TTT4136 Sound and Image Processing

Assignment 3 – Quantization of audio signals

Report by: Knut Halvor Slang

1 Music signal characteristics

1.1 The music sources

For this assigment, I have chosen excerpts of music from two CDs in my CD collection. For the solo piano music, I have chosen an excerpt from a track entitled "The Melancholy Spirit," by the American band Agalloch, taken from the album "Pale Folklore." For the pop/rock music, I have chosen an excerpt from a track entitled "Pyromancer," by the Norwegian band Sahg, taken from the album "Sahg II." The excerpt from "Pyromancer" contains guitars, bass, percussion and vocals, and should thus be a good example for this exercise, as it is rather different from solo piano music. Both tracks have been ripped to wav files using a media player called Winamp. The 20–30-second excerpts have been extracted and saved to new wav files using an audio editor called Audacity. As both music excerpts were taken directly from CD, it is clear that both excerpts have a sampling frequency of 44.1 kHz. The music excerpts are attached in the separate rar file, together with the commented Matlab code written for this assignment.

1.2 Broadband signal characteristics

In this part, the signals were broadband analyzed in Matlab, and the following parameters were measured: peak value, broadband RMS value, crest factor, probability density function. The peak value for each signal was found by taking the absolute value of the signal and then finding the highest sample value. The peak values are expressed both in integer values and in dB relative to full scale (dBFS). In our case, the source is a wav file, which stores each sound sample as a 16-bit integer with zero mean. Hence, our full scale reference is $2^{15} = 32768$. Both representations of the peak values are listed in table 1. The broadband RMS value for each signal was found by taking the square of the signal, calculating the mean of the squared signal, and finally taking the root of the mean value. The crest factor for each signal was found by taking the ratio (in dB) between the peak value and the broadband RMS value. The broadband RMS values and the crest factor values are listed in table 2. The probability density function for each signal was found by simply plotting the signal as a histogram. The histograms are shown in figures 1, 2 and 3. For further details on the computations, please see the attached Matlab code.

As we see from the results, the piano signal does not exploit the full digital range, as the peak values are relatively low. Much of the reason for this is that the piano music is an excerpt from a longer piece of music with, among other instruments,

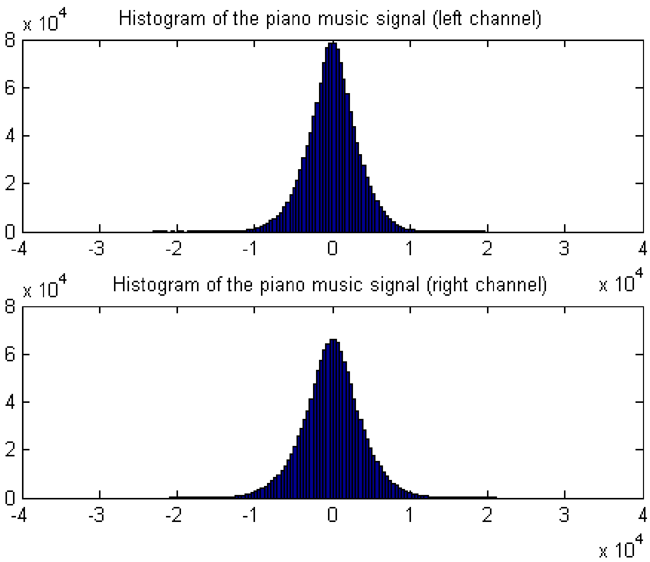

Figure 1: Histograms of the piano music signal.

