ECE 508

Project 1

SQ and VQ Design for Quantization of Speech Samples

Brankov Jovan

Code writing

It is essentially important to write a code for the MATLAB in an efficient way. The code performance (computation time) can be improved by a factor of 100. The main point is avoiding *for* loops and exploiting as much as possible matrix manipulation functions built in the MATLAB. By doing so one can achieve better performance (because of highly optimized MATLAB code) even over the code writing in C or C++. In addition, all variables should be declared in advance of its biggest size. This need to be pointed out because the MATLAB does not require neither of above, but these can lead to poor/good execution time.

Table 1. Execution time for some VQ parameter values

k\r	1	2	4	8
1	2.93 sec	5.01 sec	15.85 sec	999.31 sec
2	5.15 sec	17.98 sec	1029.18 sec	
3	8.14 sec	94.70 sec		
4	16.95 sec	930.12 sec		
8	1748.11sec			

These results were achieved on Pentium II 200 MHz with 128Mb. The memory used was less then 10Mb although all samples were converted from 2 byte *integers* to 8 byte *doubles*. (Restriction of the MATLAB.)

Training data

Training set was chosen as a collection of sample vectors placed randomly in the training speech. Each sample vector consists of *k* consecutive samples, where *k* was the vector dimension. The idea behind it was to represent the whole given training set with fewer samples but still maintain the correlation between samples and diversity of the training set. The VQ does not utilize the sample vectors correlation, but the correlation among samples within single sample vector.

E.g. in one dimension case the *pdf* optimized SQ does not utilize sample correlation, opposed to VQ. Similarly the VQ also does not utilized between sample vectors correlation.

Training ratio was chosen to be 100 though it can be adjusted regarding the size of the code book, for the reduction of the GLA execution time.

Table 2. Number of required points for training ratio of 100 if not greater then maximum available number of 491520 samples.

k\r	1	2	4	8
1	200	400	1600	25600
2	800	3200	51200	491520
3	2400	19200	491520	491520
4	6400	102400	491520	491520
8	204800	491520	491520	491520

Table 3. Training ratio for given training set vs. different *k* and *r*

k\r	**1**	**2**	**4**	**8**	**12**
1	**100**	**100**	**100**	**100**	**100**
2	**100**	**100**	**100**	3.75	0.014648
3	**100**	**100**	40	0.0097656	2.38E-06
4	**100**	**100**	1.875	2.861E-05	4.37E-10
8	**100**	0.9375	1.43E-05	3.331E-15	7.75E-25

Gray areas represent the areas where we can not achieve the required training ratio of 50.

Initialization

Because the improvement in SNR, caused by the different initialization scheme, can be about 0.1 dB I did not explore the initialization strategy from that point, but in order to obtain faster convergence of GLA. I explored just three from originally five that I was prepared to explore. Those are: initialization by uniform random variables in the region between minimum and the maximum of the training sample values, initialization from samples data by randomly assigning of partition regions and from lower order VQ code book. The last two planned algorithms were: initialization along the first Karhunun-Loeve component and by drawing the samples from data estimated multivariate Gausian/Laplacian distribution. The reason for poor interest for better initialization was in good performance of GLA code and fast computation time. In all experiments the randomly generated initial code book was used. Usually the results were obtained after less then 20 iteration. The threshold for stopping criteria was the changing in SNR less then 0.01dB.

Empty cell

Elimination of the empty cells was performed by taking point from the most populated cells with probability of 20% and reassigning to empty cell.

In the experiments beginning, to avoid the big diversity in number of points per cell, the minimum number of points per cell was set to be N/K*10%, where N is number of training points and K number of cells. Because the SNR was not significantly improved and the execution time was longer, the minimum number of samples per cell was later fixed to be 3 samples/cell.

Merit of quality

For the purpose of quantitative measurement and comparison of the obtained speech, from different quantization approach, we used the SNR defined as:

$$SNR = 10 * \log 10 \left(\frac{\sum_{j}^{N} \left(x_j \right)^2}{\sum_{i}^{N} \left(x_i - \hat{x}_i \right)^2} \right)$$

N-number of samples in the given set, x_i - ith sample in the set and $\hat{x}_i$ - ith estimate of the set point.

