

Empirical Lower Bound on the Bitrate for the Transparent Memoryless Coding of Wideband LPC Parameters

Stephen So, *Member, IEEE*, and Kuldip K. Paliwal, *Member, IEEE*

Abstract—In this letter, we determine empirical lower bounds on the bitrate required to transparently code linear predictive coding (LPC) parameters derived from wideband speech. This is achieved via extrapolation of the operating distortion-rate curve of an unconstrained vector quantizer that is trained using artificial vectors generated by a Gaussian mixture model. Memoryless coding is considered and two competing LPC parameter representations are investigated. Our results show a lower bound of 31 bits/frame when assuming high-rate linearity in the operating distortion-rate curve and 35 bits/frame for an exponential curve. We also evaluate a recent quantization scheme and compare its performance against this lower bound.

Index Terms—Immittance spectral pairs, line spectral frequencies (LSFs), linear predictive coding (LPC) parameters, wideband speech.

I. INTRODUCTION

THE quantization of linear predictive coding (LPC) parameters in CELP coders for narrowband speech (300–400 Hz) has been thoroughly investigated in the literature, where product code vector quantizers operating on vectors of ten line spectral frequency (LSF) parameters [7], generally require 24 bits/frame for transparent quality [8], [11]. Several studies have been conducted that focused on determining the minimum bitrate required to transparently code LSFs derived from narrowband speech. Such a result is mostly useful for serving as a benchmark in the evaluation of LPC quantization schemes.

Paliwal and Kleijn [12] extrapolated the operating distortion-rate curve of an unconstrained vector quantizer and showed that at least 20 bits/frame (when using partial-band spectral distortion) are required to transparently code narrowband LSFs. However, the size of the training set was relatively small, which tended to undertrain the vector quantizer codebook. Hedelin and Skoglund [6] took a different approach, where they used a high-rate optimized Gaussian mixture model (HRO-GMM) with bounded support to accurately approximate the LSF vector source. It would be expected that the resulting GMM could be used to generate enough random vectors to train a high-rate vector quantizer without the problem of undertraining. They predicted that a minimum bitrate of 23 bits/frame was required

Manuscript received October 25, 2005; revised February 12, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Patrick A. Naylor.

S. So is with the Griffith School of Engineering, Griffith University, Gold Coast Campus, Gold Coast, QLD 9726, Australia (e-mail: s.so@griffith.edu.au).

K. K. Paliwal is with the Griffith School of Engineering, Griffith University, Nathan Campus, Brisbane, QLD 4111, Australia (e-mail: k.paliwal@griffith.edu.au).

Digital Object Identifier 10.1109/LSP.2006.874444

to achieve a full-band spectral distortion¹ of 1 dB on LSF vectors, which was more conservative than the Paliwal and Kleijn result. In a similar study by Samuelsson and Hedelin [13], which investigated the recursive coding of LPC parameters, a minimum bitrate of 16 bits/frame was predicted for quantization schemes that exploited interframe memory.

In this letter, we determine empirical lower bounds on the bitrate required to transparently code 16 LPC parameters derived from wideband speech (50–7000 Hz). In addition, we compare the performance of a recently proposed quantization scheme [14] to these lower bounds. We consider only memoryless, single frame coding and two competing LPC parameter representations: LSFs [7] and immittance spectral pairs (ISPs) [2]. Similar to other studies on wideband LPC quantization, we have assumed full-band spectral distortion as an objective measure of quality. We believe that these results have not been previously reported in the literature and will serve as a useful benchmark for future studies in wideband LPC quantization schemes.

II. LPC PARAMETER QUANTIZATION EXPERIMENTS

A. Experimental Setup

The TIMIT database was used in the training and testing of the quantization experiments, where speech is sampled at 16 kHz. We have used the lower band (50–6400 Hz) preprocessing and LPC analysis of the AMR-WB speech codec (in modes of 12.65 kbps or higher), where speech is initially downsampled to 12.8 kHz, high-pass, and pre-emphasis filters are applied, and then it is divided into 20-ms frames with a 5-ms lookahead using a 30-ms asymmetric Hamming window [1]. For each speech frame, 16 linear prediction coefficients are then calculated, which are converted to LSF and ISP representation (quantization is performed on the frequency form of ISPs, i.e., ISFs). The training set consists of 333 789 vectors while the evaluation set, consisting of speech not contained in the training set, has 85 353 vectors.

B. Spectral Distortion and Conditions for Transparent Coding

For the i th frame, the full-band spectral distortion (in dB), D_i , is defined as

$$D_i = \sqrt{\frac{1}{F_s} \int_0^{F_s} \left[10 \log_{10} P_i(f) - 10 \log_{10} \hat{P}_i(f) \right]^2 df} \quad (1)$$

¹For partial-band spectral distortion, the predicted minimum bitrate was 22 bits/frame [6].