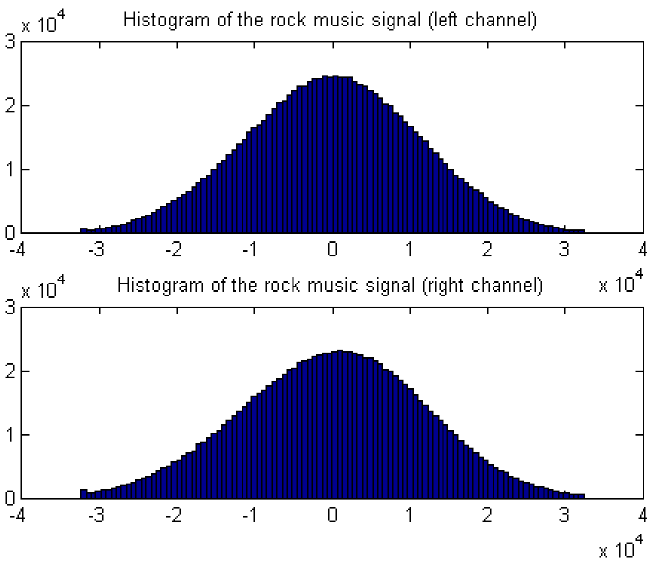

Figure 2: Histograms of the rock music signal.

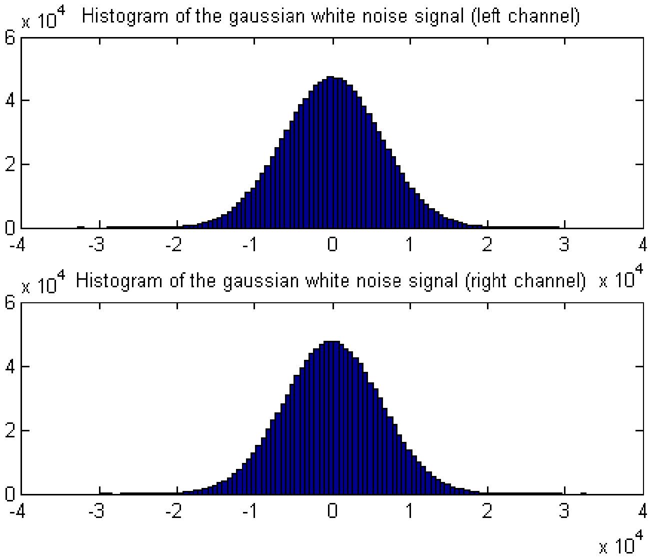

Figure 3: Histograms of the gaussian white noise.

guitars, bass and percussion. The rest of this piece of music has higher sample values than the piano excerpt, and hence it is this louder part of the music piece which has been the reference when scaling the amplitude levels. The rock signal, on the other hand, seems to exploit the digital range quite well. Contrary to the piano signal, the rock signal covers some of the loudest parts of the piece of music from which the excerpt is taken. This is much of the reason why the rock signal appears to exploit the digital range better. The gaussian white noise has been scaled to cover the full digital range, so this signal naturally reaches max. By evaluating the crest factor, we see that the rock music signal is less dynamic than the two other signals, as the crest factor of the rock music signal is significantly lower. Because of the general intensity of the rock music signal, it was expected for the rock music signal to have less dynamics. From listening to the two music signals and comparing them, it is also evident that the piano music signal is more dynamic than the rock music signal. From the results, we also observe that the piano music signal has a crest factor which is not too far from the crest factor of the gaussian white noise. By evaluating the histograms, we see that both music signals have a gaussian distribution, yet they are notably different from the distribution of the gaussian white noise. The variances of the three signals are significantly different.

1.3 Octave band limited signal characteristics

In this part, the RMS values of the signals were analyzed in the octave bands from 63 Hz to 8 kHz. The RMS values were found by filtering the signals in the different octave bands, and then computing the RMS value of the filtered signal. The results are presented in figures 4, 5 and 6. For further details on the computations, please see the attached Matlab code.

Signal	**Peak value**	**Peak value (dBFS)**
Piano, left channel	22987	-3.08 dB
Piano, right channel	21119	-3.82 dB
Rock, left channel	32400	-0.10 dB
Rock, right channel	32401	-0.10 dB
Noise, left channel	32768	0.00 dB
Noise, right channel	32617	-0.04 dB

Table 1: Peak values for the different signals.

