QUANTIZER DESIGN AND EVALUATION PROCEDURES FOR HYBRID VOICE CODERS

A. Perkis and D. Rowe

Department of electrical and computer engineering
The University of Wollongong

Digital communications group
South Australian Institute of Technology

ABSTRACT - This paper presents a complete methodology for designing Max_Lloyd optimized quantizers (Max-quantizers) for a given parameter. The main emphasis is concentrated on estimating the parameters Probability Density Function from a carefully chosen database. Examples are given throug optimization of a CELP based voice coder.

Designing a low bit rate voice codec for any given application requires two distinctly different steps: Deriving a suitable coder model and designing quantizers for the model parameters. Most medium to narrow band speech coder operating at bit rates in the range 4-8 kbit/s rely on a combination of waveform coding and analysis-by-synthesis techniques (vocoders). These so-called hybrid coding schemes will generally transmit information describing the waveform in addition to parameters obtained according to a model of the human speech production mechanism. These consist of spectral parameters and gain factors which we will denote by the coding scheme's side information. Quantization of the side information is crucial for the coders performance, and should be optimized for each individual coding scheme. Max (Max 1960) quantizers give the optimum decision levels and representation values for these model parameters based on an estimate of their Probability Density Functions (pdf).

This paper gives a thorough description of the design and performance evaluation of scalar quantization based on the Max optimization scheme. Although this algorithm is well described mathematically in a number of texts (Max 1960) and (Jayant & Noll 1984) the methodology of the design and evaluation is somewhat poorly documented. As an illustration of all the necessary procedures, the optimization of a Code Excited Linear Predictive (CELP) derived coder is used as a case study throughout the paper. However, the methods can in general be applied to any coding scheme where scalar quantization of parameters is required.

QUANTIZATION

Amplitude quantization is the procedure of transforming a given signal x(n) at time n into an amplitude y(n) taken from a finite set of possible amplitudes (Jayant & Noll 1984). In a voice codec this procedure is split into two fundamental processes; quantization and dequantization. Quantization describes the mapping of the amplitude continuous value into an L'ary number k, the quantizer index.

The signal amplitude x is specified by index k if it falls into the interval

$$F_k : \{ x_k < x < x_{k+1} \}; k=1,2, \ldots L \qquad (1)$$

As in (Jayant & Noll 1984) we will refer to x_k as the decision levels F_k, the decision interval and y_k the representation value. In general the quantization takes part in the coder resulting in a number of quantizer indices which are transmitted to the decoder, typically in binary format. At the decoder the L'ary number k is mapped into representation level y_k known as the dequantization.

Unlike the sampling process, quantization inherently introduces a loss of information, which we will refer to as the quantization noise. The quantization noise will be denoted q and is given in Eq. 2.

$$q = x - y_k \qquad (2)$$

There are two fundamentally different quantization schemes to consider uniform and non uniform quantization. Uniform quantizers have decision intervals of equal length Δ and the reconstruction levels are the midpoints of the decision interval. Although being the most common means of quantization, and also the conceptually simplest scheme for implementation, uniform quantizers are

generally not optimal in the sense of minimizing the quantization error. A smaller error variance can be achieved by choosing smaller decision intervals where the probability of occurrence of the amplitude x is high, ie. where the pdf, $P_x(x)$, is high, and by choosing larger decision intervals otherwise. This type of quantizers are often referred to as pdf-optimized quantizers or Max-quantizers. At this instance it is important to note that this procedure makes the magnitude of the quantization error depend strongly on that of the input x.

MAX-QUANTIZERS

To be able to formulate the exact design criterion for the pdf - optimized quantizer we will assume the input x to be a stationary zero mean random sequence with variance σ_x^2. For this case it can be shown that the quantization noise will also be a zero mean random process with variance σ_q^2. In (Max 1960) and (Jayant & Noll 1984) the necessary conditions for minimizing σ_q^2 by a set of x_k; k = 2,3 .. L and y_k; k= 1,2 .. L are.

$$\frac{\partial \sigma_q^2}{\partial x_k} = 0: \qquad k=2,3 \ .. \ L \tag{3}$$

$$\frac{\partial \sigma_q^2}{\partial y_k} = 0: \qquad k=1,2 \ .. \ L \tag{4}$$

This equation has no explicit solution for L>3. However the optimum values x_{kopt} and y_{kopt} can be calculated iteratively according to Eqs. 5 and 6 resulting in so called Max-quantizers.

