Università
Ca' Foscari
Venezia

Part of this material has been reproduced (with permission) from the lecture notes of the course "Multimedia systems" held by Prof. Augusto Celentano
http://www.dais.unive.it/~auce/smm2015-16/

Computer Science Applications to Cultural Heritage

Digital audio

Filippo Bergamasco (filippo.bergamasco@unive.it)
http://www.dais.unive.it/~bergamasco
DAIS, Ca' Foscari University of Venice
Academic year 2018/2019

Digital audio & CH

Dissemination and preservation of musical culture of the recent past is an important aspect of cultural heritage.

Long-term preservation of audio content demands a transfer to the digital domain.

- Acquisition of new content can be directly performed in digital with state-of-the-art equipment
- Vast selection of analogue sound recordings produced from 1898 to about 1990 (Vinyls, tapes, etc) that needs to be digitized/stored

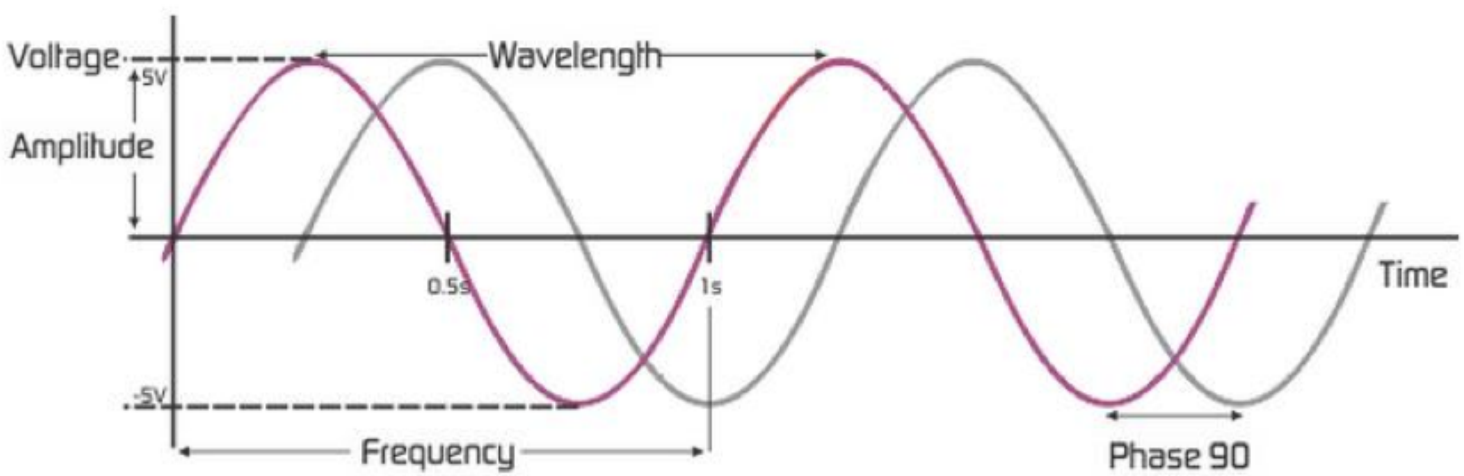
Università
Ca' Foscari
Venezia

What is sound?

Sound is composed by continuous waves of air pressure varying over time

Amplitude: Defines the perceived volume

Wavelength/frequency: Defines the sound pitch

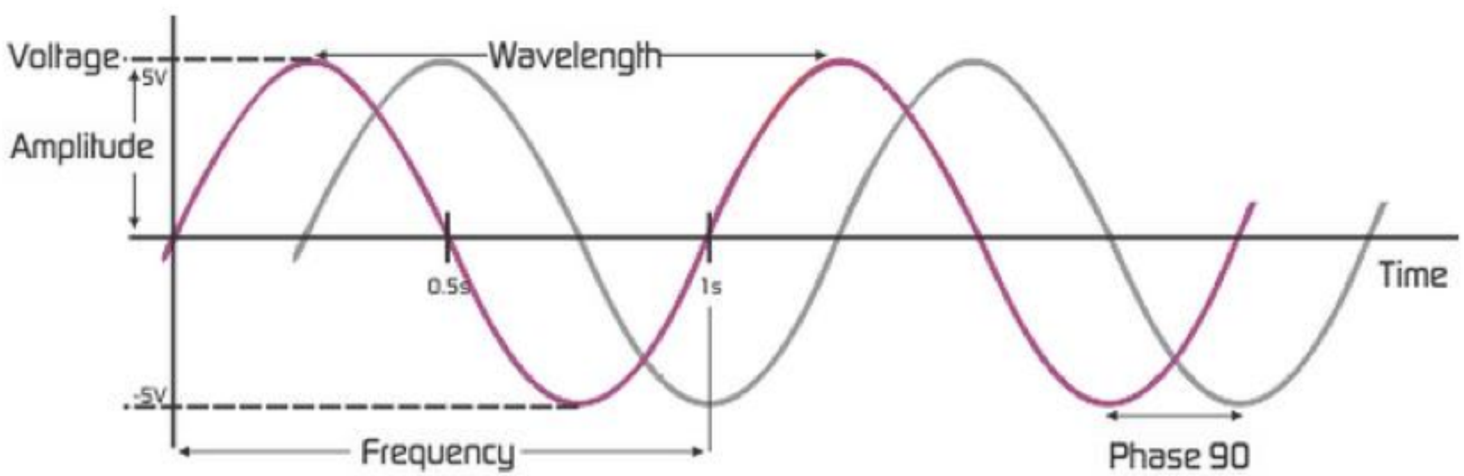


Università Ca' Foscari Venezia

What is sound?

