Signal Processing 93 (2013) 957–960

Contents lists available at SciVerse ScienceDirect

Signal Processing

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

Fast communication

Spectral envelope quantization based on conditional inter-frame prediction

Wonjin Jung [a], Moo Young Kim [a,*]

[a] Department of Information and Communication Engineering, Sejong University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 26 July 2012
Received in revised form
28 September 2012
Accepted 29 September 2012
Available online 11 October 2012

Keywords:
Quantization
Line spectral frequency (LSF)
Split-vector quantization (SVQ)
Predictive SVQ (PSVQ)
Auto-regressive (AR)
Conditional distribution

ABSTRACT

The human vocal tract system is commonly modeled by a linear predictive coding (LPC) filter whose coefficients are transformed into a line spectral frequency (LSF) vector for quantization. Predictive split-vector quantization (PSVQ) based on an auto-regressive model (AR-PSVQ), which exploits the inter-frame correlation of the LSF vectors, provides a better rate-distortion performance compared with quantization methods that only consider the intra-frame correlation. In the proposed conditional PSVQ (C-PSVQ), the conditional distribution of the current-frame LSF given the previous-frame LSF is taken into account. Compared with AR-PSVQ, C-PSVQ gains 1 bit in terms of average spectral distortion and 2 bits in terms of the number of outlier frames. Memory requirements and computational complexity of C-PSVQ are similar to those of AR-PSVQ.

1. Introduction

Speech compression is an integral component of voice storage devices such as voice recorders, text-to-speech systems (TTS), and emergency call-back systems [1]. The human speech-production system is composed of lungs, vocal cords, and vocal tract—articulators that deal with gain, pitch harmonics, and spectral envelope, respectively. An LPC filter is widely used to model the vocal-tract system. Since the minimum phase property may not be preserved in the quantized LPC coefficients, they are quantized with a transformed form such as log-area ratio, reflection coefficients, and line spectral frequency (LSF) [2]. In most speech coding standards, LPC information is quantized and interpolated in the LSF domain because the stability of LPC synthesis filter can be easily guaranteed as long as the LSF parameters are ordered and bounded within a range [3,4]. Vector quantization (VQ) produces better rate-distortion (R-D) performance than scalar quantization (SQ) since VQ has three major advantages over SQ [5]. The memory, space filling, and shape advantages correspond to better usages of correlation between samples, Voronoi-region shape, and source-distribution modeling, respectively. In this paper, a novel VQ method for LSF is proposed.

As the vector dimensionality increases, the R-D performance of VQ increases, but the required memory and computational complexity may not be practically feasible. To solve this problem, split VQ (SVQ) was proposed where an input vector is split into multiple sub-vectors that are quantized independently [6]. However, its R-D performance decreases because correlation between sub-vectors are not taken into account. In [7,8], the intra-frame correlation between sub-vectors is more carefully considered in order to improve the R-D performance.

In [9], the predictive SVQ (PSVQ) was proposed to efficiently utilize the inter-frame correlation between LSF vectors of the current and previous frames. In the quantization of LSF vectors, as inter-frame correlation is much stronger than intra-frame correlation, PSVQ produces a much better R-D performance compared with any

* Corresponding author. Tel.: +82 2 3408 3695.
E-mail address: mooyoung@sejong.ac.kr (M.Y. Kim).

0165-1684/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2012.09.023

quantization method that is based on intra-frame correlation. Although PSVQ has a drawback in channel error sensitivity, its superior R-D performance allows it to be used for voice storage devices where channel error rarely occurs. In PSVQ, the prediction coefficients can be estimated by auto-regressive (AR) models (AR-PSVQ) [9].

In this paper, conditional PSVQ (C-PSVQ) is proposed to obtain a better R-D performance compared with conventional methods including AR-PSVQ. C-PSVQ is an extended version of [7] by applying the inter-frame prediction instead of intra-frame prediction. The proposed C-PSVQ is presented in Section 2. Experimental results and conclusions follow in Sections 3 and 4, respectively.

