Audio Signal Compression

Mikołaj Leszczuk

# Exercise objectives

This section discusses the objectives of this exercise. They are: compression by using different types of audio codecs and quality models.

## Compression using various types of audio codecs

The first aim of the exercise is to compare two popular audio codecs: MP3 (*MPEG Audio Layer-3*), AAC (*Advanced Audio Coding*) and Vorbis. The comparison will concern the compression and, above all, its performance (sound quality compared to the size of the audio file.)

## Quality models

The second objective is to create simple models of quality for these signals compressed by codecs.

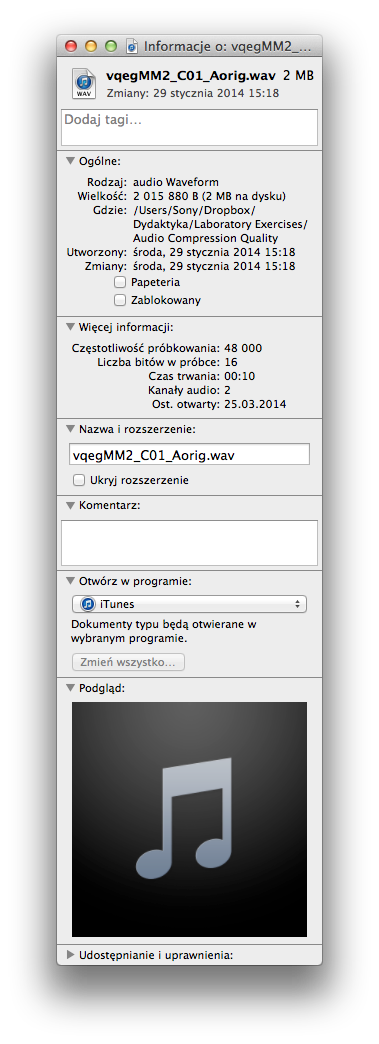
# Theory

Sound in digital form is stored in the form of samples, occurring with specified frequency, recorded for each channel (direction of arrival of sound). For example, the most popular standard called “Audio CD”, in which artists are used to sell high quality pieces, stores samples at 16 bits, taken at a frequency of 44100 Hz for each of the two channels separately. 2 independent channels are called stereo sound, as well give the impression of space. Multichannel audio is also a stereo, but it is widely accepted to use the term *stereo* only as another word for two-channel (*dual-channel*)sound for which the channels are independent of each other [[1].](#h3znysh7) Very often the sound emitted by the sound card or other stereo system is given for 5 channels. In such a case, an impression of space remains, and so-called breaks in the music scene are much smaller. The standard is that the samples occupy 24 bits each, and the frequency of the samples is 96 kHz. Such equipment extrapolates sounds of the lower parameters to these values.

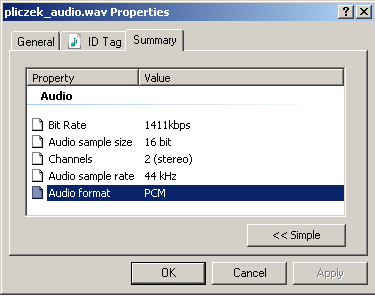
Human ear cannot recognize the direction of approaching low tones (tones whose wavelength is much larger than the size of the ear). Therefore, very often, to reduce the cost of sound, contained in the record, five channels are isolated into one bass channel (usually it is so-called “down” band of the front left channel) and five separate channels for higher frequencies (together called 5.1 channels). 5 speakers of medium-high tone deploy to roundabout *(surround)* and there is only one subwoofer. Low-tone speakers for each channel are used in the most expensive equipment (*high-end*)*.* Then the number loudspeakers are exactly equal to the number of channels. The solution with a single low-tone speaker, controlled the sum of the lower parts of the bands all five channels does not work, because the course of the sum of the signals differ from the actual course of the bass recordings.

The parameter bringing information on the quality of the sound file is the number of kilobits of data per second of the recording (called *bitrate*),in which the [kbps] unit is used. For popular audio CD the number is 2 channels \* 16 bits / sample \* 44100 samples / second, which gives 1411.2 kbps.

On OS X, the data on the quality of the audio file, can seen in the summary tab in the file inspector. An example of such information is shown here:



On Windows, the data on the quality of the audio file, can seen in the summary tab in the file properties. An example of such information is shown here:



To reduce the file size (for the stream – bit rate) audio digital encoding is used. The code is typically realized by using compression, which may be lossy or lossless. Each type of encoding of the original recording, even with the highest resolution is a lossy representation. In practice, encoders allowing reducing the size of audio information at the expense of its quality are considered lossy encoders while encoders effecting compression of digital information without loss of quality are considered lossless encoders.

