Jogos e Simulação – Audio

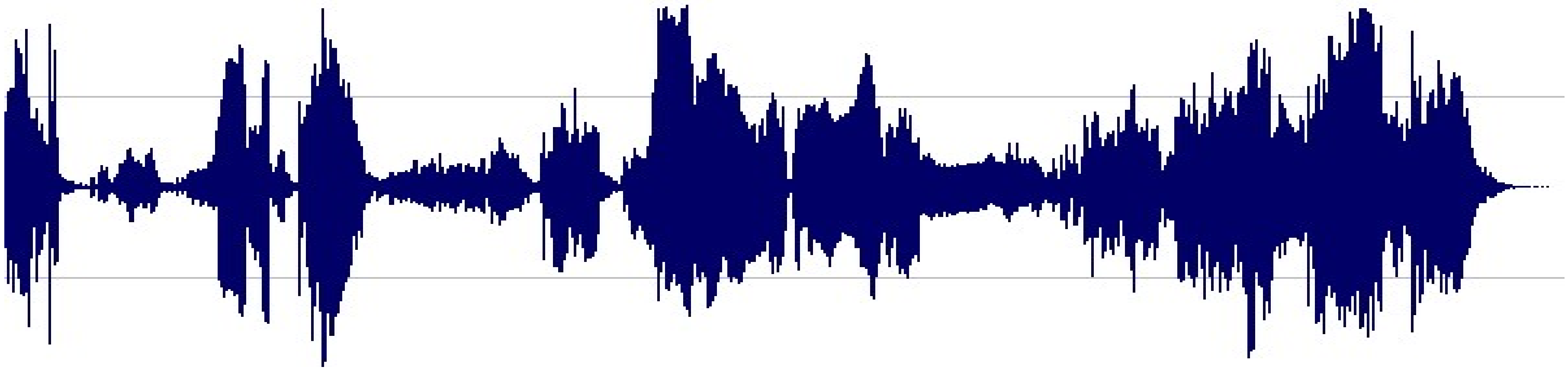
Sofia Cavaco (scavaco@fct.unl.pt)

Material

- Bibliography
  - Fundamentals of hearing, W.A. Yost, 4th edition, Academic Press, 2000
  - 3-D Sound for Virtual Reality and Multimedia, D.R. Begault: http://hdl.handle.net/2060/20010044352

- Audacity – audio editor and recording software (in http://audacity.sourceforge.net/)

Today – Introduction and Audio Analysis



Audio Analysis

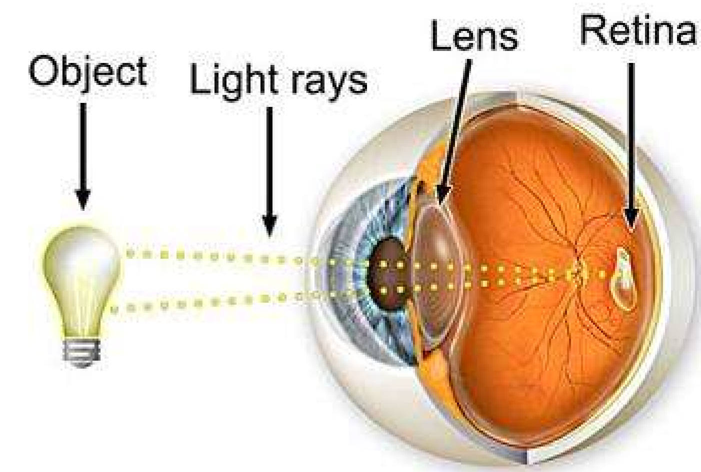
- What is sound
- Sound and hearing
- Fourier analysis and basic measurements
- Audio perception

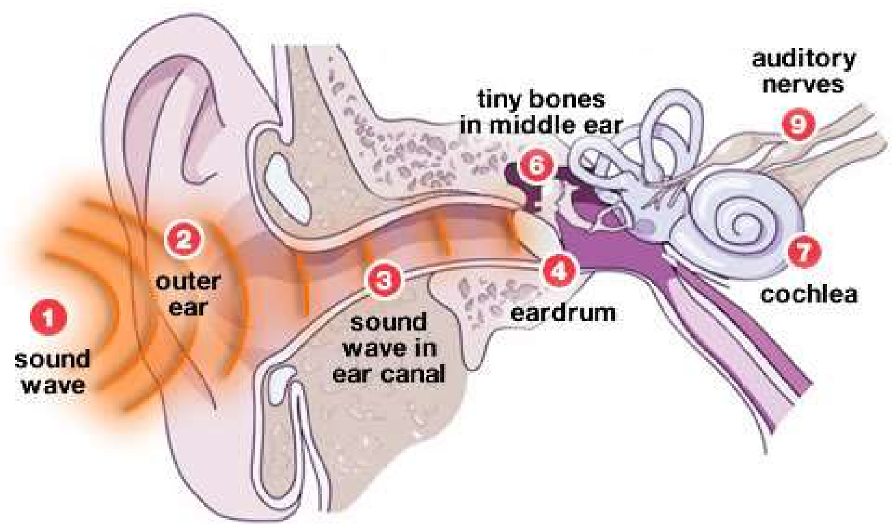
Audio in Games

- Immersion without audio? No...

- Immersion:
  - https://www.youtube.com/watch?v=8zAuSW0Lv10
- Realism
  - https://www.youtube.com/watch?v=7xzKyIq9h3s

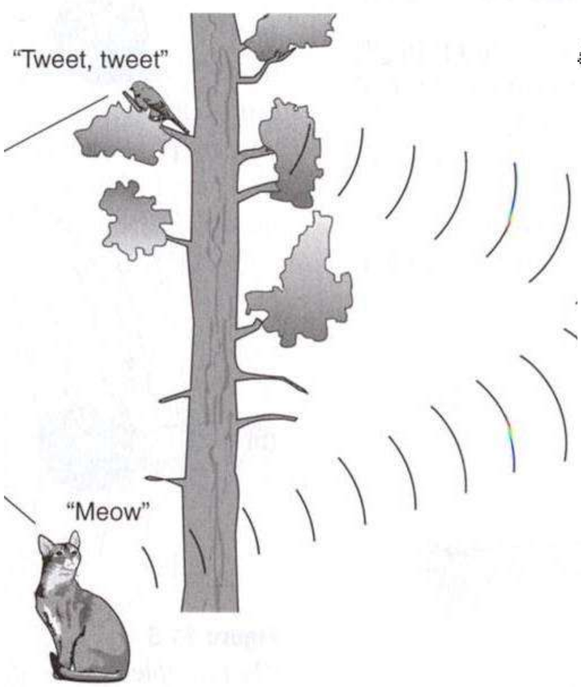
Audio and hearing

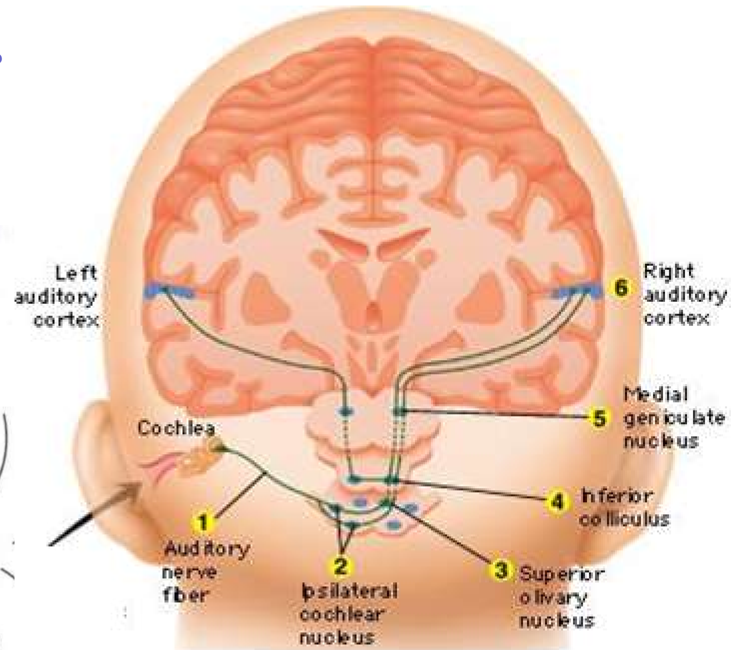




Auditory perception

- Sound – result of the movement of the air particles.

