Report for EQ2321 Speech and Audio Processing

EQ2321 Speech and Audio Processing, Project 2

Hanqi Yang hanqi@kth.se

Qian Zhou qianzho@kth.se

March 19, 2022

1 Introduction

In this project, we will mainly investigate how to design a uniform scalar quantizer, how to code speech parameters. We will also study speech waveform quantization and open-loop DPCM. We will discuss bit allocation for each parameter and evaluate SNR performance for both of them. Throughout this assignment we will use the speech data in speech8 which is sampled at 8 kHz. Speech8 will be used both for tuning (training, optimizing) and evaluating the speech coders.

2 The Uniform Scalar Quantizer

In this task we implement the uniform scalar quantizer by an encoder and a decoder. The uniform scalar encoder is entirely defined with four parameters:

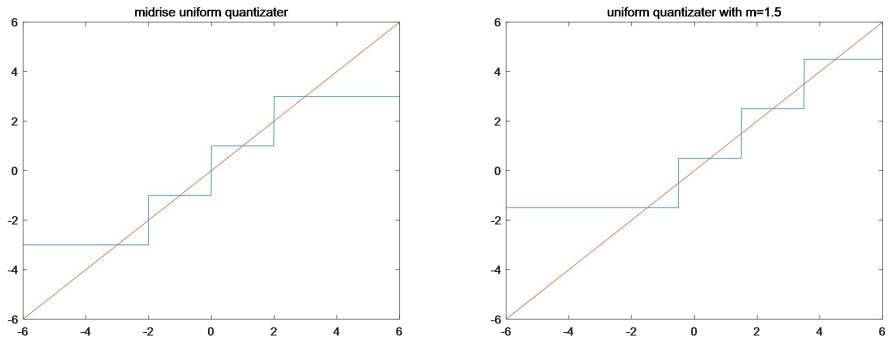
− *in,* a vector with the original speech samples;

− *n_bits*, the number of bits available to quantize one sample in the quantizer;

− x_{max} and m define the range of the quantizer from $m - x_{max}$ to $m + x_{max}$;

Therefore, the width of each quantization interval is $\Delta = 2 \times x_{max}/L$, where L is the number of quantization intervals. m defines the offset of the quantizer reconstruction levels. For example, $m = 0$ defines a "midrise" quantizer, and $m = \Delta/2$ gives a "midtread" quantizer. The encoder should return the index of the chosen quantization interval. The decoder should return the quantization value of the chosen quantization interval.

We run the encoder and decoder on a ramp signal $x = -6{:}\,0.01{:}\,6$ and implement a 2-bit quantizer with $x_{max} = 4$. We plot the quantizer output as a function of the input in Figure 2.1 when $m = 0$ and $m = 1.5$.



(a)m=0 (b)m=1.5

Figure 2.1: Uniform scalar quantizer

It can be observed that when $m = 0$, it's a midrise uniform quantizer with output level from -3 to 3. When $m = 1.5$, the output level of the quantizer is from -1.5 to 4.5.

3 Parametric Coding of Speech

In this task, we complete our design of the vocoder from the first assignment by finding the right encoding parameters to transmit the speech and quantize those parameters with appropriate bit rates.

3.1 Quantizing the Gain

We need to design a quantizer for the gain parameter. First, we plot the histogram of the gain parameter in Figure 3.1. It can be seen that the range of the quantizer is from 0 to 1700. Therefore, $m = 850$, $x_{max} = 850$. We run the vocoder with a uniform scalar gain quantizer according to the design above. We modify the parameter *n_bits* to find the rate at which we cannot hear the quantization distortion and it turns out to be 6. It can also be seen from Figure 3.2 that when $n\_bits = 6$, the gain and quantized gain have almost no difference. Hence, we should use 6 bits to code the gain for one frame.

Then we take the logarithm of the gain parameter prior to quantization. Similarly, we plot the histogram of the gain parameter in the log-domain in Figure 3.3. It can be seen that the range of the quantizer is from 0 to 7.5. Therefore, $m = 3.75$, $x_{max} = 3.75$. By repeating the same experiment we find that with $n\_bits = 5$ we cannot hear the quantization error. It can also be proved in Figure 3.4 that when $n\_bits = 5$, the log-gain and quantized log-gain have little difference.