Speech "*pdf*"

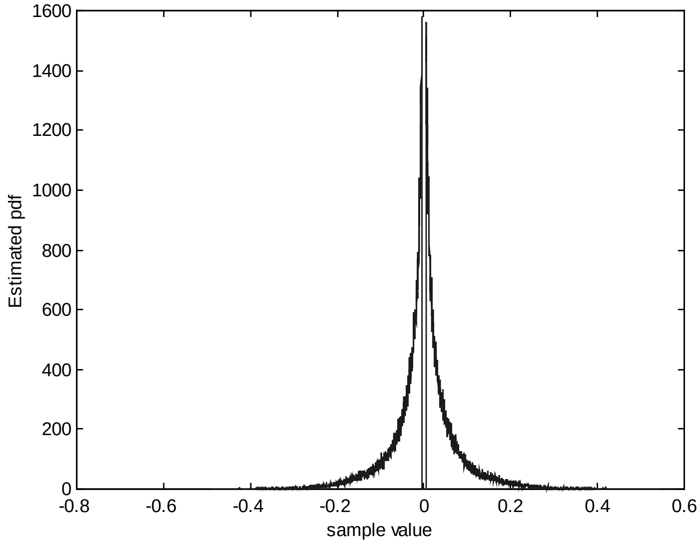

Figure 1. Estimated histogram of sample distribution for sample values >0.05

The purpose of Figure 1. was to achieve some kind of feeling for *pdf* shape of the speech for possible comparison with Code book points distribution.

VQ

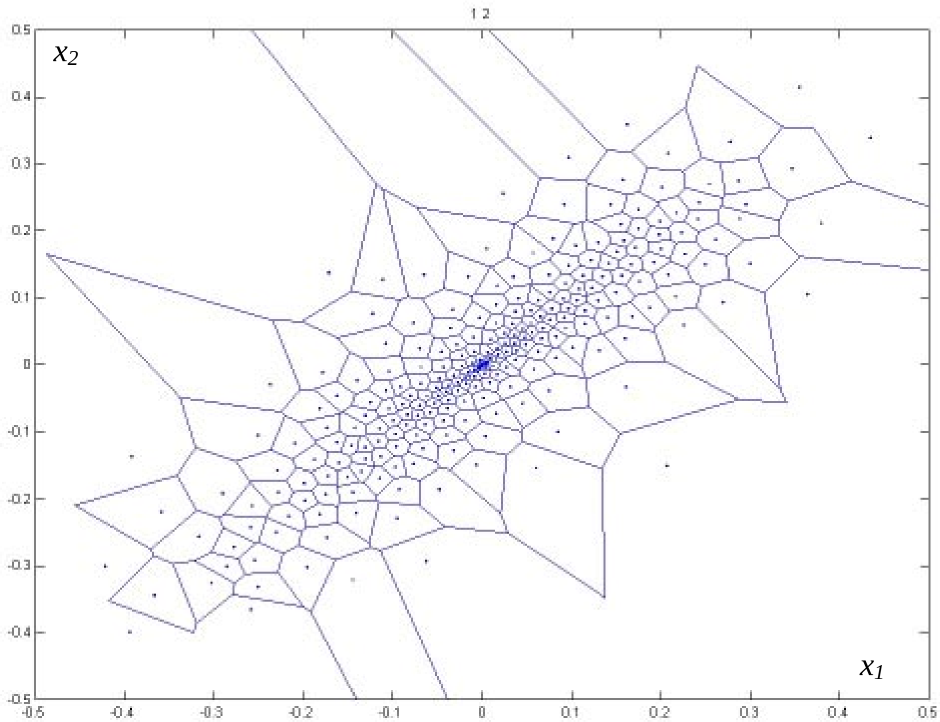

Figure 2. Voronoi cell plot for r=4, k=2

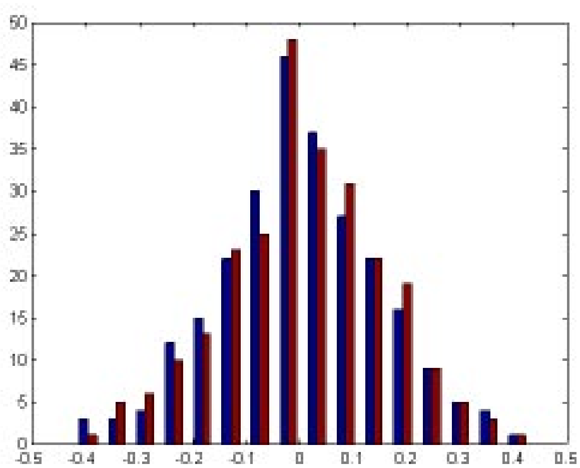

Figure 3. Estimated histogram of Code book values, x_1 x_2 for r=4, k=2

By looking at Code book points distribution over the x_1 x_2 domain one can notice the correlation between the code points. This observation is in good agreement with what we know about the speech. This can be seen also clearly in Figure 4. The correlation between the samples is weaker for samples that are separated by more samples, (compare view 1-2 to 1-4). Note that the circular shape of code points mean less correlation between points.