Signal	**Broadband RMS**	**Crest factor**
Piano, left channel	3236.7	-73.28 dB
Piano, right channel	3803.6	-75.42 dB
Rock, left channel	11025.0	-80.95 dB
Rock, right channel	11631.0	-81.41 dB
Noise, left channel	6257.0	-75.93 dB
Noise, right channel	6256.2	-75.97 dB

Table 2: Broadband RMS values and crest factors for the different signals.

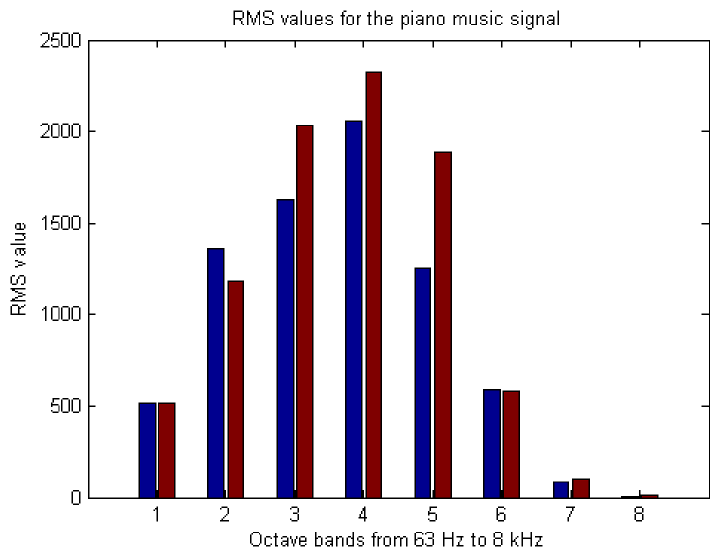

Figure 4: RMS values for the piano music signal. The left bars represent the left channel, and the right bars represent the right channel.

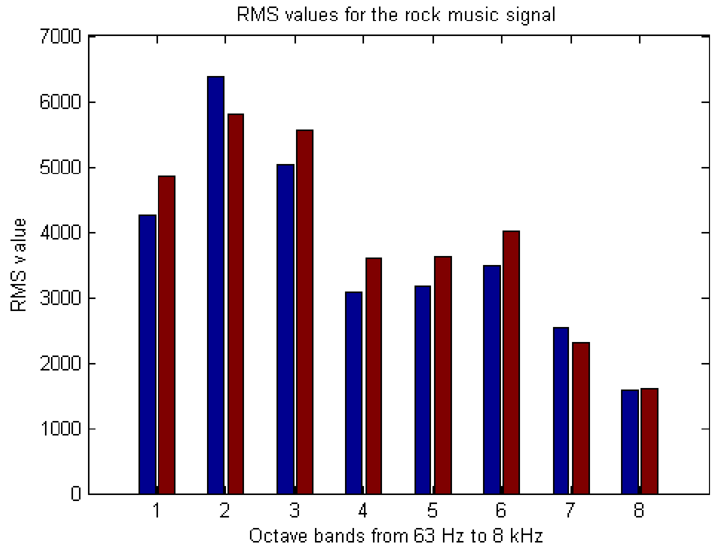

Figure 5: RMS values for the rock music signal. The left bars represent the left channel, and the right bars represent the right channel.

