UNIVERSITÀ DEGLI STUDI
DI MODENA E REGGIO EMILIA

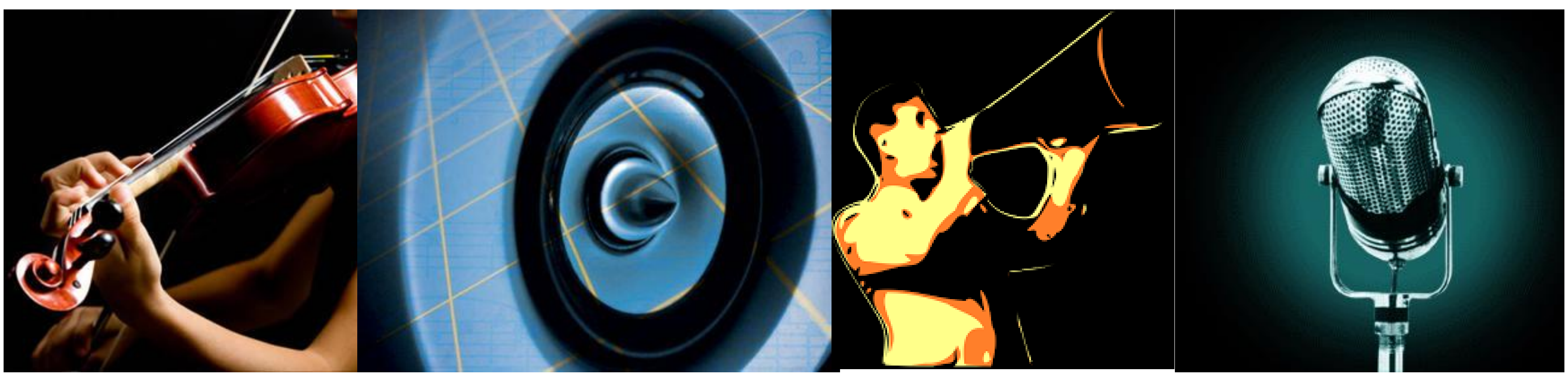
Lecture notes for
Multimedia Data Processing

Sound

Last updated on: 27/03/2020

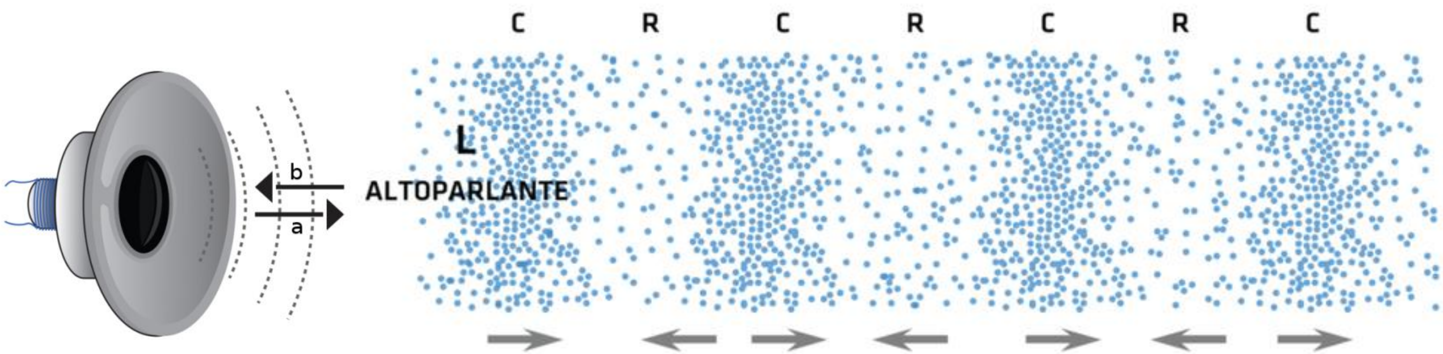
What is sound

- A first definition that we can give is that what we perceive as **sound is a cyclic variation**, with respect to a constant value, of the air pressure.
- A sound can only propagate through **a medium**.
- This means that any medium, be it solid, liquid or gaseous, can transport sound, affecting its speed depending on its density, temperature, pressure and other physical factors.



