LOW COMPLEXITY LSF QUANTIZATION FOR WIDEBAND SPEECH CODING

Stéphane Ragot, Jean-Pierre Adoul, Roch Lefebvre, and Redwan Salami

Department of Electrical Engineering, University of Sherbrooke,
Sherbrooke, Québec, Canada J1K 2R1

ABSTRACT

State-of-the-art narrowband speech coders operating from 4 to 16 kbit/s are mostly based on the code-excited linear predictive (CELP) model. They achieve a good synthesis quality usually at the expense of a high coding complexity. For example, in the 8 kbit/s G.729 coder the innovation codebook search is responsible for approximately half the total coder complexity, the latter being close to 20 MIPS in fixed-point DSP implementation. Less known is the relative part of spectral quantization, which is around 8 % of the total complexity. CELP coders are still relevant for wideband speech coding but their complexity is greater than in the narrowband case, which becomes critical for real-time implementations.

We propose in this article a two-stage algebraic-stochastic LSF quantization scheme. It combines the strengths of algebraic and stochastic techniques, namely low computation and storage cost and good performance. The generalized Lloyd-Max algorithm is adapted for optimizing lattice codebooks obtained by spherical truncation. Simulations with a Gaussian source show that the quantization method exhibits good quality/complexity tradeoffs. Several stochastic-algebraic LSF quantizers are derived and compared to a more conventional technique.

1. INTRODUCTION

Lattices have found significant applications in data transmission in the past few years. Their use in speech and audio coding has been more marginal. They have been applied mainly to residual signal quantization [1], spectral envelope quantization [7, 9] and coding in the frequency domain [10, 11]. The main advantage of a lattice quantizer is its low computational complexity and memory requirement. It generally offers a high coding efficiency despite the constraint of algebraic structure in the codebook.

In this paper we are concerned with wideband spectral envelope quantization. The envelope is modeled with linear prediction (LP) coefficients, which have been found to be best described in terms of line spectral frequencies (LSF) [6, 8]. LSF are usually quantized with stochastic codebooks. For wideband speech, an LP order of 16 is typically used. With this dimensionality, optimal quantizers reach their practical complexity limits. Lattice quantization is therefore very attractive.

This article is organized as follows. A preliminary section is included to review some concepts about lattices and to present several optimization techniques that can be applied to a lattice quantizer. The study is limited to the special case of a Gaussian distribution which allows a rigorous investigation. A novel technique consisting of radix adaptation inside a lattice codebook is proposed. The subsequent section describes the design of a fixed-rate wideband LSF quantizer incorporating these theoretical results in a hybrid multistage codebook.

2. FIXED-RATE OPTIMIZED LATTICE QUANTIZATION

A lattice is an infinite regular array that covers space uniformly. Usually, once a (good) lattice is selected, it has to be truncated, translated and scaled properly to become an optimized finite codebook.

Among possible candidates for quantizing wideband LSF up to 16 dimensions, the most interesting lattices are the Gosset and the Barnes-Wall lattices in 8 and 16 dimensions respectively. In comparison, the latter provides a smaller distortion, but is much more complex in both decoding and indexing. The Gosset lattice RE_8 is a more reasonable choice for complexity reasons and for robustness against random bit errors on a noisy channel.

The discrete-time memoryless Gaussian source is selected for primer experiments, as it is a usual benchmark in quantization theory. It is restricted here to the centered unit-variance case without loss of generality. It is indeed a legitimate first-order approximation for many real signals, including quantization residuals in a multistage quantizer. The distortion measure is furthermore restricted to be the mean square error (MSE). It may not be appropriate when the space is related to perception as with LSF, but the advantage of fast algebraic algorithms is otherwise lost. If some other measure would be used, codeword signs have to be resolved before testing every possible permutation of absolute components. In addition to low complexity provided by a mean square error criterion, the theoretical rate-distortion bound is known in the Gaussian case [2]. The signal-to-noise ratio (SNR) is upper bounded by

$$SNR_{max}(R) = 20\,R\,\log_{10} 2 \approx 6.02\ R\ \text{(in dB)} \qquad (1)$$

where R is the bitrate per sample.