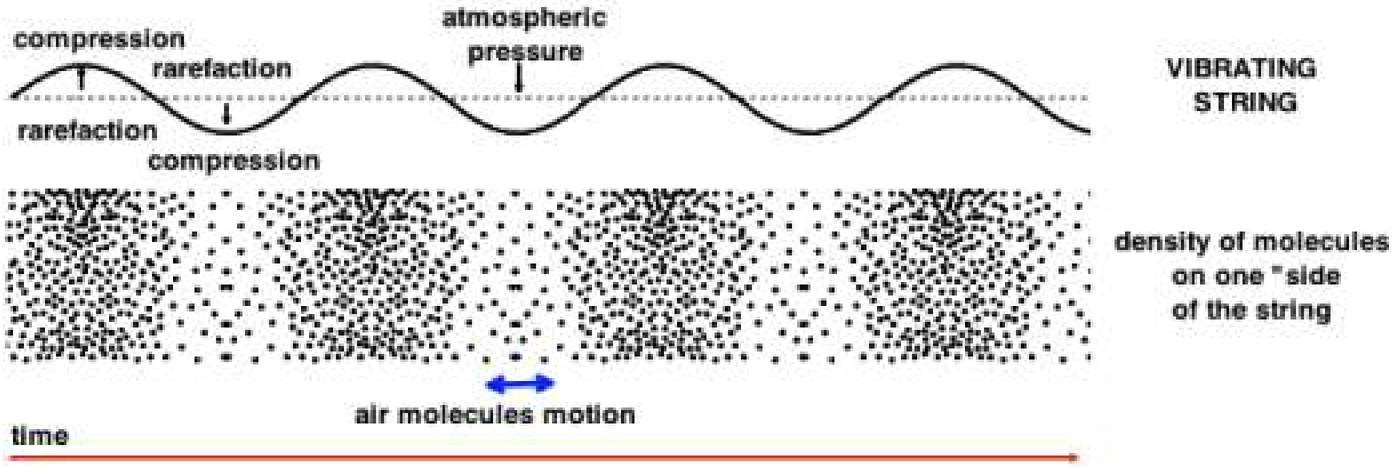




Sound wave

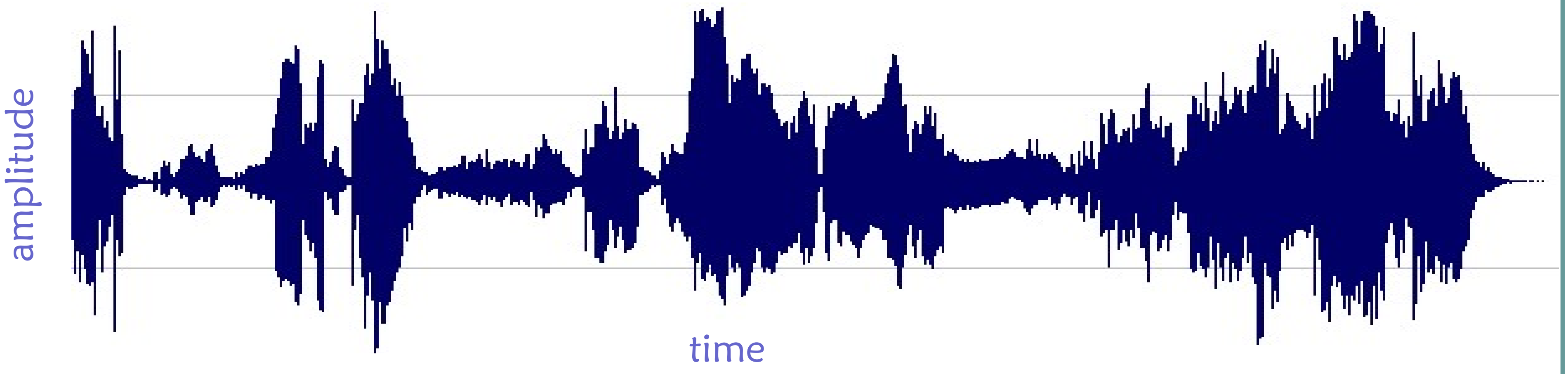
- Sound – result of the movement of the air particles.

Waveform


compression
rarefaction
atmospheric
pressure
rarefaction
compression
VIBRATING
STRING
density of molecules
on one "side
of the string
air molecules motion
time

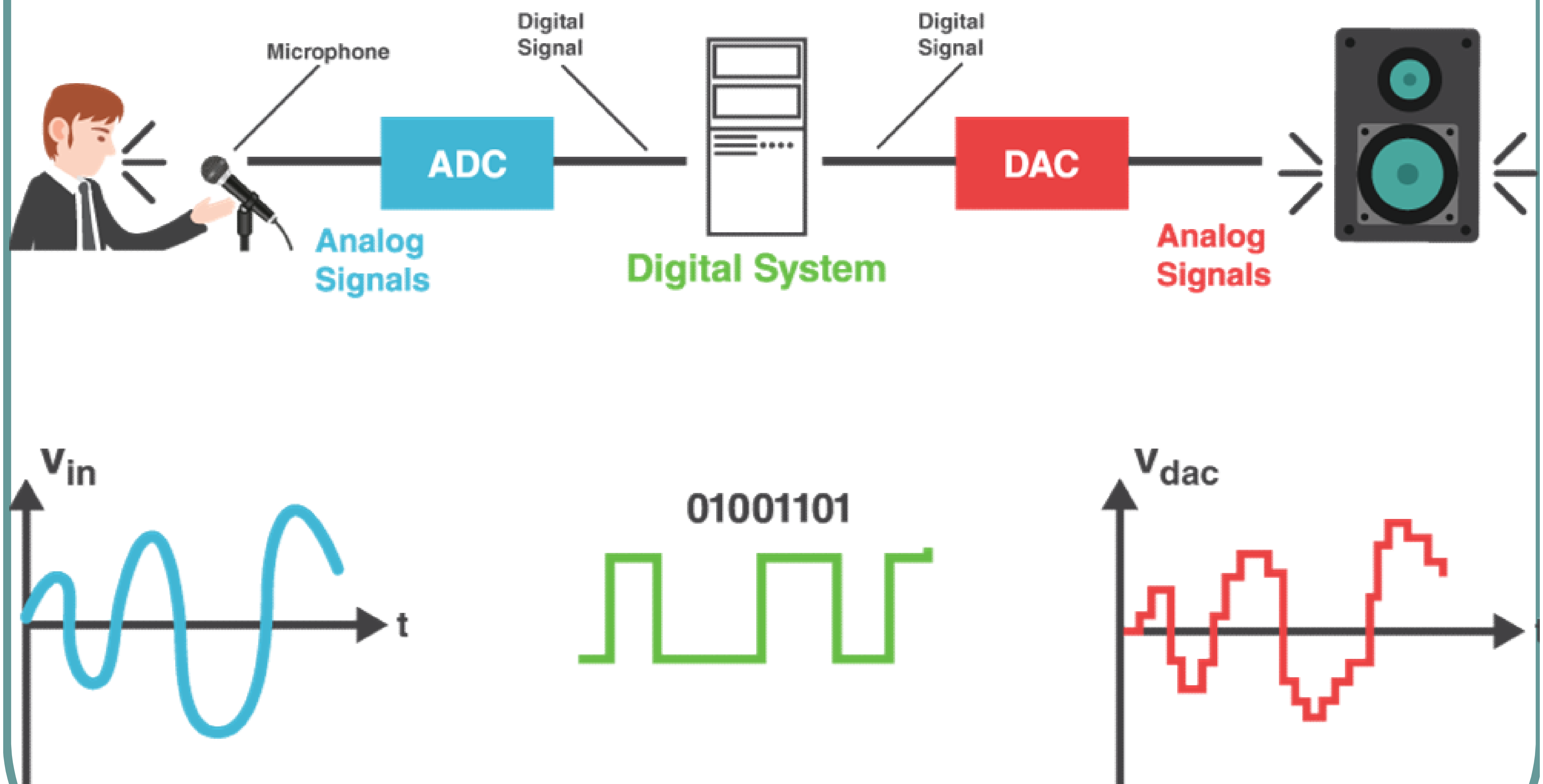
Waveform

- Waveform – shape of the sound wave against time.



- Amplitude (A) – measure of displacement. (Units: deciBel, dB).

Audio Digital Representation

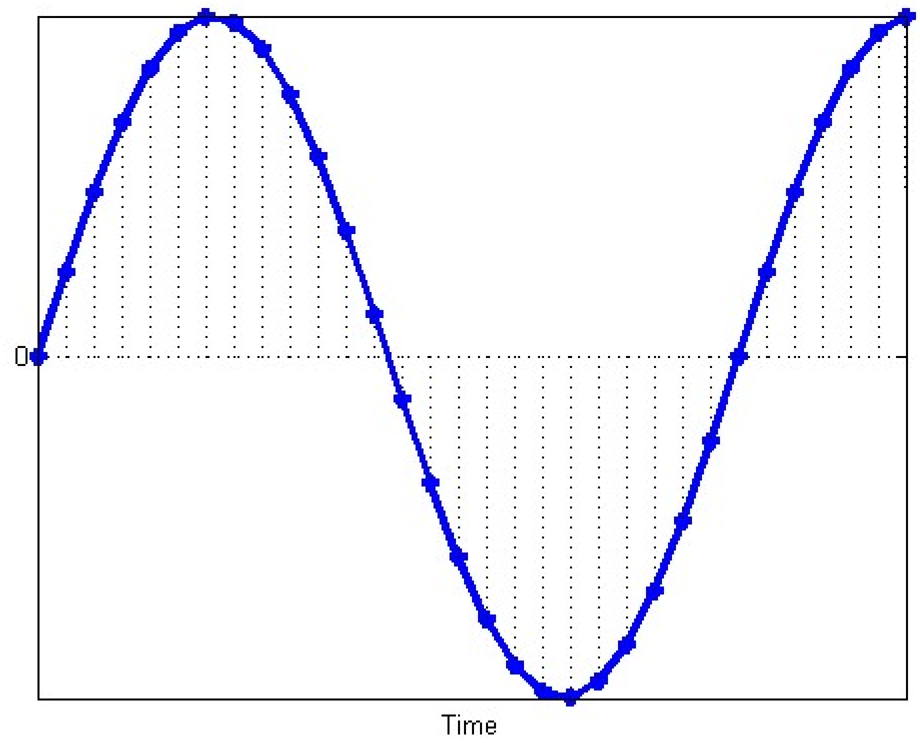


Audio Digital Representation

- Microphone – transforms air pressure variations into electric voltages.
- Analog-to-digital converter (A/D or ADC) – It converts voltages into a series of digital values, that is, it converts a continuous signal into a discrete signal (the digital representation of the sound).
- Digital-to-analog converter (D/A or DAC) – It converts digital values into continuously varying voltage, that is, it converts a discrete signal into a continuous signal.
- Speakers – transforms electric voltages into air pressure variations.

Audio Digital Representation

- (Periodic) sampling – representing a continuous signal with a series of discrete values.
- Sampling frequency (or sampling rate, f_s) – the number of samples taken per second.
  (Common: 44100 Hz,
  22050Hz)
- Sampling time – the regular intervals at which samples are taken.