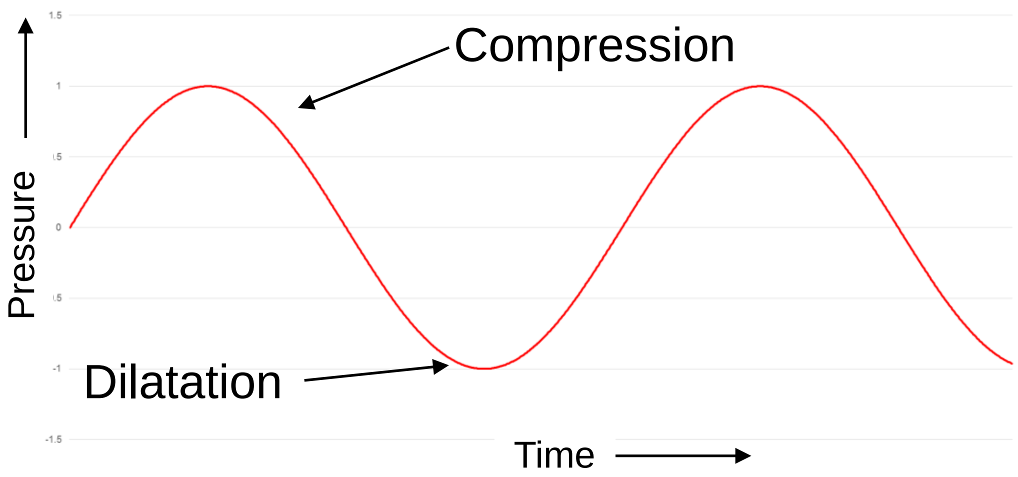
Sound propagation in the air

- Sound **propagates** through the air through **multiple collisions between particles**.
- Let's consider having a loudspeaker as our sound source.
- The membrane moves back and forth following the pattern of the electrical signal that reproduces the sound information.
- The loudspeaker moves and pushes the air particles on its right (phase a) by applying **a compression**. These, in turn, push the particles close to them, and transfer the energy they received from the speaker.
- The loudspeaker goes back and dilates to the left (phase b). Doing so, it creates **a depression** in front of it, which is filled by the air particles in the immediate vicinity.



Sound propagation in the air

- This process causes the particles to transmit energy by oscillating and not physically moving in the direction of sound propagation.
- We are easily convinced of this by thinking of a cork in a body of water where a stone is thrown. It will be observed that the cork oscillates up and down as the wave generated by the stone propagates but remains immobile with respect to the direction of propagation of the wave.
- If the loudspeaker is driven by a sinusoidal signal, the atmospheric pressure in its vicinity will have the following trend:



Sound wave

- A **wave** is a disturbance that propagates by transporting energy, but not matter. Waves are generated from a source and can propagate in a material medium.

- A **longitudinal wave** occurs when the elements of the material medium move parallel to the direction of the wave propagation.

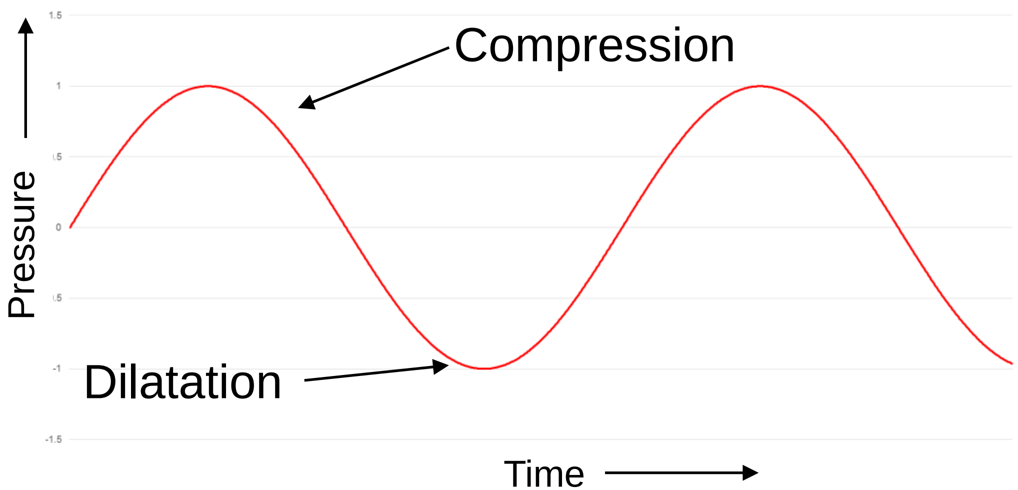
- **Sound** is a longitudinal wave generated by successive compressions and rarefactions of the medium in which it propagates: without a material medium, sound does not propagate. The source of sound is a body that vibrates.

Sound properties

- Waveforms can be very complex but fortunately any waveform can be considered (under certain conditions) as an extension of a very simple waveform: the sinusoid, expressed in its most generic form by the following formula:

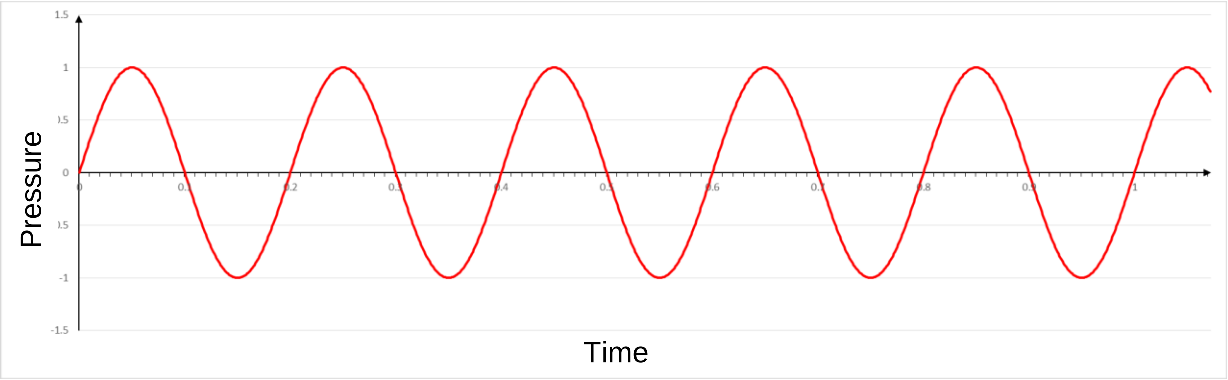
$$y = A\sin(2\pi f t)$$

- The sinusoid has a series of properties, which will be described and analyzed one by one: frequency (f), period (T), wavelength (λ), amplitude (A), phase (φ).



