digital multimedia

nigel chapman and jenny chapman

Graphics and Colour
Video and Animation
Sound
Text and Typography
Hypermedia
Flash and DOM Scripting
Multimedia and Networks

Third Edition

8

Sound

The Nature of Sound

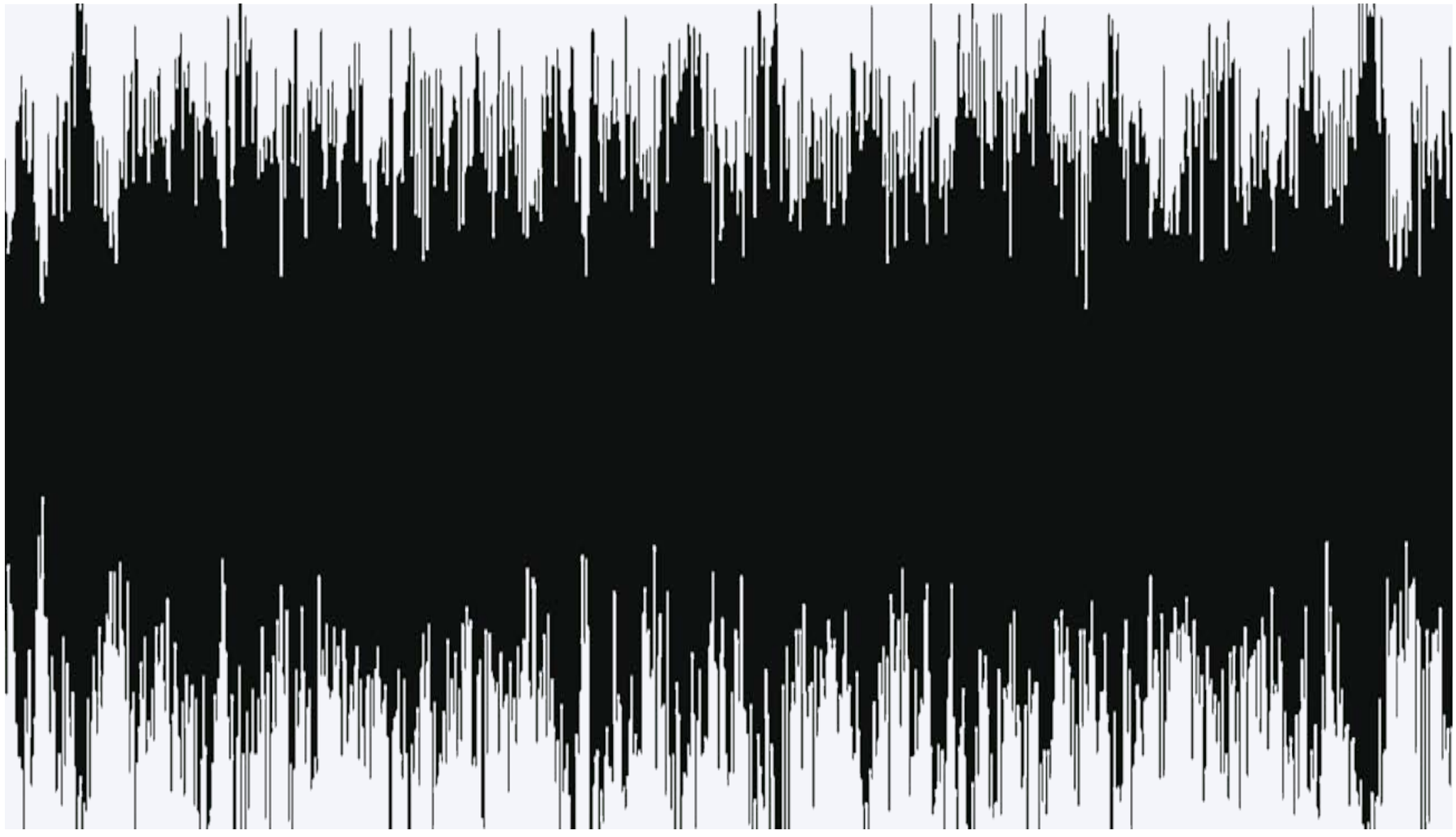
Sounds are produced by the conversion of energy into vibrations in the air or some other elastic medium, which are detected by the ear and converted into nerve impulses which we experience as sound.

A sound's frequency spectrum is a description of the relative amplitudes of its frequency components.

The human ear can detect sound frequencies roughly in the range 20 Hz to 20 kHz, though the ability to hear the higher frequencies is lost as people age.

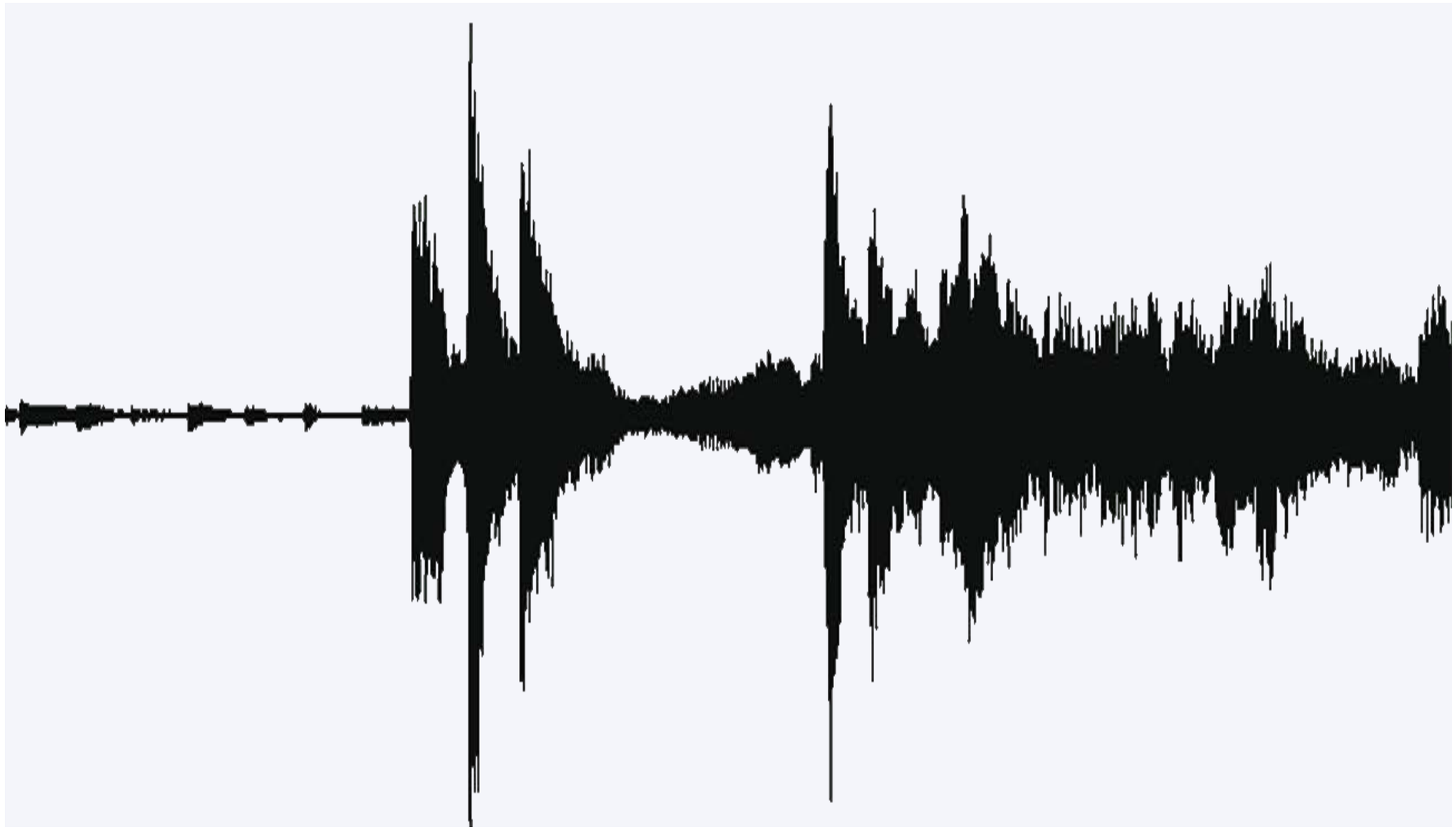
A sound's waveform shows how its amplitude varies over time.