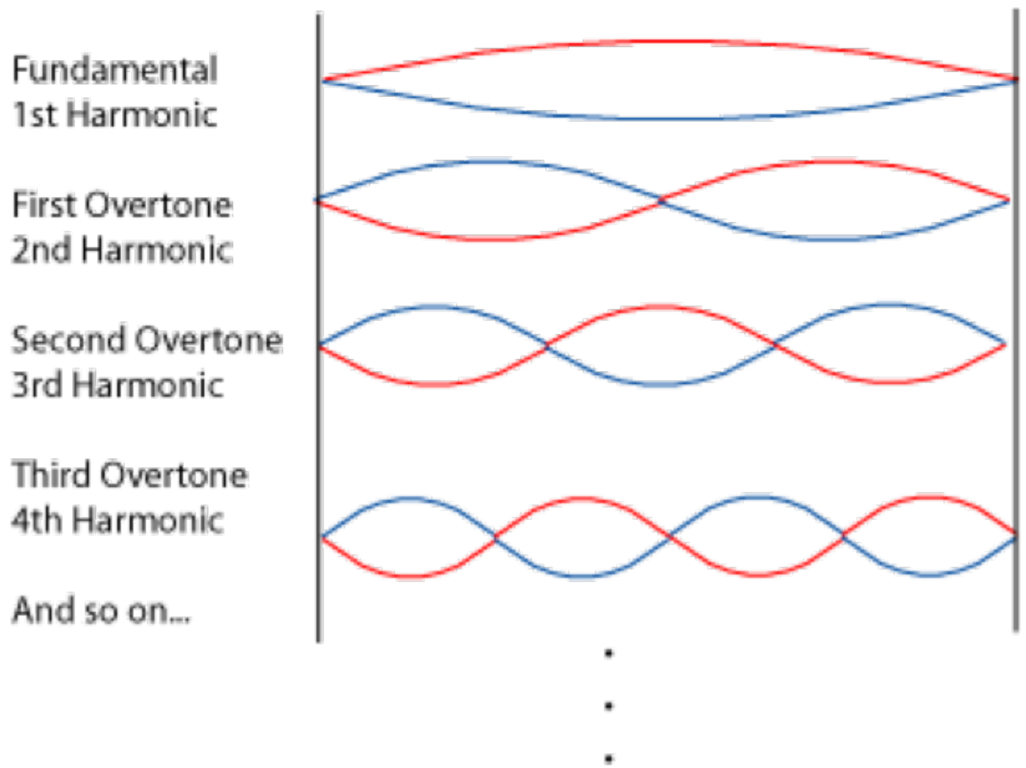
A sound is generally composed of a combination of elementary sounds (pure tones, sine waves) with variable relations of amplitude, frequency and phase, which together determine the acoustic properties

The distribution of composing frequencies (visible in the sound spectrum), defines the sound timbre



Musical sounds

Musical sounds are based on the concept of **fundamental** and **harmonic**: harmonics are frequency multiples of a fundamental frequency



Musical sounds

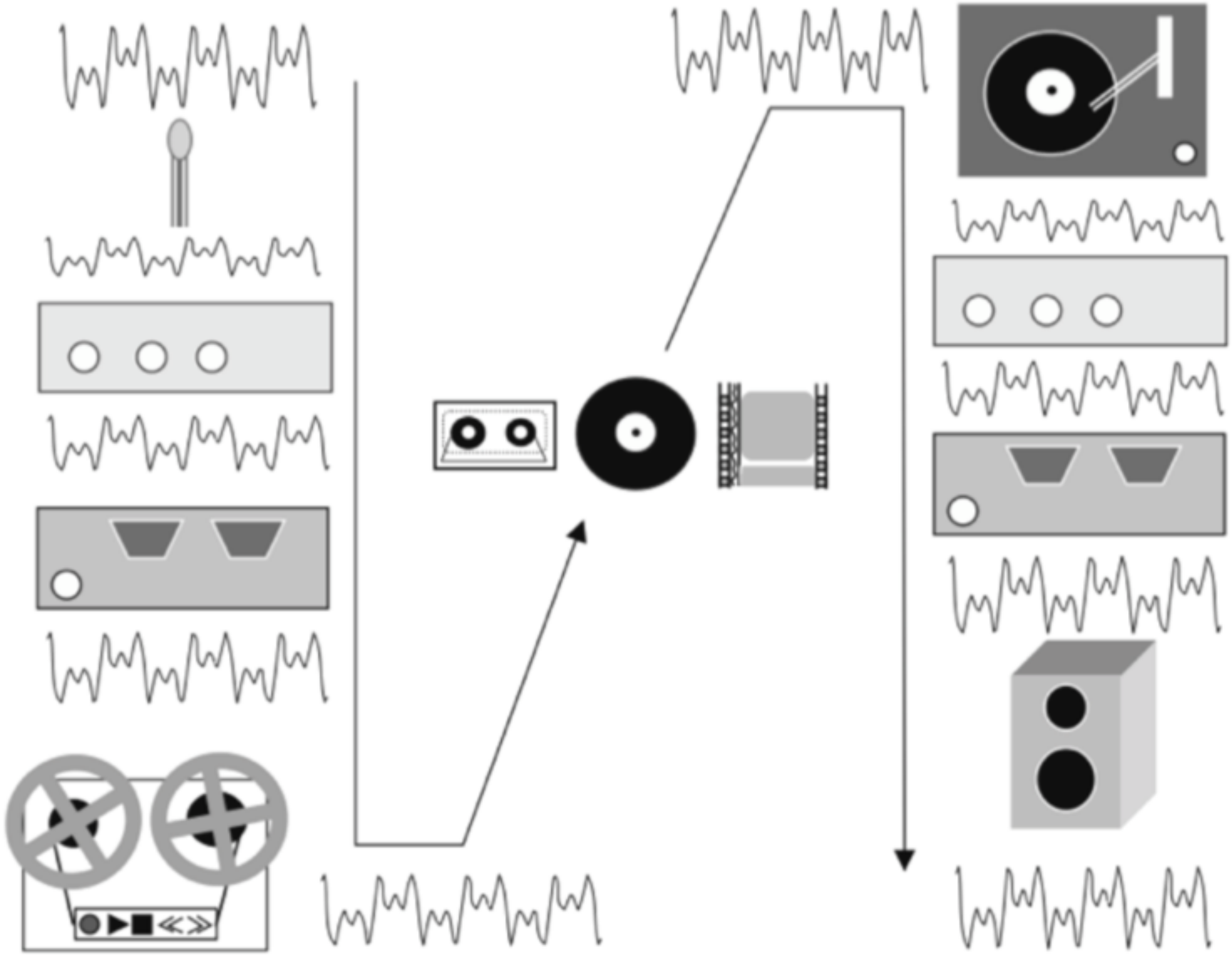
All the harmonics with a ratio of a power of two with the fundamental frequency are called **octaves**.

