EQ2320 - Speech Signal Processing

Requirement Analyse and System Design

Final Report

Félix Côte Antoine Honoré

15 mars 2016

Table des matières

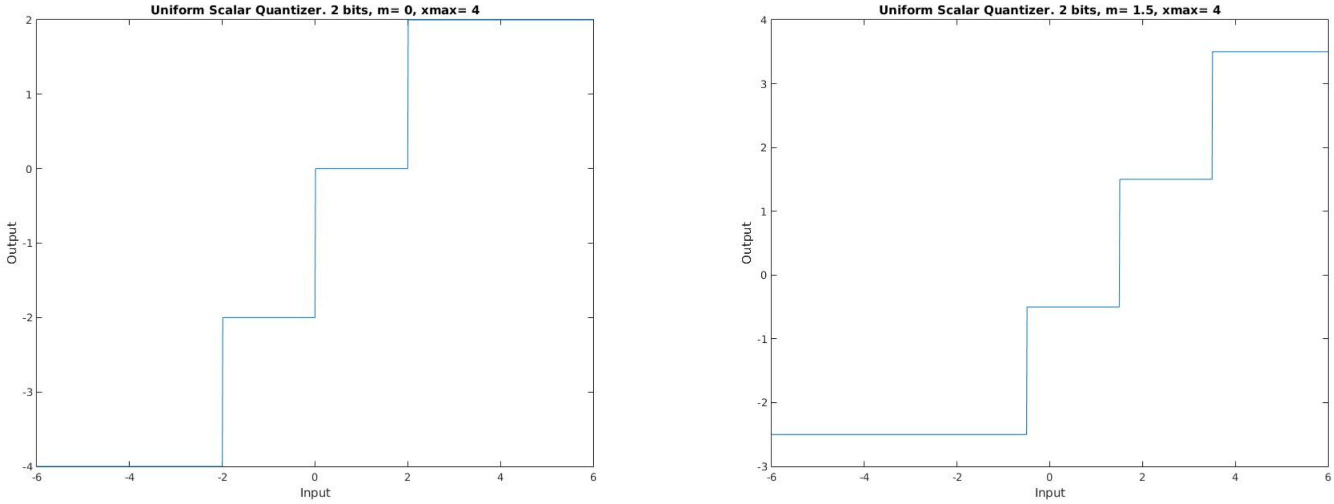
I Introduction

II Uniform Scalar Quantizer

In this part we implement the most basic quantizer. The USQ is entirely defined with three parameters :
- n_{bits}, the number of bits used to code one sample. $2^{n_{bits}}$ is the number of output value ;
- m, the mean of the output values ;
- xmax the maximum of the output values ;

In this part we tried m=0 and m=1.5. The result that we got plotting the input sigal versus the input signal is presented on figure 1.

To compare the two settings, we need to plot the distorsion-rate curve and compare the performance. This is presented on figure 2.



(a) m = 0

(b) m = 1.5

FIGURE 1 – Input vs Output

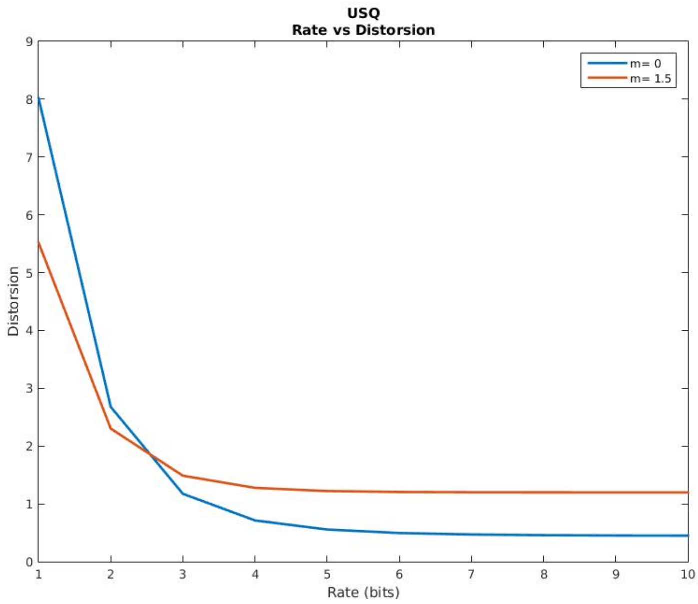


FIGURE 2 – Rate-Distorsion curve for two values of m.