Speech: "Feisty teenager"

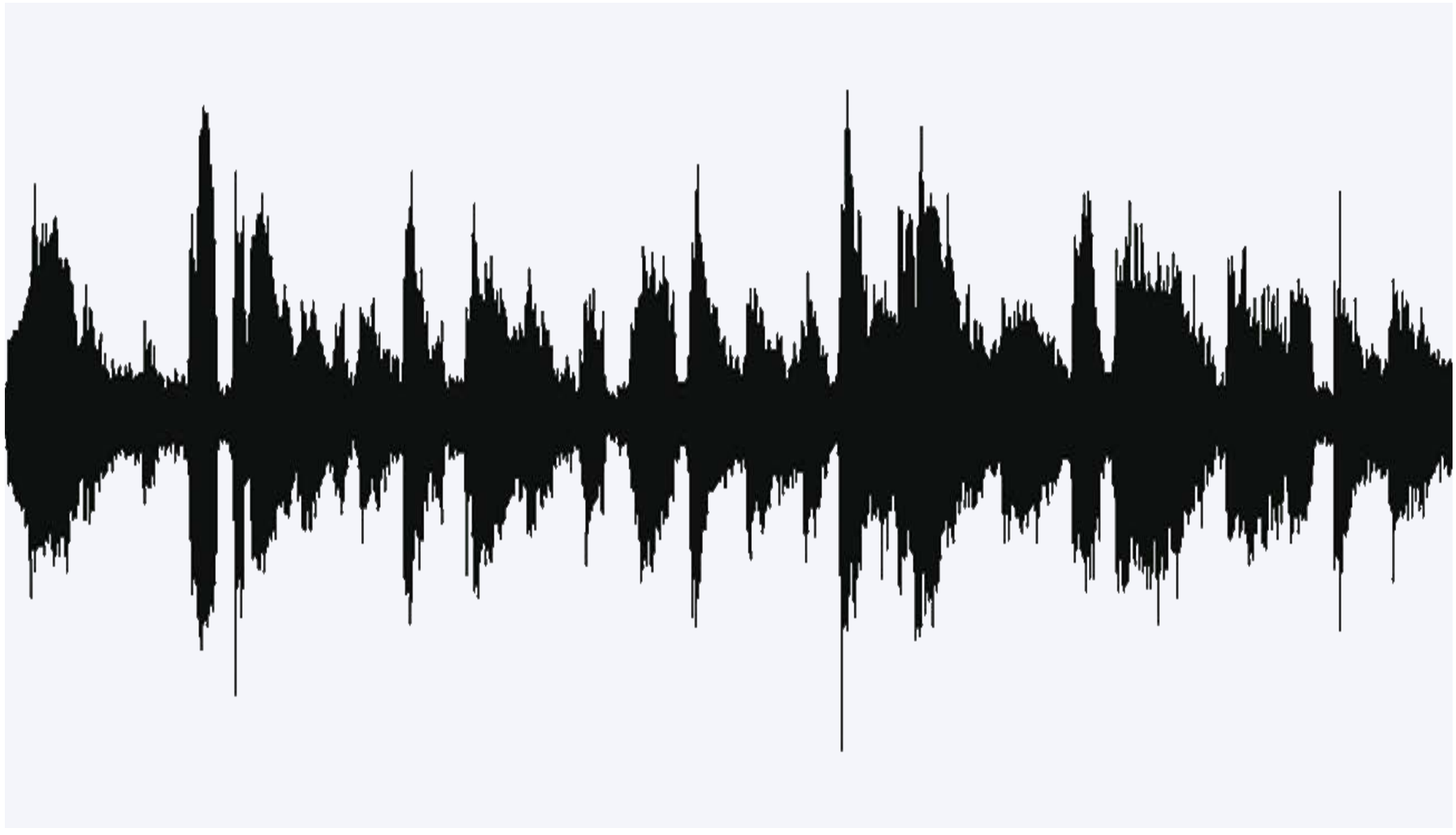


Music: Didgeridoo

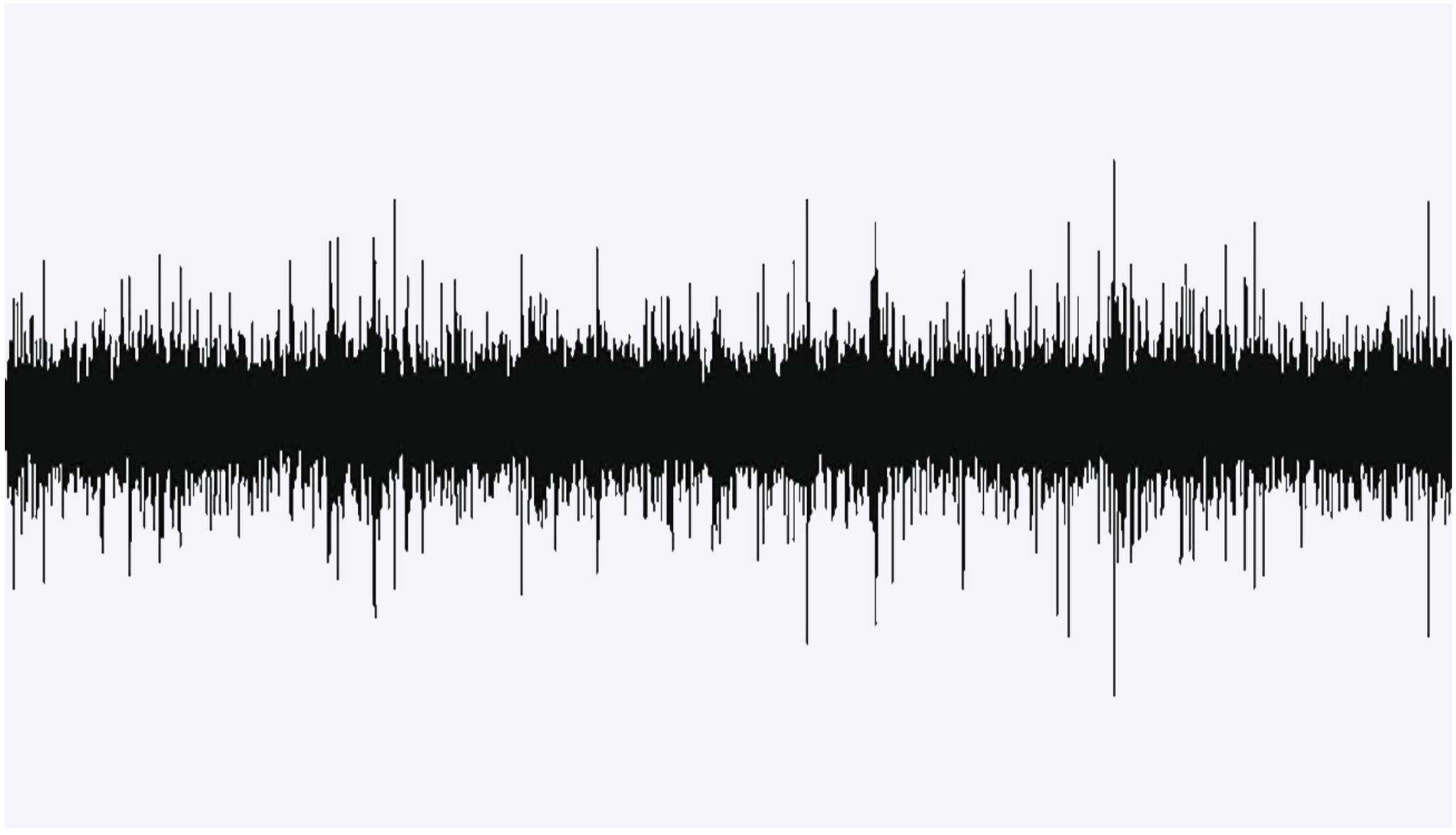
Music: Boogie-woogie



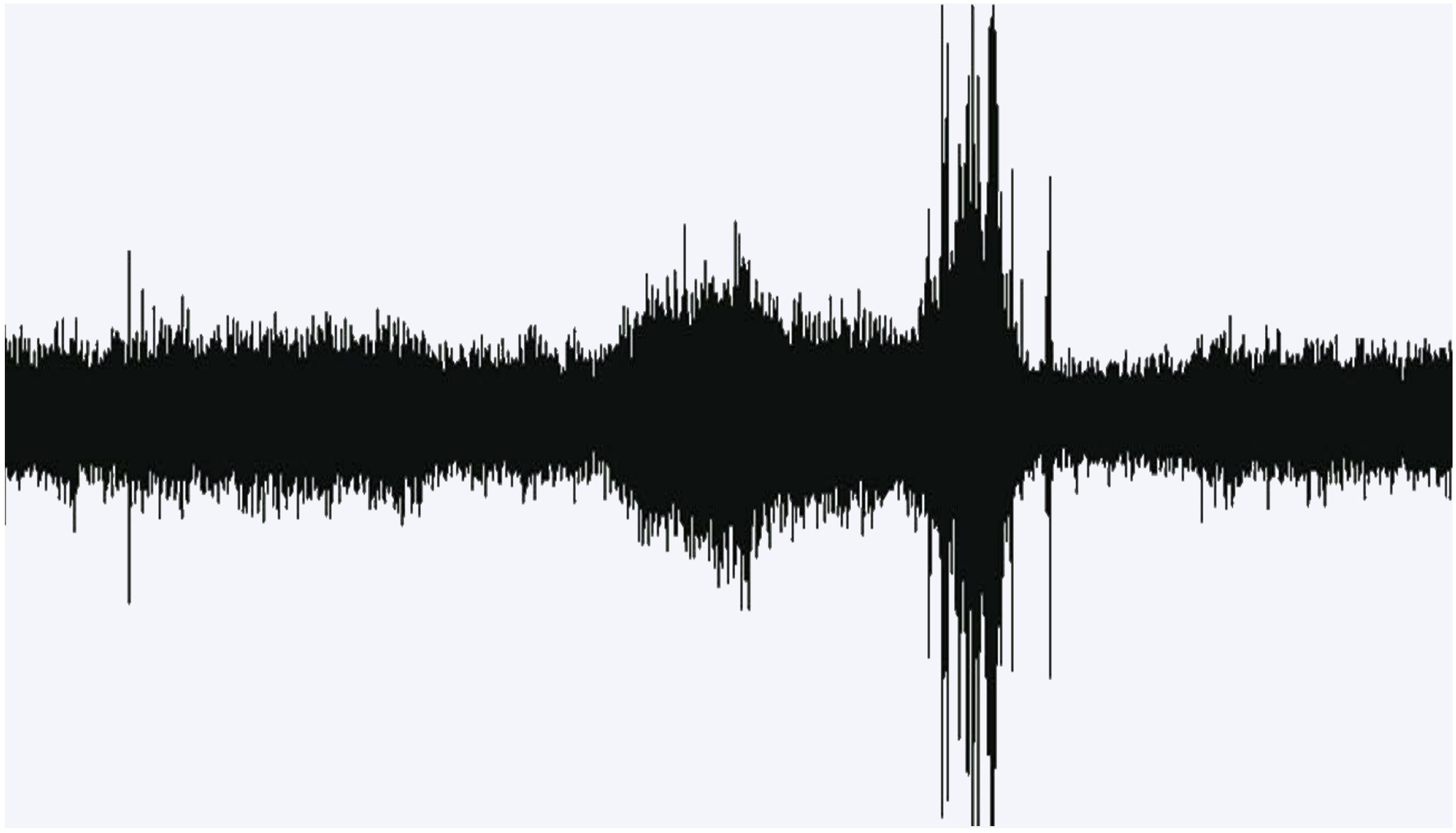
Music: Contemporary classical piece for violin, cello and piano



Singing: "Men grow cold…"



A trickling stream



The sea

Perception of sound has a psychological dimension.