Frequency

- It is defined as the number of cycles that the wave performs in one second.
- It is measured in Hertz, whose symbol is Hz and physical size is s^{-1}. A frequency wave of 1Hz completes one cycle every second.
- The following wave has a frequency of 5 Hz.



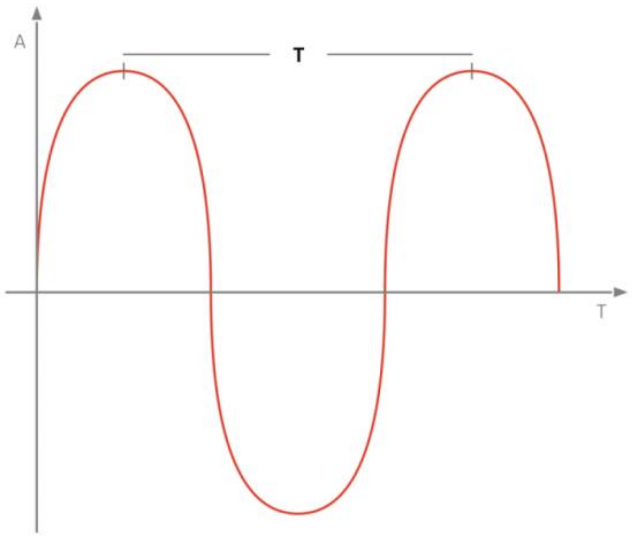
- In order for the human ear to perceive the cyclic variation of sound pressure as a sound, the variation must perform a minimum number of cycles per second (the minimum threshold is approximately 20 Hz).
- The following sound refers to a sinusoid of frequency equal to 1KHz (1000 oscillations per second):

Period

- It is the time taken to complete a whole cycle. The relation is:

$$T = \frac{1}{f}$$

- The following figure shows the duration of a sine wave period:



Wavelength

- It is defined as the distance between two corresponding points (for example two successive maxima) along the waveform. Its value can be calculated from the following formula:

$$\lambda = \frac{c}{f}$$

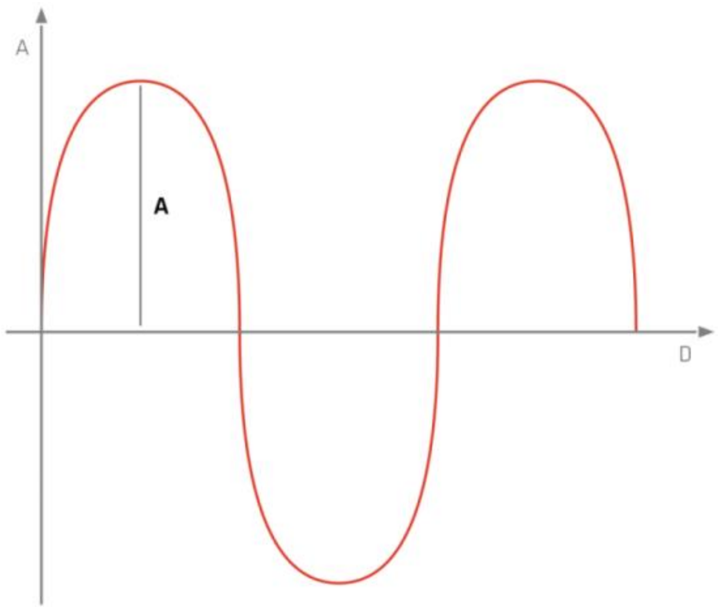
- where: c = speed of sound in the medium being considered (about 344 m/s in the air).
- Consider a 1Hz frequency wave that travels through the air. From the first formula we will have that:

$$\frac{344\ m/s}{1\ Hz} = 344\ m$$

- that is, at each cycle the wave extends for 344m, or two football stages (as we will see, the human ear begins to perceive sounds with a frequency higher than 20-30Hz, corresponding to a wavelength of 15-18 meters.)

Amplitude

- It is the measure of the maximum deviation from the equilibrium position.
- Larger amplitudes correspond to higher volumes.
- There are two types of amplitude measurement:
  - **Peak amplitude:** this quantity measures the point where the wave has maximum amplitude. It is the maximum absolute value of the difference from the signal and the reference.



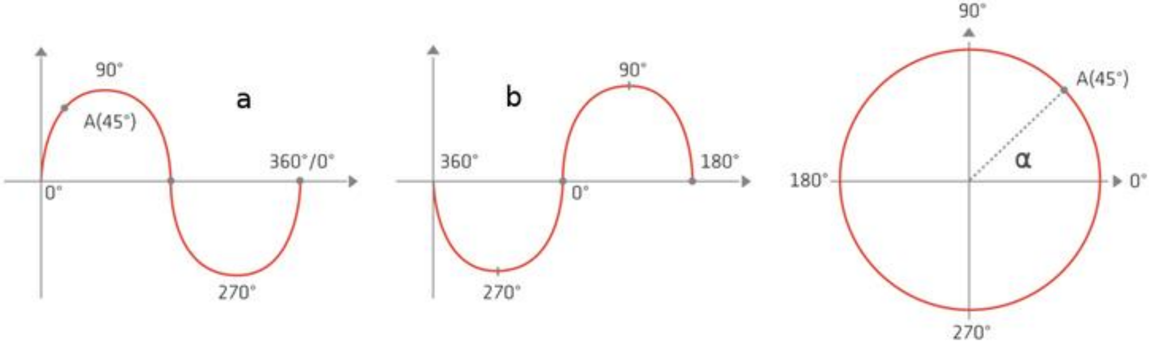
Phase

- This quantity is always a **relationship** between two waveforms having the **same frequency**.
- A **phase difference** can be described as the distance between two points which rotate at the same speed (therefore at the same frequency) but which start from different positions on the circumference. In particular, the **angle** identified by the two points is precisely the phase difference.