The degree of compression of the sound file is the ratio of the uncompressed file size to the compressed file size.

A common practice used in lossy coding is neglecting, in the first place, sound information for low-frequency sound poorly audible to the human ear.

For the (stereo) system to be able save / restore data, one must install an encoder / decoder (codec). Codecs can be hardware (expensive, fast, used in Hi-Fi system) or software (cheap or free, used in computers). Particular codecs are related to file formats that arise as a result of the process to encode audio. For example, a result of audio MPEG layer-3 encoder is formed in an *mp3* file (the file usually has such an extension as well).

The tone in digital form can be represented in the form of files or so-called *audio-stream*. The common way is to record audio-encoded bit streams in tracks on a CD, using a very simple coding involving the recording of the sample, without the use of compression. For each such disc, its file-system contains brief, 44-kilobyte files, indicating the beginnings of recordings. The contents of audio CDs is usually seen by the operating system as just these files. They have the CDA (*audio CD*)extension*.* Copying them, do not infringe the copyright of artists, because they are just shortcuts to the actual recordings. To copy the actual music one must use tool for “ripping”, literally “roaring” audio CD content to a format recognized by the operating system. A popular lossless format is uncompressed WAV (or WAVE) developed by IBM and Microsoft. WAV codec is installed by default on OS X or Windows, and the popular, free tools for “*ripping*”CDs to WMA / WMV are iTunes or Windows Media Player.

Much more efficient formats, in terms of the size of the file, are formats applying compression. At the highest rate of bits per second (so-called *bitrate*)they are perceived as lossless formats compared to the original recordings. The most famous codecs are competing:

* MP3 (MPEG Audio Layer 3) – the most popular format for audio storage, as proposed by a group of German engineers working on the “DAB” digital radio project.
* AAC (Advanced Audio Coding) – codec developed by the corporation of Dolby, Fraunhofer, AT&T, Sony and Nokia companies, presented as the seventh part of the specification of MPEG-2 and the third part of the specification of MPEG-4. Data stream encoded using an AAC encoder is often stored in a file with the M4A extension.
* Vobis – a free and open-source software project headed by the Xiph.Org Foundation (formerly Xiphophorus company). The project produces an audio coding format and software reference encoder/decoder (codec) for lossy audio compression. Vorbis is most commonly used in conjunction with the Ogg container format and it is therefore often referred to as Ogg Vorbis.
* Windows Media Audio – codec developed by Microsoft. Files encoded in WMA encoder have the WMA extension.

These formats allow recording of multiple compression ratios, both at Constant (CBR) [[2]](#h2et92p0) and Variable Bit Rate (VBR) [[3].](#htyjcwt) Sometimes ABR is also available [[4].](#h3dy6vkm)

VBR encoder samples at a constant rate while using samples of different sizes. For the passages in which the neighbouring samples slightly differ, samples are used with a greater number of bits, to provide more detail representation. If the samples show large variation, or do not show the variation at all, they can be saved with a smaller number of bits to save memory. VBR technique generally provides better ratio of sound quality to file size, but not all codecs support it. This is a problem especially for hardware MP3 codecs implemented in a cheaper playing hardware.

These codecs were often the subject of precise comparisons in tests carried out by the blindfolded audience, in soundproofed studios with a very neutral interior acoustics. Such tests are carried out according to a fixed standard. Sound is considered distinguishable if the listener can identify the better recording, in a reproducible manner, knowing how to accurately describe what, in his/her opinion, is the difference. The tester must often use for this purpose very sophisticated descriptions. Tests are repeated several times for many types of recordings.

Codecs for high degrees of compression are also subject to our (simplified) comparisons in the laboratory.

# Guidelines for the implementation of the exercise

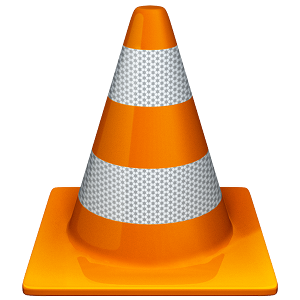
This section discusses the tasks to be performed under this exercise. They are: compression by using different types of audio codecs and quality models.

## Compression using various types of audio codecs

To perform the exercise reference sound sequences will be needed, derived from the original, lossless source. The sequences do not have to be long, but their collection should contain a variety of sound accents from the widest possible range of acoustic frequencies. To do the exercise, you can use your own recordings or recordings made available especially for this exercise.