Therefore, we find that under the same condition (i.e, no quantization error is heard), in the logarithmic domain, the coding gain requires fewer bits than in the linear domain. This means, encoding the gain in the log-domain is better.

3.2 Quantizing the Pitch and Voiced/Unvoiced Decision

The Voiced/Unvoiced decision is a binary decision so we use 1 bit per frame to encode the decision. To encode the pitch, quantization in both linear-domain and log-domain have been tested and we do not find much difference between them. To stay coherent with the gain quantization, we encode the pitch in the log domain. By running the same experiment as the log-gain quantizer, we find that with $n\_bits = 6$, we cannot hear the quantization error on the pitch. Similarly, the histogram of the log-pitch and the difference between log-pitch and quantized log-pitch when $n\_bits = 6$ is shown in Figure 3.5 and 3.6.

3.3 Quantizing the LP parameters

For the quantization of LP parameters, we use a vector quantizer (VQ). We use an order of 10 to compute the LP parameters. To encode the vector we have to find in which cluster it belongs to. We use the L2 distance and choose the cluster that has the minimum distance with the LP vector parameters. We also encode the residuals by using the same method so that the error between the LP parameters and the quantized one can be reduced.

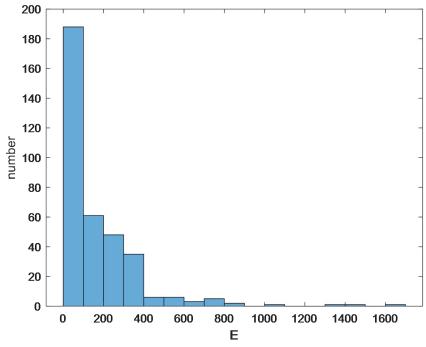


Figure 3.1: Histogram of the gain parameter

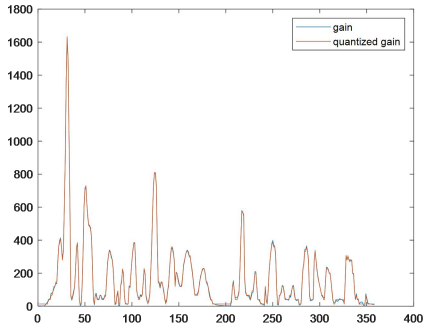


Figure 3.2: Gain and quantized gain

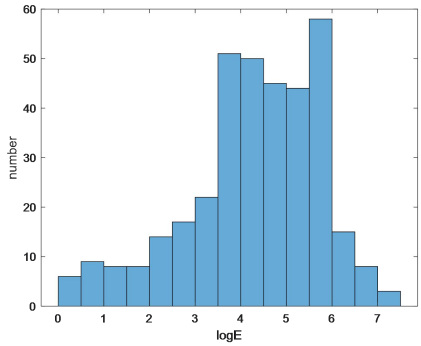


Figure 3.3: Histogram of the log-gain parameter

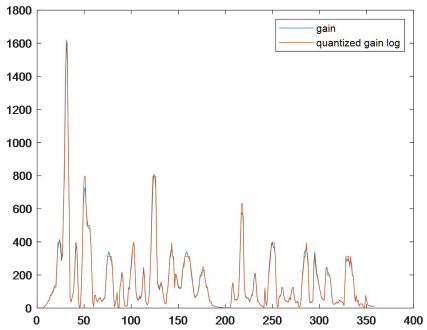


Figure 3.4: Log-gain and quantized log-gain

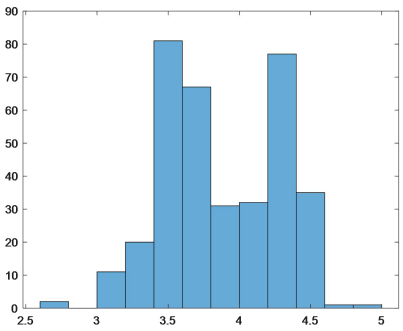


Figure 3.5: Histogram of the log-pitch

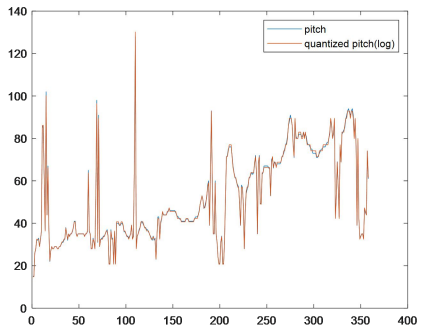


Figure 3.6: Log-pitch and quantized log-pitch

3.4 Optimizing the Bit Allocation

We find a bit allocation (i.e. the number of bits to use in each quantizer) for the gain, pitch, voiced/unvoiced quantizers, such that the quality is the same as when these parameters are unquantized. The number of bits to use for each parameter is shown in Table 1.

Table 1: The number of bits to use for each parameter

Number of bits	Bits/window	Bits/sample	Kbits/sec
Gain(E)	5	0.0391	0.3125
Voiced(V)	1	0.0078	0.0625
Pitch(P)	6	0.0469	0.375
LP(A)	20	0.1563	1.25
Total	32	0.25	2

We can evaluate the performance of our vocoder by computing the SNR using equation (3-1).

$$SNR_{dB} = 10\log_{10}\frac{\sigma_x^2}{\sigma_q^2} \tag{3-1}$$

where σ_x^2 is the power of the input signal and σ_q^2 is the power of the quantization error. We define the quantization noise by $X(n) - \widehat{X}(n)$, where $X(n)$ is the input signal and $\widehat{X}(n)$is the quantized signal.