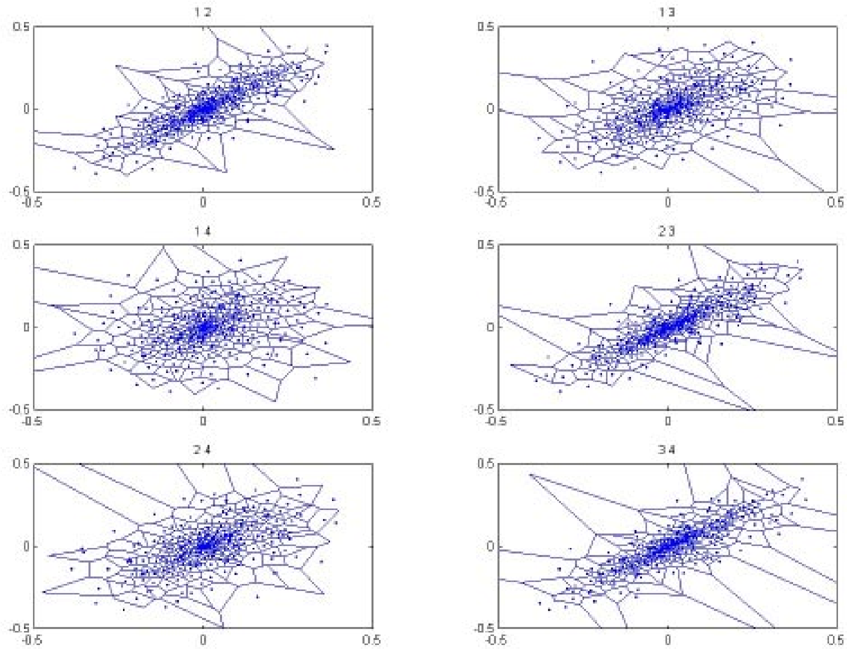

Figure 4. Estimated Voronoi plot of Code book values in pairs $x_1, x_2 \quad x_1, x_3 \quad \cdots \quad x_3, x_4$ for r=2,k=4

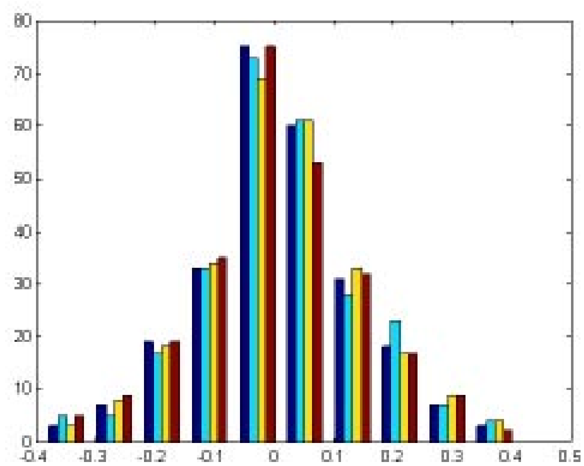

Figure 5. Estimated histogram of Code book values, $x_1 \quad x_2 \quad x_3 \quad x_4$ for r=2, k=4.

Note also that the shape of the estimated histograms of the Code book points “follow” the histogram of the speech.

Uniform SQ

To determine the step size for the given training set, and bit rate a program that perform minimization of SNR regarding the step size was written. The step size obtained by this approach was then used to obtain results for the test sample set. The result presented here, for the test set, was obtained by symmetric, mid-rise uniform quantizer.

By experimenting with mid-tread and mid-rise quantizer, the mid-rise was chosen because the quantized speech, although it is noisier, appear, subjectively, more pleasant then mid-tread. Probably the non-symmetric quantizer can be also used because the *pdf* of given speech was not symmetric.