Audio Digital Representation

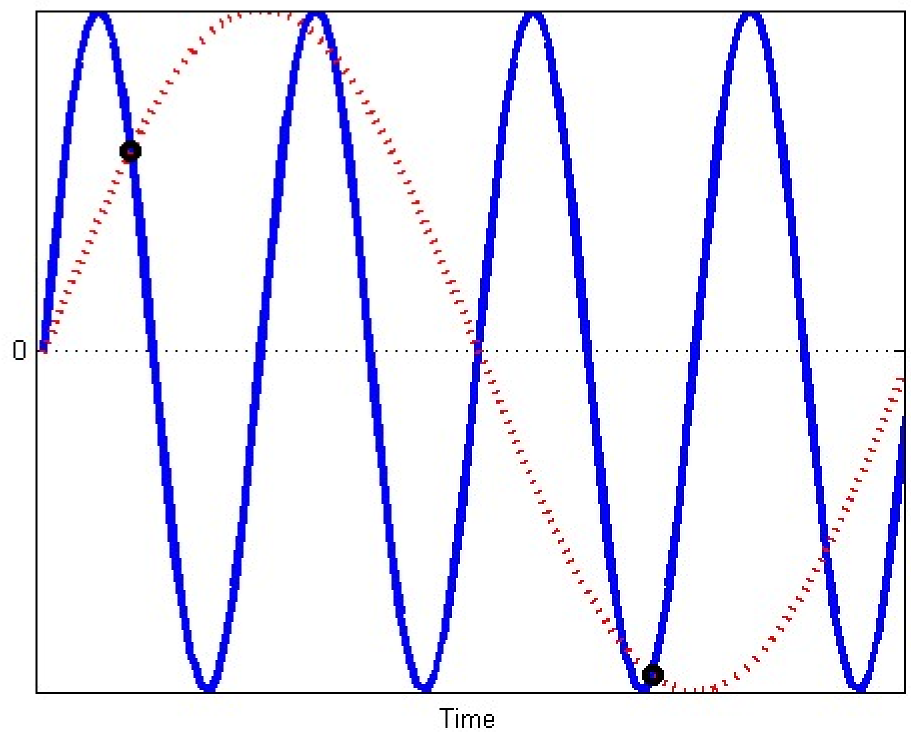
Common DA Sampling Rate

8,000 Hz	telephone, adequate for human speech
22,050 Hz	radio
32,000 Hz	miniDV digital camcorder
44,100 Hz	CD, commonly used with MPEG-1 audio (VCD, SVCD, MP3)
48,000 Hz	digital sound used for digital TV, DVD, DAT, films and professional audio
96,000 or 192,400 Hz	DVD-audio and HD-DVD audio tracks

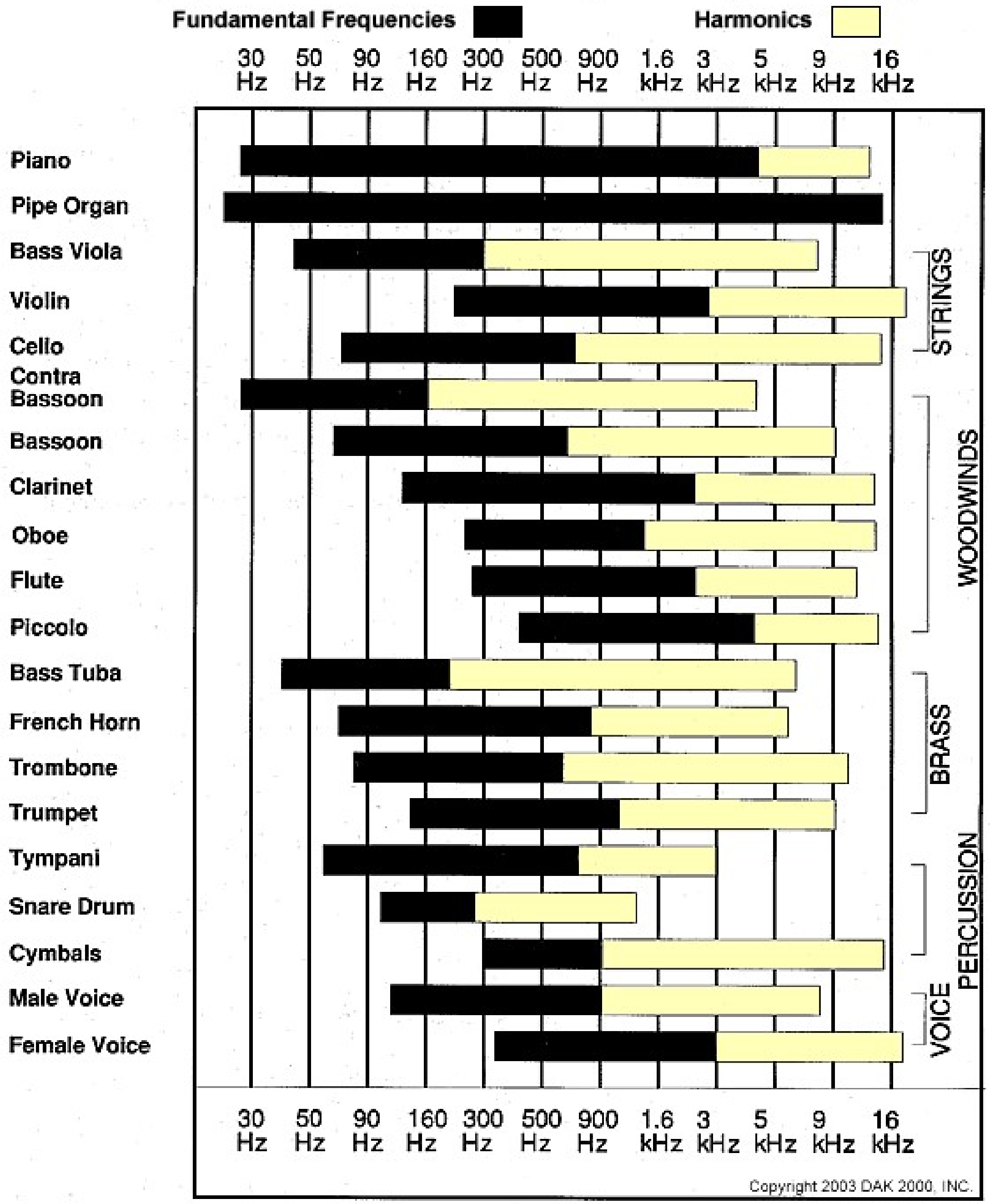
17

Audio Digital Representation

- Aliasing – signal ambiguity
- Sampling theorem (Harold Nyquist, 1928) – In order to guarantee correct reconstruction of the signal: $f_s \geq 2 \times f$
- Nyquist frequency – $f_s / 2$



Approximate Frequency Ranges



Basic measurements

Every complex wave can be decomposed into a combination of sinusoidal waves. (Fourier, 1822).

- Period (*P*) – the amount of time a sinusoid takes to complete one cycle. (Units: seconds).
- Frequency (*f*) – the number of cycles per second. (Units: Hertz, Hz).

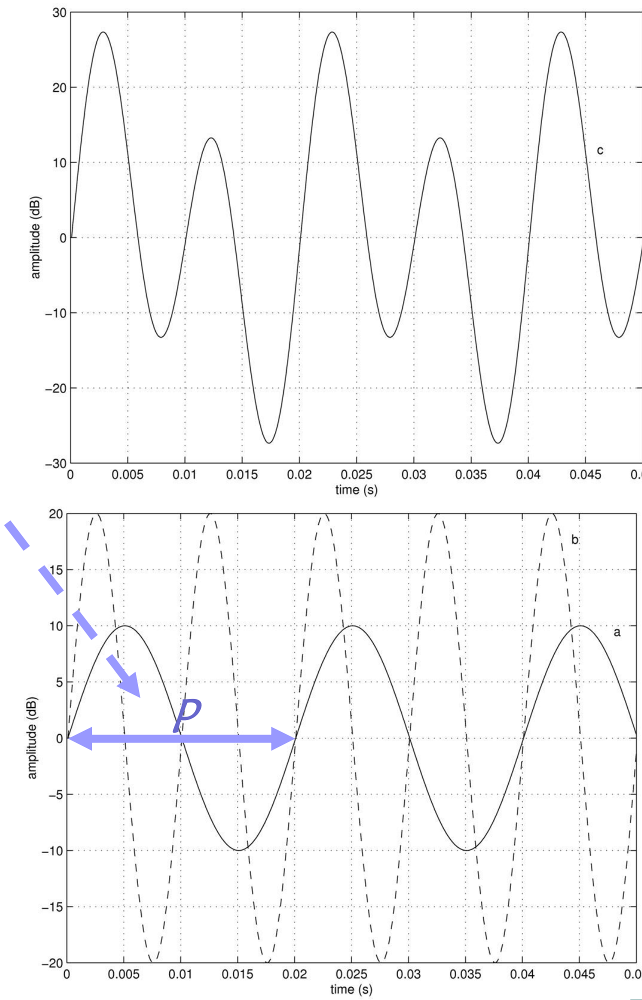
$$f = 1/P$$

- Wavelength (λ) – distance between two consecutive peaks in the sound wave. (Units: meter).

$$\lambda = c/f$$

(c = 345 meters/second – speed of sound).

Eg. c = a + b

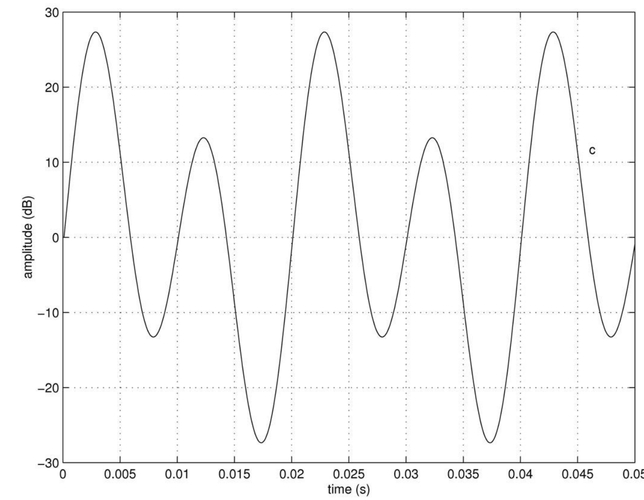


Frequency components

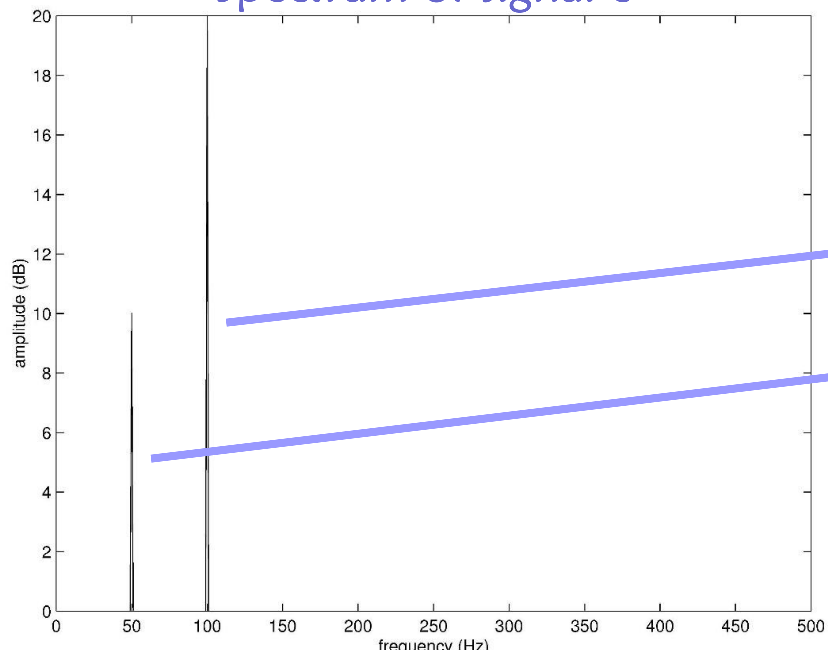
(Frequency) partials – the frequencies (sinusoids) that compose a sound.