We find SNR_{dB} = -0.899dB which means that the quantization noise is really present in the quantized signal. It does not make sense to evaluate the SNR here because we want the reconstructed signal as close as possible to the original signal.

4 Speech Waveform Quantization

4.1 Uniform Scalar Quantization of Speech

In this section, we attempt to design a uniform scalar quantizer (USQ) using equation $x_{max} = k\sigma_x$, where σ_x^2 is the variance of speech. To obtain the best SNR, we need to evaluate the optimal value of k for R=3. We calculate the SNR for different value of k from 0 to 11.5 in step of 0.01 and plot the curve in Figure 4.1. It can be found that the optimal k=3.46.

After the optimal k for R=3 is found, we can plot the SNR-rate curve using optimal k for each rate and compare it with the theoretical SNR. The result is shown in Figure 4.2. Theoretically, it is impossible for the experimental result to exceed the theoretical result due to Shannon theorem, but our goal is to be as close as possible to the theoretical limit. Assuming that the number of quantization levels is high and that the overload is negligible, the theoretical SNR is given by

$$SNR_t = 10\log_{10} 2^{2R} \approx 6.02 \cdot R \text{ dB} \tag{4-1}$$

The difference between experimental and theoretical results is because we use a uniform quantizer for a non-uniform probability distribution function.

We then listen to the original and quantized speech signals to determine which rate gives the no different result. It can be found that we can hear no difference between the two signals when R=8 bits which corresponds to SNR=31.96 dB.

We switch the point of view and listen to the quantization error signal when R=1. At R=1, we can hear that the quantization error contains a high proportion of information about the original signal, i.e. a large part of information is lost in the reconstructed signal. Increasing the rate from R=1 to R=12, we can hear that the information part is gradually reducing and finally when R=12, we can hear a perfect white noise.

According to the properties of midrise quantizer, it is known that there is a reconstruction level in the origin which means we have a non-zero level for silents and minute background noise. Therefore, compare to midrise quantizer, midtread quantizer removes the distortion caused by reconstruction level in the origin. We plot the SNR-rate curve for midrise and mistread in Figure 4.3. It can be seen that the

midtread quantizer provides a better SNR when the rate is between 3 and 10 which is the low rate. However, when the rate is below 3, i.e. R=1 and R=2, the midrise quantizer gives a better result because the offset of midtread quantizer is not negligible at such a low rate.