2.1. RE_8 decoding and indexing

The RE_8 lattice with integer coordinates is obtained by gluing together two copies of the $2D_8$ lattice [3]:

$$RE_8 = 2D_8 \cup \{2D_8 + [1\,1\,1\,1\,1\,1\,1\,1]\} \qquad (2)$$

Points in RE_8 are all located on euclidean spheres or shells of radius $\sqrt{8m}$, $m \geq 0$. RE_8 is thus made of an infinite collection of embedded spherical codes. A finite RE_8 codebook may then be easily obtained by spherical truncation, i.e., by taking only complete shells up to a given radius. Quantization is done in two explicit steps:

1. nearest neighbor search inside a finite codebook with respect to the euclidean distance, which corresponds in the coding theory to maximum-likelihood decoding in signal space ;
2. indexing of the decoded point.

0-7803-5651-9/99/$10.00

On a given shell, one can find some usefull relations between point coordinates, which leads to the important concept of absolute and signed leaders. A very efficient indexing algorithm based on table lookups is derived from this fact. Note that the embedding property makes it possible to have several RE_8 quantizers with different bitrates, which use a single indexing table and decoding template [10].

2.2. Optimization of a fixed-rate RE_8-based codebook

In this paper we consider data optimization. It is indeed a more general framework than pdf optimization, the gaussianity being assumed to get insight in the codebook design. Hence given a finite source database $\mathcal{B}$, optimization is done by minimizing the total distortion over the database with respect to all possible codebooks.

If the discrete-time source x is restricted to be the centered memoryless Gaussian source, the best lattice truncation is known to be spherical. This matches perfectly the intrinsic spherical structure of a RE_8 codebook. In this special case, the only degree of freedom in optimizing a sphere-bounded RE_8 quantizer Q is a static scale g. Its optimal value is given by

$$g_{opt} = \arg\min_{g} \sum_{x \in \mathcal{B}} ||x - g\,Q(x/g)||^2 \qquad (3)$$

If coordinate variances are not identical, each coordinate may be scaled separately as in [7] and the shape of the codebook boundary becomes ellipsoidal. Moreover when the algebraic code is not spherical, but an aggregate of several spherical codes, $Q(x/g)$ does depend on g. The optimal scale can be determined by some numerical optimization algorithm, for example by using the steepest descent method.

The radix distribution in RE_8 does not fit that of a Gaussian random variable which is χ^2-like. Lattices are indeed known to be optimal for a high-resolution uniform source only. Thus further optimization of a RE_8 quantizer may be obtained by varying the shell radix and reordering lattice shells. The quantized value of x, say $\widehat{x}$, is then a lattice point scaled by a factor depending on the shell on which it lies. The algebraic codebook can be trained iteratively as in the generalized Lloyd-Max algorithm [5]. The radius of a specific shell S is scaled by a factor, say g_S, and this factor is adapted from those database points which are in the region $\mathcal{V}_S$, where

$$\mathcal{V}_S = \{x \in \mathcal{B} \mid \widehat{x} = g_S . y\, ,\, y \in S\} \qquad (4)$$

This adaptation may be constrained so as to keep a preselected shell ordering. To some extent, the process developped for shells may be applied directly to a set of absolute leaders. The database is then partitioned by their Voronoi-like regions. Factors g_S are locally adapted in each region until global convergence.

2.3. Performances

RE_8 codebooks were optimized with a database of 100,000 Gaussian 8-dimensional vectors. Their performances were evaluated with the same number of vectors outside the learning database and are reported in Figure 1. It turns out that a simple scale optimization yields a good fidelity. The results are 1.6-1.9 dB below the Shannon bound at 16-20 bits. Shell and absolute leader optimizations can improve these results by 0.4 dB and 0.2 dB respectively.

