parameter $\beta$ is given as follows:

$$\hat{\beta} = \left( \frac{\sum (|x_i - \mu|^\rho)}{N} \right)^{1/\rho} \tag{7}$$

Formula (7) represents that the ML estimator for the scale parameter of the GLD is given by the $r$-powered mean over absolute errors between the location parameter and samples. This is a generalization of Characteristic-2.

### 4.2 The Distribution-Shape Parameter

The estimation of the distribution-shape parameter $\rho$ is the most important issue in the GLD-based modeling. A type of the EM algorithm for estimating a parameter of the distribution-shape were proposed in [6]. In this paper, we propose an estimation of $\rho$ by fitting kurtosis statistics, instead of the complicated equational method or the expensive gradient descend method.

The kurtosis given by the pdf of the GLD is:

$$\kappa(\rho) = \frac{\int x^4 \rho^{\rho-1/r} \exp \left( \frac{-|x|^\rho}{\rho \beta^r} \right) dx}{\left( \frac{\int x^2 \rho^{\rho-1/r} \exp \left( \frac{-|x|^\rho}{\rho \beta^r} \right) dx}{\rho \beta^r} \right)^2}$$

$$= \frac{\Gamma(1/r) \Gamma(5/r)}{\left( \Gamma(3/r) \right)^2} \tag{8}$$

In addition, the sample kurtosis is given by:

$$\kappa_{\text{sample}} = \frac{\sum (x - \mu)^4 / N}{\left( \sum (x - \mu)^2 / N \right)^2} \tag{9}$$

We can have the value of $\rho$ that fits (8) to (9) via the inverse function of $\kappa(\cdot)$ as follows:

$$\hat{\rho} = \kappa^{-1}(\kappa_{\text{sample}}) = \kappa^{-1}\left( \frac{\sum (x - \mu)^4 / N}{\left( \sum (x - \mu)^2 / N \right)^2} \right) \tag{10}$$

## 5. EXPERIMENTS

### 5.1 Overview

In order to prove the effectiveness of the GLD-based acoustic modeling, we carried out continuous speech recognition tests using a Japanese spontaneous speech database. First, we built a context-dependent Gaussian mixture HMM set as a baseline model using the ML-SSS algorithm [2]. Then, GLD mixture HMM sets were obtained by re-estimating the parameters in the baseline model. The major experimental conditions are summarized in Table 2.

<table>
<thead>
<tr>
<th>Table 2. Experimental conditions.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
</tr>
<tr>
<td>Acoustic model topology</td>
</tr>
<tr>
<td>Database</td>
</tr>
<tr>
<td>Search</td>
</tr>
</tbody>
</table>

In these experiments, the Euler’s Gamma function and the inverse function of $\kappa(\cdot)$ were implemented with the Lanczos’s formula and with a type of table look-up, respectively.

### 5.2 Experimental Results

#### 5.2.1 Distribution-Shape And Recognition Rate

The first experiments were carried out to examine how the recognition performance behaves with different $\rho$ values. Several values, around the initial value 2.0, were commonly set to all $\rho$ in the baseline model, then $\mu$ and $\beta$ were re-estimated with each $\rho$ values. Figure 3 shows the recognition performance with the varying $\rho$. When a common value was manually set to all distribution-shape parameters, we found that $\rho$ values smaller than 2.0 gave a better performance even though the model topology was designed based on the Gaussian distribution ($\rho = 2.0$). The best recognition performance was given at the point $\rho = 1.96$,
where the word error rate was reduced by 9.7% in comparison to the baseline model ($\rho = 2.0$).

![Figure 3. Recognition performance with varying $\rho$.](image)

### 5.2.2 Best-Fit Shape And Recognition Rate

We applied the above mentioned kurtosis-fitting estimation algorithm to the baseline model. Since the kurtosis-fitting does not guarantee anything on the gain in likelihood, we must watch the likelihood score at each step of the estimation. Figure 4 shows the estimation flow. First, the kurtosis fitting was applied, and all other parameters were re-estimated in the ML fashion. Then, we compared the likelihood score given by the estimation to the score before the kurtosis-fitting. If we had gain in the score as a result of the estimation, we went to another kurtosis-fitting estimation. Otherwise, we adopted the parameter values before the last kurtosis-fitting estimation. With this procedure, we would have different $\rho$ values by 1-dimensional pdf components. In practice, by considering the higher order statistics (kurtosis), we only estimated $\rho$ in the pdfs with larger occupation counts than a predetermined threshold. Furthermore, we constrained the $\rho$ values to be between 1.8 and 2.0 because the result shown in Figure 3 suggested that there was no advantage to drastically moving the $\rho$ value.

### Table 3. Improvement in error rate.

<table>
<thead>
<tr>
<th>Model</th>
<th>Error rate (Improvement)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>29.0%</td>
</tr>
<tr>
<td>Common-value</td>
<td>26.2% (9.7%)</td>
</tr>
<tr>
<td>Kurtosis-fitted</td>
<td>25.6% (11.7%)</td>
</tr>
</tbody>
</table>

Table 3 shows the result of speech recognition test using the best-fit values of $\rho$. We found a larger reduction (11.7%) in the word error rate than the common-value case. Thus, fitting the shape of the distributions proved to be effective.

### 6. CONCLUSION

A new base distribution called the generalized Laplacian distribution, which offers more flexibility in terms of the shape of the distribution, was proposed. Speech recognition experiments showed that we can improve the performance by fitting the shape of parameters by retraining from the Gaussian mixture baseline model. In the future, we are planning to choose the $\rho$ value that is suitable for the type of acoustic features. In addition, we are planning to study on a topological design of the acoustic model where the shape of distributions is taken into account.

### REFERENCES