The most important musical scales are typically written using 8 notes, and the interval between the first and last notes is an **octave** (ie. the frequency of the last note is the double of the frequency of the first)

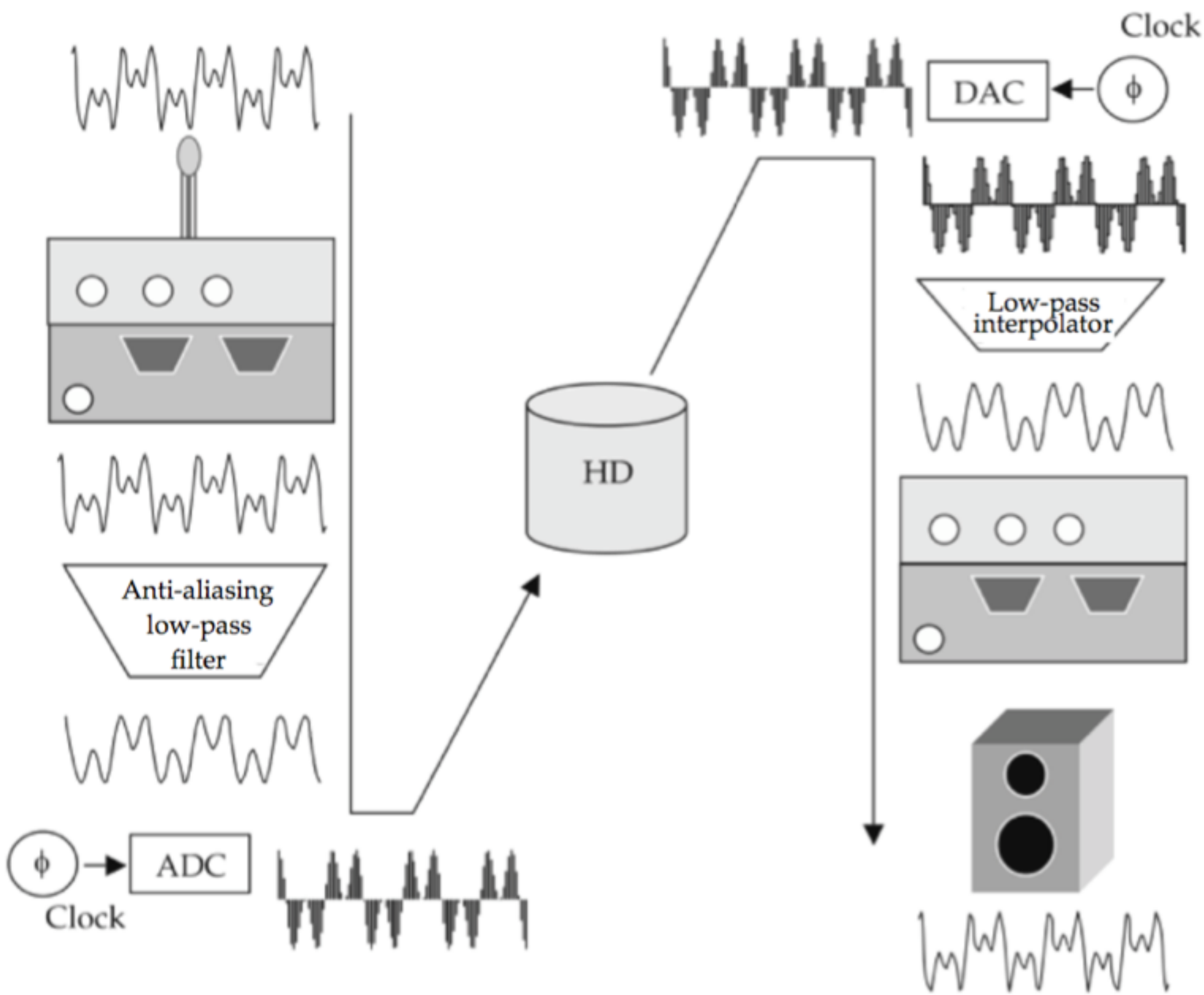
Non harmonic frequencies compose into chords and complex timbres, and can produce dissonances

