STOCHASTIC-ALGEBRAIC WIDEBAND LSF QUANTIZATION

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

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

ABSTRACT

This paper describes a fixed-rate quantizer using algebraic techniques as well as conventional stored codebooks to represent the wideband line spectral frequencies (LSF). It is based on two-stage split vector quantization with random and lattice codebooks in the first and second stage, respectively. We investigate the use of variable bit allocation in lattice quantization in order to capture outliers and reduce the overload distortion in spectral quantization, particularly for non-speech signals. Experimental results show that a variable bit allocation in split quantization yields performances slightly better than a fixed allocation at the same rate.

1. INTRODUCTION

Low bit-rate wideband coding is still an open problem. Current systems fail to achieve a good quality for both speech and audio signals at 16 kbit/s for a 16 kHz sampling frequency. One promising research axis consists of adapting speech coding techniques, which are essentially based on linear prediction, so as to represent speech and audio signals in the same framework. Typically the linear prediction coefficients are transformed into line spectral frequencies prior to quantization [1, 2]. In this context one step toward a joint coding system is to develop a LSF quantizer achieving a good quality for both speech and audio signals. Perceptually the main challenge is to handle outliers.

LSF quantization has been studied deeply for telephone-bandwidth (or narrowband) speech coding. Many solutions have been proposed yielding different trade-offs between spectral distortion, bit rate, robustness against random bit errors or frame erasures, and delay. Most of these solutions use constrained vector quantization (VQ) to cope with implementation limits, and exploit the source memory. The most popular techniques are split VQ, multistage VQ, lattice VQ, together with predictive VQ or matrix quantization. More specifically, a hybrid quantization scheme was presented in [3] for narrowband speech coding based on two-stage tree-search VQ-lattice VQ.

Hybrid structures [4] are attractive in terms of performance vs complexity trade-off for wideband LSF quantization. Based on a particular stochastic-algebraic structure, we propose in this paper a novel approach to spectral quantization by using variable-rate lattice quantization in order to handle irregular outcomes and capture outliers. The total bit rate is kept constant for the sake of simplicity.

The paper is structured as follows. We introduce first the stochastic-algebraic structure in Section 2. Lattice quantization is briefly covered in Section 3. We emphasize the design of variable-rate lattice quantization and the use of a weighted mean square error in lattice codebook search. Experimental results are presented in Section 4, and conclusions in Section 5.

2. STOCHASTIC-ALGEBRAIC STRUCTURE

A two-stage quantizer is described in Figure 1. This figure introduces also the notation used in the paper. The order of linear prediction is 16.

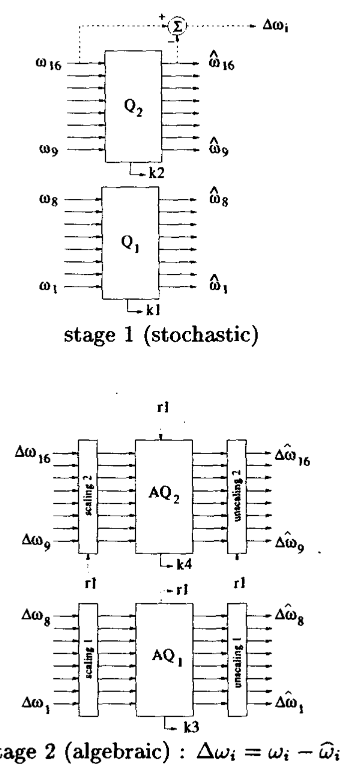

Figure 1: Two-stage split stochastic-algebraic quantizer.

2.1. Quantization strategies and bit allocation

We choose to optimize the quantizer for speech signals, so we use in training an LSF database extracted from speech material. Indeed there is more statistical structure (*e.g.* invariants) in speech than in audio signals. Since the first speech formants have bigger variances and are perceptually more important, the first stage can be biased by allocating more bits to the first split than to the second split. In order to reduce storage requirements and search complexity, 7 and 5–6 bits are typically allocated to the quantizers Q_1 and Q_2 of Figure 1. The resolution in the algebraic stage has to be high enough (typically 2 bits per sample, near 16 bits per split), so that the lattice codebooks can be effective. The main point here is that the two algebraic quantizers, AQ_1 and AQ_2, share a subset of an unscaled lattice codebook.

2.1.1. *Fixed bit allocation*

The easiest design for the stochastic-algebraic system of Figure 1 is to constrain the lattice codebooks in AQ_1 and AQ_2 to produce a fixed rate. In this case, we can apply sequential split quantization in both design and search. Note that the ordering of the reconstructed LSF values

$$0 < \widehat{\omega}_1 + \Delta\widehat{\omega}_1 < \widehat{\omega}_2 + \Delta\widehat{\omega}_2 < \cdots < \widehat{\omega}_{16} + \Delta\widehat{\omega}_{16} < \pi \quad (1)$$

and the gaps between adjacent frequencies must be checked to ensure the stability of the synthesis filter and avoid sharp resonances in the quantized spectrum.