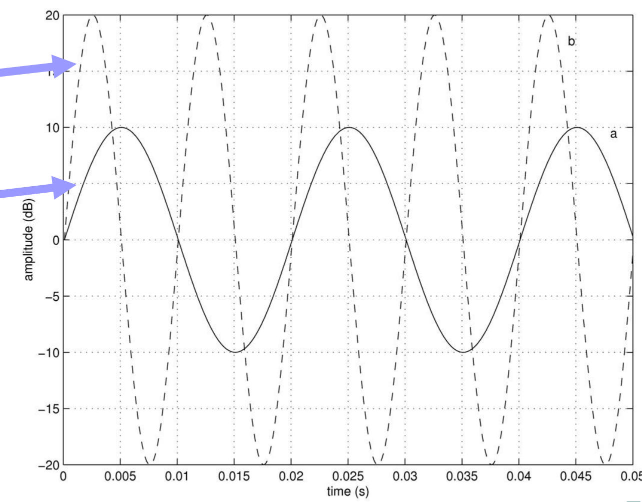
(Amplitude) spectrum – representation of the sound in the frequency domain.

Eg. c = a + b



Spectrum of signal c



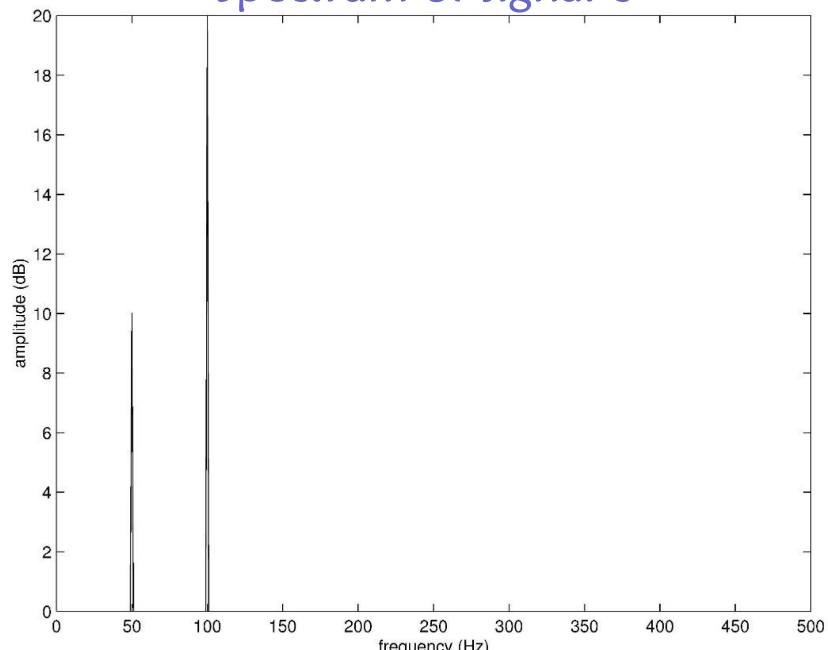


Frequency components

(Frequency) partials – the frequencies (sinusoids) that compose a sound.

(Amplitude) spectrum – representation of the sound in the frequency domain.

Spectrum of signal c

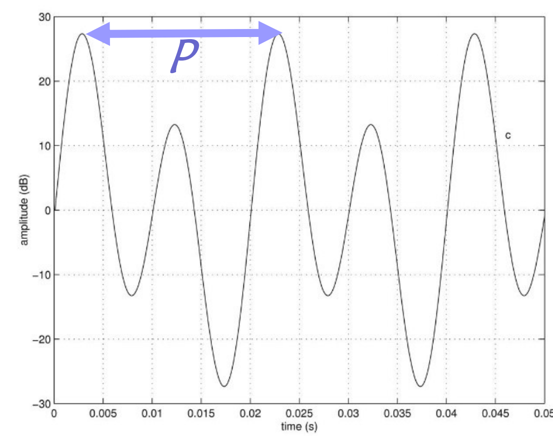


Harmonic sounds – have only partials that are integer multiples of the lowest frequency.

Harmonic – frequency partial of harmonic sound.

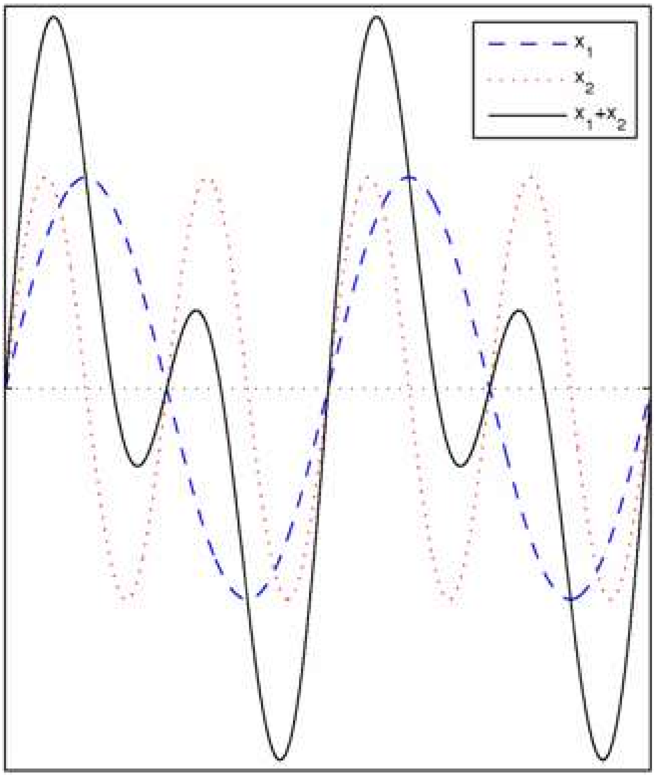
Fundamental frequency – lowest harmonic. (GCD.)

$f_0 = 1/P$



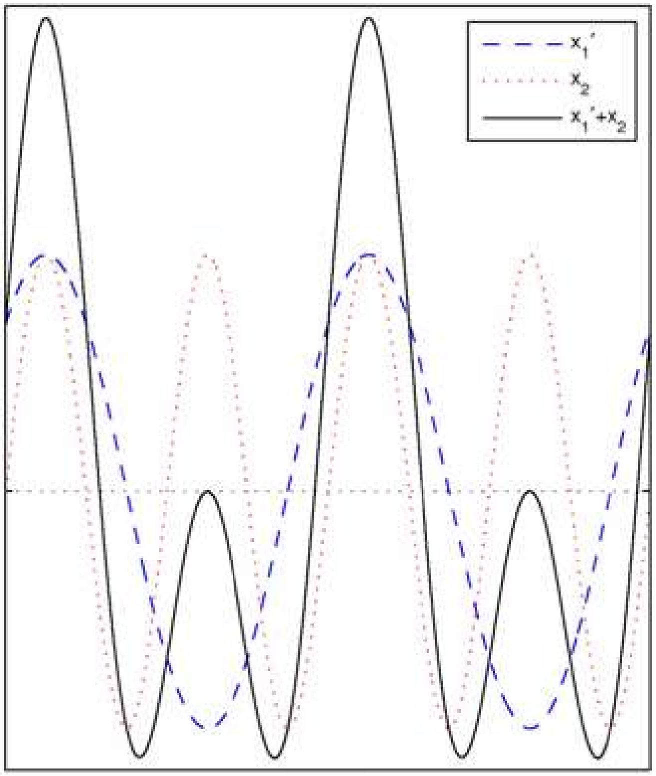
Phase

- Initial Phase (or phase, θ) – initial displacement of the sinusoid. (Units: degrees of angle).