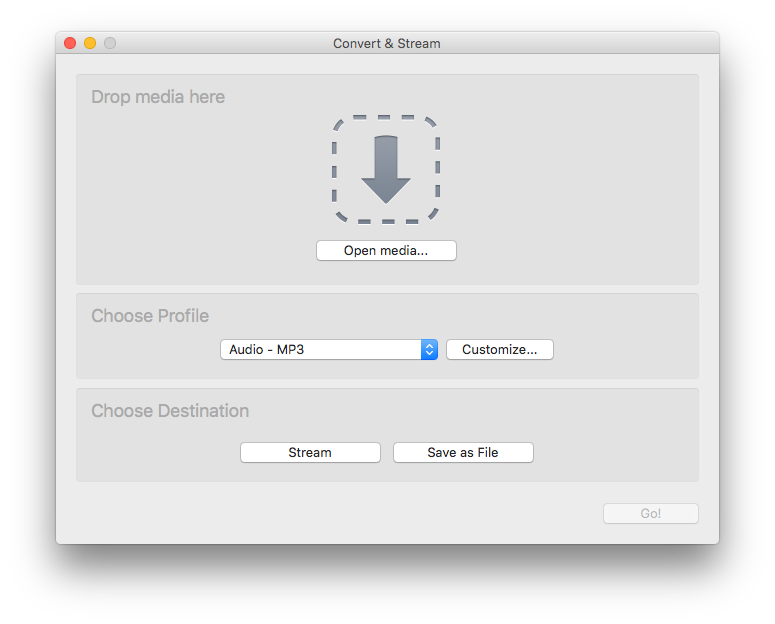
Please download test sequences (with the name "vqegMM2\_Cxx\_Aorig.wav", where xx is a sequence number). Use them as lossless (compression always start from the lossless file) standard for all students. They include a 10-second sound samples of different types of sound.

To perform the conversion to the test format, we will use the free **VLC media player** program. Its icon is characteristic:



**It can be downloaded from the** [http://www.videolan.org/.](http://www.videolan.org/)

Once enabled, select "File / Convert / Save...” – the program asks for the file(s) that you want to convert. Select the file(s) that were downloaded. Then select the format and parameters of the output file, as in the windows shown in the figure:



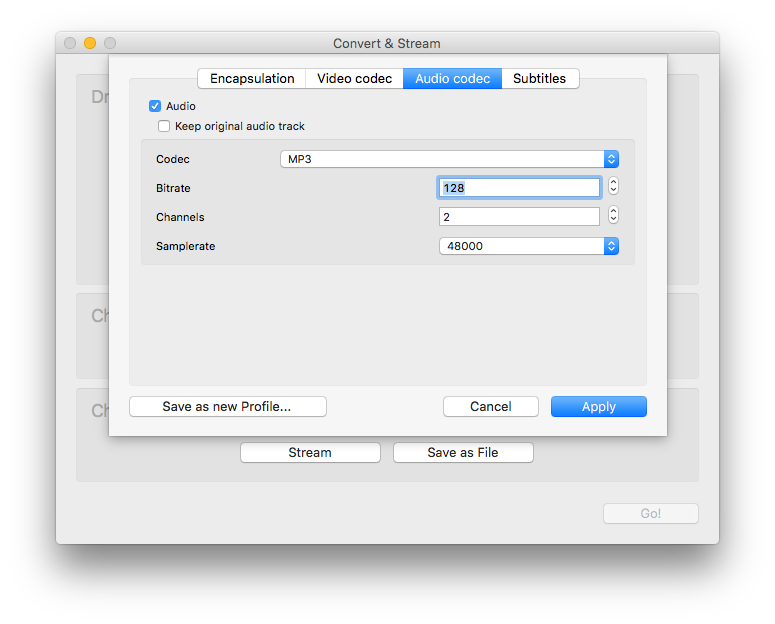
**We are interested in the output formats including MP3, AAC (MPEG-4 Audio, MP4) and Vorbis.**

If there is no profile matching your needs, create a custom one.