III Parametric coding of speech

In this part, we will experiment with parametric coding. We use the vocoder provided with the project for optimal performance. With the analysis function we are able to get crucial parameters which describe the signal we want to transmit. This way we don't have to transmit the whole signal which require a high number of bits. In order to use fewer bits, we will quantize the parameters before transmission. The goal of this part is then to find a good ratio between quantization of the parameters and quality of the reconstructed received signal by the synthesis function. The bitrate during transmission should be as low as possible while the quality of the signal on the reception side is defined by the implementation (military, artistic, communication...).

In the following sections, we will analyse each parameter of the analysis function and try to find clever ways to quantize them. We use the signal called *male8*.

III.1 Quantizing the Gain

In this section, we work on the quantization of the gain.

FIGURE 3 – Histogram of the gain with mean and boundaries

We apply a uniform quantizer on the gain with different bitrates then we synthesize the new signal. We cannot hear the distortion above a 4 bits quantization. This means that we could use **4 bits window** to code the gain and the reconstructed signal would still have the same quality as the original signal.

We now compute and plot the histogram of the logarithm of the gain.

FIGURE 4 – Histogram of the log of the gain with mean and boundaries

We can see that the histogram here is more uniform. We can expect the quantization to better represent the variations of the log of the gain compared to when used on the gain only. We apply the same test as before and find that we can go as far as quantizing the gain with 3 bits and still not hear the distortion. We can observe the effect of these two methods of quantization by plotting the results along with the original gain (Figure 5).

FIGURE 5 – Comparison of original gain and quantized gains with 3 bits

With **3 bits per window** we have 8 levels of quantization. We see that using a log quantization gives us logarithmically distributed levels of quantization, therefore we are able to describe small variations in the lowest values. The direct quantization cannot do this.

In many ways, human speech and perception work logarithmically and not linearly. The pitch and the gain are examples of this. The quantization in the log domain is better for those parameters.

III.2 Quantizing the Pitch and Voiced/Unvoiced Decision

Pitch

As said previously, the pitch perception of human beings is logarithmic. Meaning that a logarithmic scale fits human pitch better than a linear one. We will then quantize the pitch with the same method as the gain. With this method, we are able to achieve a quantization using **4 bits per window** with little to no distortion.

Voiced/Unvoiced Decision

The quantization of this parameter is very simple. Essentialy because this vector is binary by definition therefore does not need to be quantized more. We can use only **1 bit per window** to code the values of this vector → bit set to 1 to indicate a voiced sample, bit set to 0 to indicate an unvoiced sample.

III.3 Quantizing the LP parameters

The quantization of the LP paramaters is inherently different than what has been done previously. We do not use a uniform scalar quantizer this time but instead use a vector quantizer (VQ). We use an order of 10 to compute the LP parameters. This means that we have to transmit a vector of 10 scalars for each window. Instead of coding the vector directly we will find within the VQ the closest vector (linked to a specific distance) and transmit its index. This means that the coder and decoder both have to possess the VQ. We use 2 VQs : one that codes the coefficients directly, and one that codes the residual (the error between the actual coefficients and the coefficients chosen by the coding function). This way we gain precision with only a small number of bits added.

The VQs store 1024 vectors of 10 coefficients. This means that any index within the list can be coded with 10 bits. We have to transmit 2 indexes for each window. Therefore we need **20 bits per window** with this method.

We code 2 functions, *encodefilter* and *decodefilter* that both take the VQs as input. The first function finds, for each vector of coefficients, the closest vector in the VQ1 using the euclidean distance. It computes the residual and codes it using VQ2 with the same distance. It outputs the 2 indexes for each window. The second function finds, for each pair of indexes, the corresponding set of coefficients and adds them together to get a single vector.