Università
Ca' Foscari
Venezia

Analog audio processing



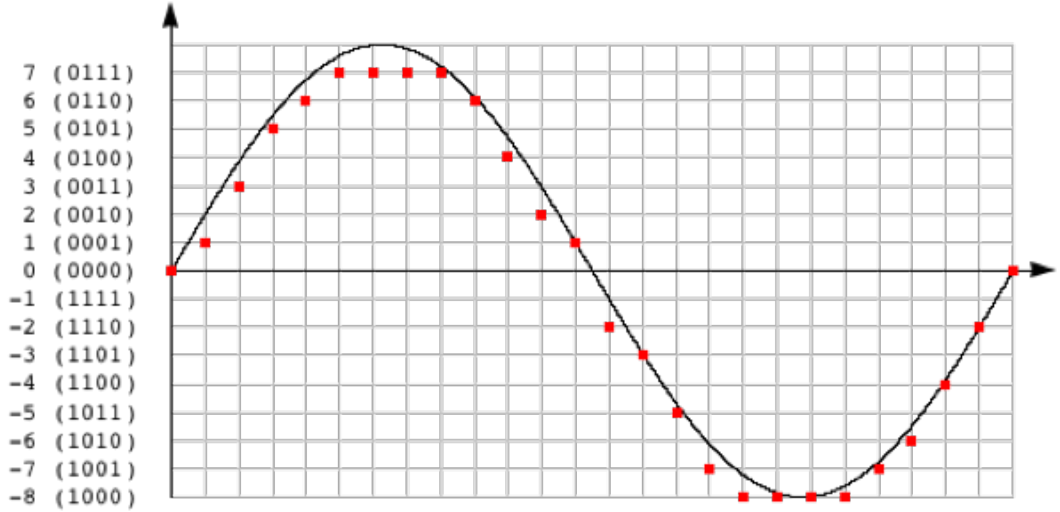
Digital audio processing



Digital audio

To be managed in a computer, analog sound must first be digitized to a sequence of discrete values

Similar to images, we can choose different sampling frequencies (temporal sampling resolution of audio data) and quantization (how many bits we use to store each numerical sample).