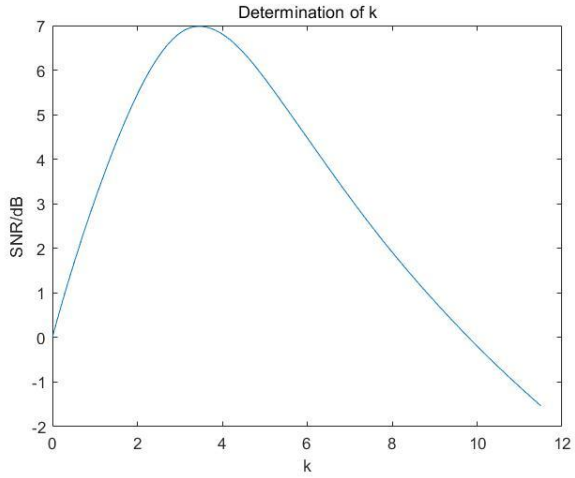


Figure 4.1: Determination of k for R=3

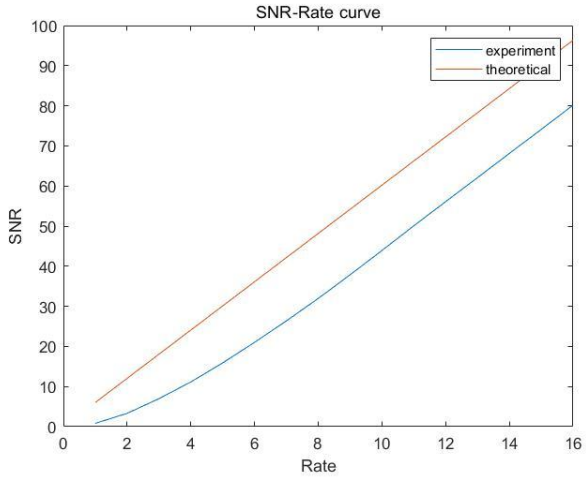


Figure 4.2: SNR-rate curve for experiment and theory

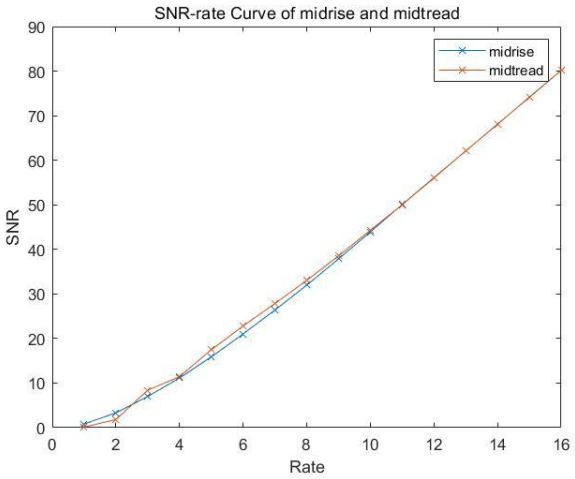


Figure 4.3: Midtread and midrise SNR-rate curve

5 Adaptive Open-Loop DPCM

In this section, we will study open-loop DPCM. The block scheme is shown in Figure 5.1.

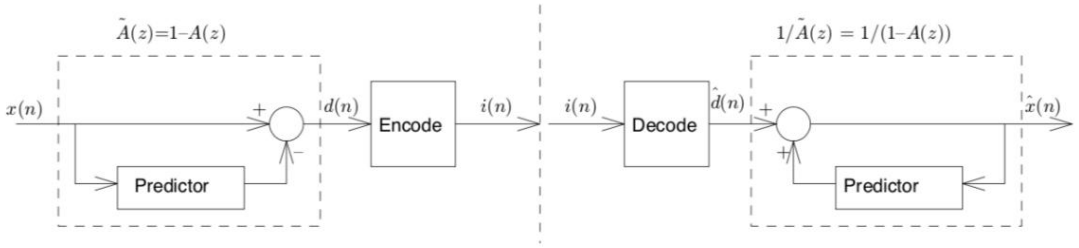


Figure 5.1: Open loop DPCM

To achieve the maximum rate, we adapt both the LP coefficients and the gain in a forward fashion. We choose the following parameters:

– Analysis frame length = 128 samples

– Update length = 128 samples

– Window function = rectangular window

– Number of bits to quantize the gain = 5 bits

– Number of bits to quantize the residual = 3 bits

To lower the compute complexity, we set *ulen* = *alen*. We choose rectangular window because there is no overlapping. The decision of bit number to quantize the gain and the residual is made by our previous problem. We use the optimal k=3.46 for R=3 from section 4.