TABLE I
COMPARING THE PERFORMANCE OF AN UNCONSTRAINED VECTOR QUANTIZER TRAINED USING REAL LSF VECTORS WITH ONE TRAINED USING ARTIFICIAL LSF VECTORS

Bits/frame	Trained using real LSFs			Trained using artificial LSFs		
	Avg. SD (dB)	Outliers (in %)		Avg. SD (dB)	Outliers (in %)	
		2–4 dB	> 4 dB		2–4 dB	> 4 dB
7	3.563	56.44	35.02	3.560	56.43	34.99
8	3.370	62.23	27.41	3.369	62.38	27.22
9	3.203	67.20	20.67	3.199	67.58	20.26
10	3.051	71.06	14.96	3.046	71.27	14.67
11	2.916	73.28	10.63	2.908	73.94	10.09

where F_s is the sampling frequency, and $P_i(f)$ and $\hat{P}_i(f)$ are the LPC power spectra of the coded and uncoded i th frame, respectively. The conditions for transparent coding of narrowband speech from LPC parameter quantization are quoted as [11] follows.

- 1) The average spectral distortion (SD) is approximately 1 dB.
- 2) There is no outlier frame having more than 4 dB of spectral distortion.
- 3) Less than 2% of outlier frames are within the range of 2–4 dB.

According to Guibé *et al.* [5], listening tests have shown that these conditions for transparency also apply to the wideband case. However, another study concluded with a different set of conditions, where a spectral distortion of 1.6 dB was considered sufficient for transparent coding [4]. For the purpose of defining a suitable reference, we will consider condition 1 (i.e., SD of 1 dB) as being sufficient for transparent coding² in this letter, though the reader should be aware of other criteria in the literature and be able to derive an alternate lower bound using the extrapolated graphs.

C. Unconstrained Vector Quantization

In this experiment, we take a similar approach to that of Paliwal and Kleijn [12], where the operating distortion-rate curve of an unconstrained vector quantizer is extrapolated to higher bitrates in order to determine the bitrate required for transparent coding (i.e., incurring a spectral distortion of 1 dB). The codebook of an unconstrained vector quantizer has no physical constraints (as opposed to product code vector quantizers), so that searching for the code-vector that minimizes distortion is done exhaustively. The Linde–Buzo–Gray (LBG) algorithm [10] is commonly used to train this codebook. In order to evaluate the vector quantizer at high bitrates based on the limited training set, we have calculated a 128 mixture component Gaussian mixture model (GMM) of this training set. Using this GMM, we then generated an artificial training set of two million vectors. In order to see how similar these artificial vectors are to the real LSF vectors, we compared the spectral distortion performance on the LSF test set of the TIMIT database via a vector quantizer trained using the real and artificial training sets (as shown in Table I). We can see that for

²We note that the distribution of the spectral distortion, as expressed in conditions 2 and 3, is also important in deciding whether or not the coding is transparent, even though we have not included it in our present investigation. Following [6], the PDF of the spectral distortion can be approximated by differentiating the cumulative distribution function for errors in high-rate vector quantizers, as given in [9].

all the bitrates considered, the artificially trained vector quantizer performs the same, if not better, than the vector quantizer that was trained using real LSF vectors. This shows that the artificial LSF vectors generated from the GMM (obtained from real training data) are adequate to train the vector quantizer.

1) *Using Linear Extrapolation:* Fig. 1 shows the distortion-rate (D-R) points of the unconstrained vector quantizer on LSF and ISP vectors. While in principle, it is possible to operate the unconstrained vector quantizer at high bitrates to determine the lower bound for transparent coding, the exponential growth of the computational complexity makes this method infeasible. We can see that at low bitrates, the D-R curve is exponential in nature, while at higher bitrates, it becomes more linear. In theory, the D-R curve of an unconstrained vector quantizer at high bitrates tends to be linear, provided that enough training vectors are available. If we make the assumption that for high bitrates (>12 bits/frame), the D-R curve follows a linear relationship, then a least-squares linear regression (as shown by the line in Fig. 1) predicts that we need 31 bits/frame to achieve an average spectral distortion of 1 dB for both LSF and ISP vectors. This lower bound is by no means a tight one for several reasons. First, the high rate linearity assumption is generally not valid in practice and will underestimate the spectral distortion (if we assume the D-R curve of a practical vector quantizer to be slightly convex-shaped). This is because the vector quantizer will become more undertrained at higher bitrates [3]. Second, the LBG algorithm that was used in the training may produce a locally optimal codebook.