2. Conditional PSVQ (C-PSVQ)

The M-dimensional LSF vector extracted from the n-th frame, $X^{(n)}=[x_1^n, x_2^n, \ldots, x_M^n]$, has not only intra-frame correlation between vector components but also inter-frame correlation with vectors in consecutive frames. In AR-PSVQ, to remove the inter-frame correlation between $X^{(n)}$ and $X^{(n-1)}$, an error vector, $E^{(n)}=[e_1^n, e_2^n, \ldots, e_M^n]$ is quantized instead of $X^{(n)}$ where

$$e_k^{(n)}=x_k^{(n)}-\alpha_k x_k^{(n-1)}, \quad (1)$$

where α_k is the k-th prediction-coefficient. In AR-PSVQ, $E^{(n)}$ is split into multiple sub-vectors and quantized independently as in [9].

In this paper, C-PSVQ is proposed based on the conditional distribution of $X^{(n)}$ for a given $X^{(n-1)}$. The original LSF vectors, $X^{(n)}$ and $X^{(n-1)}$, are assumed to be Gaussian, $N(\mu_{X^{(n)}}, \Sigma_{X^{(n)}})$, where $\mu_{X^{(n)}}$ and $\Sigma_{X^{(n)}}$ are the mean and covariance matrices of $X^{(n)}$, respectively.

Based on the Gaussian assumption, two consecutive vectors, $X^{(n)}$ and $X^{(n-1)}$ are transformed into uncorrelated Gaussian vectors as given by [10]

$$Y^{(n)}=X^{(n)}-\Sigma_{X^{(n)}X^{(n-1)}}\Sigma^{-1}_{X^{(n-1)}X^{(n-1)}}X^{(n-1)} \quad (2)$$

and

$$Y^{(n-1)}=X^{(n-1)}, \quad (3)$$

where $\Sigma_{X^{(i)}X^{(j)}}$ is the covariance matrix relating $X^{(i)}$ and $X^{(j)}$, and $(\cdot)^{-1}$ is a matrix-inverse operator. Since the uncorrelated Gaussian signals are independent, $Y^{(n)}$ is independent of $Y^{(n-1)}$ and thus $Y^{(n)}$ can be quantized separately from $Y^{(n-1)}$. The probability density function (PDF) of $Y^{(n)}$ is also Gaussian as given by $N(\mu_{Y^{(n)}}, \Sigma_{Y^{(n)}Y^{(n)}})$. We can rewrite $f(Y^{(n)})$ as a conditional PDF of $X^{(n)}$ and $X^{(n-1)}$ as given by

$$f(X^{(n)}|X^{(n-1)})=\frac{e^{\{-(1/2)(X^{(n)}-\mu_{X^{(n)}|X^{(n-1)}})^T\Sigma_{X^{(n)}X^{(n-1)}}(X^{(n)}-\mu_{X^{(n)}|X^{(n-1)}})\}}}{\sqrt{(2\pi)^M|\Sigma_{X^{(n)}|X^{(n-1)}}|}}, \quad (4)$$

where

$$\mu_{X^{(n)}|X^{(n-1)}}=\mu_{X^{(n)}}+\Sigma_{X^{(n)}X^{(n-1)}}\Sigma^{-1}_{X^{(n-1)}X^{(n-1)}}(X^{(n-1)}-\mu_{X^{(n-1)}}), \quad (5)$$

$$\Sigma_{X^{(n)}|X^{(n-1)}}=\Sigma_{X^{(n)}X^{(n)}}-\Sigma_{X^{(n)}X^{(n-1)}}\Sigma^{-1}_{X^{(n-1)}X^{(n-1)}}\Sigma_{X^{(n-1)}X^{(n)}}. \quad (6)$$

In (5) and (6), $\mu_{X^{(n)}}$, $\mu_{X^{(n-1)}}$, $\Sigma_{X^{(n)}X^{(n-1)}}$, and $\Sigma_{X^{(n-1)}X^{(n-1)}}$ are pre-calculated in the training phase. Since a matrix inversion requires much computation, instead of $\Sigma_{X^{(n)}X^{(n-1)}}$ and $\Sigma_{X^{(n-1)}X^{(n-1)}}$, $\Sigma_{X^{(n)}X^{(n-1)}}\Sigma^{-1}_{X^{(n-1)}X^{(n-1)}}$ is pre-calculated. Thus, compared with the conventional PSVQ, C-PSVQ requires a similar amount of computational complexity.