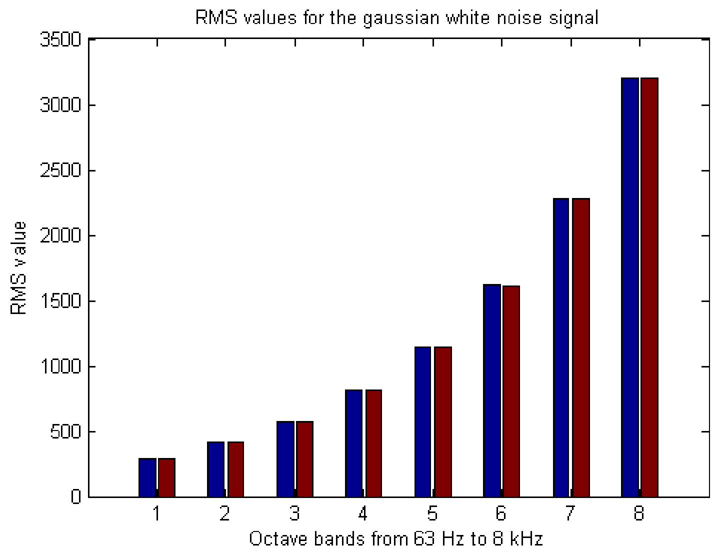

Figure 6: RMS values for the gaussian white noise. The left bars represent the left channel, and the right bars represent the right channel.

As we can see from the figures, the RMS values of the gaussian white noise increase with the octave bands, while the RMS values of the music signals do not. This is just as we would expect, as white noise has equal power for all frequencies. Higher octave bands cover larger ranges of frequencies, and this gives the gaussian white noise higher RMS values in the higher octave bands. The octave band analysis also reflects how our hearing works. Our hearing system groups the frequencies into different frequency bands which increase in range, not too different from the octave bands listed above. This is why white noise appears to us as relatively high-frequent. With this in mind, we see that the results also correspond well with what we hear when listening to the two music signals. We see from the results that the rock music signal has higher RMS values than the piano music signal both in the low and the high frequency bands. This indicates that the rock music signal both has more bass and more treble than the piano signal, which is exactly what we hear when listening to the music. Naturally, the bass guitar and the bass percussion account for much of the bass, whereas the cymbals account for much of the treble.

2 PCM (re-)quantization

2.1 Quantization with no dither

In this part, the signals were quantized with a uniform quantizer with zero probability of overload noise and no dither. The signals were quantized from the 16-bit original down to every resolution in the range from 15 to 1 bit. Following quantization, the signals were written to wav files, and played at the different resolutions. For the piano signal, I could clearly hear a difference from the original at resolutions of 7 bits and below. For the resolutions at 7 and 6 bits, the difference appeared as white noise, but at the resolutions of 5 bits and below, the difference appeared as distortion. For the rock signal, however, I could only hear a clear difference from the original at resolutions of 4 bits and below, and the difference appeared as distortion only at the resolutions of 2 and 1 bit. The reasons for this are explained in section 2.4.

2.2 The quantization error signal

In the quantization process in section 2.1, the quantization error of each signal at every resolution were also computed and written to file. The quantization error was found by subtracting the original signal from the quantized signal. In this part, these quantization error signals were evaluated. Although the quantization error signals were written to wav files for convenience, the sound was played in Matlab to avoid boosting the system volume to alarming levels. For further details on the computations, please see the attached Matlab code.

At high and medium resolutions, the quantization error signals appear only as white noise, whereas at low resolutions, they appear as a distorted signal. At very low resolutions, one can hear signatures of the original music signals. For the piano music signal, I could hear a clear non-stationary noise character at resolutions of 6 bits and below. At a resolution of 3 bits, I could hear certain signatures of the original piano music signal, and at a resolution of 2 and 1 bit the quantization error signal sounded like the original piano music signal with distortion. For the rock music signal, I could hear a clear non-stationary noise character at a resolution of 3 bits and below. At a resolution of 1 bit, I could hear very clear signatures of the original rock music signal, although much of the original signal was inaudible. By comparing with section 2.1, we see that the observations from listening to the music quantization error signals, *roughly* corresponds to what we observed from listening to the quantized music signals.

