Speech Compression and Quantization

EQ2320 Speech and Audio Processing, Assignment 2

Jingxuan Mao
jmao@kth.se

Yuqi Zheng
yuqizh@kth.se

March 18, 2022

1 Introduction

In this project, we work on the speech signal "speech8". We start from the basic midrise and mistread quantizer and realize the design of PCM and Parametric coding. We study on the quantization of parameters like gain, LP coefficients, pitch and try to explore the best strategy for decreasing the quantization error and SNR. Finally, we implement ADPCM to gain the reconstruction speech signal and figure out its unique characteristics.

2 The Uniform Scalar Quantizer

In this task, we implement the uniform scalar quantizer (USQ). Firstly, we design a uniform scalar encoder which return the index of the chosen quantization level. Then, we implement the corresponding decoder to reconstruct the input vector with the output index of the encoder.
With a 2-bit quantizer, we quantize a ramp signal $x = -6 : 0.01 : 6$. We set $xmax$ to 4. For $m = 0$ and 1.5, the quantization results are shown as Figure 1 below. We can see that the range of the quantizer is $[m - xmax, m + xmax]$ and the reconstruction levels are all the same as where they should be theoretically.

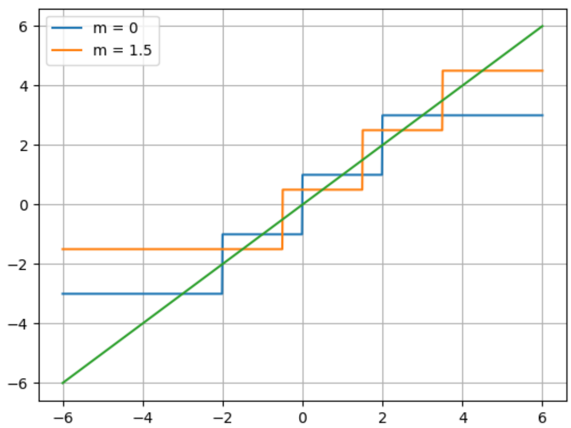


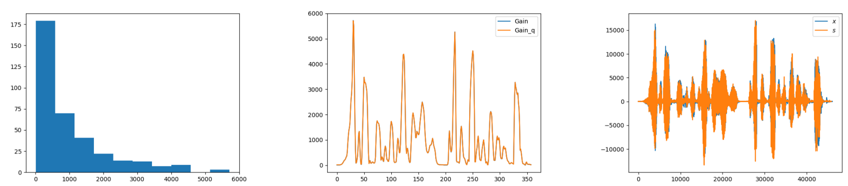
Figure 1: 2-bit Quantizer Output for $xmax = 4$, $m = 0$ and 1.5

3 Parametric Coding of Speech

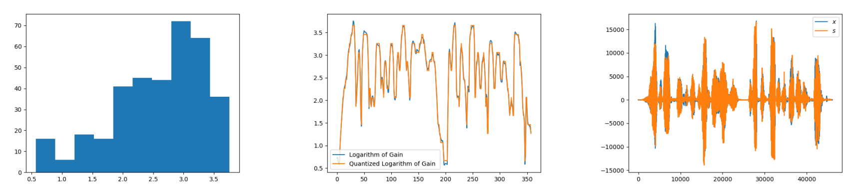
In this task, we utilize parametric coding to code the output gain, pitch, voice/unvoiced decision and LP parameters of the vocoder we designed in Assignment 1. Also, we try to find an optimal bit allocation in terms of the perceptual speech quality.

3.1 Quantizing the Gain

1. The histogram of the gain parameter is shown as Figure 2(a). We can see that the range of the quantizer should be [4, 5720], corresponding to $m - xmax$ and $m + xmax$ respectively. Hence, we know we should set m to $\frac{5720+4}{2} = 2862$ and $xmax$ to $\frac{5720-4}{2} = 2858$.
2. With the m and $xmax$ calculated above, we run the vocoder with the quantized gain. The original and quantized gain are shown as Figure 2(b). We gradually increase the rate n_bits and find that when $n\_bits = 8$, no quantization distortion is audible. Figure 2(c) shows the original speech signal x and the synthesized speech signal s.
3. We quantize the logarithm of the gain parameter instead of itself. Then, we repeat the steps above. The histogram of the logarithm of the gain parameter is shown as Figure 2(d). We can see that in this condition, the range of the quantizer becomes [0.57, 3.76], which is much smaller than the quantization range of the original gain. In the same way, we can get $m = 2.16$ and $xmax = 1.60$.
4. We implement the vocoder again with these new parameters. The original and quantized gain are shown as Figure 2(e). Since the range of quantzier is much smaller, the number of bits we need could be fewer. We find that when $n\_bits = 4$, we can obtain a signal well enough. The reconstructed signal is shown as Figure 2(f).
5. Compared to the linear domain, in log domain, with fewer quantization bits, we can obtain the perceptual speech signal of quality good enough. Therefore, the gain quantization is better in the Log domain.