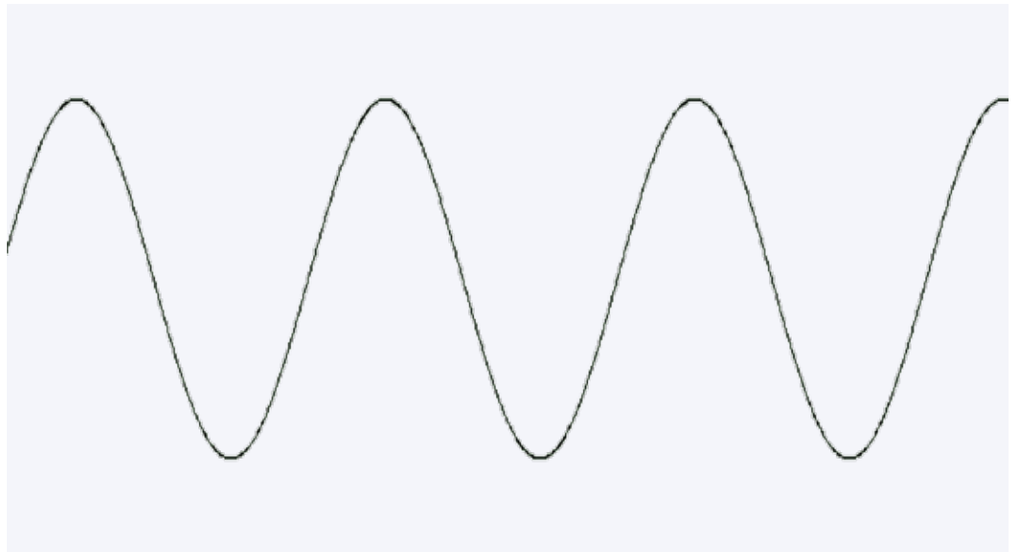
CD audio is sampled at 44.1 kHz. Sub-multiples of this value may be used for low-quality digital audio. Some audio recorders use sampling rates that are multiples of 48 kHz.

Audio sampling relies on highly accurate clock pulses to prevent jitter.

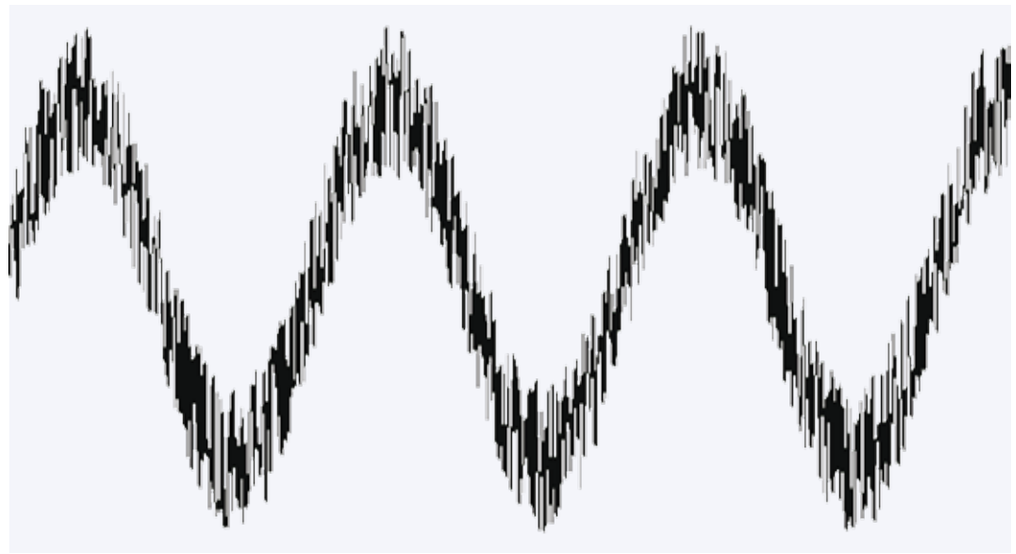
Frequencies greater than half the sampling rate are filtered out to avoid aliasing.

CD audio uses 16-bit samples to give 65,536 quantization levels.

Quantization noise can be mitigated by dithering, i.e. adding a small amount of random noise which softens the sharp transitions of quantization noise.



Undersampling a pure sine wave



Dithering

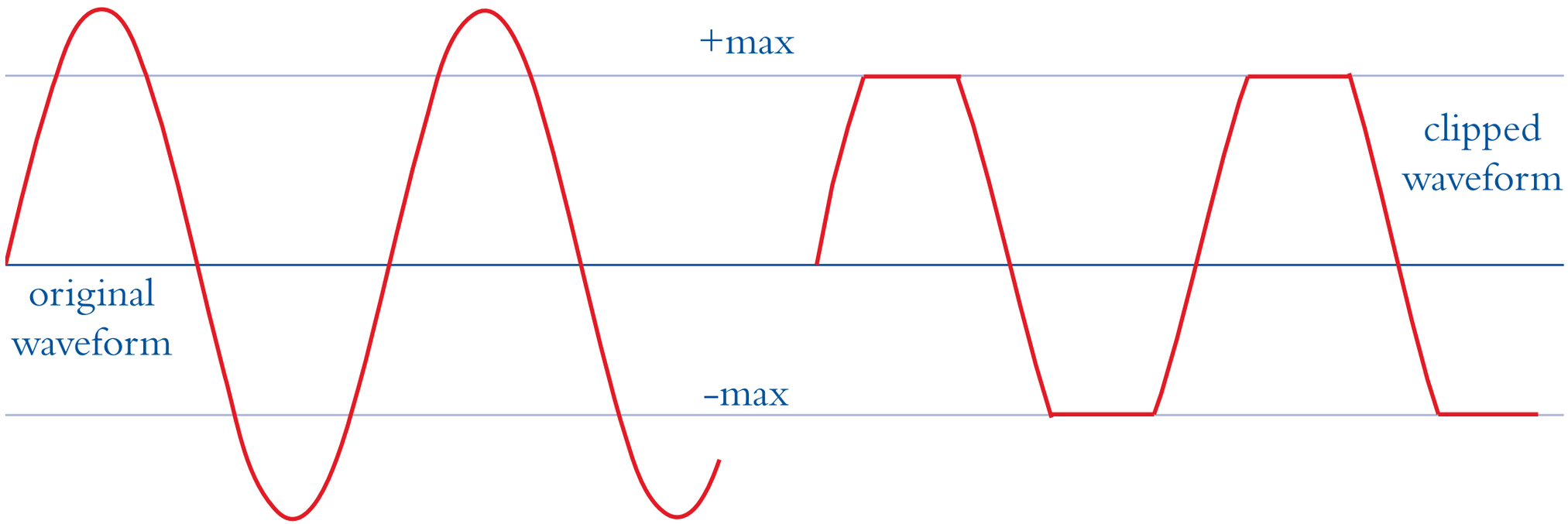
Sound may be stored in AIFF, WAV or AU files, but on the Internet the MP3 format is dominant. MP3 data may be stored in QuickTime and Flash movies.