$$\Phi = 2\pi f \Delta t$$

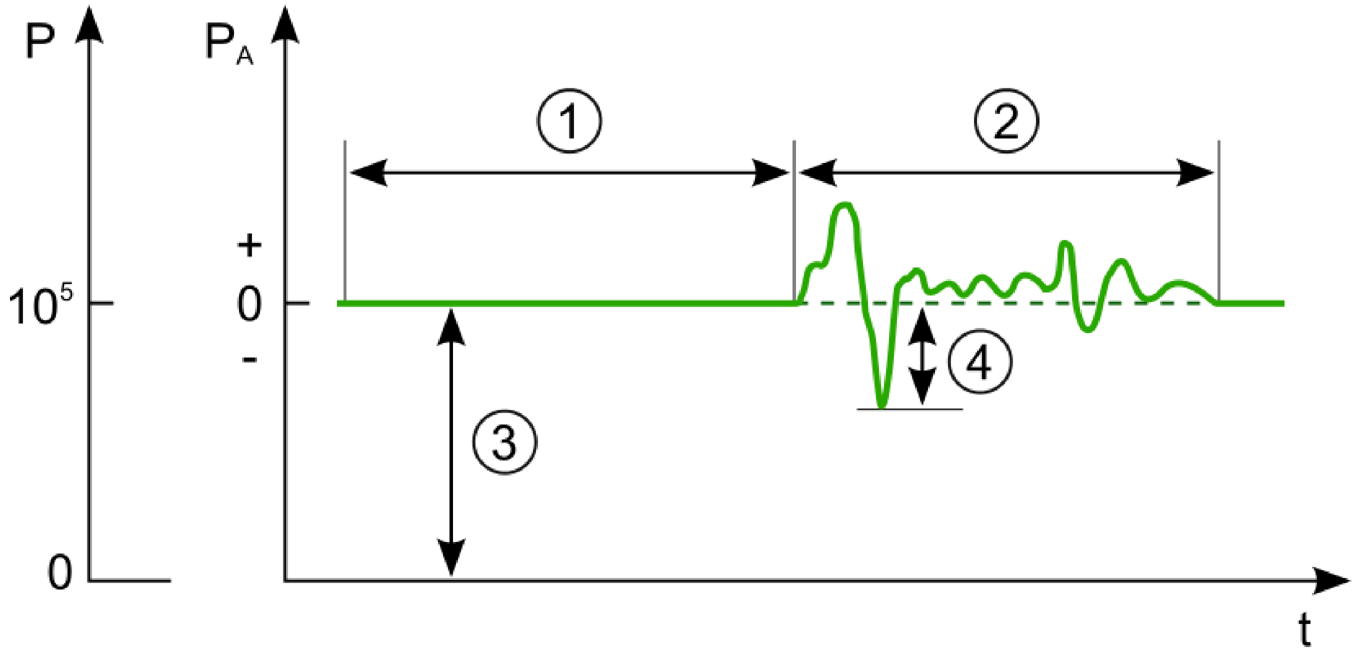
- **Example:** calculate the delay required for two 100Hz sinusoids to be out of phase by 90°:

$$\frac{\pi}{2} = 2\pi \cdot 100 \cdot \Delta t \Rightarrow \Delta t = \frac{\pi/2}{2\pi \cdot 100} s = \frac{1}{400} s = 2.5\, ms$$



Sound pressure

- The physical quantity used to describe sound is pressure.
- In sound waves the amplitude is therefore the variation of local pressure with respect to the atmospheric pressure at that point. This is called sound or acoustic pressure.
- Being it a pressure difference, it is still measured in Pascal (Pa).



Example of sound pressure diagram versus time:

1. silence
2. audible sound
3. atmospheric pressure (taken as a reference level)
4. instantaneous sound pressure

Sound pressure level

- To express a summary value of a quantity in a certain time interval, the concept of effective value is used, that is, the square root of the mean of the squares of the values (root mean square or RMS).
- Effective sound pressure p is the effective value (RMS) of pressure over a period of time.
- Sound pressure level (SPL) L_p is a logarithmic measure of the effective sound pressure of a mechanical (sound) wave with respect to a reference sound source.
- The reference sound pressure p_0 is around the hearing threshold at 1000 Hz, i.e. 0.00002 Pa, or 20 µPa.
- SPL is given by the ratio between effective and reference sound pressure, measured in dB:

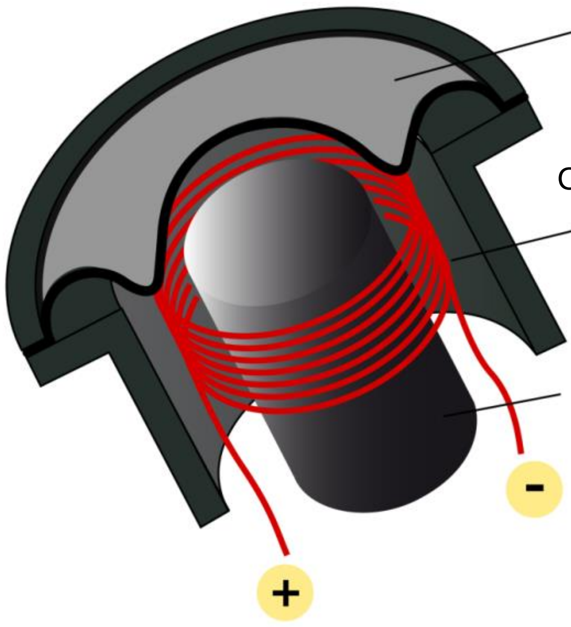
$$L_p = 20\log_{10}\left(\frac{p}{p_0}\right)\ \mathrm{dB}$$

SPL examples

Sound source	**Sound pressure**	**SPL**
Krakatoa explosion at 160 km	20 000 Pa	180 dB
Jet engine at 30 m	630 Pa	150 dB
Rifle shot at 1 m	200 Pa	140 dB
Pain threshold	100 Pa	130 dB
Hearing damage due to short term exposure	20 Pa	around 120 dB
Pneumatic drill at 1 m; discotheque	2 Pa	around 100 dB
Hearing damage due to long term exposure	0,6 Pa	around 85 dB
Heavy traffic at 10 m	0,2-0,6 Pa	80-90 dB
Passenger train in motion at 10 m	0,02-0,2 Pa	60-80 dB
Noisy office; TV at 3 m (moderate volume)	0,02 Pa	circa 60 dB
Normal conversation at 1 m	0,002-0,02 Pa	40-60 dB
Quiet room	0,0002-0,0006 Pa	20-30 dB
Rustle of leaves, relaxed human breath at 3 m	0,00006 Pa	10 dB
Audibility threshold at 1 kHz (man with healthy hearing)	0,00002 Pa	(ref.) 0 dB

Microphones and miking techniques

- Microphones are **transducers** capable of transforming acoustic energy into electric energy. **Atmospheric pressure variations** are converted into voltage variations and therefore into an electric current.



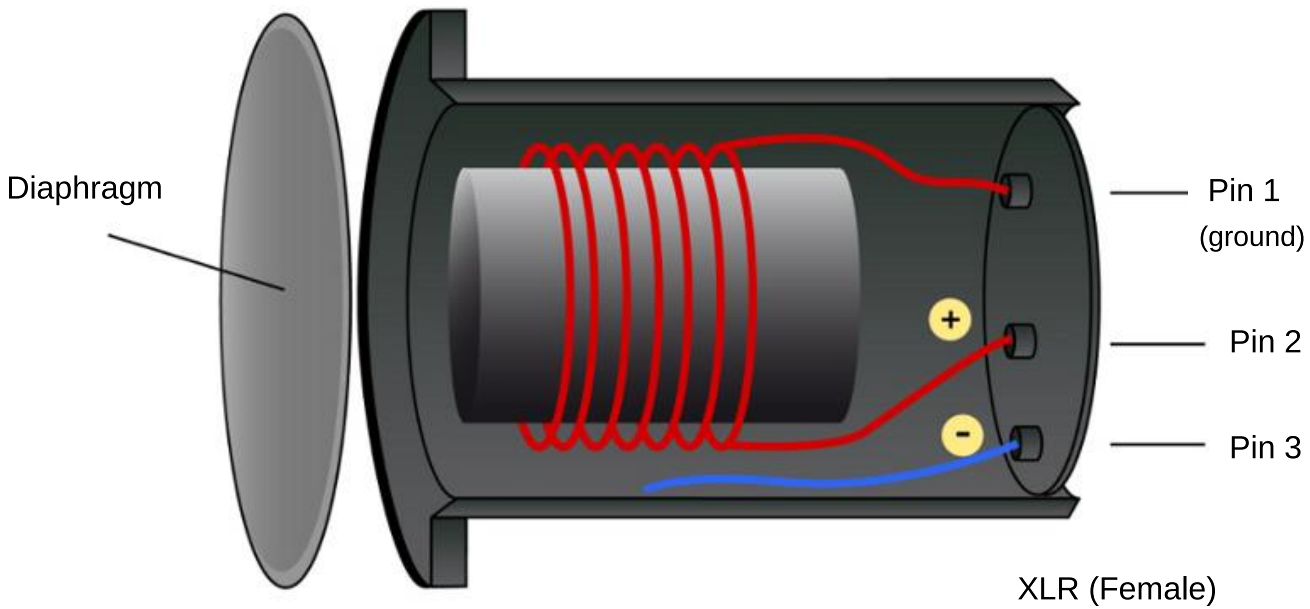
Electrodynamic microphone

- A **coil** made of a conducting material is fixed to the **diaphragm** which receives the sound wave and vibrates as a consequence.
- The coil is located within a magnetic field generated by a magnet positioned inside it.

- When the **diaphragm vibrates** the coil moves too, thus interrupting the magnetic field lines and generating a current inside the coil.

Microphones and miking techniques

- This results in the flow of an electric current inside the coil. This way **the electric signal that has been generated has the same rate as the acoustic wave** that hits the diaphragm.