The residual signal is multiplied by a time-varying gain in order to make the variance of the residual signal constant. The time-varying gain is designed as inverse proportion to the deviation of this signal. We encode the gain in log-domain and code it with 5 bits/frame from previous section. The LP coefficients are adapted with a MATLAB function *lpc()*.

Using the USQ we designed before, the reconstructed signal sounds almost identical to the original signal despite minute background noise still exists and the quantization error sounds like white noise.

An open-loop DPCM has a noise shaping property, which is because equation (5-1) is the synthesis filter that models the vocal tract, so we are expecting that the PSD of a frame of the residual signal has the same shape that the PSD of the same frame from the original signal. The DFT based spectrum of error is shown in Figure 5.2. It can be seen that the PSD of the residual signal follows the PSD of original signal.

$$H(z) = \frac{1}{1 - A(z)} \tag{5-1}$$

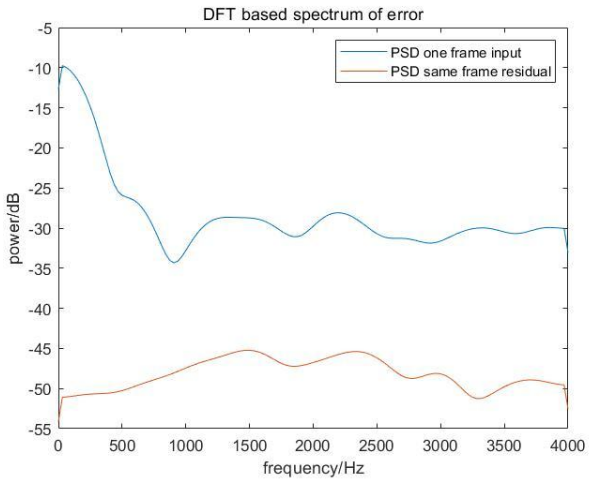


Figure 5.2: DFT of error

We measure the SNR for both PCM and ADPCM system. For PCM, the SNR is given by

$$SNR_{PCM} = 10\log_{10}\frac{\sigma_d^2}{\sigma_{qe}^2} \tag{5-2}$$

and for ADPCM, the SNR is given by

$$SNR_{ADPCM} = 10\log_{10}\frac{\sigma_x^2}{\sigma_{re}^2} \tag{5-3}$$

where σ_{qe}^2 is the variance of quantization error of residual signal and σ_{re}^2 is the variance of reconstructed error. The SNR of PCM equals to 10.1682 dB and SNR of ADPCM equals to 4.7592 dB. It can be seen that the adaptive open-loop DPCM has lower SNR than PCM at R=3. I think this is because we use open-loop DPCM in this case. One way to increase the SNR is to use a close-loop DPCM, but it will lose the noise shaping property. If we change the value of k to 0.5, the difference between PCM and ADPCM will reduce to 1 dB with ADPCM still lower than PCM.

Consider the rate, the rate of the coder is 3.1953 bits per sample and 25.563 kbits per second. In my opinion, it is better to use the quantized LP coefficients in the encoder filter than to use the unquantized LP coefficients because we transmit the index of codebook when using the quantizer, which means it is difficult to be interfered by the channel noise. However, if we transmit the unquantized LP coefficients, the detail of LP coefficients may lost during the transmission.

6 Conclusion

In this project, we have a better understanding of speech compression and quantization. At the beginning, we implement two quantizers by using the different offset. Among them, when offset = 0, it turns out to be a midrise uniform quantizer. When doing parametric coding of speech, we find that taking the logarithm of some parameters can reduce the bits used for coding. We choose an appropriate bit

allocation as 32 bits in total and evaluate the SNR as -0.899dB which indicates the existence of quantization noise. As for speech waveform quantization, we find the optimal k value as 3.46 and calculate the SNR at different rates. We can find it is better to use midtread quantizer at low rate. Finally, through the simulation of the adaptive DPCM system, the open-loop DPCM has a lower SNR than PCM because of the properties of open-loop.