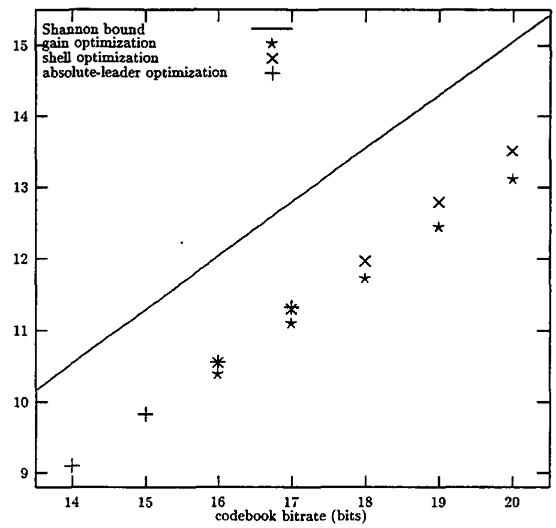

Figure 1: SNR in dB of source-optimized RE_8-based codebooks for zero-mean unit-variance memoryless (i.i.d.) Gaussian source.

The former works better at high bitrates (beyond 16 bits) and the latter at low ones (below 16 bits) where there are few absolute leaders and then few degrees of freedom. Codebooks from 16 to 20 bits were predesigned by spherical saturation of RE_8 as in [10], and completed with some points in RE_8 to achieve an integer bitrate. Bitrates lower than 16 bits were investigated in priority, since practical data memories are mainly word-oriented and the size of algebraic indexing tables is doubled when the codebook bitrate exceeds 16 bits. Codebooks at 14 and 15 bits were created by selecting properly absolute leaders in RE_8.

Note that source optimization does not affect the indexing.

3. STOCHASTIC-ALGEBRAIC LSF QUANTIZATION

In narrowband speech coding, a multistage structure [4] is commonly adopted to quantize LSF, due to its advantages in low complexity and robustness against random bit errors. It is still suitable for wideband.

This section describes a two-stage quantizer with a trained stochastic codebook in the first stage so as to fit the main statistical trends of LSF and an algebraic second stage used to reduce complexity. The M-L search algorithm with M survivors in the first stage is used for training and operation. It is suboptimal but yet very close to the optimal performance. Each stage is actually divided into two splits which both apply M-L search and fixed-rate RE_8 codebooks. This increases the robustness and allows to allocate bits in a more flexible fashion. The codebooks are optimized by iterative sequential training [4] with the generalized Lloyd-Max algorithm in the first stage and scale(+radix) adaptation in the second stage. The surviving tests in the first stage and algebraic decoding trails are performed with respect to the mean square error criterion, the winner path in the M-L search is selected by a weighted distortion measure [6] to achieve a good perceptual quality.

3.1. Statistical hypothesis tests

The use of an algebraic codebook in the second stage relies on several assumptions that have to be validated. The first-stage quantization error is reasonably centered but not its coordinates are not identically distributed as was discovered in the histogram of their estimated variances. A scale for each coordinate can then be used to compensate this fact as in [7]. Scatter plots showed that at least 6 and 5 bits have to be allocated respectively to the first and the second split in the first-stage to ensure satisfying decorrelation. This fact was reinforced by calculating the covariance matrix of the residual in each split. High-order statistics could be used to test gaussianity, but the RE_8 lattice may provide a good fidelity at studied bitrates for any uncorrelated distribution close to the Gaussian case. An histogram of the squared norm of the first-stage 8-dimensional residuals depicted a χ^2-like shape. Consequently the Gaussian model is quite meaningful as a first-order approximation when the bitrate of the first stage is high enough.

3.2. Quantization results

The LSF database used in the experiments was made of 74,000 vectors in the training sequence and 24,000 vectors in the test sequence. It was built with the front-end of a wideband CELP coder, which uses a 20 ms frame. Linear prediction was performed on a hybrid Hamming-cosine window of 30 ms. Lag windowing, white noise correction and a preemphasis factor of 0.75 were used to enhance the analysis.

Several bit allocations have been tested. The hybrid stochastic-algebraic quantizers were benchmarked with a low-complexity two-stage stochastic quantizer with respect to split-by-2 (9-7) and split-by-5 (3-3-3-3-4) in the first and the second stage. The inverse harmonic mean measure [7] was used for path selection during the M-L search in the hybrid cases, and the square error in the stochastic case. The spectral distortions were calculated by a 512-point fast Fourier transform (FFT) and are depicted in Table 1. The bins corresponding to 0 Hz and the 7-8 kHz band were discarded. The spectral distortion was not weighted due to the lack of a standard reference and meaningful transparency criteria in wideband.

