# 10 – Digital recording, compression, and Internet delivery

At present and in the future there is practically no and will not be any other sound recording than digital recording. Digitized sound can be easily converted between different digital format including lossy and lossless compression. Compressed sound files can be easily delivered via the Internet in the form of download and streaming. Related questions will be discussed in this chapter.

# Analog / digital conversion

Replacing the wave form by a sequence of <u>heights</u> sampled at the <u>sampling frequency</u> and stored in the <u>binary</u> format (base 2)

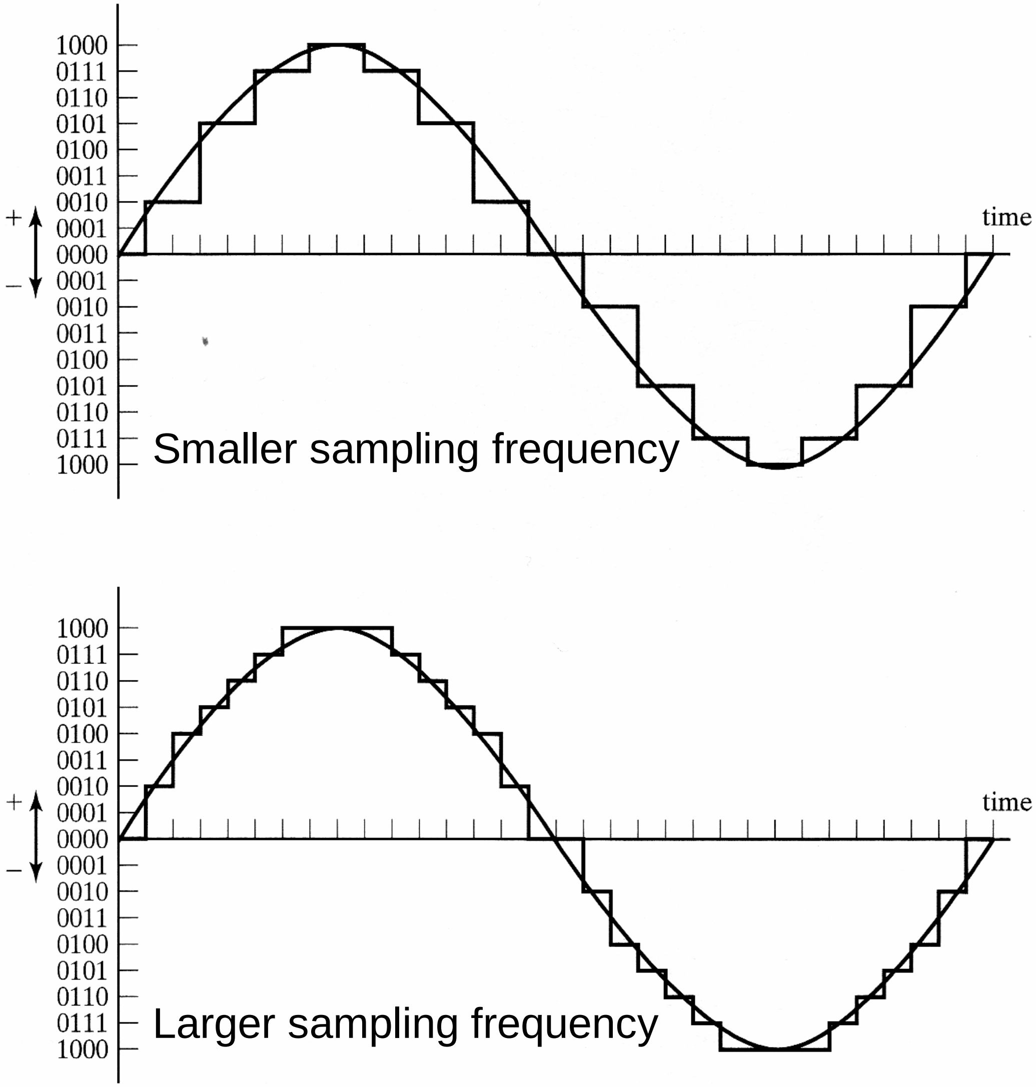


**Table 7-1** Numbers up to 32 in Base 10 (Decimal) and Base 2 (Binary)

| Base 10 | Base 2 |
|---|---|
| 0 | 000000 |
| 1 | 000001 |
| 2 | 000010 |
| 3 | 000011 |
| 4 | 000100 |
| 5 | 000101 |
| 6 | 000110 |
| 7 | 000111 |
| 8 | 001000 |
| 9 | 001001 |
| 10 | 001010 |
| 11 | 001011 |
| 12 | 001100 |
| 13 | 001101 |
| 14 | 001110 |
| 15 | 001111 |
| 16 | 010000 |
| 17 | 010001 |
| 18 | 010010 |
| 19 | 010011 |
| 20 | 010100 |
| 21 | 010101 |
| 22 | 010110 |
| 23 | 010111 |
| 24 | 011000 |
| 25 | 011001 |
| 26 | 011010 |
| 27 | 011011 |
| 28 | 011100 |
| 29 | 011101 |
| 30 | 011110 |
| 31 | 011111 |
| 32 | 100000 |

## Sampling frequency

To faithfully store the sound in the digital form, the sampling frequency should be at least two times greater than the maximal frequency in the sound signal. Since human ear hears signals with frequencies up to 20 KHz, the CD standard uses the sampling frequency 44.1 KHz. Another sampling frequency of 48 KHz was used by higher-quality recorderd as DAT recorder.