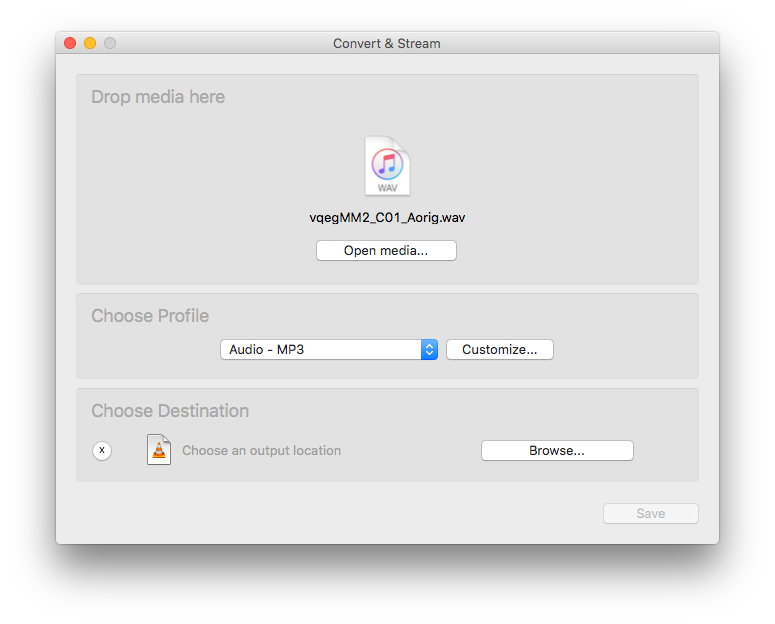
|  |
| --- |
| If VLC cannot encode MPEG 4 Audio (AAC) / (MP4A), please use the following workaround: <https://forum.videolan.org/viewtopic.php?t=111641> |

The program allows you to change the compression parameters: rate (so-called *bit-rate* or *data-rate*), the number of channels and sample rate. **However, majority of the parameters can be (at least at the beginning) left in the default settings. Please mainly modify the bit-rate parameter (defined per stream).**

Please carefully consider the number of the measured quality levels (minimum 5), to succeed in 90 minutes to do a full lab experiment. It is best to choose the same files and bit-rates (data-rates) for different codecs. You must remember that you must examine all the codecs for (ideally) all the source sequences and several qualities (bit-rate) levels, reaching both below and above the default 128 kbit/s. **You have to go up and down with the bit-rate and rather do not touch anything other settings, like the number of channels (**<http://en.wikipedia.org/wiki/Audio_channel>**) or sampling rate (**<http://en.wikipedia.org/wiki/Sampling_(signal_processing)>**)**:



Please compress each file one by one, because then they have to be independent files. Effect of any compression should be saved to a separate file, and then, after listening, assign it a subjective, individual assessment of the MOS scale. For each file compressed please browse and choose different destination (location):



During this conversion, the file can not be played, and after the conversion, you must listen to the files.

## Models of quality

Past gathering data, create models of quality for each of the studied codecs. To complete this part of the exercise MATLAB will be helpful. Please run it, and then create a new variable by clicking on the „Workspace” on the „New variable”:



The program will ask you to name the new variable; the typed-in name does not matter, it could be, for example, the name of the currently tested codec (in the instructions the default „unnamed” name is left):



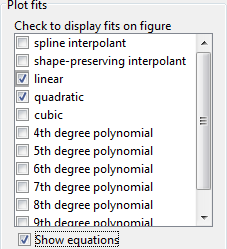
Please double-click on the newly created variable



Variables editor opens then. In the editor, in the first column we will type-in the bitrate of stream for a given codec, and in the second column – the own opinion score (1-5). After entering all the data available for the codec, please select the entire area of the data entered and plot: „MATLAB Scatter / Bubble Plot”:



Then, please run the tool (“Tools”) for a simple function matching (“Basic Fitting”). Please try to match the function to the data, preferably the best linear (“linear”) or square (“quadratic”) one. It is possible to display the pattern matching functions (“Show equations”):



For which values of bit-rate, the first audible distortion (4) is appearing? When these distortions begin to be annoying (3)?

For comparative purposes, please repeat the modelling for each codec. What are the differences between them?

## Mean Average Error

The exercise is performed using MATLAB.

The *compare\_audio.m* file required to run the simulation is in the folder of *Compare*, so you must download it. You should run the *compare\_audio.m* file. Please copy examples of wav and compressed files, which were mentioned in the previous exercise parts, to the *compare\_audio.m* directory. In one of the first lines please subsequently enter file names of wav and compressed, which will be compared:

src = audioread('vqegMM2\_C01\_Aorig.wav');

pas = audioread('vqegMM2\_C01\_Aorig.mp3');

|  |
| --- |
| **WARNING! MATLAB may have problems with paths where there are spaces. For this reason, you should remember to not to have SPACES in names of wav and compressed files and folders.** |

Some MATLAB installations fail to read mp3 files, saying:

Error using **audioread**  
Audio file I/O requires Microsoft(R) Media Foundation.  
Install this on your system and restart MATLAB.

Others fail to read ogg, saying:

Error using **audioread**  
File could not be read due to an unexpected error. Reason:  
File contains data in an unimplemented format.