2.3 The quantization error probability density function

In this part, the probability density functions of the quantization errors from section 2.2 were evaluated. The probability density functions were found by plotting histograms of the quantization error signals in the same way as described in section 1.2. From the plotted histograms, we see that the probability density function is rather uniform for high, medium and low resolutions. For very low resolutions, the distribution has an inverse gaussian-shaped envelope with the minimum at zero, and the maxima at the left and right boundaries. This means that for very coarse quantization, we have higher probability of bigger deviations than for quantization of higher resolutions. There are noticeable differences, however between the probability distributions of the quantization errors of the two different music signals. For the piano music quantization error signal, the inverse gaussian-shaped envelope is visible for resolutions of 3 bits and below, whereas for the rock music quantization error signal, the inverse gaussian-shaped envelope is only visible at the resolution of 1 bit. This implies that we have more audible distortion for the quantized piano music signal than for the quantized rock music signal – something we have already confirmed in section 2.1. For simplicity, and in order to make the report more readable, only the most illustrative histograms have been included in this document. The selected histograms are shown in figures 7, 8, 9 and 10. All histograms can easily be generated by running the attached Matlab code.

2.4 Signal-to-quantization-noise ratio

In this part, the broadband signal-to-quantization-noise (SQNR) ratios of the quantized signals computed in section 2.1 were evaluated. The SQNR values were found by expressing the ratio between the original signal and the quantization error in dB. For further details on the computations, please see the attached Matlab code. The results are shown in tables 3 and 4. As we can see from the results, the SQNR decreases with the resolution, just as expected. The SQNR values of the rock music signal is generally roughly 10 dB higher than the SQNR values of the piano music signal. This corresponds well with the observations from section 2.1, and partly explains why we hear noise and distortion at higher resolutions for the piano music signal than for the rock music signal. Yet, the SQNR values do not explain everything.

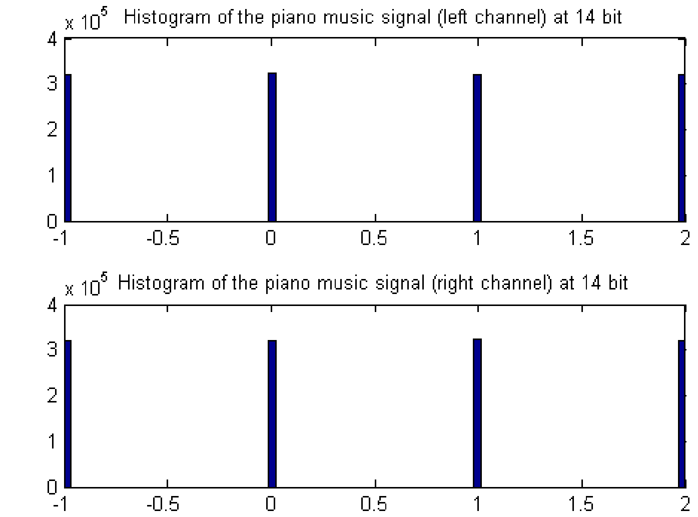

Figure 7: Histograms for the quantization error of the piano music signal quantized to a resolution of 14 bits.

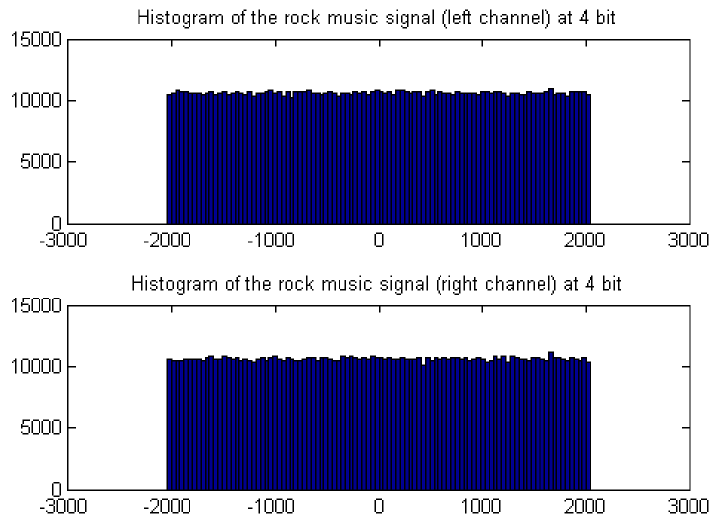

Figure 8: Histograms for the quantization error of the rock music signal quantized to a resolution of 4 bits.