(a) Histogram of the Gain in Linear Domain (b) Original and Quantized Gain in Linear Domain (c) Original and Reconstructed Speech Signal in Linear Domain

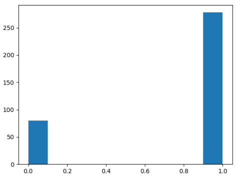


(d) Histogram of the Gain in Log Domain (e) Original and Quantized Gain in Log Domain (f) Original and Reconstructed Speech Signal in Log Domain

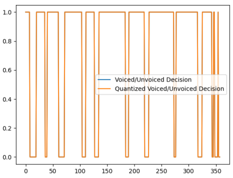
Figure 2: Quantizing the Gain in Linear and Log Domain

3.2 Quantizing the Pitch and Voiced/Unvoiced Decision

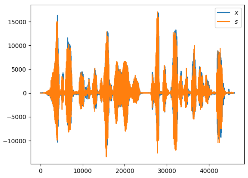
In this task, we quantize the pitch and voiced/unvoiced decision parameters in log domain. The procedure is the same as that for gain quantization.
Since the voice/unvoiced decision parameters only have two values, 0 and 1, we use 1 bit to quantize them. The quantization results are shown as Figure 3(a) to Figure 3(c).
For the quantization of the pitch, we set the bit rate to 5 as we find that the synthesized speech is perceptually good enough at this point. The quantization results are shown as Figure 3(d) to Figure 3(f).



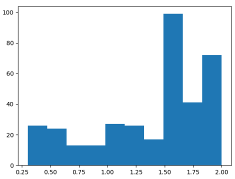
(a) Histogram of Voice/Unvoiced Decision



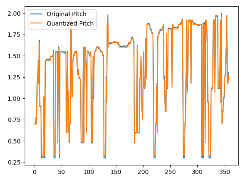
(b) Original and Quantized Voice/Unvoiced Decision



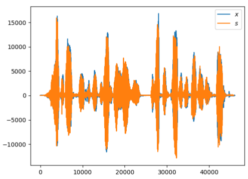
(c) Original and Reconstructed Speech Signal with Quantized Voice/Unvoiced Decision



(d) Histogram of Pitch



(e) Original and Quantized Pitch



(f) Original and Reconstructed Speech Signal with Quantized Pitch

Figure 3: Quantizing the Voiced/Unvoiced Decision and Pitch

3.3 Quantizing the LP Parameters

In this task, we no longer use the uniform scalar quantzier as before. Instead, we utilize a 2-stage vector quantizer to quantize the LP parameters. Firstly, we quantize the corresponding 10-dimension linear spectrum frequenciy (LSF) vectors of the LP coefficients with the first stage, 10 bit, of the vector quantizer. Then, we quantzie the residual part with the second 10-bit stage. These two steps give the index of the optimal codeword in the codebook, which maps the incoming parameters to the colsest codeword. Next, we input the index to the decoder to get the quantzied LSFs and the corresponding LP coefficients. The reconstructed siganl with the vector quantized LP parameters is shown as Figure 4.

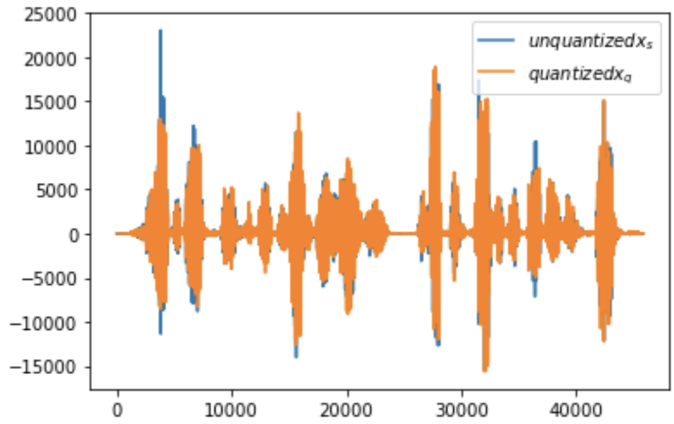


Figure 4: Reconstructed Speech Signal with Vector Quantized or Unquantized LP Parameters

3.4 Optimizing the Bit Allocation

In this task, we try to find the optimal bit allocation to reduce the perceptual quantization distortion, such that the quantized speech signal sounds the same as the unquantized one.
1. We synthesize the speech signal with all the parameters quantized as above. The SNR is $-2.91dB$.
2. We use 4 bits for the gain, 1 bit for the voice/unvoiced decision and 5 bits for the pitch quantization.
3. The rate in bits per sample is $\frac{4+1+5+20}{128} = 0.234$ bits/sample. In bits per second, the rate is $\frac{4+1+5+20}{\frac{128}{8000}} = 1.875$ kbps.
4. We get a small value of SNR here, but the synthesized speech signal is still intelligible. Actually, it does not make sense to evaluate SNR here. Because in parametric coding, an exact signal reproduction is not the main objective. In particular, phase estimation is not considered. We just extract and quantize the main characteristics of the original signal and use them to synthesize a speech signal, as the principle of the source filter. Therefore, the perceived quality of the synthesized speech signal cannot be quantified by objective distortion measures such as SNR. Instead, the speech quality depends on our subjective perception.