Processing Sound

For a sampling rate of r Hz and sample size of s bits, each second of digitized sound will occupy $rs/8$ bytes.

For CD quality, $r = 44.1 \times 10^3$ and $s = 16$, so each second occupies just over 86 kbytes (for a mono signal).

If the recording level is too high, clipping will occur, causing distortion.

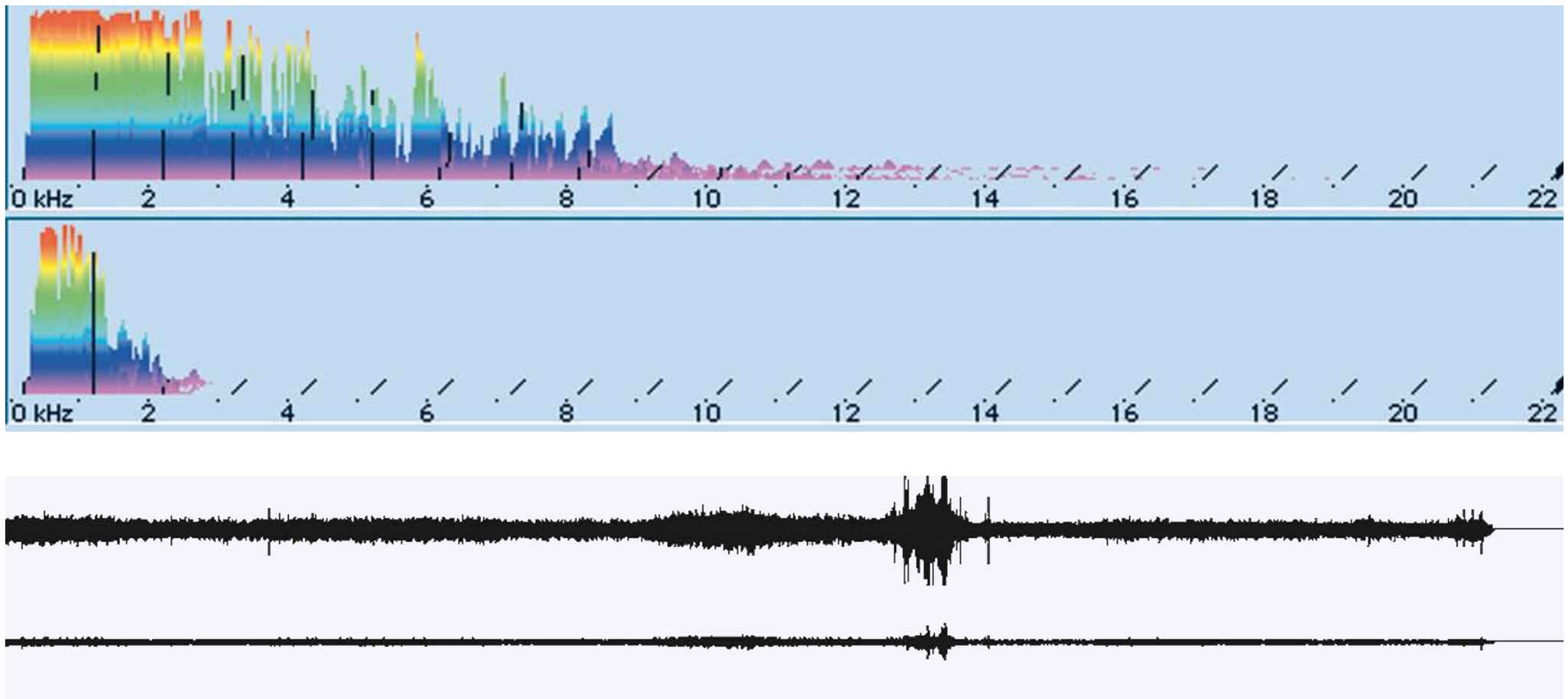


Clipping

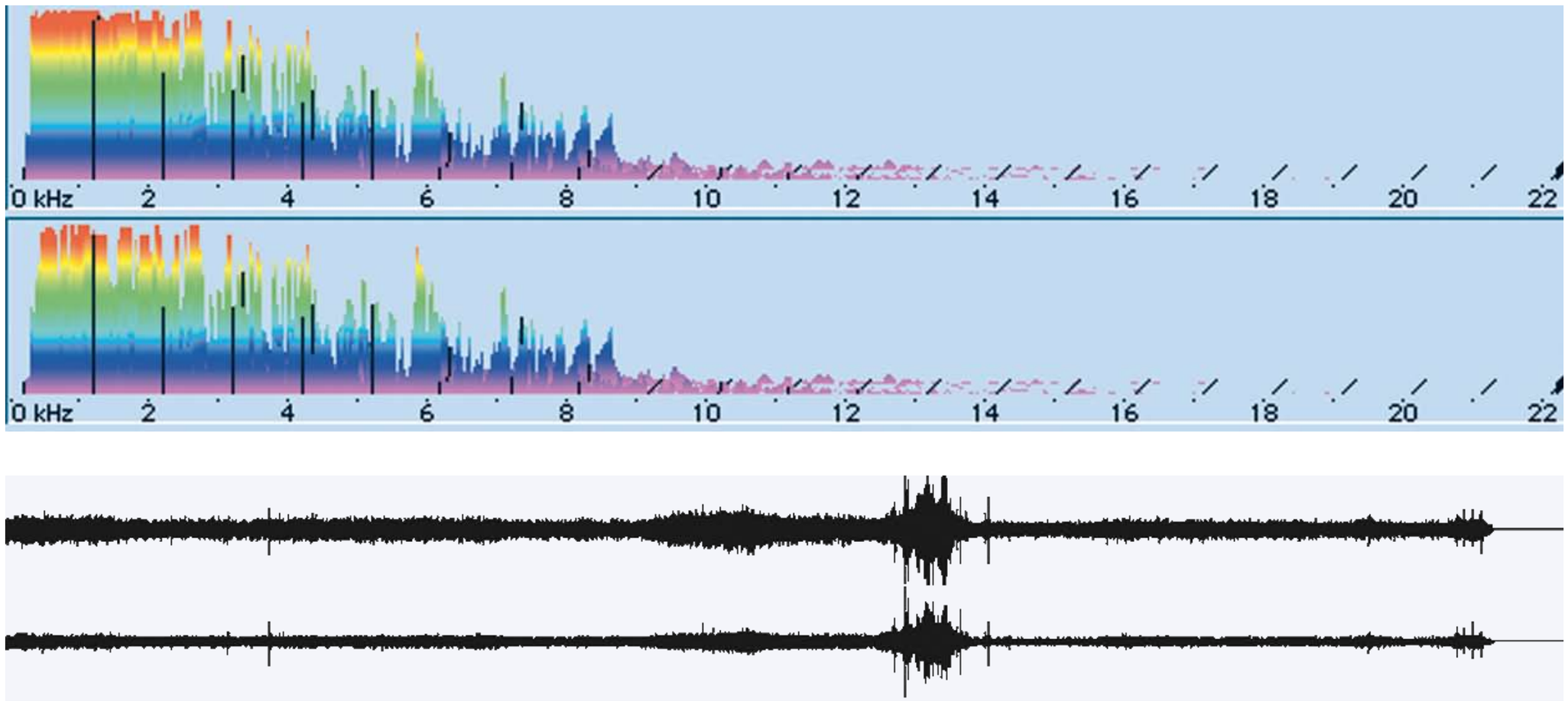
Sound editing programs use a timeline interface, with multiple tracks (usually displayed as waveforms), which are mixed down to produce a stereo or mono output.

Short loops may be used to create voices for samplers; longer loops may be combined (e.g. in GarageBand) to build songs from repeating sections.

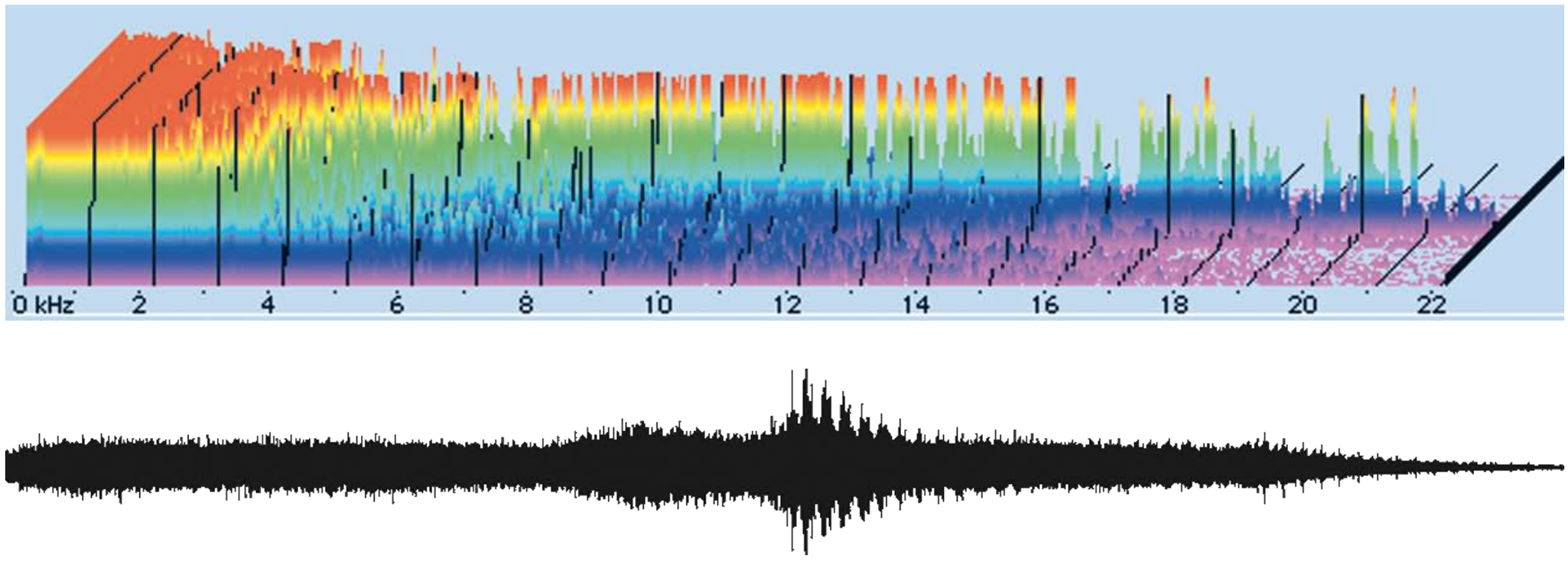
Filters and gates are used to correct defects (e.g. remove noise) or to enhance or modify sounds (e.g. reverb).



Low pass filtering



High pass filtering



Echo reverb

Time stretching (slowing down and speeding up) and pitch alteration are more easily applied to digital audio than they were to analogue audio.

They are used for synchronization and for matching (e.g. when combining separately recorded loops).

Sound can be combined with pictures in a video editing program: sound tracks are displayed on the same timeline as video tracks, where they can be synchronized.

Timecode is just a fiction when working with sound, owing to the high sampling rate, but it is valuable for synchronization.

If sound and video are physically independent in a movie, synchronization may be lost, especially when it is sent over a network.