If the LSF vector in the previous frame, $X^{(n-1)}$, is given, the PDF of $X^{(n)}$ is changed into $N(\mu_{X^{(n)}|X^{(n-1)}}, \Sigma_{X^{(n)}|X^{(n-1)}})$ as shown in (4). In the proposed C-PSVQ, $X^{(n)}$ is quantized based on the assumption that $X^{(n-1)}$ is given and the PDF of $X^{(n)}$ follows $N(\mu_{X^{(n)}|X^{(n-1)}}, \Sigma_{X^{(n)}|X^{(n-1)}})$. To remove the inter-frame correlation between $X^{(n)}$ and $X^{(n-1)}$ the conditional PDF is transformed into $N(0,1)$ as given by

$$X^{(n)}_{\mu,\Sigma}=\Sigma^{-1/2}_{X^{(n)}|X^{(n-1)}}(X^{(n)}-\mu_{X^{(n)}|X^{(n-1)}}), \quad (7)$$

where the conditional mean, $\mu_{X^{(n)}|X^{(n-1)}}$, varies on a frame-by-frame basis according to $X^{(n-1)}$ but the conditional covariance, $\Sigma_{X^{(n)}|X^{(n-1)}}$, is a constant matrix that can be calculated in the training phase as shown in (5) and (6). Since a constant transform does not change the R-D performance, instead of (7), the mean removal process is applied before quantization as given by

$$X^{(n)}_{\mu}=X^{(n)}-\mu_{X^{(n)}|X^{(n-1)}}. \quad (8)$$

Fig. 1 illustrates the block diagram of C-PSVQ. For the input vector $X^{(n)}$, $X^{(n)}_{\mu}$ is calculated for quantization in an SVQ block. Since the conditional mean $\mu_{X^{(n)}|X^{(n-1)}}$ in (5) cannot be obtained at the decoder, instead of $X^{(n)}_{\mu}$, $\overline{X}^{(n)}_{\mu}$ is calculated as given by

$$\overline{X}^{(n)}_{\mu}=X^{(n)}-\hat{\mu}_{X^{(n)}|X^{(n-1)}}. \quad (9)$$

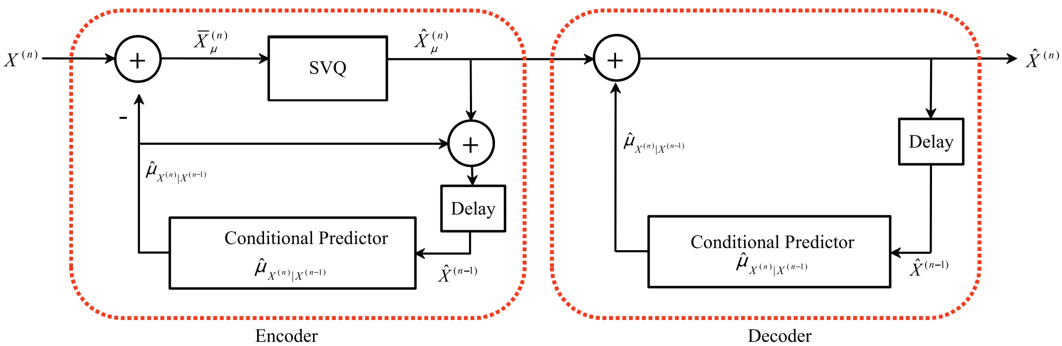

Fig. 1. Block diagram of the proposed conditional PSVQ.