4 Speech Waveform Quantization

In this task, we design our uniform scalar quantzier with $xmax = k\sigma_x$. We choose appropriate value of k and the rate R to maximize SNR.

4.1 Uniform Scalar Quantization of Speech

1. When $R = 3$, in the range [2.0, 5.0], we find the k to maximize the SNR. The optimal k should be 3.46.
2-3. The experimental SNR can be calculated by $10\log_{10}2^{2R}$. The theoretical SNR and the SNR obtained with k and R are shown as Figure 5.

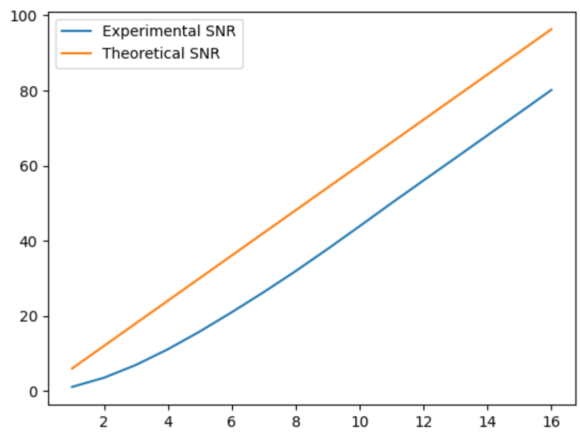
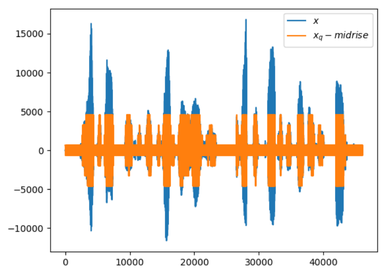
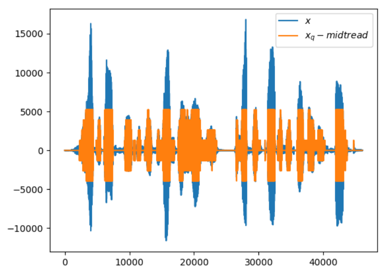


Figure 5: Theoretical SNR and Experimental SNR vs. Rate

4. When R=8, we cannot tell the difference between the original and the quantized speech signal. At this point, the SNR is 32dB.
5. When R=1, we can hear a portion of the original speech from the quantziation error signal. We measure the correlation between the original speech signal and the quantization error signal with their covariance. We find that as the rate increases, the covariance between them decreases. Besides, the original signal is gradually inaudible and the quantization error sounds more and more like a white noise. When R=12, the covariance becomes significantly smaller and we can hear a white noise clearly.
6. For low rates, we find that the midtread quantizer performs much better than the midrise quantizer. Because the midtread filter will quantize the unvoiced parts of the signal to zero, while the midrise filter will not, which introduces more errors. We can see the quantization results of the speech signal with these two kinds of quantizers as Figure 6(a) and Figure 6(b). The SNR obtained with the midrise quantzier is 6.98dB, while that obtained with the midtread quantzier is 8.94dB.



(a) Speech Signal Quantized with Midrise Quantizer



(b) Speech Signal Quantized with Midtread Quantizer

Figure 6: Quantization Results with Different Quantizers

5 Adaptive Open-Loop DPCM

In this task, we realize the open-loop ADPCM by designing the encoder and decoder. We study and compare the influence of quantization of $d(n)$, gain and LP parameters on the SNR. Further, we compare the performance between ADPCM and PCM to prove the advantage of ADPCM over SNR.

5.1 The performance of ADPCM

The length of each frame is set as 256 samples with no overlap between different frames. We use 4 bits, 20 bits, and 3 bits to quantize the gain g, prediction error $d(n)$, and the LP coefficients separately. From Figure 7 it can be seen that the reconstructed signal x_{hat} of ADPCM is sufficiently close to the original input speech signal x. The quantization error sounds like white noise.