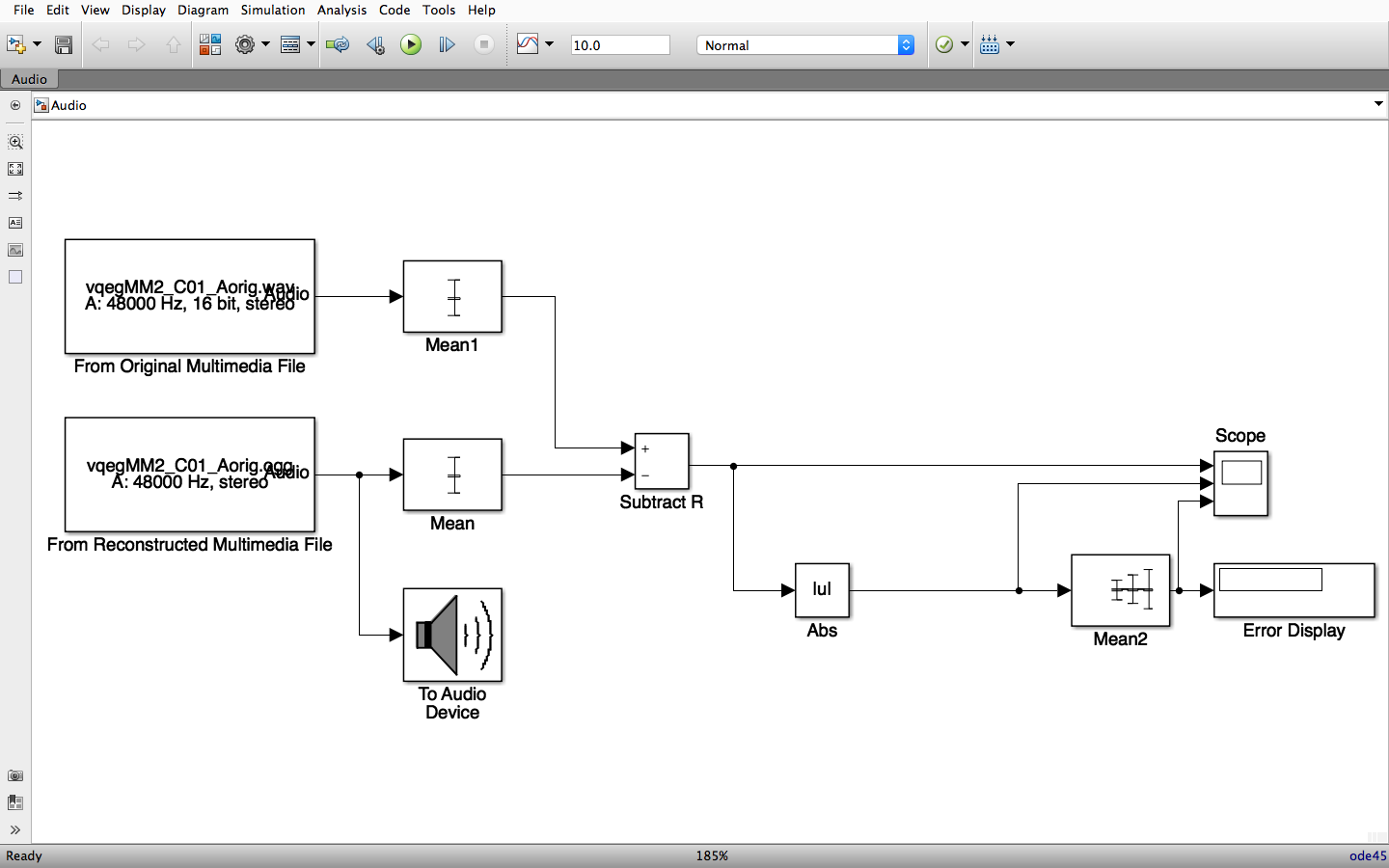
In both cases a workaround is to convert mp3 or ogg files into the (lossless) wav format.

The of *compare\_audio.m* reads wav and compressed files in the directory. The next step is to compare the uncompressed file to the output file using Mean Average Error (MAE). The results are shown in the console.