$$x_{kopt} = 1/2(y_{kopt} + y_{k+1opt}); \qquad k=2,3, \ .. \ L \tag{5}$$

$$x_{1opt} = \infty, \ x_{Lopt} = -\infty$$

$$y_{k,opt} = \frac{\int_{x_{kopt}}^{x_{k+1opt}} x \, p_x(x) \, dx}{\int_{x_{kopt}}^{x_{k+1opt}} p_x(x) \, dx} \tag{6}$$

Using Eqs. 5 and 6 quantizers can be designed by using a database of the parameter in question. The first equation states that the decision levels are half way between neighbouring reconstruction values, and the second equation states that the reconstruction value should be the centroid of the pdf in the appropriate interval.

The rest of this paper will deal with the methodology of designing such quantizers, and will cover such issues as how to estimate the pdf, evaluate the results and finally lead to an engineering approach of fixing the bit rate for a given voice coder.

PRACTICAL QUANTIZER DESIGN

As an illustration of the necessary procedures for designing Max-quantizers the optimization of our Bit Error Masked - Code Exited Linear Predictive coder (BEM-CELP) (Rowe & Perkis 1990) will be used as a case study.

Coder specifications

The speech coder model is based on Linear Prediction, and an analysis-by-synthesis approach for optimizing the innovation sequence for the residual signal. The coder transmits five distinct parameter groups representing the long term spectral information ($\{a_i\}$), the fine spectral structure (L,β) and the residual signal (α,i) as indicated in Figure 1.

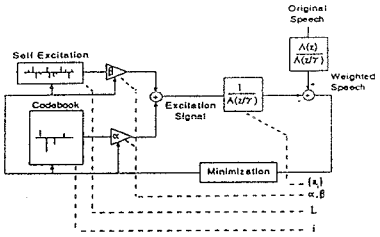

Figure 1. Block diagram of the BEM-CELP

The SESTMP is a block based coder, where the parameters are calculated for each input frame of speech, in our case consisting of 240 samples (30 msec.). For each frame a set of LPC parameters is estimated giving the filter coefficients {ai}. This frame is then split into six equally sized subframes of 40 samples (5 msec.), each of which are matched in an analysis-by-synthesis manner by a self excitation sequence and a codebook vector.

Training and test databases

The design of Max-quantizers will be based on the iterative solution given in Eqs. 5 and 6. A crucial design parameter in this is the pdf of the parameter in question. In order to be able to estimate this pdf we need to design a database of actual values of our parameter. To get an accurate estimate it is important that our database, which we will denote the training database, be large enough, and also representative of the real data we intend to code.

The criterion of choosing a large enough training database is met by ensuring that we have a large amount of values resulting in statistically valid estimates. To give a rough idea of how many large enough would be, consider the two crucial parameters to be decided in estimating the pdf, M, the total number of values, and N, the number of decision intervals, in our histogram. As a rule of thumb it is considered necessary to have at least 1000 entries in each interval, requiring a total of 1000•N values. However, considering that N should be around 200, in practical designs this is not always feasible and smaller subsets will be sufficient. As an example consider the 5th LSP frequency of our BEM-CELP coder. Two different training databases for this parameter have been obtained by running two different sets of speech through the coder logging the value for each frame. Set number 1 consist of app. 22 minutes of speech (44.000 frames, 22 speakers) while set number 2 only consists of 24 seconds (800 frames, 4 speakers). The estimated pdfs, using N=100, for the two training databases are shown in Figs 2 and 3 respectively.

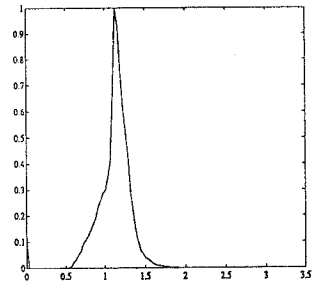

Figure 2. Estimate of pdf using training set 1

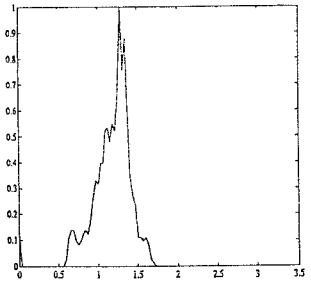

Figure 3. Estimate of pdf using training set 2

As we can see the smaller training set gives a poor estimate of the parameters pdf and thus is unsuitable for quantizer design.