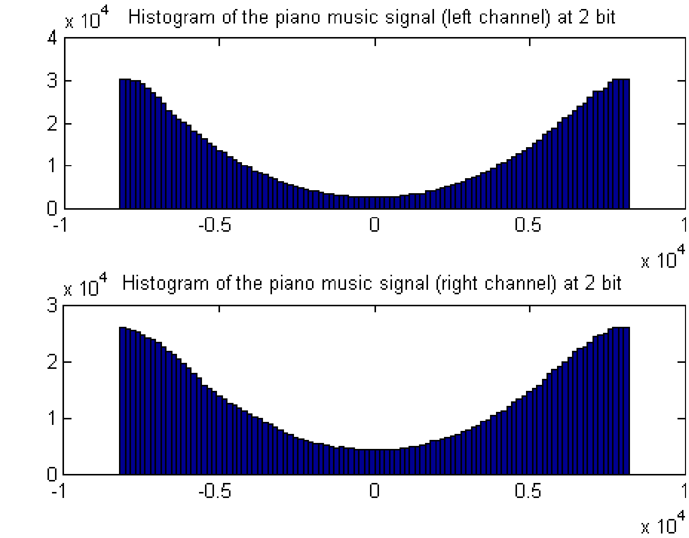

Figure 9: Histograms for the quantization error of the piano music signal quantized to a resolution of 2 bits.

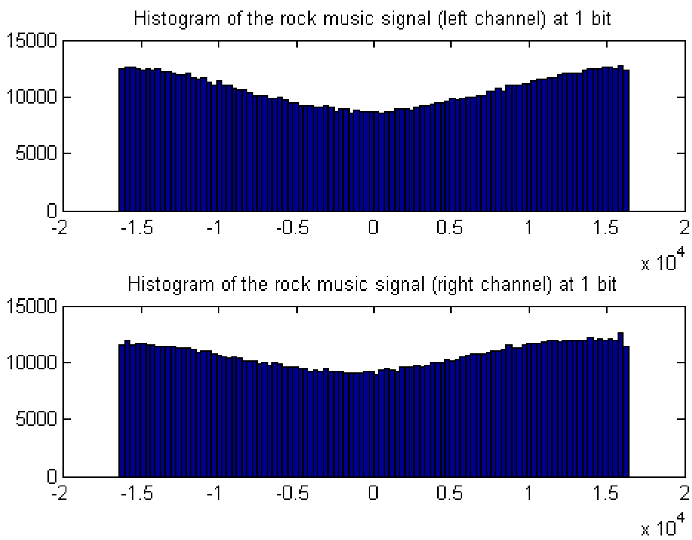

Figure 10: Histograms for the quantization error of the rock music signal quantized to a resolution of 1 bit.

If we compare the resolution levels of the two music signals for a given SQNR value, we see that the piano music signal generally is roughly two resolution levels above the rock music signal. However, if we look at what we found in section 2.1, we see that the quantization error was audible at resolutions of 7 bits and below for the piano music signal, and at resolutions of 4 bits and below for the rock music signal. Correspondingly, the quantization error appeared as distortion at resolutions of 5 bits and below for the piano music signal, and at resolutions of 2 bits and below for the rock music signal. These values have a resolution difference of 3 bits, whereas the SQNR values only explain a resolution difference of roughly 2 bits. This last bit resolution difference can perhaps be explained by the genre and nature of the music signals. For instance, in the rock music signal, several instruments (guitars, bass, percussion and vocals) are played at the same time, helping to mask noise and distortion with a more intense and chaotic signal, compared to the piano music signal. In addition, the guitars in the rock music signal are played with overdrive effects, which further help masking noise and distortion.