## Error Visualisation

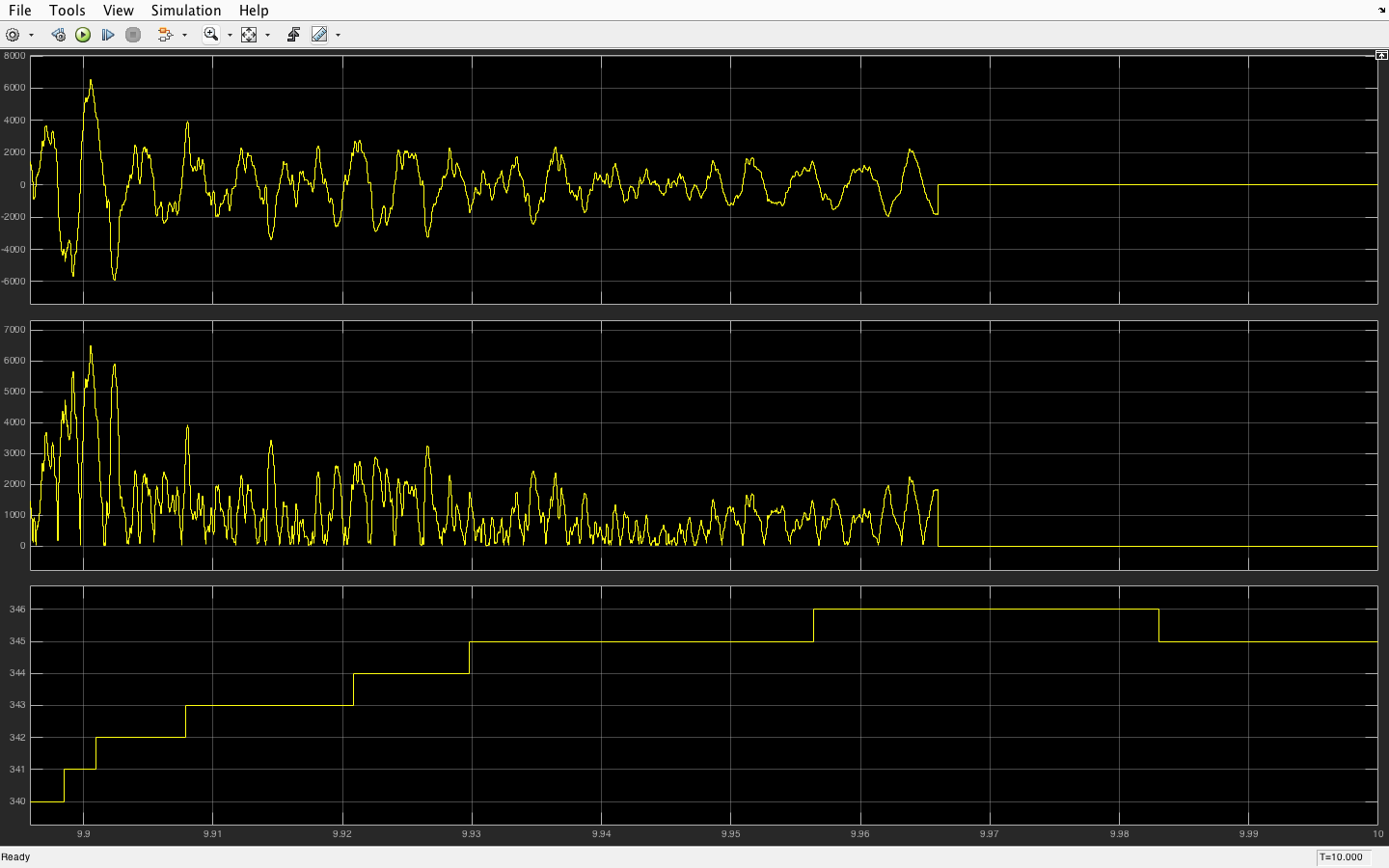
The exercise is performed using Simulink.

The *Audio.slx* file required to run the simulation is in the folder of *Audio*, so you must download it. You should run the *Audio.slx* file (please always try to use the newest model version). Please copy examples of wav and compressed files, which were mentioned in the previous exercise parts, to the *Audio* directory. In the “From Multimedia File” blocks subsequently enter file names of wav and compressed, which will be compared:



|  |
| --- |
| **WARNING! MATLAB may have problems with paths where there are spaces. For this reason, you should remember to not to have SPACES in names of wav and compressed files and folders.** |

The of *Audio.slx* reads wav and compressed files in the directory. The next step is to compare the uncompressed file to the output file and observe the comparison using the “Scope” block:



Please use the “Scale X & Y Axes Limits” button to scale the window area where results are shown.

## Comparison of Charts

The exercise is performed using MATLAB.

