# Audio Compression Techniques

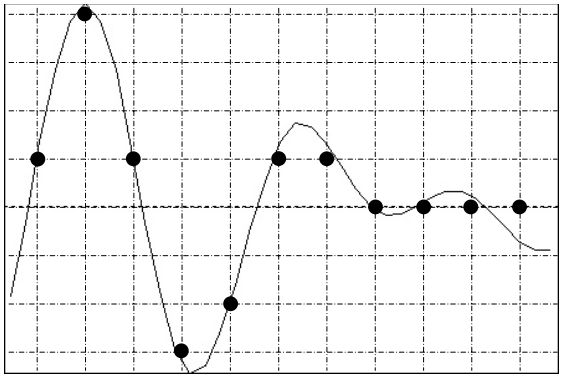
This guide describes important audio compression techniques and the effects they can have on the quality of the resulting sound files.

**General Framework:**

- In archiving programmes such as DOBES it is important to understand the mechanisms that increase or decrease the quality of the documentation.
- A general principle for archives is that the best possible quality should be generated, since we don't know how future generations will use the resources.
- In particular the popular MiniDisc and MP3 technologies imply compression. Therefore, we argue that they should only be used in exceptional cases.

## 1. High Quality Digital Audio

A high quality digital audio representation can be achieved by directly measuring the analogue waveform at equidistant time steps along a linear scale. This procedure is indicated in figure 1. The distance between the grid markers on the time axis is defined by the sample frequency, and on the amplitude axis by the maximal amplitude divided by the number of bits. Therefore, higher sample frequencies give narrower grids along the time axis and more bits (narrower grids) along the amplitude axis. The individual values to be stored on computers and representing the sound wave can only be located on the cross points of the underlying grid. Since the actual sound pressure will never be exactly on these cross-points, errors are made, i.e. digitization noise is introduced. It is obvious that the narrower the grids are the smaller the error will be. Therefore, a higher sample frequency and a larger number of bits will increase the quality.
However, for speech signals we know that the highest frequency a child's voice can create is in the order of 8 kHz. According to the Nyquist rule, a double as high sample frequency is sufficient to represent a frequency component nicely, so 20 kHz is sufficient. Our human ear is said to be sensitive up to 20 kHz, therefore the HiFi norm requires sampling frequencies above 40 kHz. Also for normal speech signals and singing, a dynamic range of 96 dB is said to be sufficient which is achieved by 16 bits. For certain pieces of classic music, however, 96 dB is not sufficient to hear the low sound level signals as well as the high level ones. Therefore, for classic music 24 bits are recommended that yield a dynamic range of up to 144 dB.



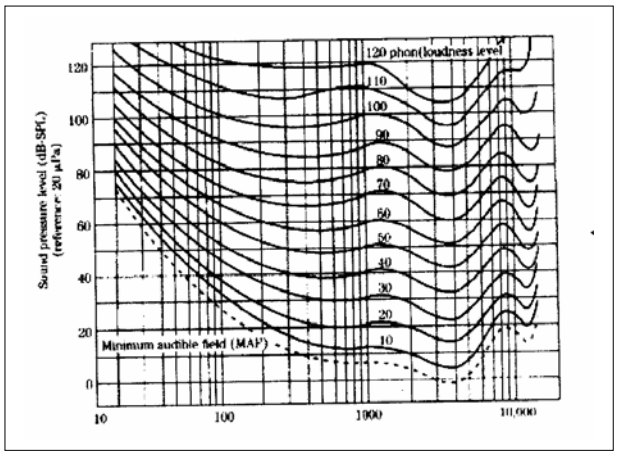
Linear PCM (pulse code modulation) is a direct digitization technique for sound waves. At equidistant moments in time defined by the sample frequency and equidistant points at the amplitude scale, samples are taken, i.e. digital numbers are generated that represent the original wave form. While digitizing, an error is introduced that is the smaller the narrower the grids on both axis are. For speech signals it is said that 20 kHz and 16 bit is sufficient, therefore the generally applied 44.1/48 kHz and 16 bit are acceptable for field work material. For special music sounds etc it may be necessary to operate with 96 kHz and 24 bits.