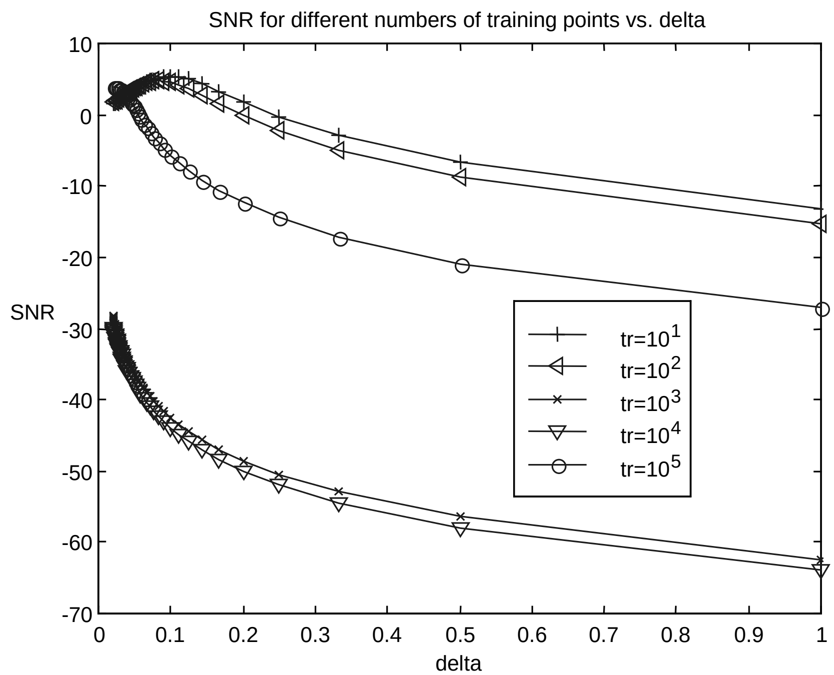

Figure 6. Importance of sufficient training points for optioning good delta. Uniform quantifier with *r*=2

In Figure 6. the values of SNR vs. the steps size of the uniform quantized, with the training ratios as a parameter, are shown. It can be seen that the training rate should be maintained as high as possible to obtain accurate results.

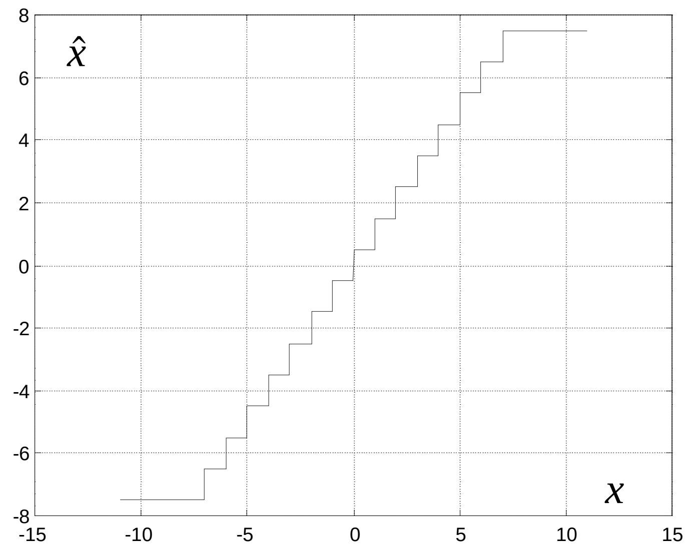

Figure 7. Mid-rise uniform quantizer for delta=1, r=4

Table 4. The uniform quantizer step for different *r*

r	**1**	**2**	**4**	**8**	**12**
delta	0.0552	0.053	0.0288	0.0034	2.33E-04

Table 5. SNR of uniform SQ using delta from Table 4.

r	**1**	**2**	**4**	**8**	**12**
SNR [dB]	1.32	3.96	11.44	27.29	32.71

Note that by using the step size different than the one obtained from training set, one can obtain better results for high-resolution uniform SQ. This comes from difference in the training set and testing set sample distribution.

Table 6. Improved SNR for *r*=8,12

r	**8**	**12**
Step	0.0044	2.97E-04
SNR [dB]	34.13	58.02

For mu-low 8 bit quantized speech the SNR is 37.21dB.