The files required to run the simulation are in the folder of *FFT*, which you must download. The *fft\_audio.m* uses other files, so you should only run it. Please copy examples of wav and compressed files, which were mentioned in the previous exercise parts, to the *FFT* directory. In one of the first lines please subsequently enter file names of wav **WITHOUT EXTENSION** and compressed **EXTENSION**, which will be compared:

file\_name = 'vqegMM2\_C01\_Aorig';

ext = 'ogg';

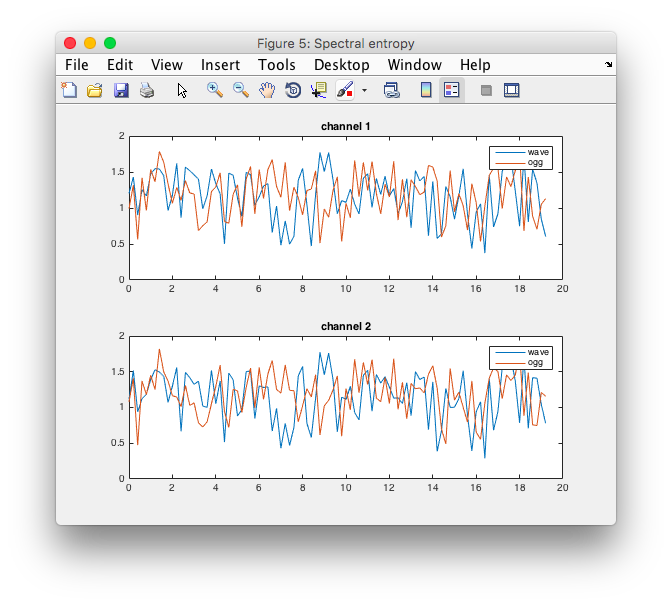
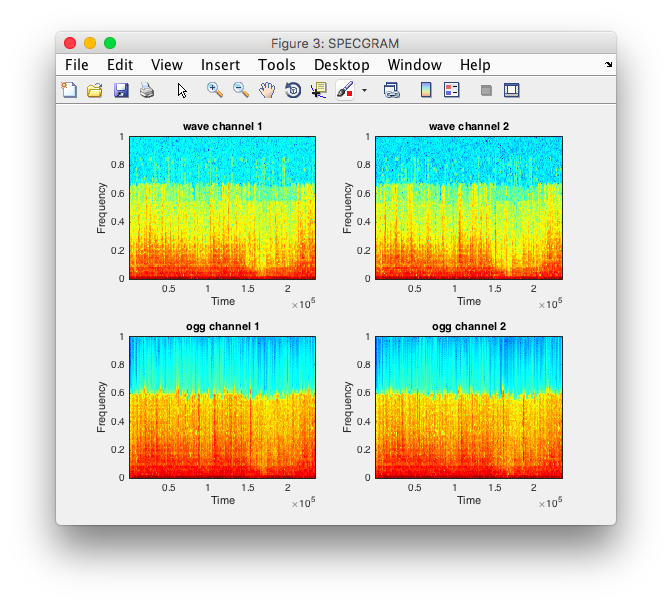
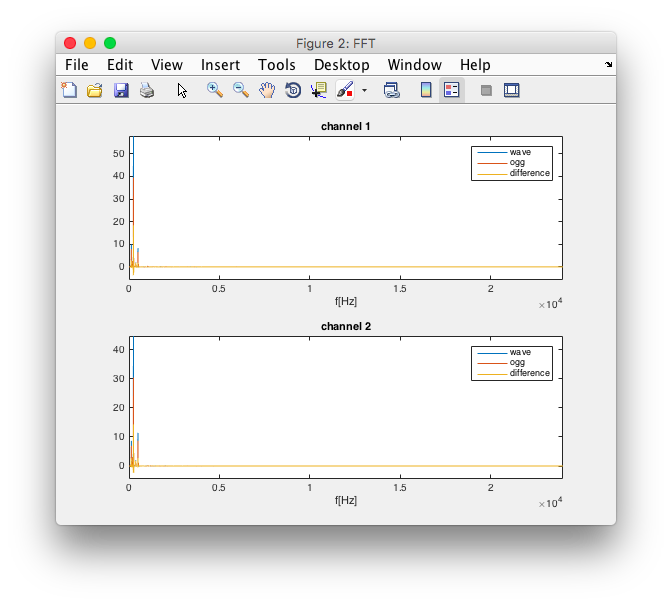
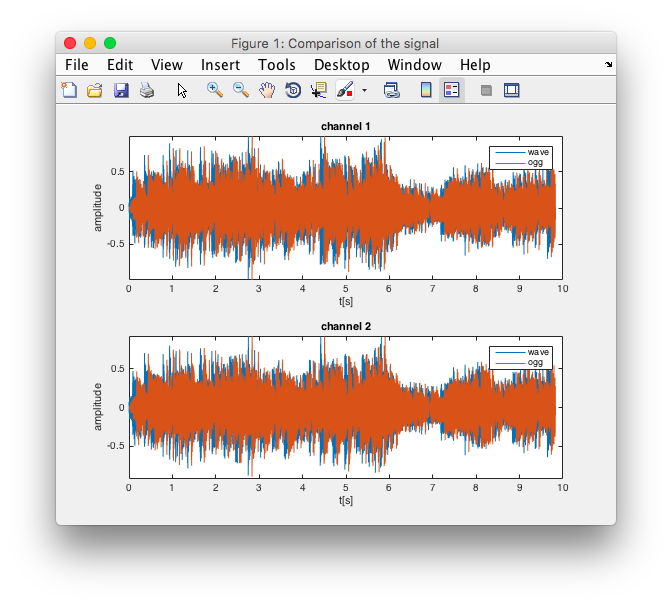
|  |
| --- |
| **WARNING! MATLAB may have problems with paths where there are spaces. For this reason, you should remember to not to have SPACES in names of wav and compressed files and folders.** |