time

x_1 has been displaced by 45°



time

Fourier Analysis

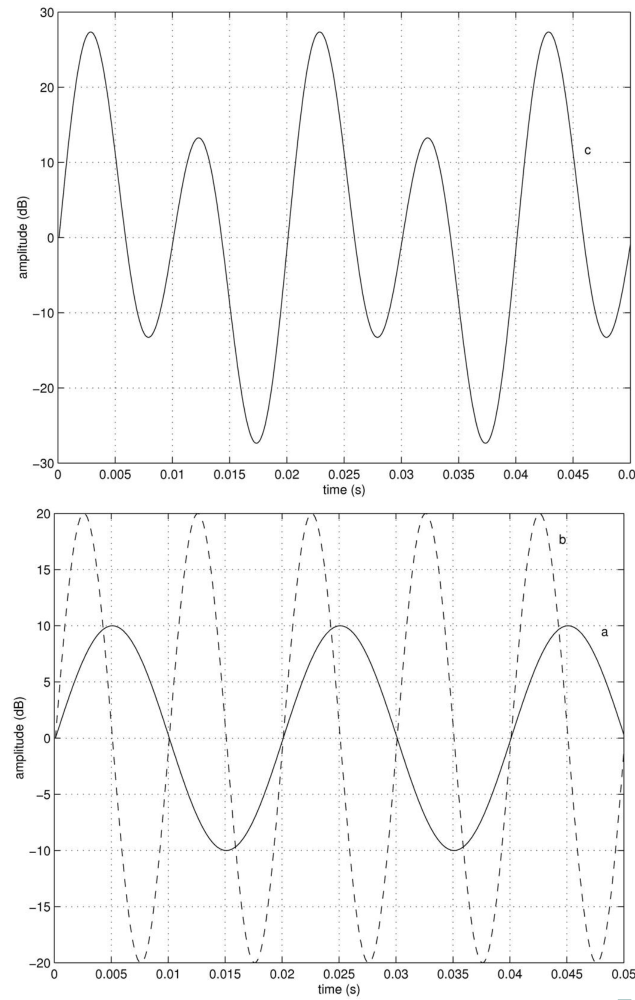
- Pure tone – sinusoid

  $$s(t) = A \sin(2\pi f t + \theta)$$

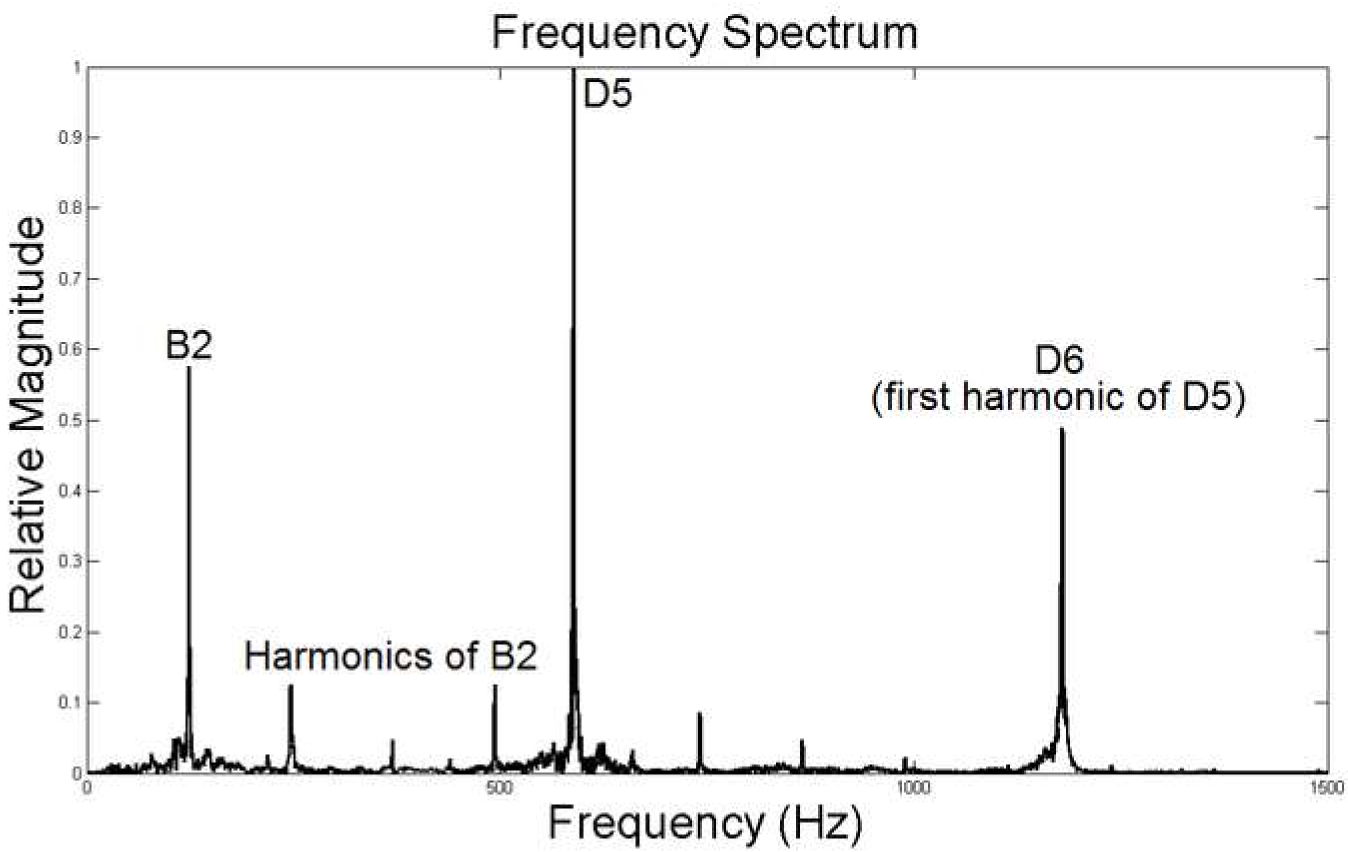
  $s(t)$ – instantaneous amplitude
  A – maximum amplitude
  f – frequency
  t – time
  θ – starting phase

- Fourier analysis – analysis of complex waves in terms of sinusoids (pure tones).
- The Fourier Transform converts the signal from the time domain into the frequency domain.
- The amplitude (A) and phase (θ) of the frequency components can be derived from the Fourier coefficients.
- Fast Fourier Transform (FFT) – efficient algorithm to implement the Discrete Fourier Transform (DFT).

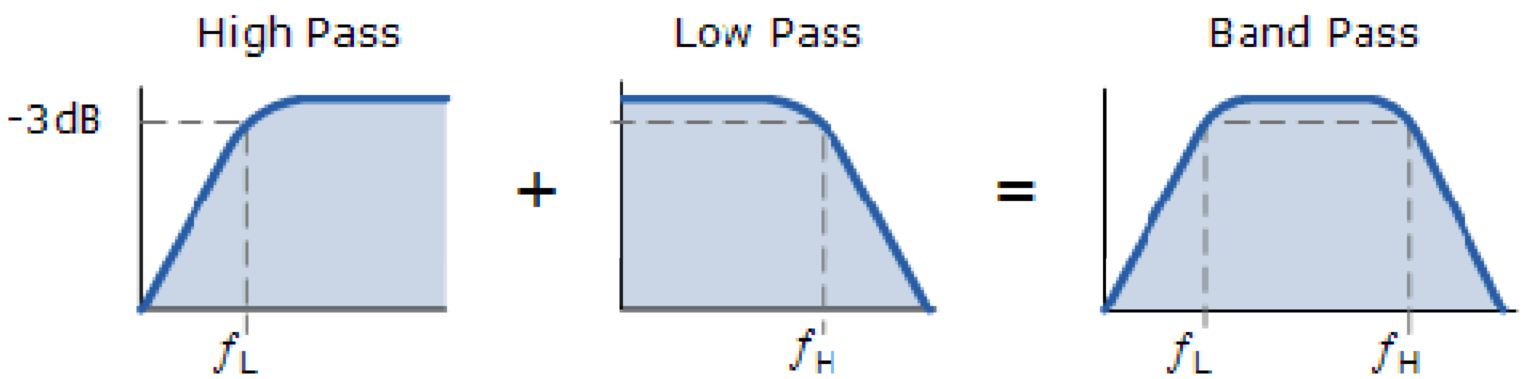
Eg. c = a + b



Fourier Analysis



Filters



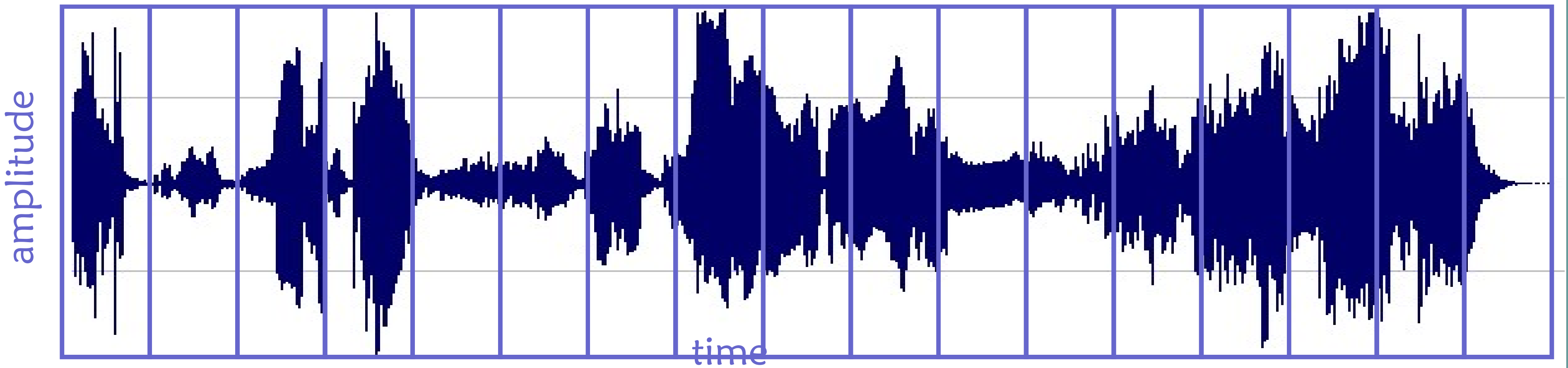
Cuttoff frequency – frequency at which attenuation starts
Roll-off – rate of attenuation
Bandwith = $f_H - f_L$

Example:
https://www.youtube.com/watch?v=ZmZBfictzSM

Fourier Analysis

What about if we want to see the evolution of the spectrum through time?

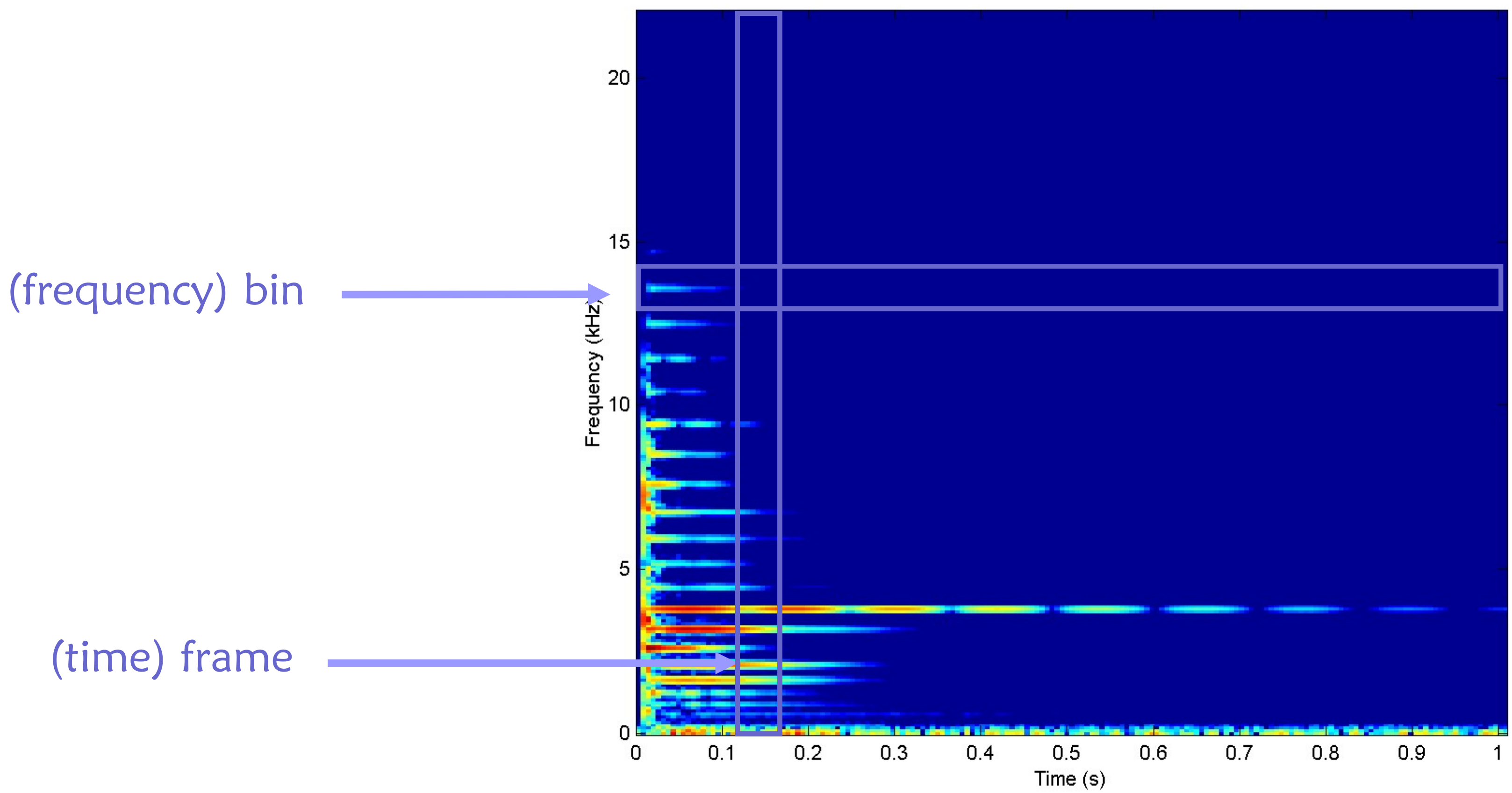
- Find the magnitudes and frequencies (with FT) of the signal at small time windows.