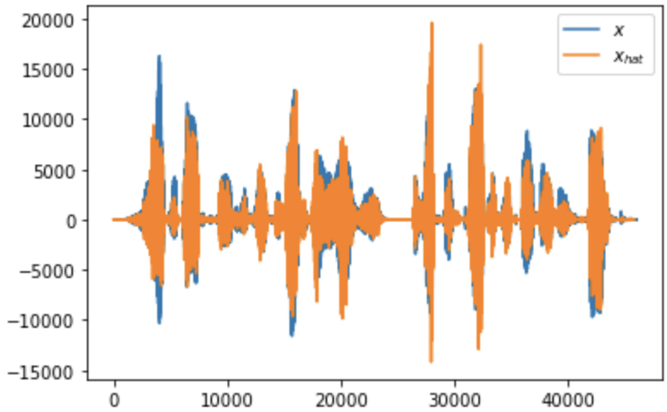
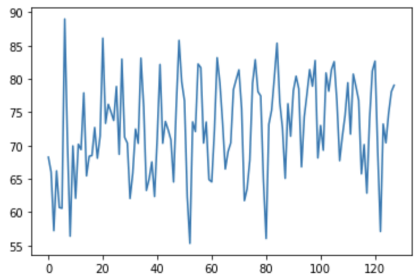


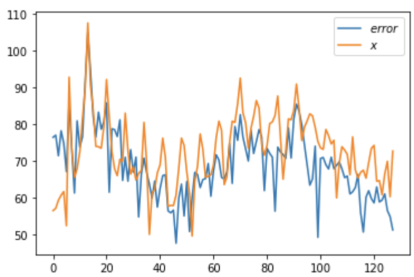
Figure 7: Original and reconstructed speech

5.2 The reconstruction and quantization error spectrum

From the theory, the reconstruction error of the open-loop DPCM on Z-domain is $q(z)/(1 - A(z))$, where q(z) is the quantization error. Given that q(z) should have the spectrum which is close to that of white noise, the spectrum of reconstruction error must therefore have the same envelope as that of the speech signal. In Figure 8 and Figure 9, we select the voiced frame $x[10000 : 10255]$ and draw thecorresponding spectrum of it. The figures prove that the practical results are all compatible with the theoretical deduction.



(a) The quantization error spectrum



(b) The reconstruction error spectrum

Figure 8: The reconstruction and quantization error spectrum

5.3 The SNR of ADPCM and PCM

We set $x_{max} = 2\sigma_d$ for the quantizer of PCM. As result, we find that the SNR of ADPCM is $3.68dB$ while the SNR of PCM is $2.10dB$, which leads to the conclusion that ADPCM outperforms PCM in the value of SNR. It is consistent with the theory since ADPCM is adaptive and distinguishes between voiced and unvoiced frames, while PCM does not update the coefficients frame by frame and merely quantizes the speech.

5.4 The total rate of ADPCM

In each frame, we use 4 bits and 20 bits for the quantization of gain and LP coefficients. For the 256 samples in each frame, we implement 3 bit for the quantization of the prediction error. Therefore, intotal we need $3 + 24/256 = 3.09$ bits for each sample, which is $8000(3 + 24/256) = 2.475$ kbits per second.

5.5 The influence of quantization on LPCs

We use the unquantized LP coefficients and gain the reconstruction signal shown in Figure 9. Compared with the result of using quantized LP coefficients in Figure 7, it can be seen that the reconstruction signal generated by unquantized LP coefficients is closer to the original speech signal. Also, its SNR is increased to $7.23dB$, which is about twice of the former value of $3.68dB$.

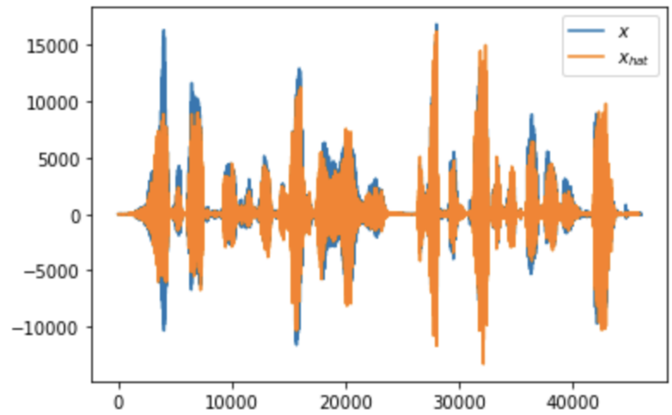


Figure 9: Original and reconstructed speech using unquantized LPCs

6 Conclusions

In this project, we work on the speech signal "speech8". We start from the basic midrise and mistread quantizer and realize the design of PCM and Parametric coding. We study on the quantization of parameters like gain, LP coefficients, pitch and try to explore the best strategy for decreasing the quantization error and SNR. Also, we test on the influence of quantization on the effect of decoding, finding the optimal quantization coefficients based on the numerical criterion or the straight feedback from our ears. Finally, we implement ADPCM to gain the reconstruction speech signal and figure out its unique characteristics. By comparing it with PCM, we prove its advantage over PCM on the value of SNR.