The of *fft\_audio.m* reads wav and compressed files in the directory of *FFT*. The next step is to compare the uncompressed file to the output file using several sub-functions. The results are shown in the graphs:

1. Comparison of the signal (both channels)
2. FFT – Fast Fourier Transform
3. Specgram – The purpose of this function is to analyse the frequency time-varying signal. The result of the function is an image in which the horizontal axis has the time information, and frequency information is on the vertical axis.
4. Energy entropy
5. Spectral entropy – a measure of spectral entropy
6. Spectral Roll Off– This feature tells you how high in the spectrum lies some energy
7. Short time energy
8. Zero crossing rate

## Tasks

1. You should check the impact of compression parameters on the size of output files (best to see if a file occupies disk space), and compare the results for depending on the file size of the parameter of bitrate **(number of measurement points should be selected by the student so that it was possible to identify reliable trends)** for the same the sampling frequency and the same number of channels for different codecs. It should be noted that there are situations in which after the re-compression, minor changes to the length (in seconds) of the audio sequence is applied. Without a thorough verification code codec, it is not possible to clearly specify the reason, but it can be expected that the cause is rounding frames of sound or simply inaccurate readings of players.
2. For a number of selected sets of parameters, it should be evaluated subjectively, which codec works best. To minimize the impact of sound and acoustics to assess the efficiency of coding, a set of encoding parameters should illustrate the high degree of compression [5]. To listen to all formats it is best to use VLC. **Scale mapping has to be the scale (figure) of MOS [6].** For listening, we recommend using headphones (please pay attention to hook the handset to the appropriate output on your computer).
3. You should compare several types of wav and compressed files in MATLAB, observing differences arising on the charts and briefly commenting on them.



Do not use the settings for the multi-channel audio, or settings for samples greater than 16 bits, because the sound system in laboratory settings such does not support it.

With the technique of VBR, the reference rate is generally not equal to the bit rate resulting. This is because in the VBR technique, we can request only approximately desired rate. The actual average rate of the resulting stream depends on the characteristics of the current waveform of the original sound. In order to obtain clear differences in quality recordings encoded with varying degrees of compression, as well as to determine the exact bit rate, it is best to use the exercise mainly CBR.

There is no clearly defined minimum number of files and data points to be carried out during the exercise. You probably will not be able to test all due to the lack of time – the objective should be to maximize the number of combinations (including combinations of CBR and, if time permits, also VBR). It should also be noted that not all combinations of parameters are all technically possible.

# References

1. http://en.wikipedia.org/wiki/MP3
2. <http://en.wikipedia.org/wiki/Constant_bitrate>
3. <http://en.wikipedia.org/wiki/Variable_bitrate>
4. <http://en.wikipedia.org/wiki/Available_Bit_Rate>
5. <http://en.wikipedia.org/wiki/Data_compression_ratio>
6. h[ttp://en.wikipedia.org/wiki/Mean\_Opinion\_Score](http://en.wikipedia.org/wiki/Mean_Opinion_Score)
7. Austerberry D.: *The Technology of Video and Audio Streaming*. Focal Press, Oxford 2002
8. Halsall F.: *Multimedia Communications, Applications*, Networks, Protocols and Standards. Addison-Wesley, Essex 2001
9. Griwodz C., Halvorsen P.:” Media and User Behaviour”. INF 5070 – *Media Servers and Distribution Systems*, 2005
10. Thomas Wiegand: „New Techniques for Improved Video Coding”