- To avoid discontinuities – do not use rectangular window, use Hanning windows, Hamming windows… and overlap.

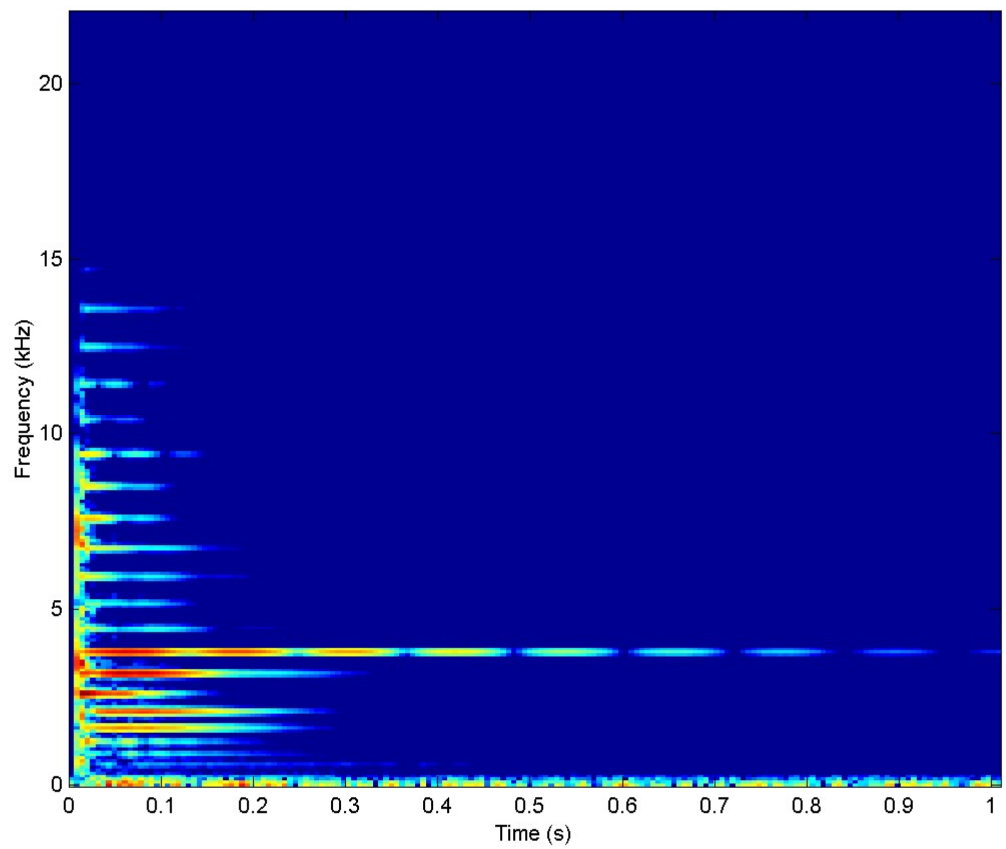
Fourier Analysis

- Short Time Fourier Transform (STFT) or Spectrogram – representation of the sound in frequency and time.



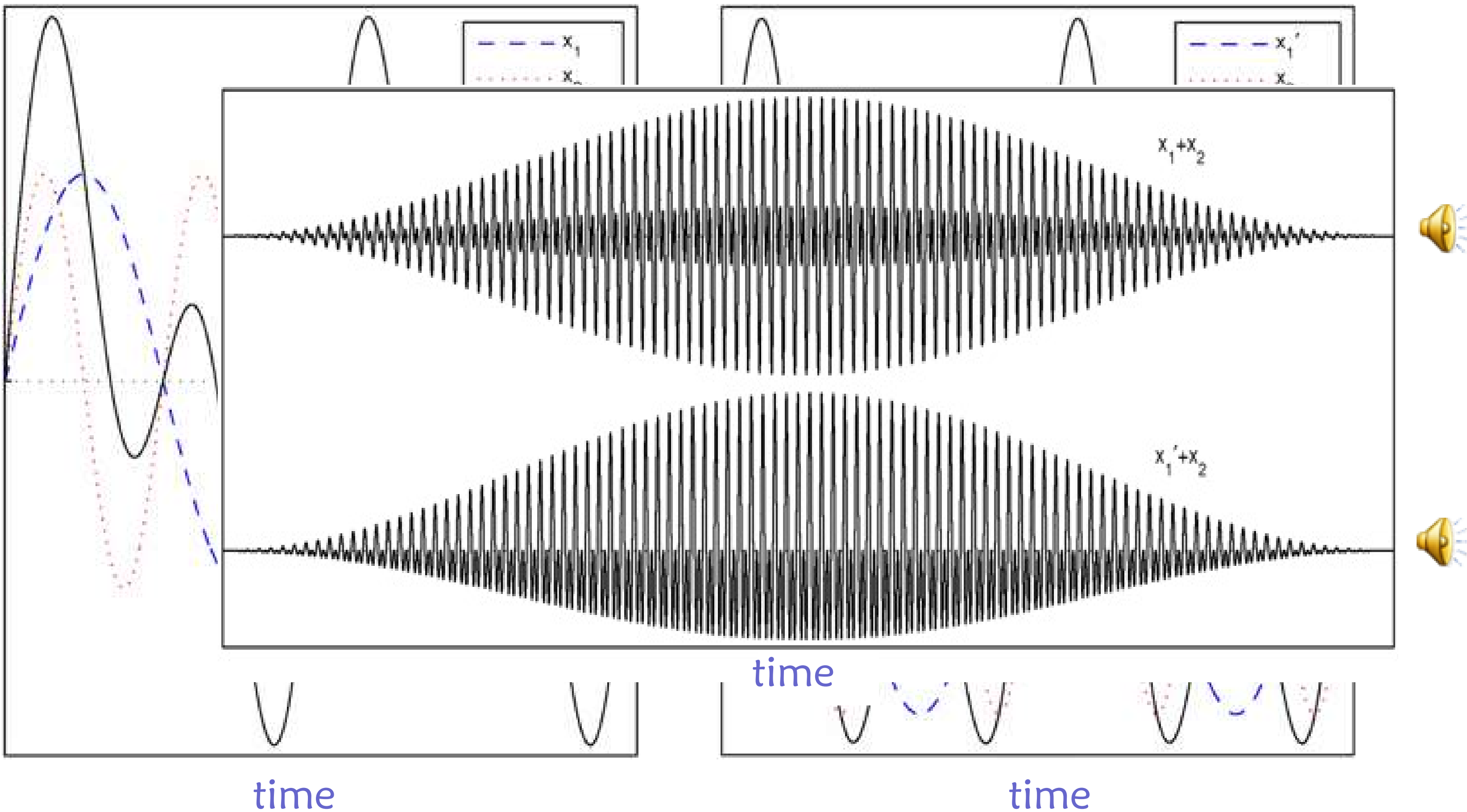
Sinusoidal and Transient Sounds

- Transients – characterized by abrupt changes in energy.
  - Detected by inspecting energy variations in the signal (rapid and strong changes in energy).
- Pure tones – have a longer duration than transients.
  - Look for bins in STFT that have energy in several continuous frames.
- Sinusoidal sounds – consist of pure tones.