Choosing a representative training database will reflect on the performance of our quantizers. In designing a voice codec the predominant input signal is obviously going to be speech. In other words the training database should reflect the variety of speakers one would expect to use the system. It is therefore important that both sexes and a wide range of different speakers and utterances are represented in the database.

The test database is needed for evaluation purposes. Ideally the test database should consist of both different speakers and different utterances from that of the training database. As this is often difficult to achieve a good compromise is to use a test database with a subset of the same speakers but different utterances. The size of the test database can be significantly smaller than the training database.

Design of a practical training and test database

The speech material used for designing the training database consist of;

22 speakers, 10 male, 10 female, a male child and a female child
60 seconds each of phonetically balanced Harvard speech sentences

Giving a total of 44.000 sets of LPC parameters and 264.000 gain values.

In the design a 10 th order LPC analysis was used. Note that new quantizers have to be designed for every change in the coder model, as this will change the values in the training database and hence the pdf.

Quantizer evaluation

Two standard ways of evaluating the designed quantizers are considered; The log RMS spectral distortion (SD) and evaluating the distribution of the quantization error. For spectral quantizers it is generally accepted that the SD, given in Eq. 7

$$SD = \sqrt{\frac{1}{\Pi}\int_0^{\Pi} [10\log\frac{S(\omega)}{\hat{S}(\omega)}]^2 \, d\omega} \quad ; \; S(\omega) = |H(\omega)|^2 \qquad (7)$$

where H(w) is the short term speech spectrum, is the best measure for spectral distortion. By ensuring that the SD is below 1 dB, the just noticeable limit, the degradation caused by quantization will be non-detectable. Thus for quantization of the LPC parameters we will only allocate the sufficient amount of bits to get the SD below 1 dB.

For the gain quantizers we will allocate bits according to the pdf of the quantization error. The quantization error is given in Eq. 3 . For uniform quantizers with decision intervals Δ the quantization error q will be bounded by $\pm\Delta/2$. Assuming that we use enough decision levels it seems reasonable to assume that all amplitudes within $\pm\Delta/2$ will occur with the same probability. This means that the pdf of the quantization error will be constant over the interval indicating that the quantization error will in fact be white noise., In (Jayant & Noll 1984) a much more rigourous treatment of the matter is given and the theory is generalized to hold for non-uniform quantizers. However, for a practical design this means that choosing enough bits for your quantizer means ensuring that the estimated pdf of the quantization error (using the test database) should be maximally flat.

SPECTRAL QUANTIZATION

Line Spectrum pairs are an alternative representation of the LPC parameters, and are thoroughly described in (Soong & Juang 1984). The LSP coefficients have some important properties such as limited dynamic range and natural ordering. In addition the LSPs provide efficient control with the shape of the speech spectrum, and allow a simple check for filter stability. Using the training database, 2,3,4 and 5 bit Max-quantizers have been designed for each LSP frequency, based on their individually estimated pdfs. Using the SD measure the chosen bit allocation of 3,3,4,4,3,3,3,3,3,3 bits per frequency gives a 1.01 dB distortion.

GAIN QUANTIZERS

In the BEM-CELP the excitation signal is built up as a codebook vector with a stochastic gain factor, α, and a self excitation sequence with gain factor, β. The resulting code vector models the overall shape of the residual signal, while the gain factors give the correct energy matching. In order to design Max-quantizers for the two gains, their pdfs are estimated based on our training database. The pdfs are shown in Figures 4 and 5 respectively. Included in the figures are the pdf estimates of the test database. The resemblance of the two prove that the training database consists of a representative cross section of speech.