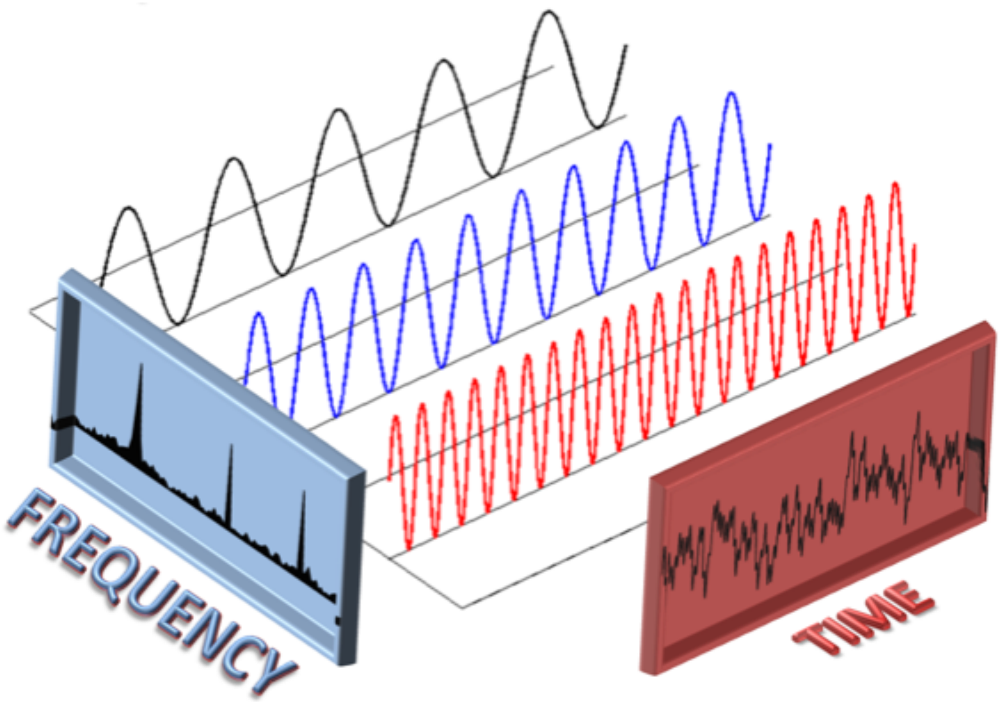


Ex: CD Audio sampling rate is 44.1khz at 16bit

Università Ca' Foscari Venezia

Fourier signal theory

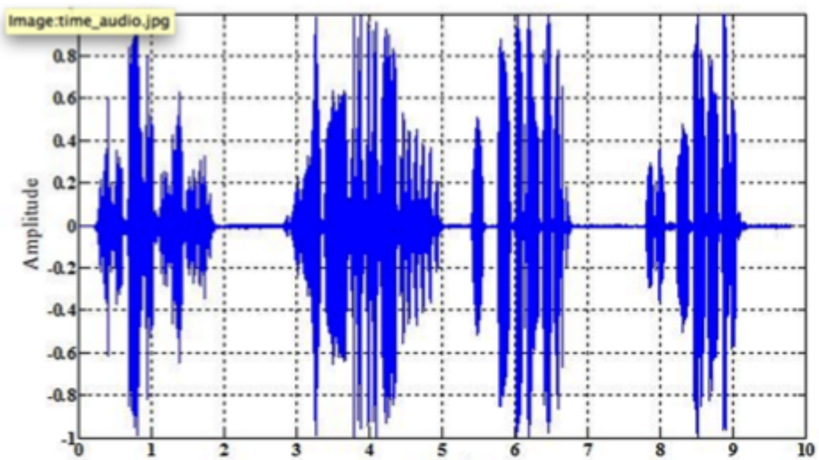
Any periodic signal can be obtained by composition of sine waves with different amplitude and phase.

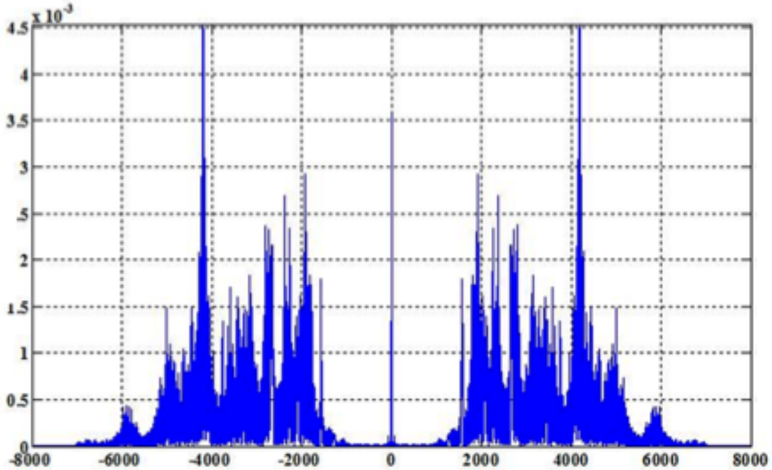


We can represent and analyze the same signal in time or frequency domain, depending by the application

Fourier signal theory

The Fourier transform is a powerful tool to transform a signal between the two domains





Time domain

Frequency domain

How sound amplitude changes as a function of time

how the amplitude of sound spectral components change as a function of frequency in a specified time interval

Sound features in time domain

Volume (loudness):

Root mean square value of the sound signal over a specified time interval

Zero crossing rate:

How frequent the signal change its sign (positive/negative or vice-versa)

Silence ratio:

Portion of a sound fragment whose amplitude is below a defined threshold threshold

Sound features in frequency domain

Spectrum:

Distribution of frequency contributions (computed on frames of short duration to capture its evolution over time)

Pitch:

Greatest common divisor of frequency peaks (also called fundamental harmonic)

Brightness:

Centroid (mean) of the spectrum distribution

Sound features in frequency domain

Bandwidth:

Difference between the highest and lowest frequency in a given signal (definition may vary)

Harmonicity:

Deviation of the sound frequency spectrum w.r.t. a pure harmonic sound (sound composed only of frequencies multiple of the fundamental harmonic)

Speech vs. Music

Feature	Voice	Music
Bandwidth	0-7 KHz	0-20 KHz
Brightness (spectral centroid)	Low	High
Silence ratio	High	Low
Zero crossing	Variable	Less variable

Audio coding

Digital coding and decoding of audio signal has more problems than images:

- audio has a temporal structure that cannot be modified or downgraded (e.g., frequency)
- audio information is variable in time
- the required playback quality is much better than understandability threshold
- data size is proportional to quality and is usually very large for music applications (hence the need for compression)

Quality/Size

Quality	Sample Rate (Khz)	Bits per Sample	Mono / Stereo	Data Rate (uncompressed) (kB / sec)	Frequency Band (KHz)
Telephone	8	8	Mono	8	0.200-3.4
AM Radio	11.025	8	Mono	11.0	0.1-5.5
FM Radio	22.05	16	Stereo	88.2	0.02-11
CD	44.1	16	Stereo	176.4	0.005-20
DAT	48	16	Stereo	192.0	0.005-20
DVD Audio	192 (max)	24(max)	6 channels	1,200 (max)	0-96 (max)

Common audio formats

Uncompressed:

- Waveform Audio File (wav). Used on PCs
- Audio Interchange File Format (aiff). Used on Macs

Lossless:

- FLAC (lossles free format and codec mainly used for archive purposes)
- Apple Lossless Encoding (designed by Apple for iTunes-based applications)
- Windows Media Audio (can be lossless or lossy. Designed by Microsoft)

Common audio formats

Lossy:

Mpeg-1 audio

- compressed format for variable quality encoding at constant bitrate
- multiple stage compression algorithm based on psychoacoustics
- Different bit-rate targets called layers. Usually used in layer 3, commonly referred as **mp3**

Mpeg-4 audio

- standard with MPEG-4 H.264 video coding for network delivered digital audio/video
- introduced in consumer market by Apple iTunes product line (commonly referred as **aac**)
- Better quality than mp3

Common audio formats

Lossy:

OGG/Vorbis

- lossy free format and codec designed as replacement for MP3/AAC
- Claims superior fidelity wrt MP3 but requires more resources in decompression and introduces some audible artifacts

How does the lossless and lossy audio compression works?