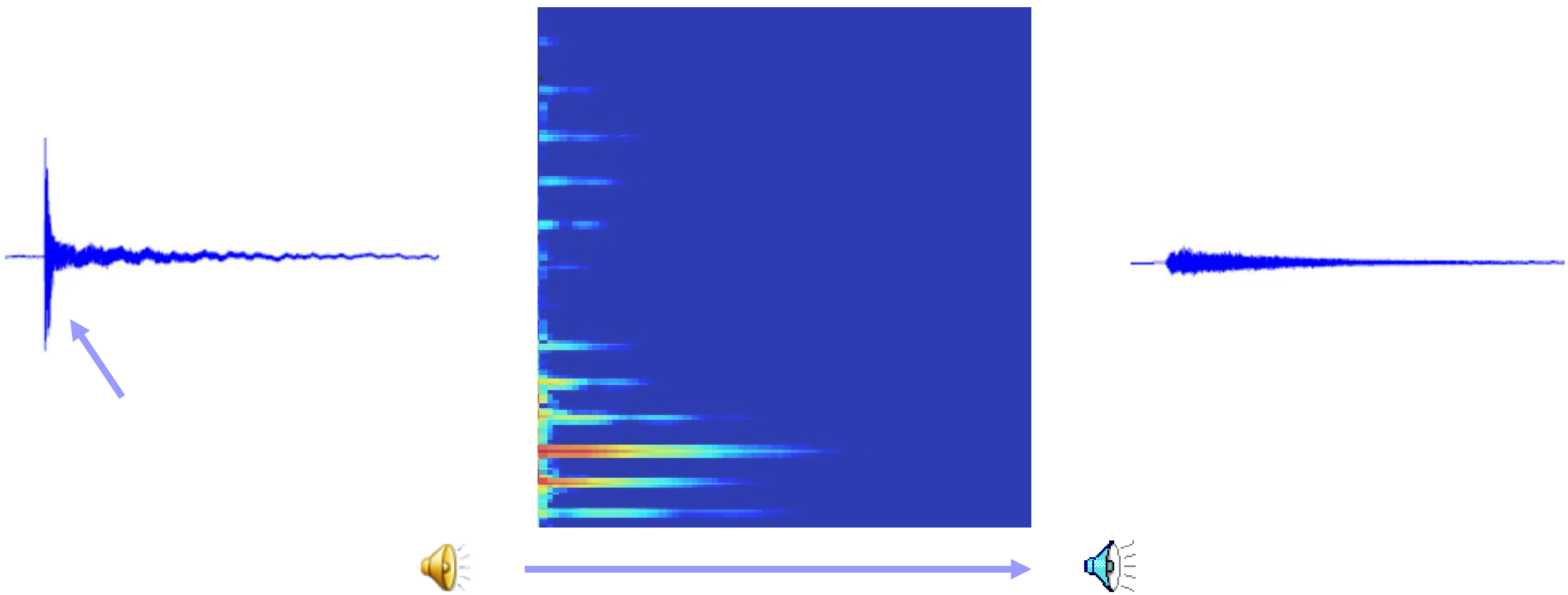


Audio Perception – Phase

- Phase is not perceptually relevant in sinusoidal sounds



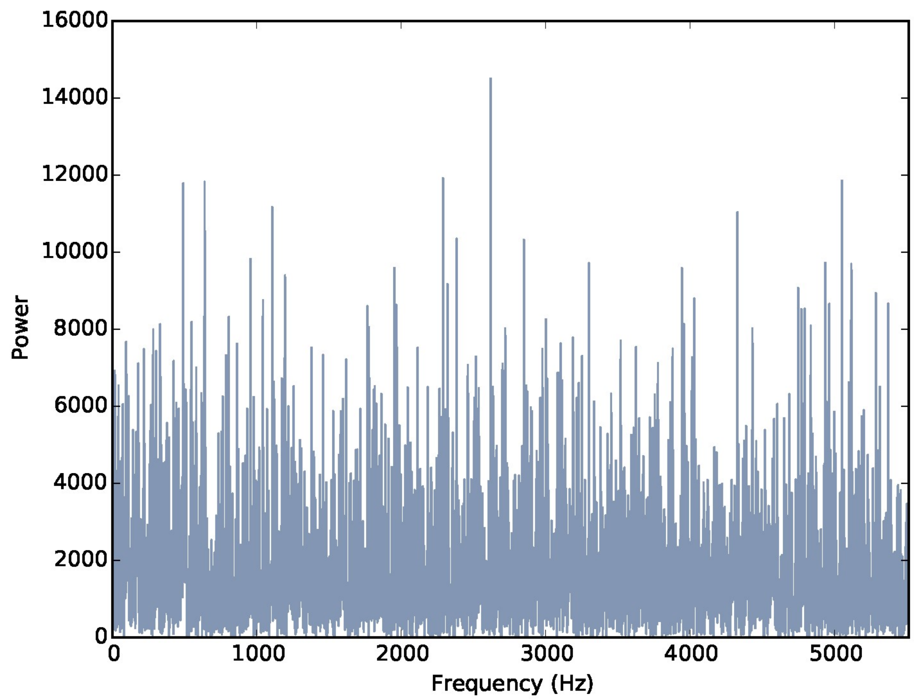
Audio Perception – Phase



- Transients – perceptually important (for instance, in the recognition of musical instruments)

Noise

- Noise (in signal processing) – a signal with energy spread all over the frequency spectrum and without periodic structure.



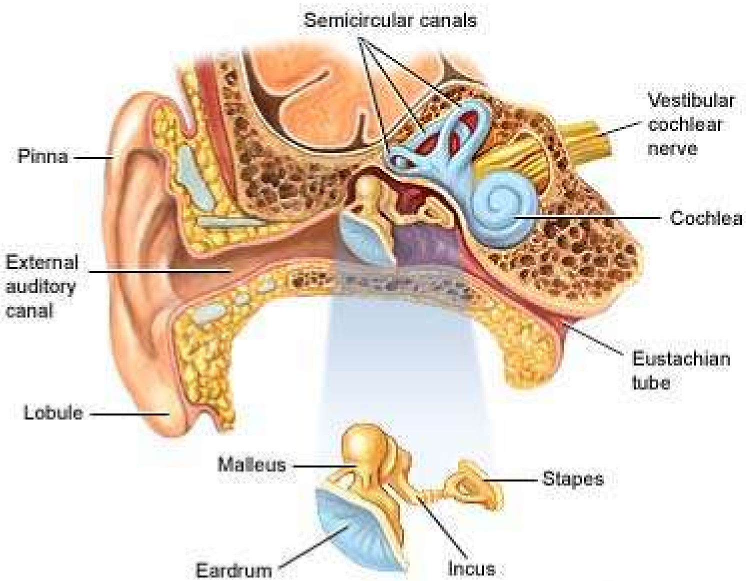
Video with noise: https://www.youtube.com/watch?v=3vEDZ-_iLNU

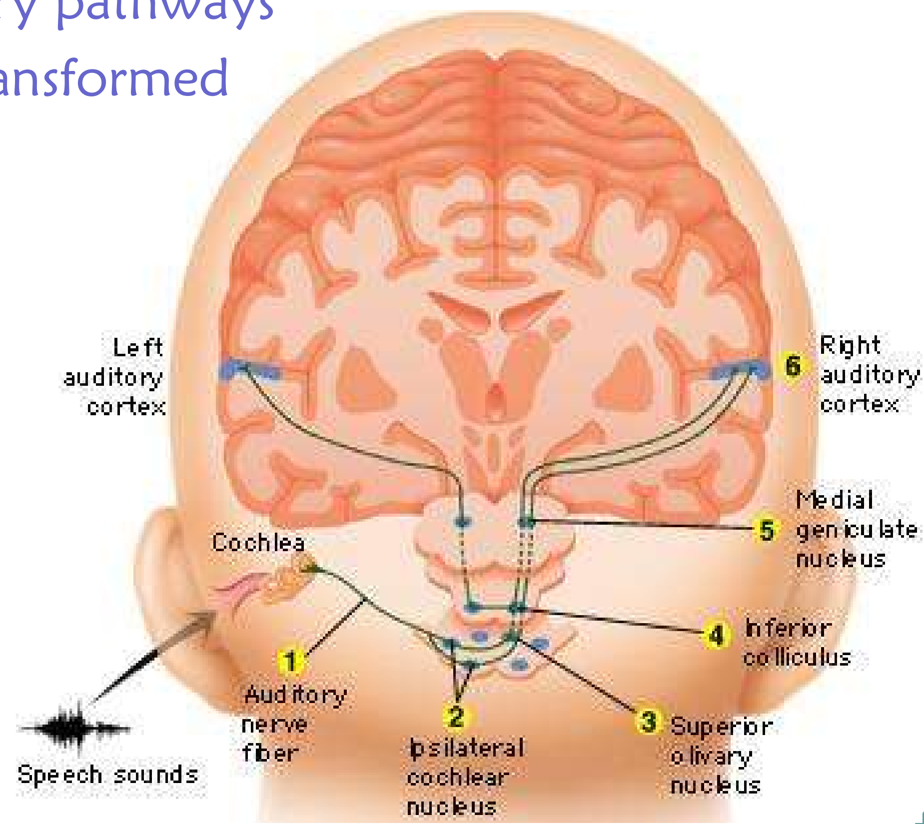
Audio Perception

- Sound perception – sound is our interpretation of sound waves, just as images are our interpretation of light waves.
- The ear is an organ which transforms the sound waves arriving on its surface into neural stimulus while the auditory pathways

  lead the information captured and transformed

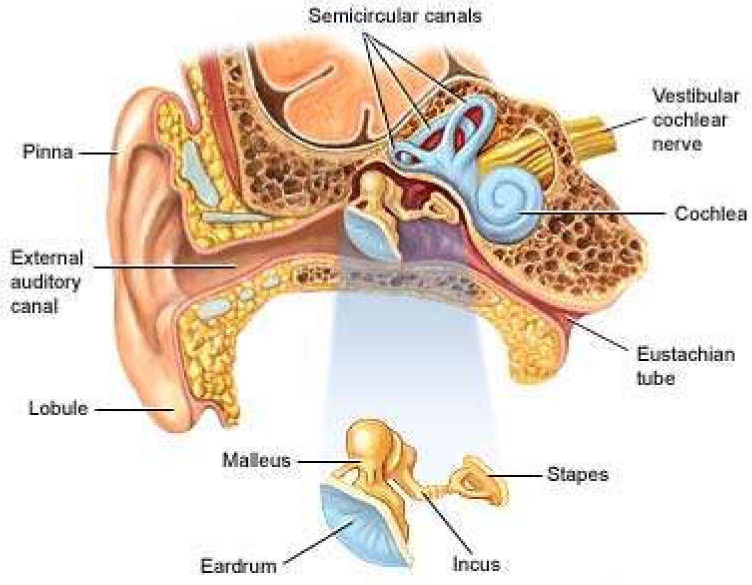
  by the ear into the brain.

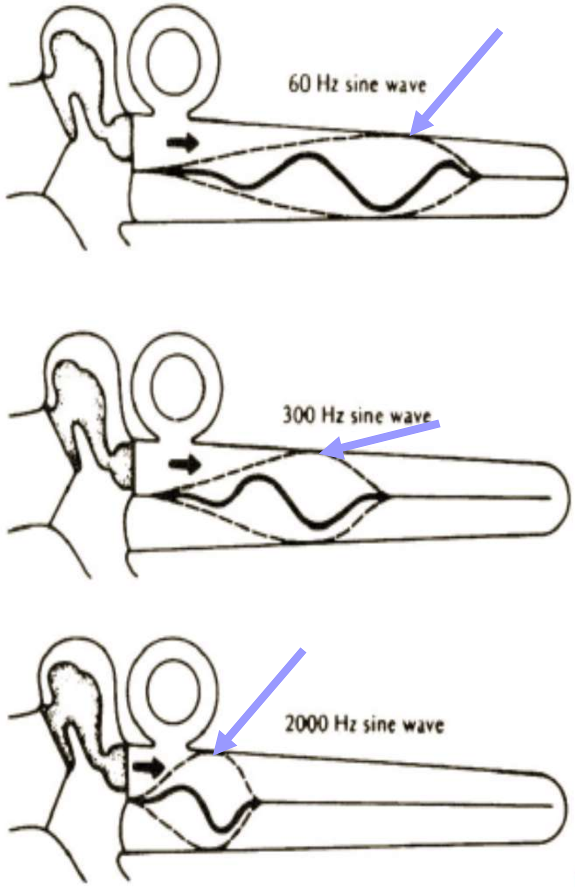




Audio Perception

- Vibration in basilar membrane depends on the frequencies of the sound waves.
- Maximal vibration occurs at different locations depending on frequency.
- Our ears are sensitive to frequencies between 16-20 Hz and 20000 Hz.