Modern high-resolution digital sound recordings use sampling frequencies of 96 KHz or even 192 KHz that can be found in good DVD players, too. Vinil records store signals with frequencies up to 50 KHz that are unaudible to humans. Nevertheless audiophiles say that such high-frequency components, although unaudible directly, contribute to the „athmosphere“ of the recordings. That is why vinily records sound better than CDs. High-resolution recordings at 96 and 192 KHz can capture these high frequencies. However, the resulting file sizes are pretty large.

For recording signals limited frequency range, such as speech, lower sampling frequencies can be used, such as 32, 16, or even 8 KHz.

# Sample depth

Sample depth or bit depth characterizes how the signal heights are converted into binary numbers (the same applies to videos and images.) The CD standard uses the 16-bit A/D conversion, so that, with the sampling frequency, the CD standard is 44.1/16. 16-bit sample depth means that the heights (from zero to the maximal height) are discretized and replesented by $2^{16}$=65536 different numbers (the minimal nimber being 0 and the maximal number being $2^{16}–1$). In the binary system, these numbers are 16-figures long. This means that any of these numbers occupies 16 bits storage space, on the computer or any other digital carrier. Discretizing the hights of the wave form helps to reduce the amount of the stored information.

One could ask why such a great accuracy, discretizing the heights by $2^{16}$=65536 different numbers is needed. The answer is that our ear is a very sensitive device. Discretizing the heights with the sample depth 8 bit, that is, by $2^8$=256 different numbers would result in a discretized wave form that would be indistinguishable from the original wave form with the eye. However, our ear easily recognizes a 8-bit digital signal as coarse.

To get a high-fidelity recording, it is more important to increase the sampling depth than the sampling frequency above the CD standard. This increases the dynamic range of the recording, that is, the difference between the quietest and loudest recorded signals. In the 16-bit recording, some very quiet sounds are coded with 0000000000000000, that is, as the absolute (digital) silence. This means that these sounds are lost. 24-bit recording uses $2^{24}$=16777216 different numbers for the digital coding of the signal. Thus very quiet sounds do not get lost.

Another problem is digital noise arising in the editing process, especially amplification of the signal. This can lead to emerging of artificial very quiet sounds out of the digital silence. Many editing steps result in the growth of this digital noise. Increasing the bit depth to 24 or even to 32 bits eliminates this problem.

## Sample depth and the dynamic range

Dynamic range mentioned above is the difference between the quietest and the loudest recorded sound in decibels. For the sample depth SD the quietest sound above the digital silence has the amplitude $I_0$=1 (in relative units), while the loudest sound has the amplitude $I_{\max}$=$2^{\mathrm{SD}}-1$. The ratio of their intensities is

$$\frac{I_{\max}}{I_0} = \left( \frac{2^{\mathrm{SD}} - 1}{1} \right)^2 = \left( 2^{2\mathrm{SD}} - 1 \right)^2$$

so that for the dynamic range one obtains

$$\mathrm{DR} = 10 \log_{10} \left( \frac{I_{\max}}{I_0} \right) = 20 \log_{10} \left( 2^{\mathrm{SD}} - 1 \right) \cong 20 \log_{10} 2^{\mathrm{SD}} = 20\mathrm{SD} \log_{10} 2 \cong 6\mathrm{SD}$$

The CD standard is SD=16, so that one obtains for the dynamic range $\mathrm{DR}=96\ \mathrm{dB}$. This is still smaller than the dynamic range of the ear 120 dB. For the 24-bit recording one has $\mathrm{DR}=144\ \mathrm{dB}$ that covers the dynamic range of the ear. These calculation explain why CDs sound not so good as LPs. This is because of the sample depth reather than because of the sampling frequency.

# Bit rate and file size

Bit rate BR is the amount of information in bits transferred by a digital signal during one second. Bit rate is the product of the sampling frequency $f_S$ and the sample depth SD. For stereo signals one has to multiply everything by the number of channels 2. Thus one obtains the formula

$$\mathrm{BR} = 2\mathrm{SD} f_S$$

Bit rate of the CD signal is thus 2×16×44100=1411 kbit/s. We usually write kbps instead of kbit/s.

Byte (octet) consists of 8 bits, so that the transfer rate of the CD signal is 1411/8 = 176 KByte/s.