2.1.2. *Adaptive bit allocation*

Variable-rate lattice quantization can be used in AQ_1 and AQ_2. In this case a side information r_1 is produced, and the first and second splits are quantized jointly. The parameter r_1 indicates how many bits are allocated to AQ_1 within a certain allowed range. Implicitly AQ_2 receives the remaining bits in the total fixed-rate budget and its scaling is adapted. This approach can be expected to fit real situations in a more flexible fashion than the fixed allocation.

2.2. Codebook training and optimization of scaling

Some algorithms can be found in [5] for training a multistage structure, and in [6] for a stochastic-algebraic structure. We use here an iterative sequential algorithm. It consists of optimizing sequentially each stage until the global convergence of quantization distortion. One stage is updated at a time, while the other is kept static. The two stages are optimized differently. In the algebraic stage, only a few scaling parameters are tuned approriately, *e.g.*, by a steepest descent method based on estimated gradients. The random codebooks in each split of the stochastic stage are trained with a modified version of the generalized Lloyd-Max algorithm [7] which takes into account the second stage.

Based on the model of Figure 2, the adaptation of each codeword $\theta_k[n]$ at iteration n is given by

$$\theta_k[n+1] = \arg\min_{\widehat{\omega}\in\mathbb{R}^8} E\left[\|\omega - (\widehat{\omega} + \Delta\widehat{\omega})\|^2 : \omega \in V_k[n]\right], \quad (2)$$

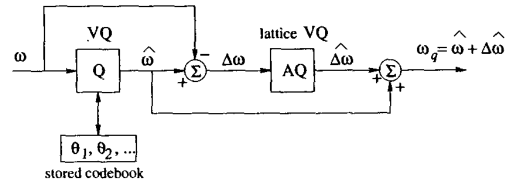

Figure 2: Two-stage model for training one split ($\omega \in \mathbb{R}^8$).

where $V_k[n]$ is the Voronoi region associated to $\theta_k[n]$:

$$V_k[n] = \left\{\omega \in \mathbb{R}^8 : Q(\omega) = \theta_k[n]\right\}. \quad (3)$$

For instance the model can be applied to the first split with

$$\begin{array}{rcllll} \omega & = & [& \omega_1 & \cdots & \omega_8 &]^T, \\ \widehat{\omega} & = & [& \widehat{\omega}_1 & \cdots & \widehat{\omega}_8 &]^T, \\ \Delta\omega & = & [& \Delta\omega_1 & \cdots & \Delta\omega_8 &]^T, \\ \Delta\widehat{\omega} & = & [& \Delta\widehat{\omega}_1 & \cdots & \Delta\widehat{\omega}_8 &]^T. \end{array} \quad (4)$$

It can be shown that the centroid $\theta_k[n+1]$ is given by

$$\theta_k[n+1] = E\left[\omega - \Delta\widehat{\omega} : \omega \in V_k[n], \Delta\widehat{\omega} = AQ(\omega - \theta_k[n])\right]. \quad (5)$$

A similar derivation is obtained if the squared error $\|.\|^2$ is weighted.

2.3. Codebook search

The M-L search algorithm [5] is chosen for codebook search in each split. We extract M survivors in the first stage with respect to mean square error (MSE), apply M parallel second stage quantization trials, and then select the codewords giving the best reconstruction with respect to the weighted mean square error (WMSE). The resulting performance is sub-optimal, but close to that of full search for reasonable values of M.

The codeword selection in each split relies on an *ad hoc* quadratic measure, which approximates the spectral distortion due to the difference between ω and $\omega_q = \widehat{\omega} + \Delta\widehat{\omega}$:

$$d_W^2(\omega, \omega_q) = (\omega - \omega_q)^T W (\omega - \omega_q) \quad (6)$$

The weighting matrix W is a diagonal here, so we obtain

$$d_W^2(\omega, \omega_q) = \sum_{i=1}^{8} W_i \left(\Delta\omega_i - \Delta\widehat{\omega}_i\right)^2. \quad (7)$$

A similar measure is used for the search in the second split. Intuitively each weight W_i should measure the spectral sensitivity of the original value ω_i. The choice of weighting is usually based on comparative tests between several practical measures, although some theoretical results [8] are available.

3. LATTICE QUANTIZATION REVISITED

Lattice quantization generalizes uniform scalar quantization and is efficient for high-resolution sources. A lattice populates the whole space with an infinite array of points [9].