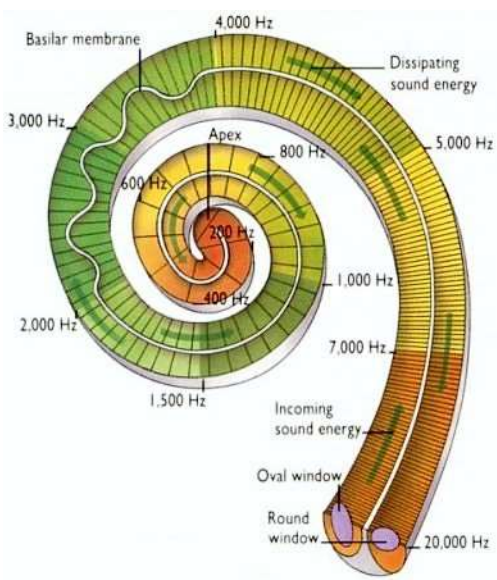


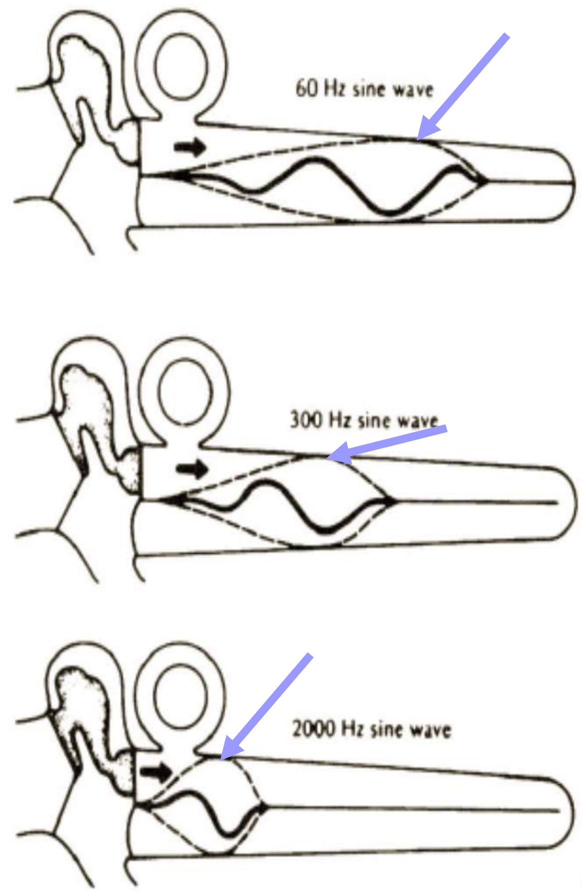


From Yost, 2000

Audio Perception

- Vibration in basilar membrane depends on the frequencies of the sound waves.
- Maximal vibration occurs at different locations depending on frequency.
- Our ears are sensitive to frequencies between 16-20 Hz and 20000 Hz.





From Yost, 2000

Audio Perception – Loudness

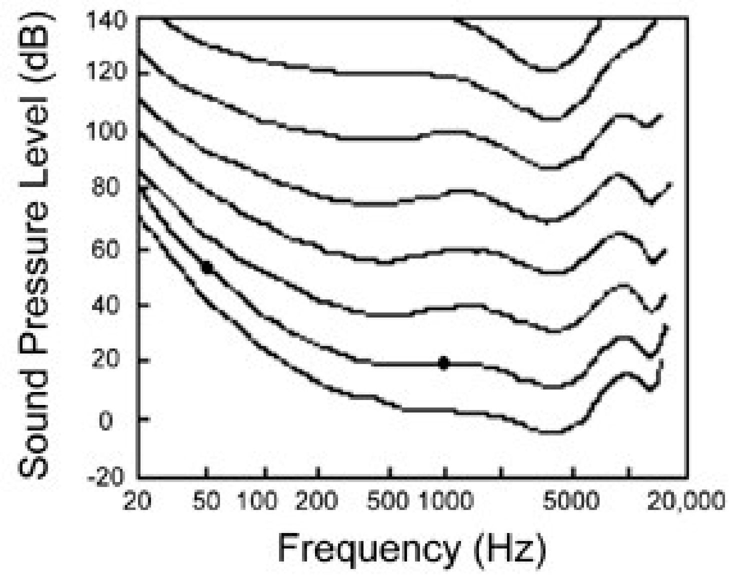
- Loudness – (subjective) measure of auditory sensation that is related to the amplitude of the waveform.

- Loudness depends on frequency.

  (Eg. Play two sinewaves with the same amplitude but different frequencies and notice the perceived loudness of each sound.)

- Equal loudness contour –

  All pairs (frequency, pressure level) in the same curve have the same loudness.



Decibel

- Dynamic range of hearing – we can hear 10^{14} different levels of intensity!
- 1 decibel (dB) = $10 \log (I_1/I_2) = 10 \log (P_1/P_2)$
- 1 decibel (dB) = $10 \log (A_1^2/A_2^2)$,   P – power, proportional to A^2

  $= 20 \log (A_1/A_2)$

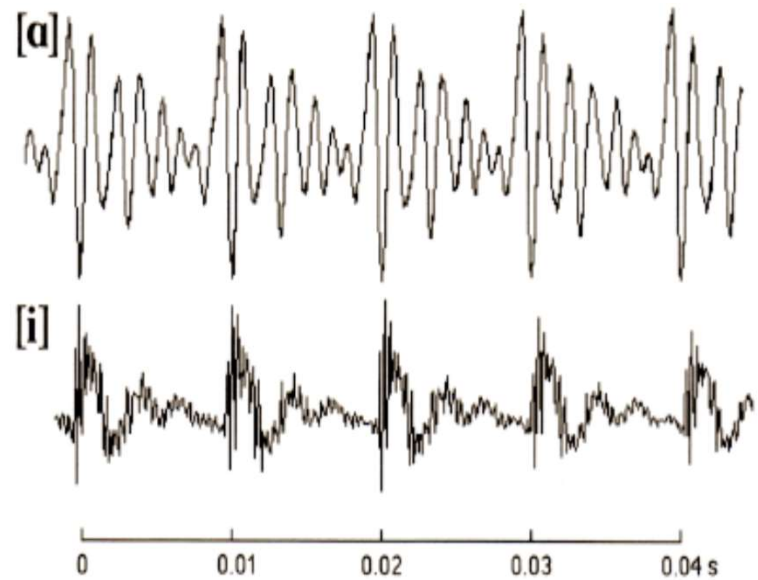
- Dynamic range of hearing = $10 \log 1/ 10^{14}$ = 140 dB.
- Decibel – relative (not absolute) measurment. It is the ratio of two quantities.
- Decibel conventions:
  - Sound Pressure Level (SPL) – decibels expressed in relation to a pressure of 20 micropascals (P_2).
  - Sensation Level (SL) – decibels expressed in relation to the least intensive sound a particular subject can detect in an experiment.

Sound pressure

Sound pressure	
150 dB	Jet take-off (eardrum rupture)
140 dB	Painfull
120 dB	Sensation threshold
96 dB	Subway noise
72 dB	Factory noise
60 dB	Conversation in restaurant
50 dB	Conversation at home
20 dB	Whisper
10 dB	Breathing
0 dB	Hearing threshold

Audio Perception – Pitch

- Pitch – (subjective) measure of auditory sensation that depends on frequency.
- Pitch is perceived only for fast repetitions (>20-50 Hz)
- Repetitions below 20-50 Hz are perceived as distinct sounds. Eg.:
  - sequence of phonemes
  - sequence of notes
- In complex sounds it often depends on the fundamental frequency, that is, it depends on the frequency of repetition of the complex wave.



Two vowels with the same pitch

From CP08, Lewicki (CMU)

Pitch → Perceptual Measurement

Fundamental frequency → Physical Measurement

Audio Perception – Pitch

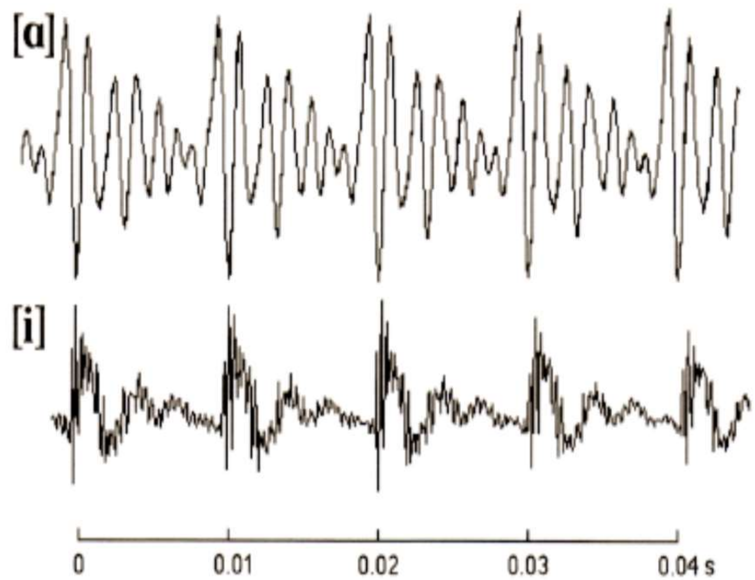
- Pitch of the missing fundamental – sometimes pitch can be perceived as the frequency of f_0 even when f_0 is missing but other times this does not happen,
- Pitch shift of the residue – pitch does not correspond to f_0

Determining the pitch is still a challenge for hearing scientists!

Pitch also depends on level, bandwidth…

Audio Perception – Timbre

- Timbre (pronounced "timber") – (subjective) measure of auditory sensation that allows us to distinguish sounds that have the same pitch and loudness.
- Transient sounds can have timbre (but not pitch)
- Sounds with different timbres have different waveforms.
- The inverse is not necessarily true due to phase deafness for sinusoidal sounds.



Two vowels with the same pitch
From CP08, Lewicki (CMU)

Audio Perception – Timbre

Timbre depends on the harmonic content of the signal.