2) *Using Exponential Extrapolation:* In practice, vector quantizers are often trained based on a limited number of training vectors. At high bitrates, the ratio of the number of training vectors to the number of code-vectors becomes low, and the efficiency of the codebook degrades—a condition known as undertraining [3]. Hence, as mentioned in the previous section, the operating D-R curve of vector quantizers used in practice will tend to flatten out [3]. If we make the assumption of an exponential D-R curve of the form

$$D = a2^{bR} \quad (2)$$

then a least-squares regression of $a = 4.939$ and $b = -0.0671$ (as shown by the line in Fig. 2) shows that at least 35 bits/frame are required to achieve a spectral distortion of approximately 1 dB for LSFs. For ISPs, the least-squares regression of $a = 4.933$ and $b = -0.0657$ also predicts that 35 bits/frame are required to achieve 1 dB of spectral distortion. We can see from Fig. 2 that,

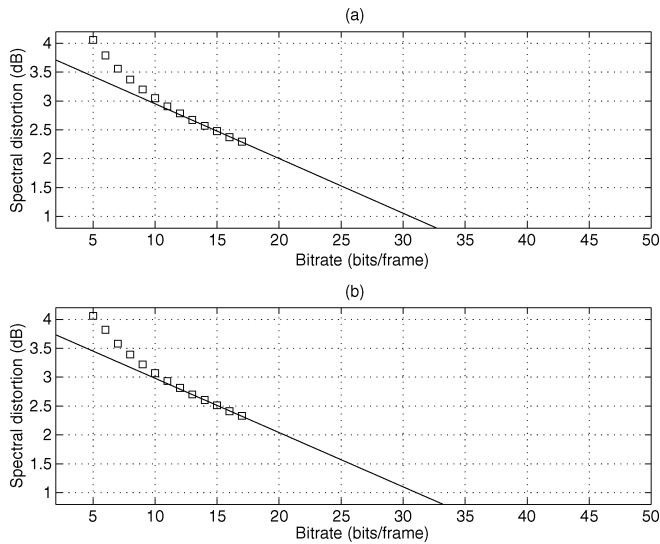


Fig. 1. Linear extrapolation of the operating D-R curve of an unconstrained vector quantizer trained using an artificial training set. (a) For LSF vectors. (b) For ISP vectors.

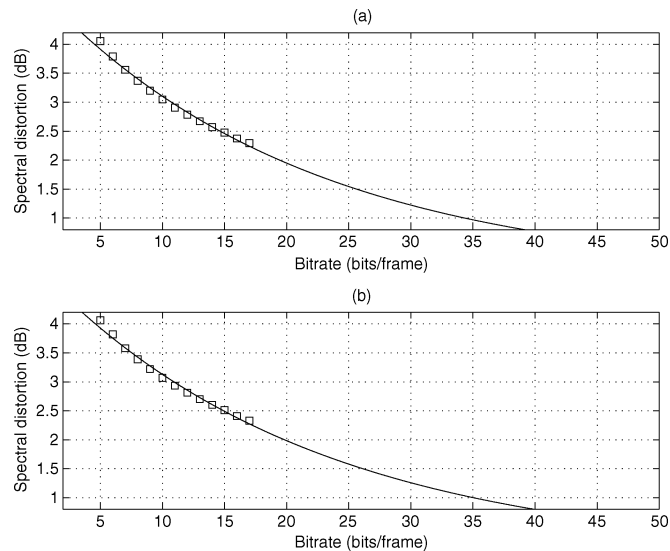


Fig. 2. Exponential extrapolation of the operating distortion-rate curve of an unconstrained vector quantizer trained using an artificial training set. (a) For LSF vectors. (b) For ISP vectors.

apart from the first point, the exponential curve fits the operating D-R points reasonably well.

D. Comparison With a Recent LPC Quantization Scheme

The GMM-based block quantizer [14] has been used as a low complexity alternative to vector quantization for coding speech LSFs. This scheme involves approximating the source probability density function using a GMM and designing an optimal Karhunen–Loève transform coder for each of the m mixture components. In order to code a frame, it is first quantized using all m transform coders, and the distortion incurred by each coder is calculated. The transform coder that incurs the least distortion is then chosen. The main advantages of this scheme are