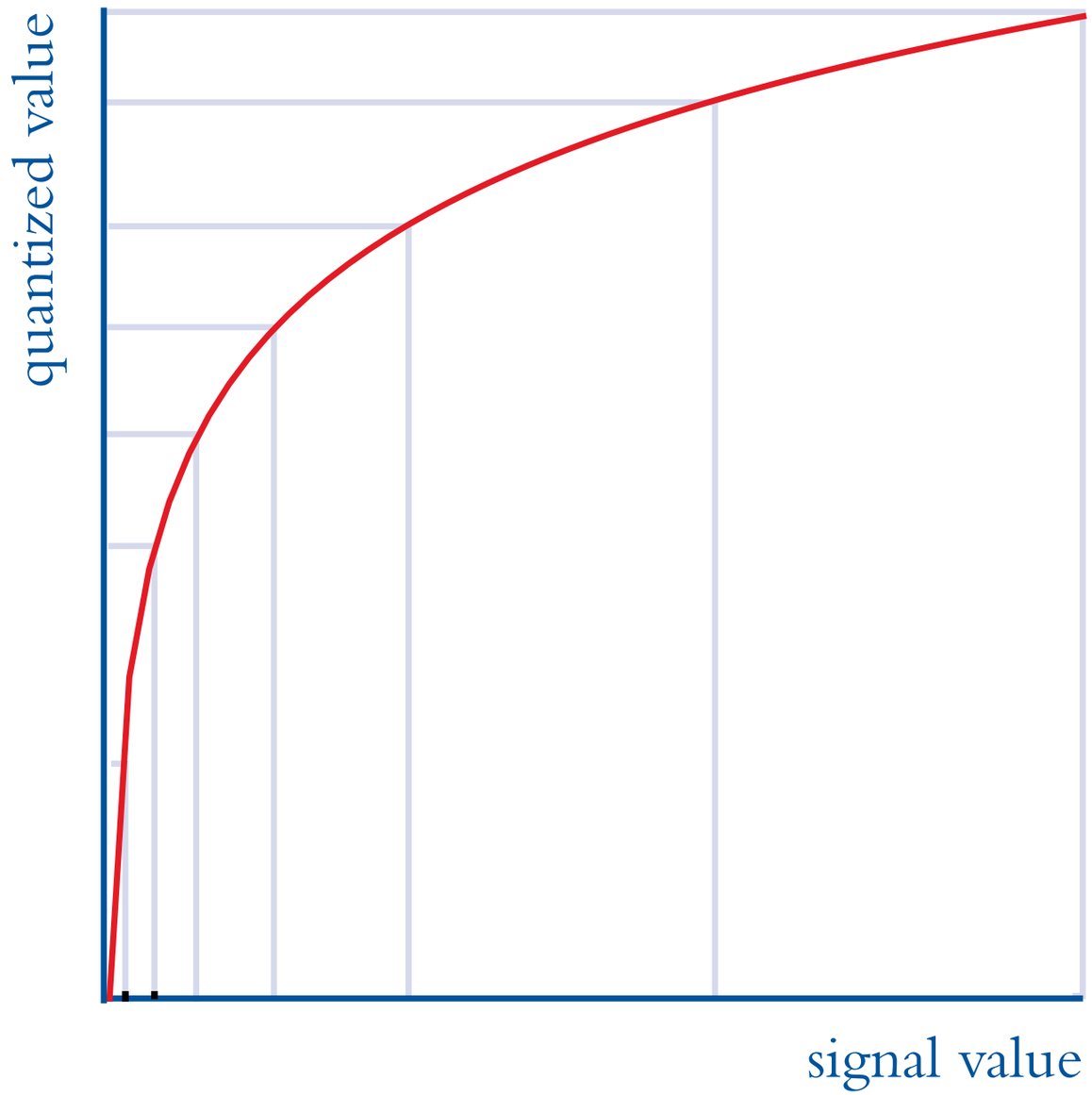
Compression

Sound is difficult to compress using lossless methods, except for special cases.

Some compression of audio can be obtained by run-length encoding samples that fall below a threshold that can be considered to represent silence.

Companding uses non-linear quantization to compress speech.

μ-law and A-law companding are used for telephony.

