Signal Processing 114 (2015) 219–224

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Fast communication

Line spectral frequencies modeling by a mixture of von Mises–Fisher distributions

Zhanyu Ma [a,*], Jalil Taghia [b], W. Bastiaan Kleijn [c], Arne Leijon [d], Jun Guo [a]

[a] Pattern Recognition and Intelligent System Laboratory, Beijing University of Posts and Telecommunications, Beijing, China
[b] Neural Information Processing Group, Technische Universität Berlin, Germany
[c] School of Engineering and Computer Science, Victoria University of Wellington, New Zealand
[d] Communication Theory Laboratory, School of Engineering, KTH-Royal Institute of Technology, Stockholm, Sweden

ARTICLE INFO

Article history:
Received 23 September 2014
Received in revised form
4 February 2015
Accepted 16 February 2015
Available online 24 February 2015

Keywords:
Speech coding
Line spectral frequencies
Vector quantization
Von Mises–Fisher distribution
Mixture modeling

ABSTRACT

Efficient quantization of the linear predictive coding (LPC) parameters plays a key role in parametric speech coding. The line spectral frequency (LSF) representation of the LPC parameters has found its applications in speech model quantization. In practical implementations of vector quantization (VQ), probability density function optimized VQ has been shown to be more efficient than the VQ based on training data. In this paper, we present the LSF parameters by a unit vector form, which has directional characteristics. The underlying distribution of this unit vector variable is modeled by a von Mises–Fisher mixture model (VMM). An optimal inter-component bit allocation strategy is proposed based on high rate theory and a distortion-rate (D-R) relation is derived for the VMM based-VQ (VVQ). Experimental results show that the VVQ outperforms the recently introduced Dirichlet mixture model-based VQ and the conventional Gaussian mixture model-based VQ, in terms of modeling performance and D-R relation.

1. Introduction

Quantization of the line predictive coding (LPC) model is ubiquitously applied in speech coding [1,2]. The line spectral frequency (LSF) [3] presentation of the LPC model is most commonly used for quantization of the LSF [4,1], because of its relatively uniform spectral sensitivity [5]. The linear predictive polynomial can be expressed as the mean of a palindromic polynomial and an antipalindromic polynomial. The LSF representation of the LPC model consists simply of the location of the roots of the above-mentioned polynomials. Efficient quantization methods for the LSF parameters have been studied intensively in the literature (see e.g., [4,6–8]). Among these methods, the probability density function (PDF)-optimized vector quantization (VQ) scheme has been shown to be superior to those based on training data [6,7]. In PDF-optimized VQ, the underlying distribution of the LSF parameters is described by a statistical parametric model, such as a Gaussian mixture model (GMM) [6]. Once this model is obtained, the codebook can be either trained by using a sufficient amount of data (theoretically infinitely large) generated from the obtained model or calculated theoretically. Thus PDF-optimized VQ can prevent the codebook from overfitting to the training data, and hence the performance of VQ can be significantly improved [6,7].

Statistical modeling plays an important role in PDF-optimized VQ; hence, several studies have been conducted to seek an effective model to explicitly capture the statistical properties of the LSF parameters or its corresponding transformations. A frequently used method is the GMM-based VQ (GVQ), which models the LSF parameters' distribution with a GMM [6,7]. The LSFs represent one method to ensure stability of the all-pole synthesis filter. For a linear predictive model with order K, the LSFs are

* Correspondence author.
E-mail address: mazhanyu@bupt.edu.cn (Z. Ma).

http://dx.doi.org/10.1016/j.sigpro.2015.02.015
0165-1684/

interleaved on the unit circle [3] as $0=\omega_{q_0}<\omega_{p_1}<\omega_{q_1}<\cdots<\omega_{q_{K/2}}<\omega_{p_{(K/2)+1}}=\pi$. By recognizing the bounded property (all the LSF parameters are placed in the interval $(0,\pi)$ and $\text{LSF}_i<\text{LSF}_{i+1}$), Lindblom and Samuelsson [9] proposed a bounded GVQ scheme by truncating and renormalizing the standard Gaussian distribution. In [10], the LSF parameters were linearly scaled into the interval $(0,1)$. The authors introduced a beta mixture model (BMM)-based VQ scheme, which took into account the bounded support nature of the LSF parameters. As the LSF parameters are also strictly ordered, a Dirichlet mixture model (DMM-based VQ (DVQ) scheme was recently presented to explicitly utilize both the bounded and the ordering properties [11,8]. In the DVQ scheme, the LSF parameters were transformed linearly to the ΔLSF parameters [11]. Modeling the underlying distribution of the ΔLSF parameters with a DMM yields a better distortion-rate (D-R) relation than those obtained by modeling the LSF parameters with a GMM [7,12] and a BMM [11]. Hence, the practical quantization performance was also improved significantly [8,13].Previous studies suggest that transforming the LSF parameters into some other form and applying a suitable statistical model to efficiently describe the distribution can potentially benefit the quantization [8,10,11].