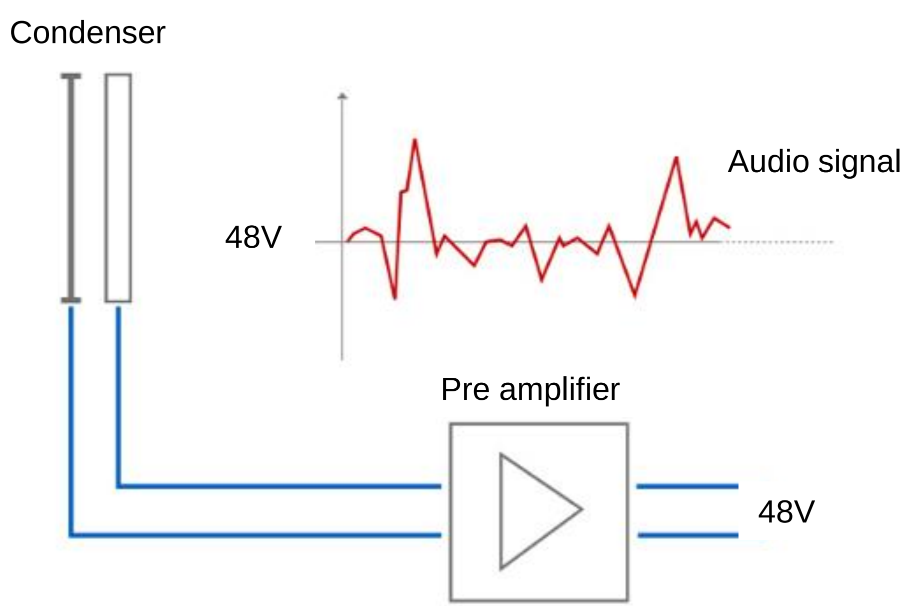


Microphones and miking techniques

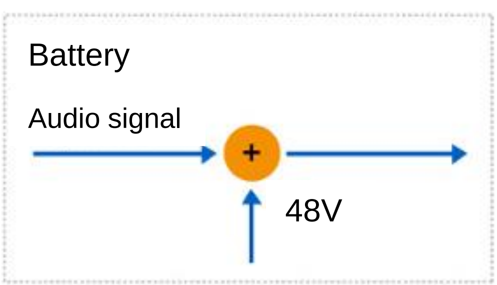
- Short list of the main characteristics of **electrodynamic microphones**:
- They are the most **resistant** and are therefore often used for **live performances** in which energetic singers can express their exuberance without damaging them.
- Their resonance frequency is about 2.5 KHz, this makes them particularly **ideal for the correct reproduction of vocals and guitars**.
- They are capable of enduring very high sound-pressure.

Condenser microphone

- This kind of microphone, also known as *electrostatic*, **has a condenser inside it**.
- One of the condenser two plates is the microphone diaphragm, which vibrates when the acoustic wave hits it.
- The plate's **vibrations** produces **a variation in the distance** between the two plates, therefore altering the capacity value. Charge takes time to vary, so the voltage must change instantly.