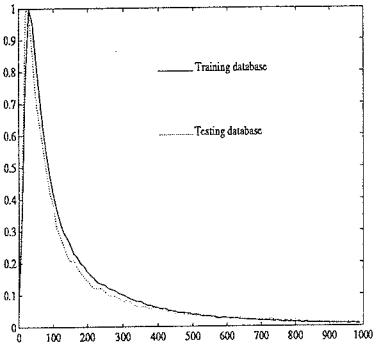

Figure 4 Probability density function for α

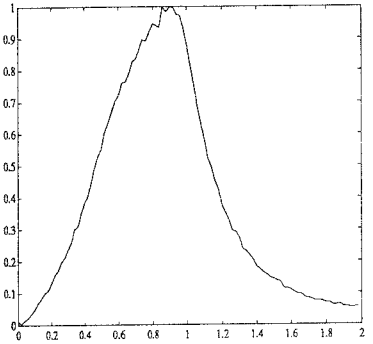

Figure 5 Probability density function for β

The pdf for β allows for straight forward design of 2,3,4 and 5 bit Max-quantizers while the pdf for α illustrates a classic trade off problem in its large dynamic range. In allowing for the large dynamic range we would require a large number of bits, or get a poor overall performance. If however, we allow for some overload distortion, ie. clipping the pdf, we could design a quantizer with a better overall performance. Unfortunately the latter case introduces a theoretical problem. In order to satisfy the conditions of finding a global optimum using Eqs. 5 and 6, the pdf must be univariate (have a single peak), which means clipping in reality introduces the possibility of designing sub-optimal quantizers.

For a practical design a good approximation to a univariate distribution would be if the magnitude of the tails in the pdf after clipping are sufficiently small in comparison to the peak of the pdf. Figure 6 shows the pdf of α clipped at three different levels with; a) too small and b) correct dynamic range. Choosing the clipping of Figure 6b), 2,3,4 and 5 bit Max-quantizers have been designed for α.

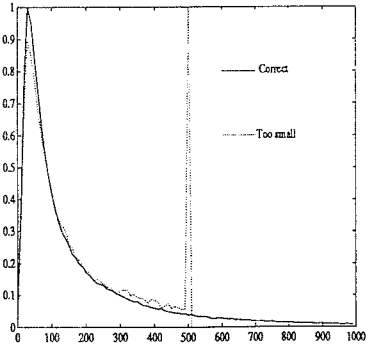

Figure 6. Various levels of clipping for α

Figure 7 shows the error distribution for 3,4 and 5 bit quantization of the test database.

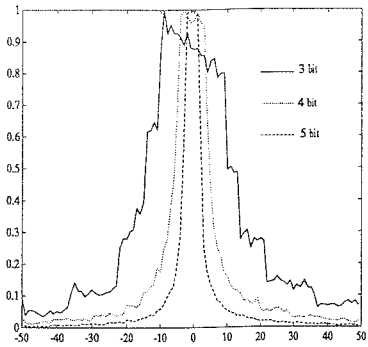

Figure 7. Distribution of the quantization error for 3,4 and 5 bit quantization for α.

To obtain a reasonably uniform error distribution function for α, indicating sufficient number of levels in the quantizer, it seems necessary to use at least 5 bits per frame . Similarly we need to use 3 bits per frame for β. From the figure we notice a slight slant in the error distribution, especially for the 3 bit quantizer. This is due to the prominent offset between the training and the test database as can be seen in Figure 6. With this fact recognized we accept the pdf of the quantizer to be flat for the 5 bit quantizer. Marginally we might also accept the 4 bit quantizer to give a flat distribution. The final decision however, has to be based on some subjective criterion, ie. actually listening to the coded speech.

CONCLUSIONS

In this paper we have presented a complete methodology for designing Max-quantizers for various model parameters in low bit rate voice coders. As a case study we have fully quantized a CELP based coder, designing quantizers for the spectral parameters (LSPs) and two gain factors (α and β). The chosen bit allocation for the LSP frequencies and the two gain factors, together with the 12 bits for the residual signal gives a coder operating at 5333 bits per second. The unquantized coder has a segmental SNR of 8.20 dB while the quantized version gives 7.20 dB giving a reduction of 1 db resulting in hardly no noticeable degradation, proving the success of our quantizer design.

REFERENCES

Jayant, N.S. and Noll, P. (1984) *Digital coding of waveforms,principles and applications to speech and video*, Prentice Hall, New Jersey.

Max, J. (1960) *Quantizing for minimum distortion*, IRE Trans. on Information Theory, pp 7-12.

Rowe, D. and Perkis, A. (1990) *Error masking in a real time voice codec for mobile satellite communication* This conference

Soong, F. K. and Juang, B. H. (1984) *Line Spectrum Pairs (LSP) and data compression* Proc. ICASSP.