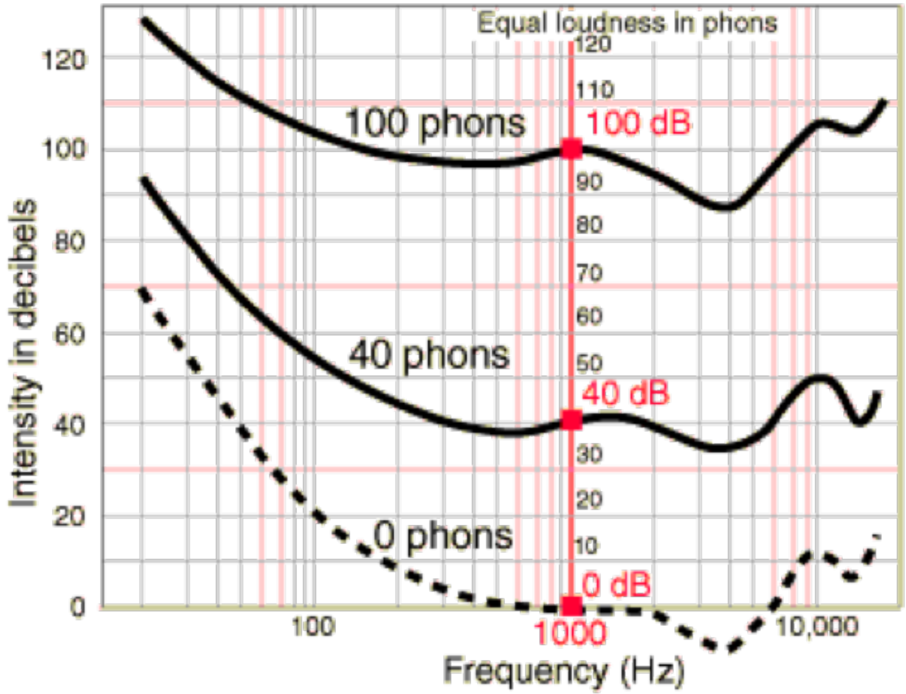
Lossless compression strategies

- Detect silence as consecutive samples below a defined threshold. Same idea than the run-length encoding
- Adaptive Differential Pulse Code Modulation:
  - Store the difference between consecutive samples instead of the signal itself
  - Particularly efficient at low frequencies
- Linear predictive coding
  - Fit a defined audio model to the signal
  - Transmit the model parameters together with the difference between the actual signal and the signal predicted by the model

Lossy audio compression

Compression is based on psychoacoustics

Human ear sensitivity is variable through the audio spectrum



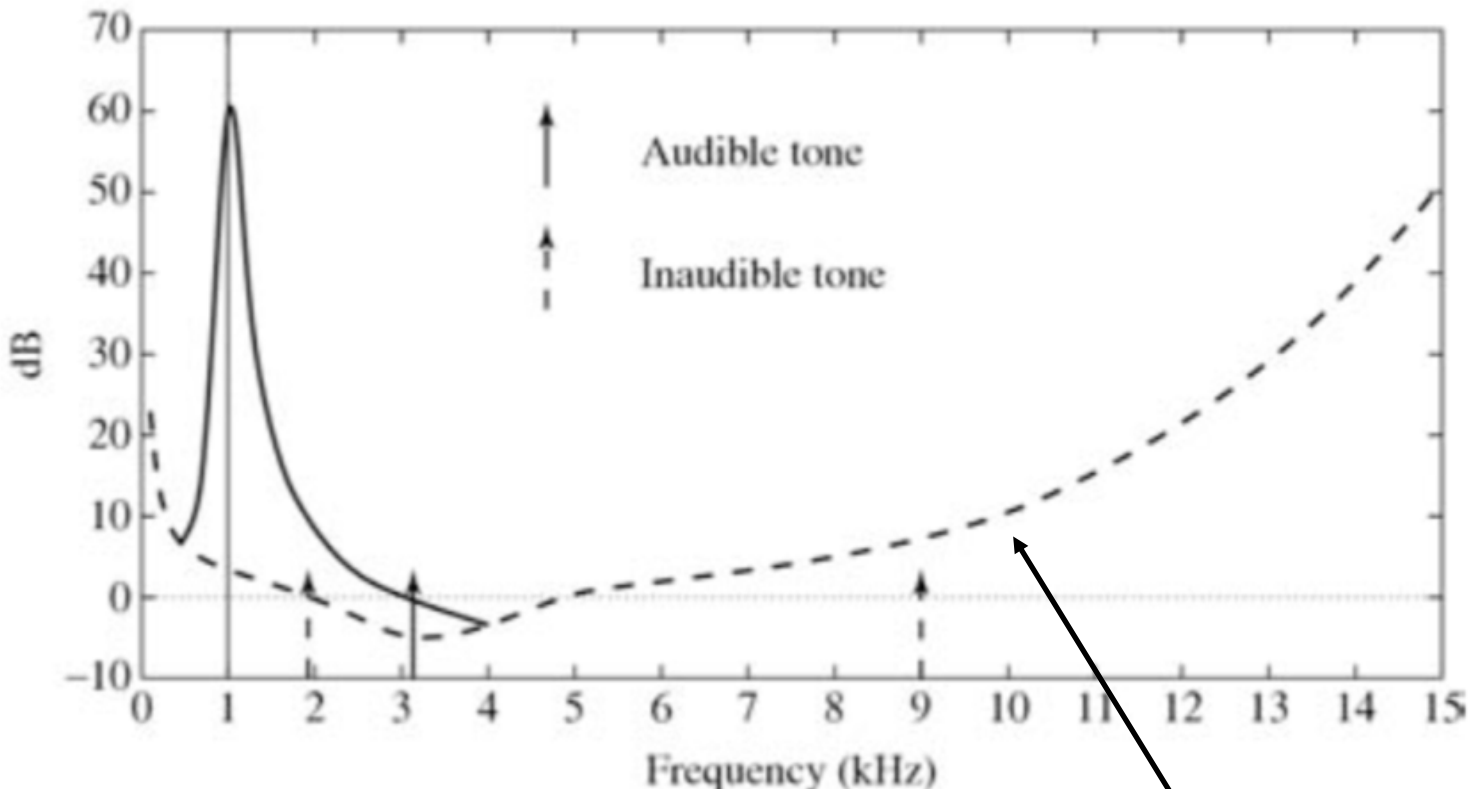
Example equal - loudness curves for the human ear.