$$V = \frac{Q}{C}$$

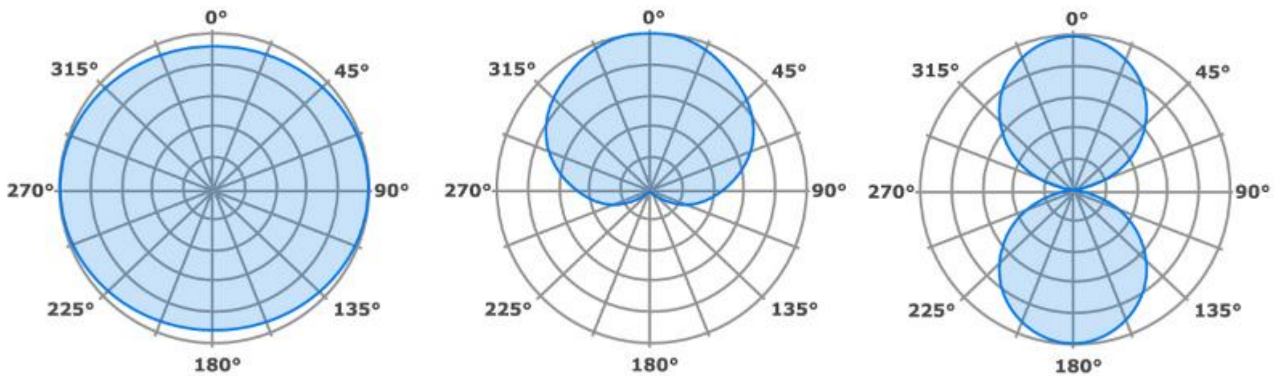


Condenser microphone

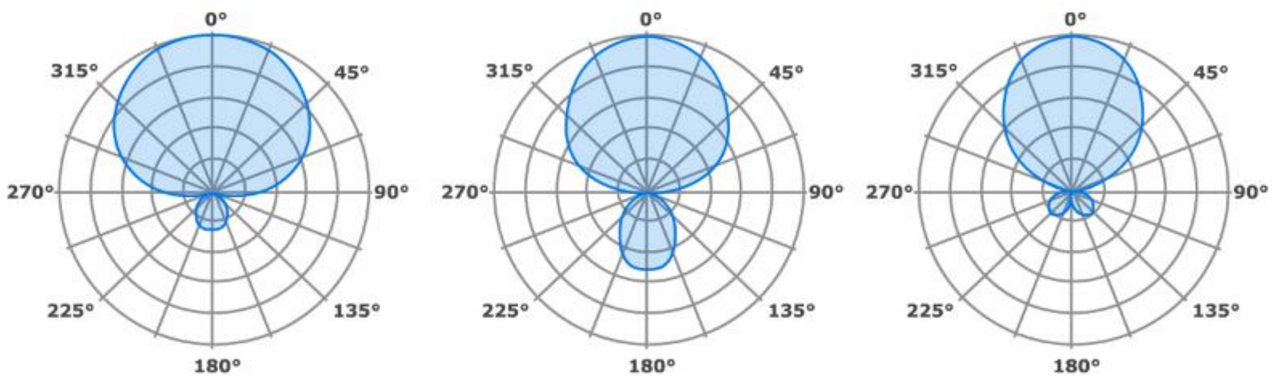
- Initially it is necessary to apply some **voltage in order to polarize** the condenser.
- This voltage is called **phantom power** and is usually supplied by the mixer to which the microphone has been connected, or by a battery.
- **High sensitivity**, which allows the microphone recording of sound **sources placed at a distance** from the microphone. This feature makes it particularly suitable for the use of **stereophonic shooting** techniques.
- **Very thin diaphragm** which allows good reproduction even of higher frequencies.
- Easily **damaged** if exposed to very high sound-pressure.
- Very delicate, therefore not ideal for live performances. More suitable for **recording studio work**.

Polar pattern of a microphone

- The sensitivity's rate in relation to the direction of the sound source is illustrated on a graph called *polar pattern*.



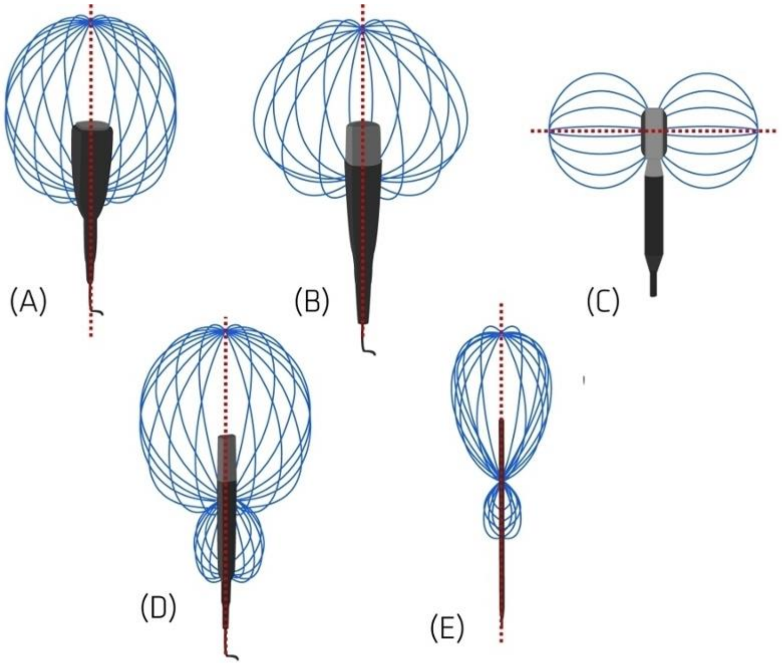
(A) Circular (B) Cardioid (C) Figure of 8



(D) Super cardioid (E) Hyper cardioid (F) Shotgun

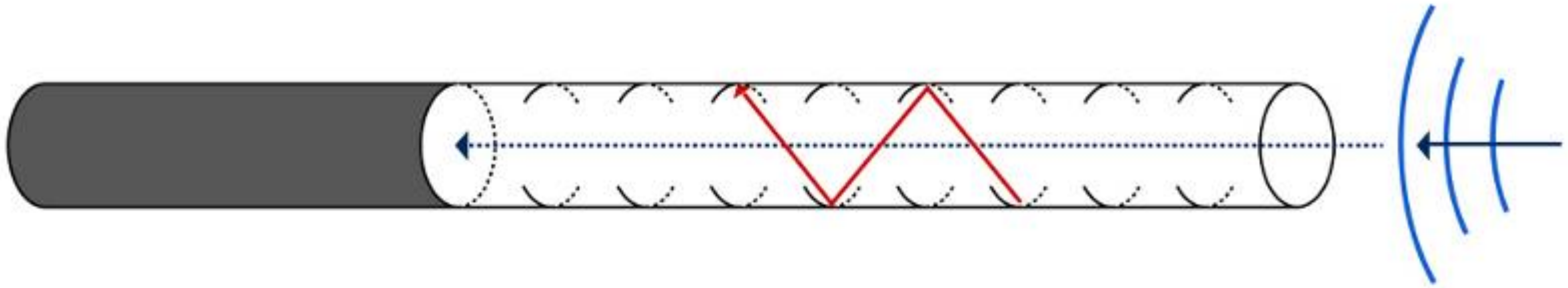
Polar pattern of a microphone

- Three-dimensional perspective
- **Super cardioid (D):** the same as the cardioid diagram but with more pronounced directional characteristics.
- However, **to tighten the front-side pattern there is an inevitable insurgence of a little posterior lobe**. This entails a slight increase in sensitivity to sounds issuing from behind the microphone.