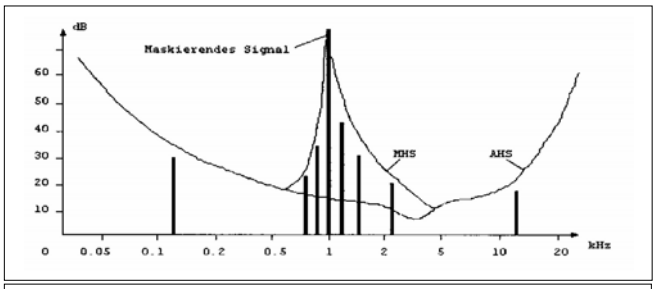
## 2. Compression Principles

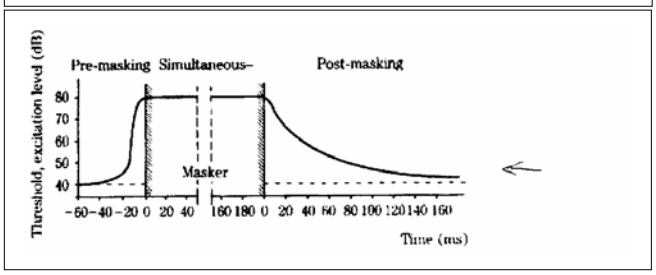
Modern chip makers invested time to work out compression mechanisms that reduce the amount of storage space needed and nevertheless produce intelligible sounds. The sample frequency and the number of bits cannot be reduced, therefore they were looking to other methods. The designers of the ATRAC (used in MiniDisc recorders) and the MP3 compression algorithms both used some characteristics of the human sound perception system to reduce the capacity needed. The algorithms are "psycho-acoustically" based compression techniques, since they claim to reduce information in the sound signal that cannot be heard by the human ear anyhow.

Peter Wittenburg, MPI for Psycholinguistics, 3. April 2005

# Audio Compression Techniques 2




The algorithms both focus on the representations of those parts where the ear is more sensitive which is around 1000 Hz (linear PCM treats all components equally). This means that fricatives for example are not represented as well as some of the formants of, for example, vowels. Further, the algorithms use two masking effects: (1) The spectral masking says that if there is a dominant peak in the spectrum somewhere (perhaps introduced by a bird or so) the ear will filter out all neighboring frequency components that are below an envelope function derived from psychoacoustic experiments. Thus relevant frequency components of the speech signal could be filtered out. (2) The temporal masking says that if a loud sound occurs the human ear will filter out other sounds that are below a certain threshold function also derived from psychoacoustic experiments, whereby the post-masking effect is much stronger than the pre-masking effect. Also here useful sound components over time could be filtered out.

## 3. Measurable Effects

As far as we know just a few people (von Son, Campbell, Wittenburg) have analyzed the results of MP3 and ATRAC compression. Some major findings can be summarized as follows:

- there is almost no difference in Pitch, Spectrum, Formant extraction except at high frequency signals such as they occur for fricatives
- there are only small consequences even for the calculation of production parameters
- changing the microphone type can have more severe consequences than this type of compression on the quality of the sound representation
- for “normal” quality speech signals, there were no significant differences in hearing tests
- the effect of time filtering after loud sounds is noticeable

## 4. Recommendations

The Technical Committee of IASA (International Association of Sound Archives) has made a clear statement that no compression techniques should be used to create and store sound material when long-term preservation is intended. The best possible quality has to be maintained. DOBES and similar archives have to follow these recommendations.
The compression format leads to another difficulty: we don’t know how long the algorithms will be supported and understood. A direct digitization is preferable in this respect as well.

**Only create MiniDisc and MP3 recording where there is no opportunity for making uncompressed high quality recordings.**
**Information that is lost by compression cannot be recovered.**