III.4 Optimizing the Bit Allocation

Now that we have quantized each parameter separately we can see the effect of quantizing all of them on the reconstructed signal. To sump up, here is a table of the bitrate for each parameter.

Number of bits	bits/window	bits/sample	kbits/sec
Gain (E)	3	0.0248	0.1987
Voiced/Unvoiced (V)	1	0.0083	0.0662
Pitch (P)	4	0.0331	0.2650
LP Parameters (A)	20	0.1656	1.3248
Total	28	0.2318	1.8547

FIGURE 6 – Table of bitrates for *male8* (length = 25242 samples = 3.15 sec)

We can see that the LP Parameters take indeed the largest portion of the bitrate.

We compute the SNR and get $SNR = 3.9dB$. SNR is a great tool to describe the distortion that additive noise add to a signal or to describe the distortion implied with quantization. However here we have to point out that we have quantized a set of parameters which served to synthesize a new signal. Eventhough we want this signal to be as close as possible to the original signal, it might have some clear differences (delay, amplitudes, periodicity...). Therefore the SNR here is not necessarily relevant to the quality of the signal.

IV Speech Waveform Quantization

IV.1 Evaluation of the optimal k

In this part we design a Uniform Scalar Quantizer (USQ) that is adapted to the speech signal. For that we define

$$xmax = k\sigma_x$$

where σ_x^2 is the variance of the speech signal. This kis to be calculate for everybit rate we choose. The value of k for a bitrate R = 3 is 3.4. The figure 8 shows a plot of SNR = f(k). The SNR has been evaluate with the following formula

$$SNR = \frac{\sigma_x^2}{\frac{1}{N}\sum_{n=1}^{N}(x_n - q_R(x_n))^2}$$

where N is the size of the input speech signal, x_n is the input speech signal, $q_R(x_n)$ is the input signal quantized with a bitrate R. Here R=3.

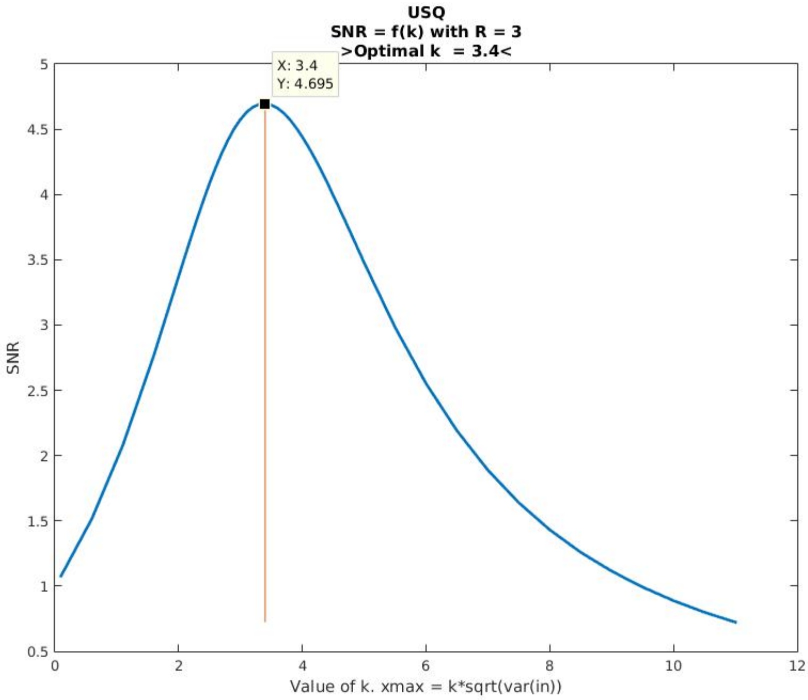


FIGURE 7 – SNR = f(k) for a bitrate R = 3

IV.2 Rate - SNR curve

The curve ploted on figure **??**, represents the rate versus the SNR_{dB}. The SNR_{dB} has beenevaluated with the following formula :

$$SNR_{dB} = 10log_{10}\frac{\sigma_x^2}{\frac{1}{N}\sum_{n=1}^{N}(x_n - q_R(x_n))^2}$$

where N is the size of the input speech signal, x_n is the input speech signal, $q_R(x_n)$ is the input signal quantized with a bitrate R.