Once scaled, truncated and shifted, it produces a lattice codebook. Truncation is also called shaping. This paper focuses only on the Gosset lattice [9]

$$RE_8 = 2D_8 \cup \{[1\,1\,1\,1\,1\,1\,1\,1] + 2D_8\}, \tag{8}$$

where

$$D_N = \left\{ (x_1, \cdots, x_N) \in \mathbb{Z}^N : \sum_{i=1}^{N} x_i \text{ is even} \right\}. \tag{9}$$

3.1. Fixed-rate lattice quantization

Truncating or shaping a lattice is not trivial in a fixed-rate lattice codebook design. Ideally the codebook boundaries should wrap the stationary signal distribution without too many outliers. However, a sophisticated shaping technique may have a prohibitive complexity, so the easiest solution is often to apply a "hard" shaping which enables fast search and indexing (for instance a pyramidal, spherical or Voronoi shaping). Spherical shaping is suitable when the distortion measure is quadratic, and it allows to employ an elegant indexing algorithm based on the concept of absolute and signed leaders [10].

3.2. Variable-rate lattice quantization

A "good" lattice yields a low granular quantization distortion within the shaping boundary, assuming that the resolution is high enough. In some applications such as transform coding or spectral quantization, we may break the constraint of fixed-rate quantization in order to distibute a constant total bit budget over different components. Embedded algebraic vector quantization (EAVQ) [10] is a self-adaptive quantization technique, and it is well-suited here.

Table 1: Lattice codebook rates in adaptive strategy.

A: 4-bit increment

r_1	rate of AQ_1 [bits]
00	12
01	16
10	20
11	unused

B: 2-bit increment

r_1	rate of AQ_1 [bits]
00	14
01	16
10	18
11	20

Using the embedding principle [10], we can design RE_8 codebooks having a bit rate multiple of b. The choice of the integer bit increment b depends on how finely bits should be allocated. Table 1 describes the bit rate of AQ_1 in the adaptive strategy. For instance in encoding A, AQ_1 can receive 12, 16 or 20 bits. If the total bit budget for lattice quantization is 32 bits, AQ_2 receives implicity the rest of bits (20, 16 or 12 bits respectively).

3.3. Lattice quantization for a weighted measure

A weighted mean square error is used in spectral quantization to approximate perceptual distortion, while lattice codebook search usually minimizes the Euclidean distance. We propose here an algorithm inspired from algebraic bounded-distance decoding which aims at reducing this target mismatch.

Given a positive definite weighting matrix W, an input value $\Delta\omega$ and a lattice codebook $\mathcal{C}$,

1. Find the nearest neighbor $\Delta\overline{\omega}$ of $\Delta\omega$ in $\mathcal{C}$ with respect to the Euclidean distance and compute the weighted squared error

$$d_W^2(\Delta\omega, \Delta\overline{\omega}) = (\Delta\omega - \Delta\overline{\omega})^T W (\Delta\omega - \Delta\overline{\omega}). \tag{10}$$

   This procedure is equivalent to a hard decoding or a preselection phase.

2. Select the winner $\Delta\hat{\omega}$ in $\mathcal{C}$ with respect to the weighted distortion d_W^2 inside an Euclidean sphere $\mathcal{S}$ centered at $\Delta\overline{\omega}$, *i.e.* search around $\Delta\overline{\omega}$ for a better candidate $\Delta\hat{\omega}$.

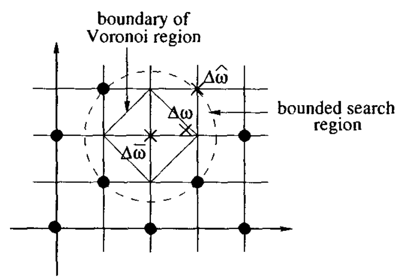

Figure 3: An example of suboptimal D_2 lattice search : preselection $\Delta\overline{\omega}$ and corrected point $\Delta\hat{\omega}$ for $\Delta\omega = [2.8, 2.1]^T$. The weighting is $W_1 = 2$ and $W_2 = 1$.

The bounded search region $\mathcal{S}$ is restricted here to the 240 points in the first shell of RE_8, but it can be easily extended to several low-energy RE_8 shells. These 240 points are generated by all permutations of the signed coordinates of $[2\,2\,0\,0\,0\,0\,0\,0]$ and $[1\,1\,1\,1\,1\,1\,1\,1]$. Algebraic properties can be used to reduce the additional search complexity. The indexing is unchanged.

4. EXPERIMENTAL RESULTS

The stochastic-algebraic quantizer shown in Figure 1 was trained with a database containing 74,000 vectors. The predictive coefficients were extracted every 20 ms with frames of length 30 ms, a Hamming windowing, a pre-emphasis factor of 0.75, a white noise correction factor 1.0001 and a 60 Hz bandwith expansion. The inverse harmonic mean weighting was used for codeword selection as in [3, 4].