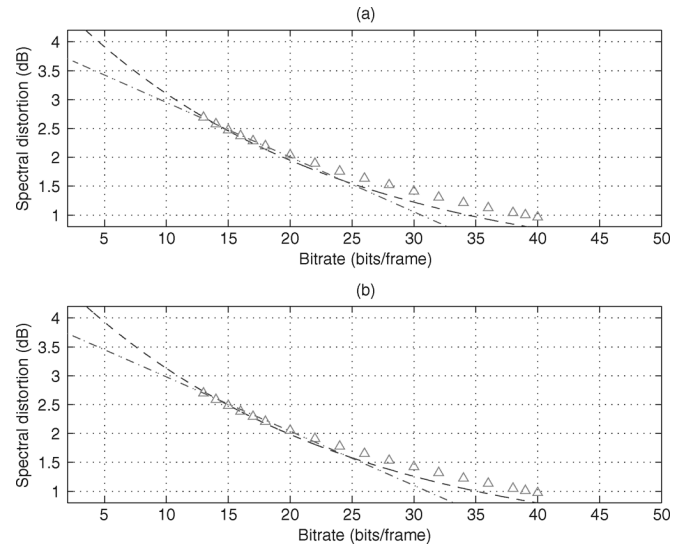


Fig. 3. Comparing the spectral distortion performance of the GMM-based block quantizer (shown as triangles) with the predicted performance of the unconstrained vector quantizer (linear extrapolation is represented by the dash-dot line; exponential extrapolation is represented by the dashed line). (a) For LSF vectors. (b) For ISP vectors.

bitrate-independent complexity and bitrate scalability, since the GMM is a compact and complete description of the source [14].

We have designed a GMM-based block quantizer using 128 mixture components and quantized the LSF and ISP vector test sets. The spectral distortions of the GMM-based block quantizer on LSFs and ISPs are presented in Fig. 3 as triangles and are overlaid with the extrapolated curves of the unconstrained vector quantizer. Note that we have not evaluated this scheme at lower bitrates since up to 7 bits are required for mixture component identification. Operating this scheme at lower bitrates would lead to a shortage of bits to be allocated to each transform coder. The bitrate required for 1 dB of spectral distortion is 39 bits/frame for both the LSF and ISP representation. We can see that at the lower bitrates, the GMM-based block quantizer incurs similar spectral distortion to the vector quantizer. At higher bitrates, however, it starts to diverge away from the extrapolated curves. This is expected as the GMM-based block quantizer is generally suboptimal when compared with the “best” vector quantizer. This suboptimality amounts to a difference of roughly 8 bits/frame when compared against the lower bound predicted by the high-rate linear extrapolation or 4 bits/frame when compared against the lower bound predicted by the exponential extrapolation.

III. CONCLUSION

In this letter, we have determined lower bounds for the bitrate required to transparently quantize wideband LSFs and ISPs. If we assume the D-R curve for an unconstrained vector quantizer to be linear at high bitrates, then a lower bound of 31 bits/frame is determined. If the D-R curve is assumed to follow an exponential trend, then the lower bound is determined to be 35 bits/frame. We have also evaluated a recently proposed memoryless quantization scheme and compared its performance to

the determined lower bounds. The GMM-based block quantizer was determined to be suboptimal by 4–8 bits/frame.

REFERENCES

- [1] B. Bessette, R. Salami, R. Lefebvre, M. Jelínek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Järvinen, "The adaptive multirate wideband speech codec (AMR-WB)," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 8, pp. 620–636, Nov. 2002.
- [2] Y. Bistriz and S. Peller, "Immittance spectral pairs (ISP) for speech encoding," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1993, pp. II-9–II-12.
- [3] J. S. Collura and T. E. Tremain, "How good is your β ?—observations on VQ training ratios," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1995, pp. 744–747.
- [4] M. Ferhaoui and S. Van Gerven, "LSP quantization in wideband speech coders," in *Proc. IEEE Speech Coding Workshop*, Jun. 1999, pp. 25–27.
- [5] G. Guibé, H. T. How, and L. Hanzo, "Speech spectral quantizers for wideband speech coding," *Eur. Trans. Telecommun.*, vol. 12, no. 6, pp. 535–545, 2001.
- [6] P. Hedelin and J. Skoglund, "Vector quantization based on Gaussian mixture models," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 4, pp. 385–401, Jul. 2000.
- [7] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Amer.*, vol. 57, p. S35, Apr. 1975.
- [8] W. P. LeBlanc, B. Bhattacharya, S. A. Mahmoud, and V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ for LPC parameters for 4 kb/s speech coding," *IEEE Trans. Speech Audio Process.*, vol. 1, no. 4, pp. 373–385, Oct. 1993.
- [9] D. H. Lee and D. L. Neuhoff, "Asymptotic distribution of the errors in scalar and vector quantizers," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 446–460, Mar. 1996.
- [10] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, no. 1, pp. 84–95, Jan. 1980.
- [11] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. Speech Audio Process.*, vol. 1, no. 1, pp. 3–14, Jan. 1993.
- [12] K. K. Paliwal and W. B. Kleijn, "Quantization of LPC parameters," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Amsterdam, The Netherlands: Elsevier, 1995, pp. 443–466.
- [13] J. Samuelsson and P. Hedelin, "Recursive coding of spectrum parameters," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 492–502, Jul. 2001.
- [14] A. D. Subramaniam and B. D. Rao, "PDF optimized parametric vector quantization of speech line spectral frequencies," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 2, pp. 130–142, Mar. 2003.