Digital audio

In the past century, sound was mostly recorded on analogue media, such as reel tapes or cassette tapes. Nowadays, audio recordings tend to be born digital.

How are digital audio files created?

First, in order to record live sound, a microphone is required for capturing small fluctuations of air pressure into changes in voltage that can travel within the audio cable. There are many kinds of microphones, but these devices are usually not designed to digitise sound, only to "grab" it, i.e., to function as sound "sensors".

If one wishes to record sound into a format that can be used in a mobile device or in a computer, the sound needs to be digitised. In the digitisation process, discrete **samples** or measurements are taken of the sound (that was converted into electric signal by the mic) at very, very short but very, very steady intervals. The device that takes care of the measurements is called an A/D converter (Analog to Digital). The A/D converter can be located, e.g., in an audio recorder, in the audio circuit of a mobile device, or in the sound card of a computer.

The digitization only needs to be performed once. In case the recording is intended for research use, it is important to use a high-quality device for digitization (or digital recording), so as to ensure that the sampling is performed as reliably as possible. For instance, the audio recording features in a regular office laptop are usually much worse than those in digital audio recorders or in external audio interfaces that can be plugged into the computer.

The digitized sound is stored as a sound file that can be copied, converted into various formats and played back in a computer or in digital players. The original sound file and all its copies will remain identical (i.e., all of them will contain the same exact samples) unless the file is transformed by downsampling or resampling the contents or compressed with a lossy method (and unless the hard drive, flash drive or some other media is damaged, making the files unreadable). In other words, the properties of the A/D converter in your device do not play any role after the original sound has been digitised and the sound file has been saved.

Sampling sound

When digitizing, samples could be taken from the sound at a rate of, e.g., 22050 times per second. In this case, the sampling frequency or **sample rate** of the audio file would be 22050 Hz (hertz), i.e., 22.05 kHz (kilohertz). In the computer, each of these samples is represented by one number that is expressed in bits (since we are talking about computers!). The **sample size**, or bit depth, represents the amplitude resolution. On the basis of the steady sequence of numbers, the waveform of the sound signal can then be reconstructed.

The sample rate indicates the frequency resolution of the digital recording: the higher the sample rate, the denser sound waves, i.e., the higher frequencies can be included in the recorded signal. The highest frequency that can possibly be represented by a given audio signal (the so-called *Nyquist frequency*) is exactly half of the sample rate of the signal. That is, in case the sample rate is 22050 Hz, the digitized sound can potentially include frequencies up to 11025 Hz – not higher than that. When recording speech for research purposes, you should use the sample rate of at least 22050 Hz but preferably 44050 Hz, and the sample size of at least 16 bits. Even higher resolutions can of course be selected when required.

Lossy and lossless compression

Audio files tend to take up lots of disk space, especially when high accuracy is required, i.e., when the samples have been collected at a very high rate. Audio compression methods have been developed in order to save space.

Uncompressed sound file formats include, for instance, WAV developed by Microsoft, and AIFF, developed by Apple. Out of these, WAV is used more widely and it can be generally recommended. In research use, it may be a good idea to keep at least one uncompressed copy of the original material in WAV format.

MP3 files were originally designed for storing and listening to music in mobile devices, and they use *lossy compression*. When creating an MP3 file, the general idea is to only represent those details of the sound that human listeners are known to be able to hear, while throwing away those details that will not significantly affect the listening experience. For many purposes, MP3 can be a convenient format. However, lossy compression is an issue for speech researchers, since there is no way to find out which properties of the original sound have been modified or excluded or which time spans the changes may have affected. Moreover, MP3 is a proprietary format. Since Praat is open source, it cannot be used for creating MP3 files, but they can be opened and played back in Praat.

FLAC is one of the currently existing *lossless compression* methods. FLAC files can usually be opened in Praat, or they can be uncompressed back into WAV files, for instance. However, I have run into some FLAC files, possibly saved in older devices, that did not open in Praat directly but only after uncompressing them with the original FLAC decoder. However, at least for larger speech corpora, FLAC can provide an opportunity for saving some disk space.

Stereo or mono?

Stereo recordings provide the listener with an impression of space that can add to the experience when listening to music, for instance. When listening to a stereo recording through loudspeakers or headphones, the small timing differences between the sound events occurring in the left and right audio signals will give you the impression of three-dimensional space. The stereo experience can be created just by recording the same show or sound source with two microphones in separate channels. Of course, sound can also be recorded in multiple channels, mixed and post-processed, or the entire recording can be created digitally to begin with. So, the listener’s perception can be manipulated in many ways.

Audio recorders, mobile devices or sound editing tools might record in stereo by default. This may not be a problem. Nevertheless, you should remember that sound files take up disk space. Storing one minute of CD quality audio takes about 10 MB (megabytes). If you convert an uncompressed stereo file into a single track file, i.e., mono, it will only require half of the original space.

In speech research, the spatial impression is often not useful at all – rather, the production of a given speaker should be recorded as directly and “cleanly” as possible. If it is desirable or possible to use only one, single-channel microphone for recording, a single channel is sufficient in the file as well. A stereo file or a multichannel file can be useful if there are several speakers. When each participant in the conversation is captured with an individual microphone and when the recorded tracks can be separated, each speaker’s voice can be played individually or together with the other speakers. This will make transcription work much easier, and acoustic methods can be used for analysing the material in useful ways.