Resolution	**SQNR, left channel**	**SQNR, right channel**
15 bits	76.22 dB	77.62 dB
14 bits	69.24 dB	70.64 dB
13 bits	63.01 dB	64.40 dB
12 bits	56.93 dB	58.33 dB
11 bits	50.90 dB	52.30 dB
10 bits	44.87 dB	46.27 dB
9 bits	38.85 dB	40.25 dB
8 bits	32.84 dB	34.23 dB
7 bits	26.80 dB	28.21 dB
6 bits	20.79 dB	22.19 dB
5 bits	14.77 dB	16.15 dB
4 bits	8.66 dB	10.11 dB
3 bits	2.09 dB	3.71 dB
2 bits	-5.46 dB	-3.69 dB
1 bit	-12.75 dB	-11.13 dB

Table 3: SQNR values for the piano music signal

3 Dithering

3.1 Uniform dither

In this part, uniform dither was added to the signals, and the signals were then quantized in the same way as described in section 2.1. In addition, the quantization error signals were computed in the same way as described in section 2.2. The uniform dither was created by adding white noise with uniform distribution in the amplitude range $[\ -\Delta/2,\ \Delta/2\ ]$ to the signals. Here, Δ denotes the quantization step. For further details on the computations, please see the attached Matlab code. The reason

Resolution	SQNR, left channel	SQNR, right channel
15 bits	86.87 dB	87.33 dB
14 bits	79.88 dB	80.35 dB
13 bits	73.64 dB	74.11 dB
12 bits	67.57 dB	68.04 dB
11 bits	61.54 dB	62.01 dB
10 bits	55.52 dB	55.98 dB
9 bits	49.49 dB	49.95 dB
8 bits	43.48 dB	43.94 dB
7 bits	37.45 dB	37.92 dB
6 bits	31.43 dB	31.90 dB
5 bits	25.41 dB	25.88 dB
4 bits	19.40 dB	19.86 dB
3 bits	13.38 dB	13.84 dB
2 bits	7.35 dB	7.84 dB
1 bit	0.88 dB	1.48 dB

Table 4: SQNR values for the rock music signal

why we add dither to a signal before quantization, is to transform distortion into white noise. From listening to the quantized signals and the quantization errors, we can clearly hear the effect of dithering. For the piano music signals quantized with uniform dither, I could hear no distortion at any of the resolution levels, although the power of the gaussian white noise seemed rather high. For the piano music quantization error signals, I could hear no non-stationary noise characters at any of the resolution levels. For the rock music signals quantized with uniform dither, I could only hear distortion at a resolution of 1 bit. Correspondingly, I could only hear a non-stationary noise character for the rock music quantization error signal at a resolution of 1 bit.

3.2 Triangular dither

In this part, what was done in section 3.1 was repeated, only with triangular dither this time. The triangular dither was created by adding white noise with triangular distribution in the amplitude range $[\ -\Delta,\ \Delta\ ]$ to the signals. Just like above, Δ denotes the quantization step. For further details on the computations, please see the attached Matlab code. From listening to the music signals quantized with triangular dither and the corresponding quantization errors, the triangular dither appeared to have the exact same effect as the uniform dither. At a resolution of 1 bit, there was still distortion at the rock music signal quantized with triangular dither. In addition, there was still a non-stationary noise character at the corresponding quantization error signal. From the results, the triangular dither appeared to do the same as uniform dither. However, according to the theory, the triangular dither should have given slightly better results than the uniform dither.

4 Attachments

- File entitled `code.rar`
- Requantized sound files at:
  http://folk.ntnu.no/slang/skole/lyd&bilde/assignment3/