Let $t_R$ be the total time of the recording in seconds. Then the file size FS in bytes is given by

$$\mathrm{FS} = t_R \mathrm{BR} / 8$$

One second of the CD recording is 176 kbytes and one minute of the CD recording is 60×176 = 10584 KB = 10.5 MB. If the CD capacity is 700 MB, one can store 700/10.5 = 67 minutes music on it.

For high-resolution audio the bit rate is higher than for the CD-quality audio.

| | | |
|---|---|---|
| SD = 16 bit | $f_S$ = 44.1 KHz | BR = 2×16×44.1 = 1411 kbps |
| SD = 24 bit | $f_S$ = 44.1 KHz | BR = 2×24×44.1 = 2116 kbps |
| SD = 24 bit | $f_S$ = 96 KHz | BR = 2×24×96 = 4608 kbps |

At 24bit/96KHz, a 700 MB CD can store only $700\times10^3/(4608/8\times60)$ = 20 minutes of music

# Internet delivery of digital sound; Streaming

Bit rate is an important quantity for the data delivery via the Internet. For instance, one can calculate how much time $t_{tr}$ will it take to transfer 1 minute of music in the CD format via a DSL line with the transfer rate of TR=1200 kbps. The solution is the following. The file size in bits is given by FS = $t_R$ BR. Thus the time needed to transfer this file with the transfer rate TR is given by

$$t_{tr} = \frac{\mathrm{FS}}{\mathrm{TR}} = t_R \frac{\mathrm{BR}}{\mathrm{TR}}$$

For BR=1411 kbps one obtains $t_{tr}$ =1.18 min > $t_{R}$..

Streaming is delivery of music via the Internet that plays as it is being transfered. Evidently streaming requres TR>BR, otherwise the music stream will be interrupted. Streaming of music in the CD format is possible if the Internet connection is fast enough, TR>1411 kbps. Even in this case it will be a pretty heavy load on the Internet connection that should be avoided. o reduce the file size, compression of audio files is being done.

# Audio compression

Audio files can be compressed without any data reduction, just like any other files. There are several lossless encoders optimized for audio. Probably the most efficient of them is Monkey's Audio (files with extension *.ape). MA compresses files with complex music such as symphonic by a factor of about two and simpler music such as smooth piano by factors of up to four. The bit rates of the compressed files are reduced by exactly the same amount as their sizes. The resulting files have bit rates of about 500 kbps that is comparable with the average video files and they can be streamed. The files are delivered in the compressed (encoded) form and decoded on the fly on the recieving side.

In addition to the lossless compression (or encoding) there is lossy encoding. The first and most well known of them is the MP3 encoding. All algorithms of lossy encoding use the masking phenomenon considered earlier. The sounds that are masked by louder sounds with close frequency are removed from the recording, reducing the amount of information and thus the file size. When a lossy encoded file is decoded (decompressed), its file size becomes the same as that of the initial uncompressed file. The amount of information in it is, of course, smaller than in the original file. This is not a contradiction. One can digitally record silence (zero information) and the file size will be the same as that of the file with a complex music of the same duration.

The main characteristic of the lossy compressed files is their bit rate. Average MP3 files of 128 kbps are by a factor of  about 10 smaller than uncompressed CD-quality audio files. Their quality is generally good enough, although one can hear compression artefacts, especially in fast percussive piano music that sounds somewhat coarse. MP3s at 256 kbps are already practically indistinguishable from the original.

MP3 remains the most popular lossy audio compression format, although many other formats such as WMA or MP4 audio (M4A or AAC) are much better. The reason is, probably, conservatism of hardware manufacturers who do not bother much to support other formats. The most advanced and open (nonproprietary) format AAC is practically boycotted, probably as an opposition to the dominance of Apple who is basing on this format. M4A files at 128 kbps sound indistinguishable from the original, unlike MP3 at 128 kbps.

Note: 1 min CD-quality audio uncompressed  takes ~10 MB
………………… compressed at 128 kbps…    ~1 MB

There are three variants of lossy compressed files: Constant bit rate (CBR), variable bit rate (VBR) and average bit rate (ABR). As music can be more or less complicated at different moments of time, different bit rates is required to reach a consistently good quality. Places with simpler music can be encoded with a smaller bit rate while those with more complex music require a higher bit rate. The ABR encoding seems to be the best as it redistributes the number of available bits over the whole recording while keeping the average bit rate fixed. CBR encoding is less intelligent and it does not make such a redistribution. VBR encoding does not aim at a particular bit rate but it rather aims at a desirable quality level. VBR encoded recordings of a simple music are smaller than those of a complex music, at the same quaity level. This is not bad, in principle. However, is some cases the quality estimation fails for some reasons and VBR encoding produces poor results. Since checking all resulting VBR encodings before delivery is tiresome, it is better to use ABR encoding.