IV.3 Quality of the quantized signal

A bit rate of 8 bits provides a good quality for the quantized signal.

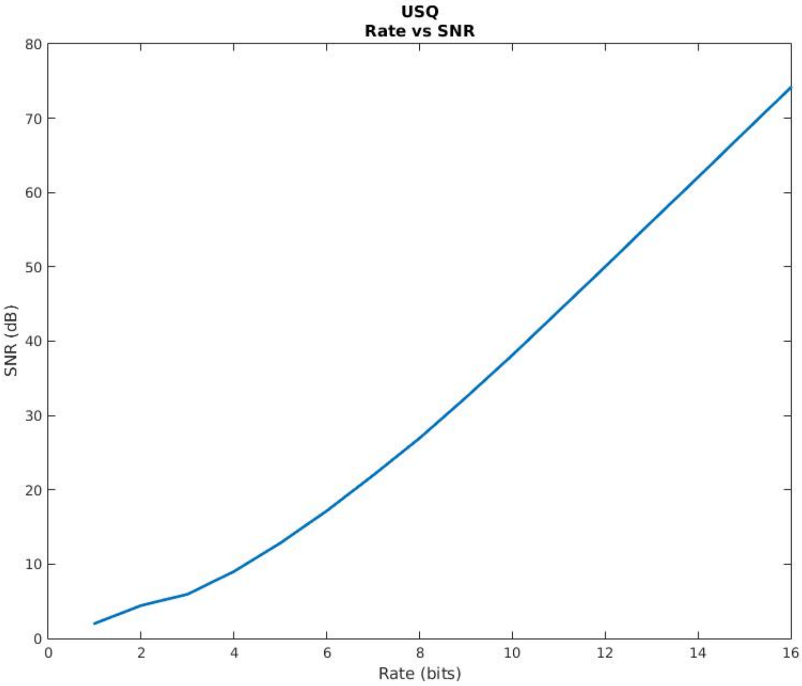


FIGURE 8 – Rate versus SNR curve, for rate = {1,2,...,16}

IV.4 Error signal

For a bit rate of 1 bit, the error signal contains of the information. It is easier to unserstand the message by listening to the error signal than listening to the quantized signal. The error signal is then highly correlated with the input signal. That compromises one of the fundamental assumptions when we want to remove an additive noise on a signal. On the otherhand when the rate is high, say 11 bits, the error signal sounds like a white noise and is consequently totally decorrelated with the input signal.

IV.5 OPTIONAL :

With a midtreat quantizer, the message is more understandable a low bitrate and at high bitrate there are no differences. The figure 9 shows the selected levels for a 2 bits midtreat quantizer.