RE_8-based codebook optimization	bit allocation in each stage		M	average SD (dB)	SD 2-4dB (%)	SD > 4 dB (%)
	1st	2nd				
global scale	6-6	16-16	4	1.14	3.15	0.02
			8	1.11	3.05	0.01
	7-6	16-16	4	1.04	2.23	0
			8	1.10	2.29	0
	7-7	16-16	4	1.09	2.40	0
			8	1.02	2.33	0
			128	1.01	2.29	0
scale by coordinate	6-6	16-16	4	1.11	2.81	0.03
			8	1.09	2.88	0.03
	7-6	16-16	4	1.09	2.03	0
			8	1.02	1.99	0
	7-7	16-16	4	1.06	1.82	0
			8	0.99	1.74	0
scale by absolute leader	7-5	15-15	4	1.11	4.16	0.01
			8	1.10	4.03	0
stochastic codebook	7-5	6-7-7-6-6	4	1.13	2.50	0
			8	1.06	1.85	0

Table 1: Spectral distortion (SD) statistics at different bitrates and the number of survivors M in the M-L search algorithm.

Prediction was not used in quantizers for the sake of simplicity. Frame erasures might also cause severe quality degradation otherwise. Robustness was not tested but it is expected to be close to that of conventional techniques. Informal listening with a simplified coder using an unquantized LP residual and LSF interpolation validated the good quality of the quantizers. Note that the stochastic benchmark requires more storage than the hybrid quantizers using optimized RE_8 codebooks.

4. CONCLUSION

Algebraic codebooks are promising candidates for the quantization of high-dimensional sources within practical complexity limits. The quantization of line spectral frequencies in wideband speech coders is one such example where they can be applied efficiently.

There is still room for improvements in the proposed hybrid quantization scheme. The bit allocation in the algebraic stage may be variable by using the embedding property of the lattice codebook. Prediction may save few bits. Several distortion measures may be tested. Perception topics are actually more important in wideband speech coding than in the narrowband case.

ACKNOWLEDGMENTS

The authors wish to express their thanks to Vesa T. Ruoppila for his careful reading and some helpful comments about this paper.

REFERENCES

[1] J.-P. Adoul, C. Lamblin and A. Leguyader, "Baseband speech coding at 2400 bps using spherical vector quantization," ICASSP, pp. 1.12.1–1.12.4, 1984.

[2] T. Berger, *Rate Distortion Theory*, Prentice-Hall, 1971.

[3] J. Conway and N.J. Sloane, *Sphere Packings, Lattices, and Groups*, Springer-Verlag, 1993.

[4] W.F. LeBlanc, B. Bhattacharya, S.A. Mahmoud, and V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding," *IEEE Trans. Speech and Audio Process.*, vol. 1, no. 4, pp. 373–385, 1993.

[5] Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–95, 1980.

[6] K.K. Paliwal and W.B. Kleijn, "Quantization of LPC Parameters," in *Speech Coding and Synthesis* (W.B. Kleijn and K.K Paliwal, eds.), pp. 433–466, Elsevier Science Publishing, 1995.

[7] J. Pan, "Two-stage vector quantization-pyramidal vector quantization and application to speech LSP coding," ICASSP, vol. 2, pp. 737–740, 1996.

[8] F. Soong and B. Juang, "Line Spectrum Pair (LSP) and speech data compression," ICASSP, pp. 1.10.1–1.10.4, 1984.

[9] M. Xie and J.-P. Adoul, "Fast and low-complexity LSF quantization using algebraic vector quantizer," ICASSP, vol. 1, pp. 716–719, 1995.

[10] M. Xie and J.-P. Adoul, "Embedded algebraic vector quantization (EAVQ) with application to wideband audio coding," ICASSP, vol. 1, pp. 240–243, 1996.

[11] L. Mainard, M. Lever, "A Bi-Dimensional Coding Scheme Applied to Audio Bitrate Reduction," ICASSP, vol. 2, pp. 1017–1020, 1996.