In this paper, we study the high rate D-R performance of the LSF parameter by using the recently proposed square-root ΔLSF (SRΔLSF) representation [14,15]. This representation is obtained by taking the positive square-root of the ΔLSF parameters. By concatenating a redundant element to the end of the SRΔLSF parameter vector, a unit vector that contains only positive elements is obtained. Geometrically, this unit vector has directional characteristics and is distributed on the hypersphere with the center at the origin. For such a unit vector, the von Mises–Fisher (vMF) distribution is a widely used statistical model to describe the underlying distribution [16]. One application domain of the vMF distribution is information retrieval where the cosine similarity is an effective measure of similarity for analyzing text documents [17]. Other application domains are bioinformatics (e.g., [14,15,17]) and collaborative filtering (e.g., [18]) in which the Pearson correlation coefficient serves as the similarity measure. More recently, Taghia et al. [14,15] proposed a text-independent speaker identification system based on modeling the underlying distribution of SRΔLSF parameters by a mixture of vMF distributions.

Here, we model the underlying distribution of the SRΔLSF parameters by a VMM and propose a VMM-based VQ (VVQ) scheme. The modeling performance obtained by VMM is superior to both those obtained by DMM and GMM. Based on the high rate quantization theory [20], the D-R relation can be analytically derived for a single vMF distribution with constrained entropy. Finally, the D-R performance for the overall VVQ is derived. Compared with the recently presented DVQ and the conventionally used GVQ, the VVQ shows convincing improvement for the D-R relation.

The remaining parts of this paper are organized as follows. In Section 2, different representations of the LSF parameters are introduced. We briefly review the vMF distribution and the corresponding parameter estimation methods in Section 3. A PDF-optimized VQ based on VMM is proposed in Section 4 and the experimental results are shown in Section 5. Finally, we draw some conclusions and discuss future work in Section 6.

2. LSF, ΔLSF, and SRΔLSF

2.1. Representations

The LSF parameters are widely used in speech coding. The LSF parameters with dimensionality K are defined as

$$\mathbf{s}=[s_1,s_2,\ldots,s_K]^{\mathrm{T}}, \tag{1}$$

which are interleaved on the unit circle [3].

By recognizing that the LSF parameters are in the interval $(0,\pi)$ and are strictly ordered, we used a particular representation of LSF parameters called ΔLSF [11] for the purpose of LSF quantization [8]. The ΔLSF parameters are shown by $\mathbf{v}$ and represented as [11]

$$\mathbf{v}=\varphi(\mathbf{s})=[s_1,s_2-s_1,\ldots,s_K-s_{K-1}]^{\mathrm{T}}. \tag{2}$$

Another representation of the LSF parameters was introduced in [14,15], which uses the square-root of the ΔLSF parameters. The K-dimensional SRΔLSF $\mathbf{x}$ parameter vector is then obtained as

$$\mathbf{x}=\phi(\mathbf{v})=\mathbf{v}^{1/2}=[\sqrt{v_1},\sqrt{v_2},\ldots,\sqrt{v_K}]^{\mathrm{T}}. \tag{3}$$

2.2. Likelihood and distortion transformation

We study the likelihood and distortion transformation between the SRΔLSF and the ΔLSF spaces in this section.

2.2.1. Likelihood transformation

Denote the PDFs of $\mathbf{v}$ and $\mathbf{x}$ as $g(\mathbf{v})$ and $f(\mathbf{x})$, respectively. By the principles of integration by substitution, the PDF transformation from $\mathbf{v}$ to $\mathbf{x}$ is described by the following relation:

$$f(\mathbf{x})=g\big(\varphi(\mathbf{x})\big)|\det\mathbf{A}(\mathbf{v})|. \tag{4}$$

In the above equation, $\mathbf{A}(\mathbf{v})$ is the Jacobian matrix of $\varphi(\mathbf{x})$ w.r.t. $\mathbf{x}$ as

$$A_{i,j}(\mathbf{v})=\begin{cases}\dfrac{\partial\varphi(\mathbf{x})_i}{\partial x_j}\Big|_{\mathbf{x}=\varphi^{-1}(\mathbf{v})}=2\sqrt{v_i} & i=j\\ 0 & i\neq j\end{cases}, \tag{5}$$

which can also be expressed in matrix form as

$$\mathbf{A}(\mathbf{v})=2\begin{bmatrix}\sqrt{v_1} & & & \mathbf{0}\\ & \sqrt{v_2} & & \\ & & \ddots & \\ \mathbf{0} & & & \sqrt{v_K}\end{bmatrix}. \tag{6}$$

Thus, the transformation between the PDFs is

$$f(\mathbf{x})=g(\mathbf{v})2^K\sqrt{\left(\prod_{k=1}^{K}v_k\right)}. \tag{7}$$