Based on:

- Frequency masking
- Temporal masking

Frequency masking

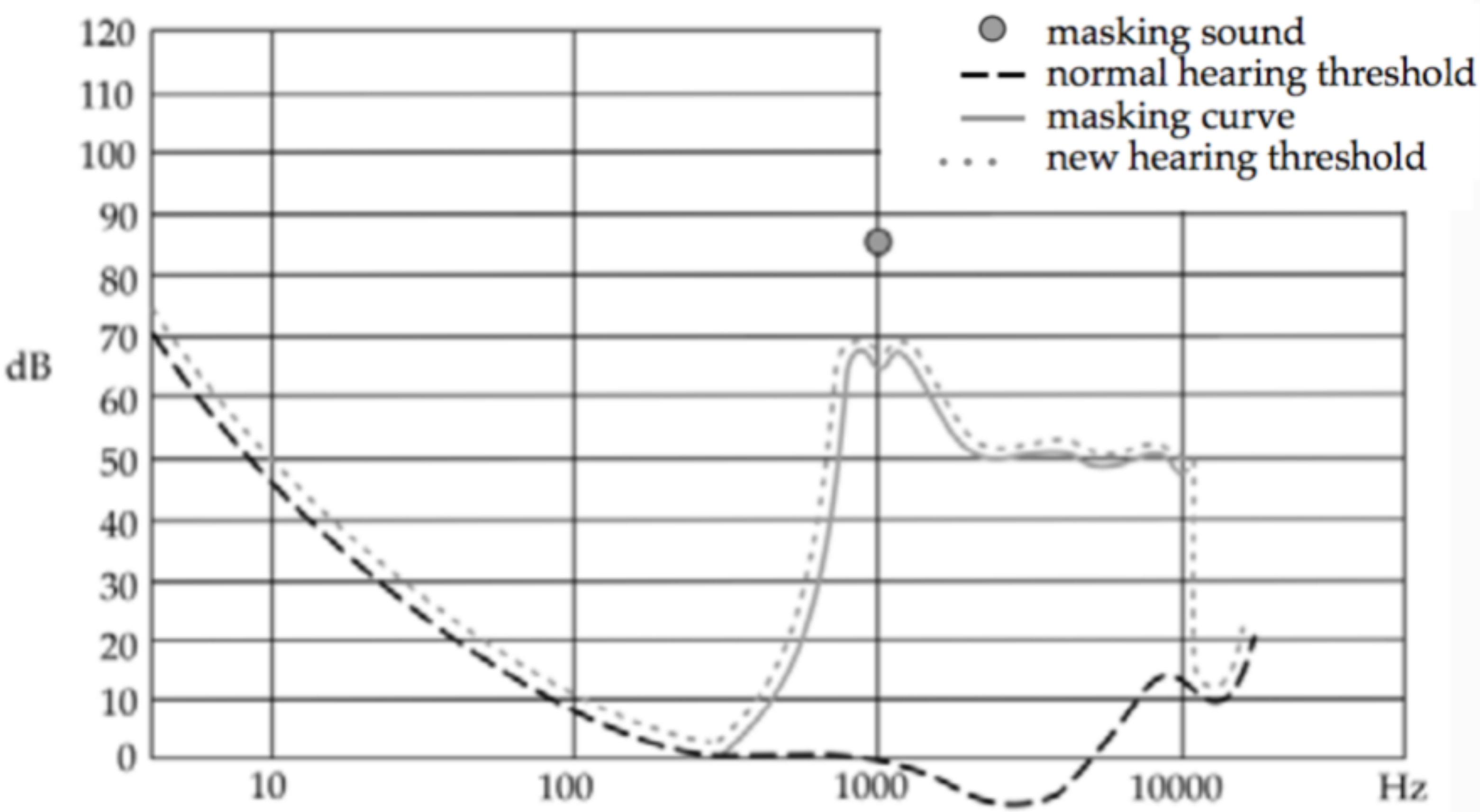
A sound can mask another sound with a similar frequency and lower level