Comparison

Table 7. VQ SNR of test set for different k, r

k\r	1	2	4	8
1	2.1277	6.443	15.682	26.5794
2	5.3143	9.2324	18.1099	
3	5.6	10.7851		
4	6.2236	11.555		
8	7.8528			

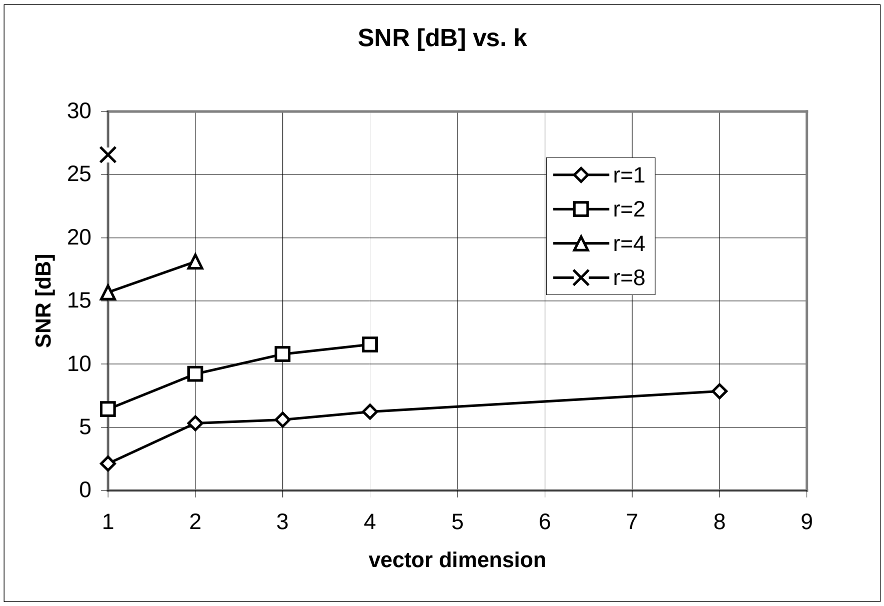

Figure 8. SNR for test set vs. *k*-vector dimension for VQ
the *r*-bit rate is a parameter

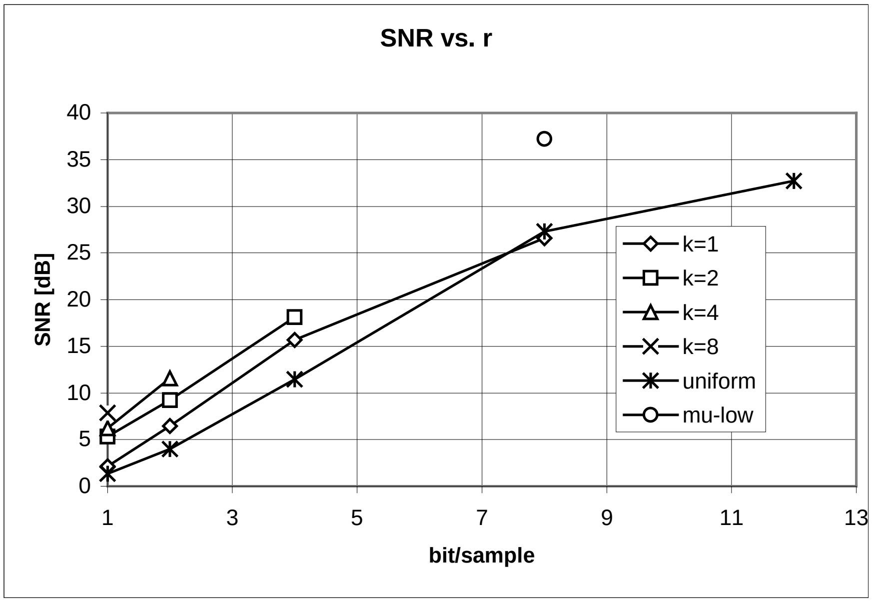

Figure 9. SNR of the test set vs. *r* for VQ and SQ (uniform and mu-low) the *k* is a parameter

For low bits rate uniform SQ is approaching the VQ for *k*=1, which is *pdf* optimized quantizer, because of optimization procedure in choosing of the delta by minimization SNR. Also one can notice that for high resolution there is also small difference between the VQ *k*=1 and uniform SQ.

By listening the speech one can qualify the difference, for the high-resolution quantizer, in the good agreement with the SNR. For the low resolution this is not straightforward. Probably different measurement that accounts for the human perception characteristic should be applied.

The difference between the estimated and the original speech becomes more uniformly distributed as the resolution and SNR increases.

For the training set the results in term of SNR are much better. This is the consequence of the optimization procedure for that particular sample set.

Image example

Figure 10. Image VQ for k=3 (RGB colors) and different r.
Upper left r=1/3, upper right r=1, lower left r=2 and lower right original

The SNR obtained for figure above were 14.1dB, 21.7dB and 29.06 dB respectively.