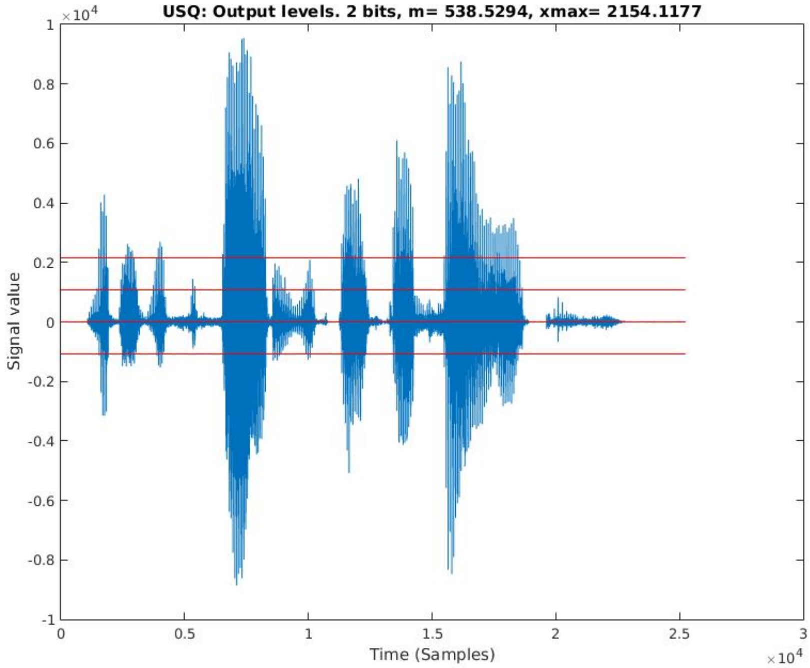


FIGURE 9 – Midtreat quantizer for R = 2

V Adaptive Open-Loop DPCM

In this section we have to implement an a Differential Pulse Code Modulator. The functional scheme is presented on figure 10.

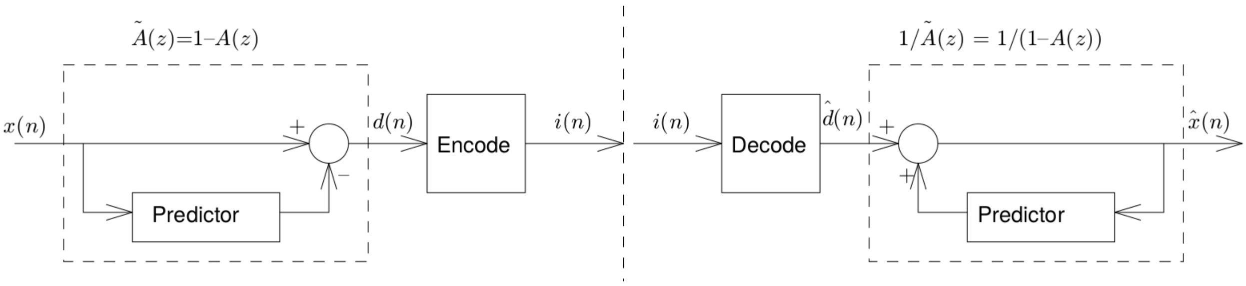


FIGURE 10 – Functional scheme of a DPMC

The idea here is to perform a linear prediction on an input signal, and to transmit the error signal with the prediction coefficient. It is not efficient to transmit a quantized input because the frames are strongly correlated between each other, as we can see on figure 11.

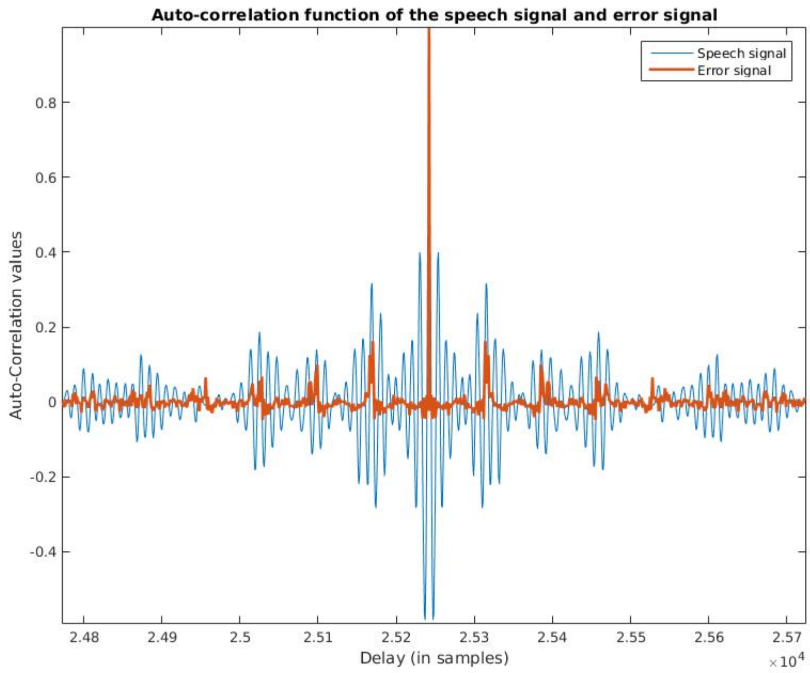


FIGURE 11 – Correlation function of the input speech signal.