Masking produced by a 1kHz tone

Frequency masking

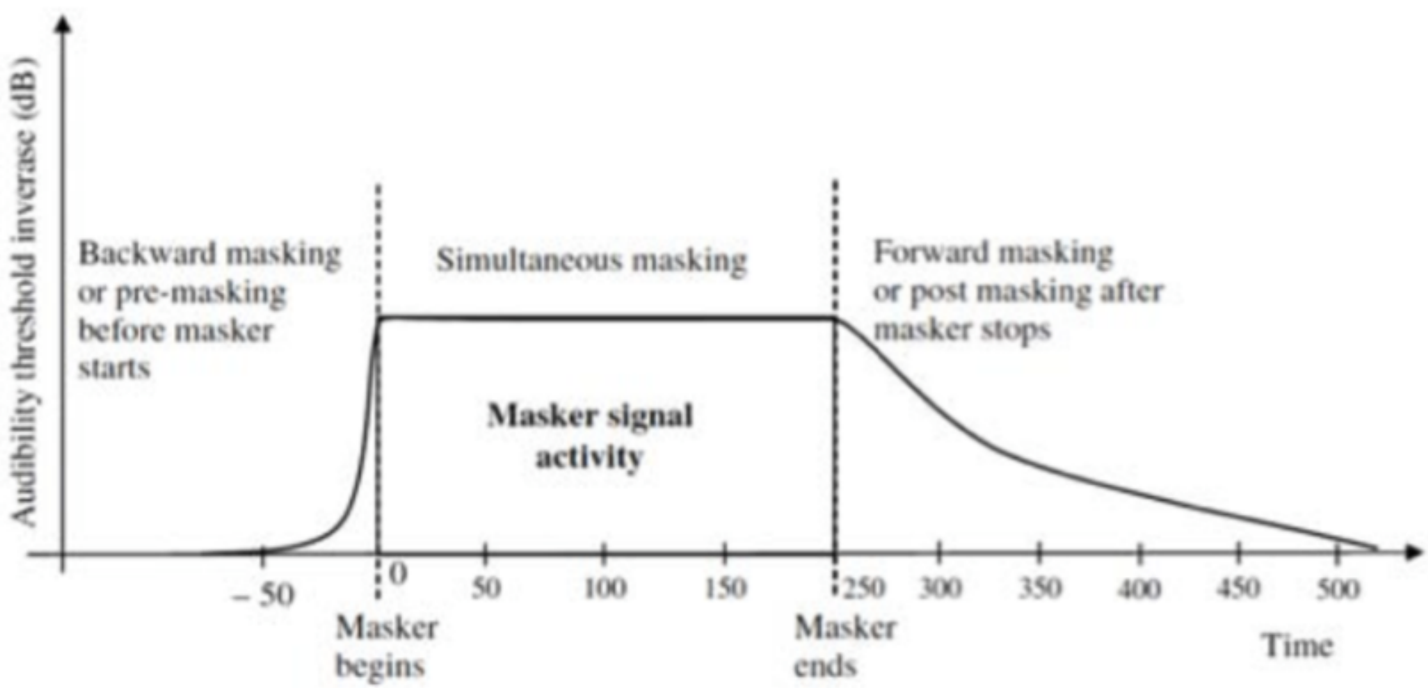
Masking function depends by the frequency and the intensity of the sound. If the sound has a single frequency is called tonal masking



Temporal masking

A sound can mask another sound that happens shortly after.

Post-masking hides next low sounds in 50 to 200 ms time range



Università Ca' Foscari Venezia

Mp3

MPEG-1 layer 3 (MP3) is the standard for consumer audio applications.

Compression is made in four steps using psychoacoustic model properties

- split the audio signal into 32 frequency sub-bands
- determine amount of masking for each band (frequency+temporal masking combined)
- a band whose power is below the masking threshold is not encoded. Otherwise, determine the number of bits needed to the signal such that quantization noise is below the masking effect (1 bit ~ 6 dB of noise)
- format the bitstream

Mp3

- common bit rates from 32 kbits/sec to 384 kbits/sec (uncompressed CD audio is >1.4 Mbits/sec)

- compression rate 6:1 (256 kbits/sec) is scarcely distinguishable from original for a non professional user

- 96 to 192kbit/sec optimal for consumer applications

- several sampling frequencies (32, 44.1 and 48 kHz)

- monophonic, dual, stereo, joint stereo channels

AAC

Improves over MP3 quality with a new coding algorithm

- extended range of sampling frequencies: 8-96 KHz
- up to 48 independent channels
- arbitrary bit-rate
- filtering algorithms simpler and more efficient, based on DCT
- higher coding efficiency for stationary and transient signals
- more flexible joint stereo processing (variable with frequency)
- defines several profiles according to different compression parameters and re-filtering algorithms