The performance of spectral quantization is presented in terms of spectral distortion [2]

$$SD = \sqrt{\frac{1}{\delta\omega} \int_{\omega_-}^{\omega_+} \left(10 \log_{10} A(\omega)^2 - 10 \log_{10} A_q(\omega)^2\right)^2 d\omega} \tag{11}$$

where $A(\omega)$ and $A_q(\omega)$ refer to the original and quantized linear predictive spectra, ω_+ and ω_- correspond to 50 Hz and 7000 Hz respectively, and $\delta\omega = \omega_+ - \omega_-$. It provides a basis for benchmarking, even if it is a flat measure in the frequency domain and does not match the behavior of the human ear. The results are summarized in Table 2 and were obtained from a speech test database of 24,000 vectors. M_1 and M_2 indicate the number of survivors used in M-L search in the first and second split, respectively. Note that a LSF reordering forced the stability of all the synthesis filters if needeed, therefore no "unstable" filter was discarded in the statistics.

Table 2: Distortion statistics over the 50 Hz–7 kHz band.

Fixed bit allocation ($M_1 = 8$, $M_2 = 4$)

Rate [bits]	Code sizes [bits] k_1, k_2, k_3, k_4	$\overline{SD}$ [dB]	2–4 dB [%]	> 4 dB [%]
44	$7, 5, 16, 16$	1.05	2.07	0.008
45	$7, 6, 16, 16$	1.02	1.37	0.004
46	$7, 6, 17, 16$	0.98	1.13	0
47	$7, 6, 18, 16$	0.95	0.94	0

Adaptive bit allocation ($M_1 = 8$, $M_2 = 4$)

Rate [bits]	Code sizes [bits] $k_1, k_2, k_3 + k_4, r_1$	$\overline{SD}$ [dB]	2–4 dB [%]	> 4 dB [%]
47	$7, 6, 32, 2$ (A)	0.97	0.91	0
47	$7, 6, 32, 2$ (B)	0.95	0.62	0

It was found that a lattice codebook search with respect to WMSE brings a negligible improvement (near 0.01–0.02 dB in $\overline{SD}$) over MSE search, even for extended bounded-search regions. Therefore their mismatch is not significant.

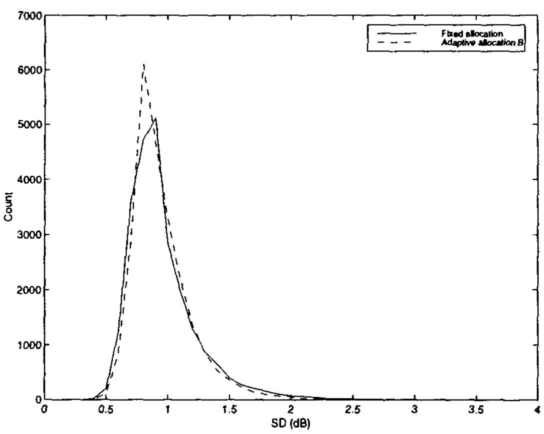

Figure 4: Histograms of spectral distortion at 47 bits.

Figure 4 shows that for speech signals the adaptive allocation reduced slightly the amount of outliers. However, the potential advantage of variable-rate lattice quantization was not fully exploited because of some severe constraints, as constant total rate and limited range in bit allocation.

5. CONCLUSIONS

A hybrid two-stage split quantizer scheme was presented for the wideband LSF parameters. Although its design was focused on speech signals, the use of algebraic codebooks made the system versatile. A variable allocation between splits improved the quantization performance slightly. It would be interesting to extend the algebraic quantizers to cover a wider range of bit allocations, or to allow the whole quantizer to produce a variable rate.

For moderate-to-high rates and dimensions, algebraic codebooks enable less computations and less storage than optimal random codebooks, but they require more program lines. They have another advantage which was exploited here: scalability.

ACKNOWLEDGEMENTS

The authors would like to thank Vesa T. Ruoppila for reviewing carefully this paper and for some helpful comments.

REFERENCES

[1] F. Itakura, "Line spectral representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Amer.*, vol. 57 Suppl. 1, pp. S35 (A), 1975.

[2] 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.

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

[4] S. Ragot, J.-P. Adoul, R. Lefebvre and R. Salami, "Low complexity LSF quantization for wideband speech coding," *IEEE Workshop on Speech Coding*, Porvoo (Finland), June 1999.

[5] 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.

[6] J. Pan and T. Fischer, "Two-stage vector quantization-lattice vector quantization," *IEEE Trans. Information Theory*, vol. 41, pp. 155–163, 1995.

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

[8] W.R. Gardner and B.D. Rao, "Theoretical analysis of the high-rate vector quantization of LPC parameters," *IEEE Trans. Speech Audio Processing*, vol. 3, no. 5, pp. 367–381, 1995.

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

[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.