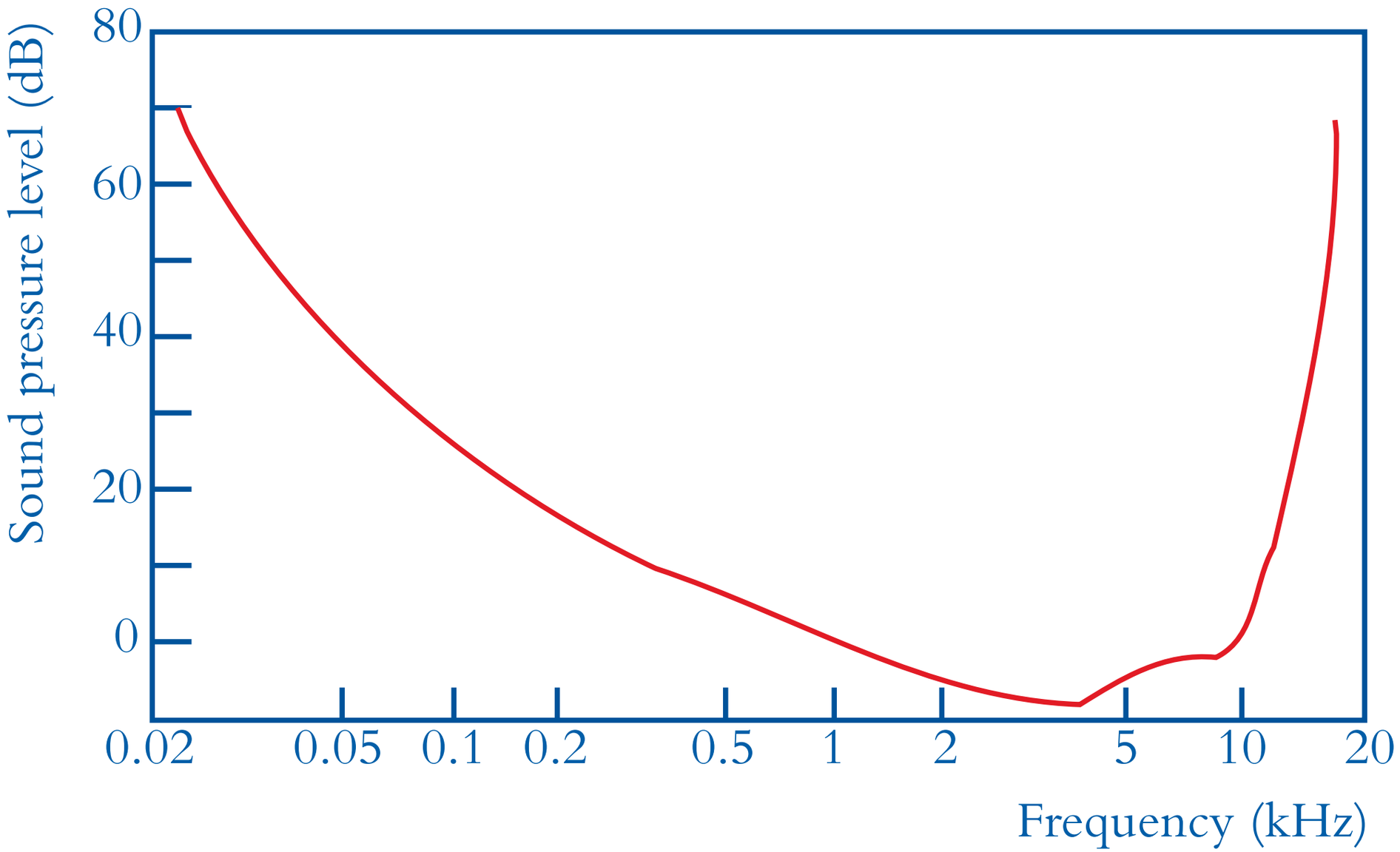


Non-linear quantization

Adaptive Differential Pulse Code Modulation (ADPCM), which works by storing information about the difference between a sample and a value predicted from the preceding sample, is also used in telephony.

Perceptually based compression discards inaudible sounds.

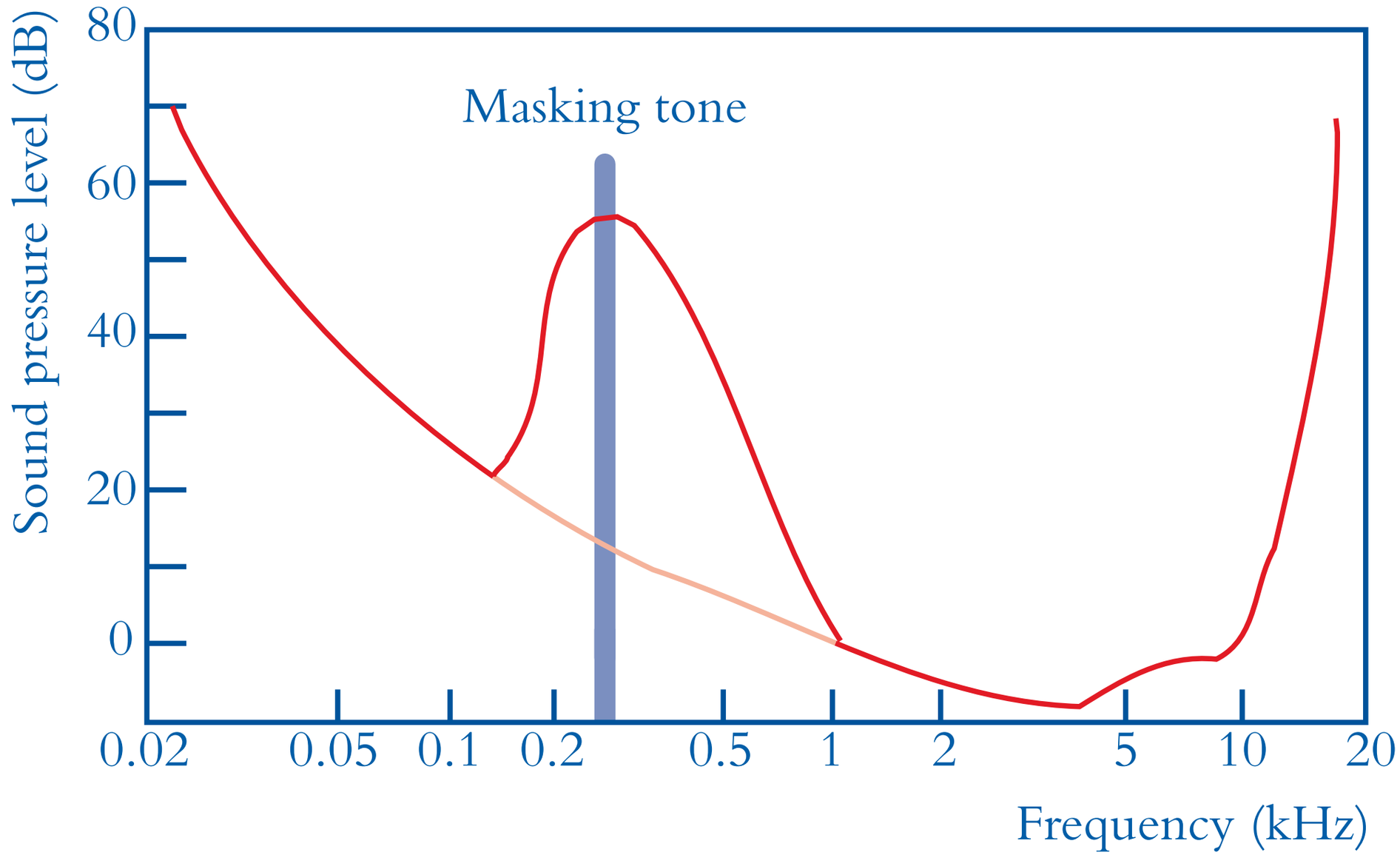
A psycho-acoustical model describes how the threshold of hearing varies non-linearly with frequency.



The threshold of hearing

Masking is a modification of the threshold of hearing curve in the region of a loud tone.

The threshold is raised in the neighbourhood of the masking tone.



Modification of the threshold of hearing by a masking tone

Filters are used to split a signal into 32 bands, and a masking level for each band is computed. Signals that fall below the level can be discarded.

Practical implementations of perceptually based compression are the basis of MP3 and AAC compression.

MIDI

MIDI provides a way of representing music as instructions describing how to produce notes, instead of as a record of the actual sounds.

MIDI provides a standard protocol and hardware interface for communicating between electronic instruments, such as synthesizers, samplers and drum machines, allowing instruments to be controlled by hardware or software sequencers.

A computer can control instruments through a MIDI interface, synthesize notes on a sound card or play back samples from disk in response to MIDI instructions.

MIDI messages are instructions that control some aspect of the performance of an instrument.

Each instruction has a status byte, indicating the type of message, and two data bytes, providing the values of its parameters. (e.g. Note On + note number + key velocity.)

Running status allows the status byte to be omitted if it is the same as in the preceding message.

General MIDI is a standard association between 128 Program Change values and voice names. (There is no guarantee that identical sounds will be produced for the same voice names on different instruments.)

QuickTime incorporates MIDI-like functionality. MIDI tracks can be combined with audio, video or any of the other media types supported by QuickTime.

MIDI software allows recording from a MIDI device, input as musical notation or on a "piano roll", and editing, often integrated with sound editing.