Special microphone: shotgun

- This microphone consists in a **diaphragm placed at the end of a slitted tube**.
- The peculiarity of this kind of microphone is that any sound not coming from the pointing direction enters through the slits and, **due to the tube's length, it undergoes countless reflections** which pretty much nullify each other.
- The sounds coming from the pointing direction on the other hand, pass through the tube without any obstacles. This microphone is used to pick up a specific sound source within the space, even at great distances.



Microphone sensitivity

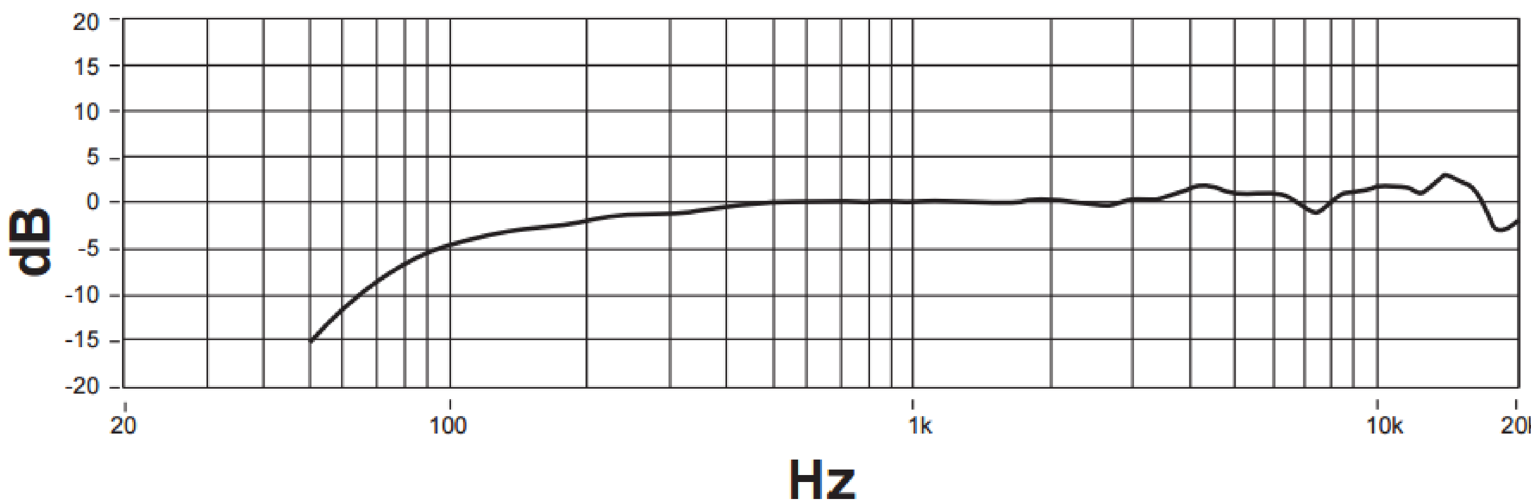
- Sensitivity is a microphone's ability to accurately convert an acoustic wave-form into an electric signal.
- Sensitivity is therefore measured in V/Pa. However, considering the quantities involved, it is more appropriate to use mV/Pa.
- Sensitivity is measured by exposing the microphone to a sine wave with a frequency of 1 kHz and an SPL of 94 dB (i.e. an effective sound pressure of 1 Pa) at a distance of 1 m.
- Distance is important, since sound pressure decreases linearly with distance.
- Another way of expressing the sensitivity value is to use dBV. Again, a reference is needed. This was chosen at 1000 mV / Pa, an unattainable value for any microphone:

$$Sensitivity_{\mathrm{dBV}} = 20\log_{10}\left(\frac{Sensitivity_{\mathrm{mV/Pa}}}{1000\,\mathrm{mV/Pa}}\right)\,\mathrm{dBV}$$

- Sensitivity in dBV will always be negative.

Frequency response

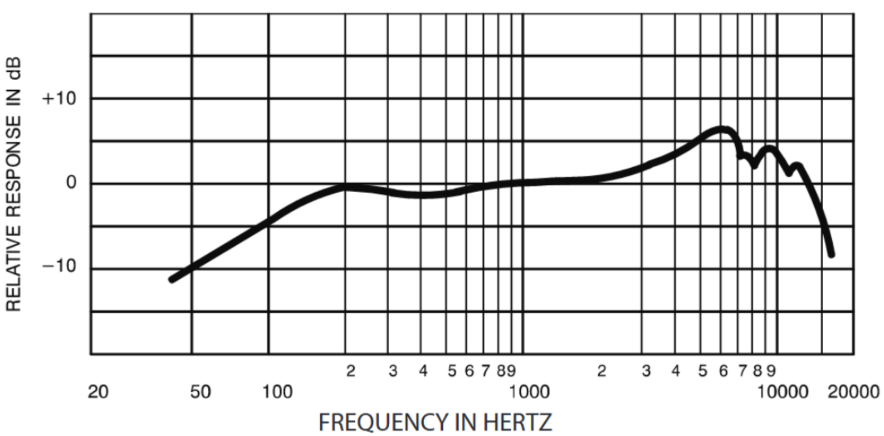
- The sensitivity of a microphone is not constant as the frequency changes.
- To describe its behavior, a Bode plot in logarithmic scale on the horizontal axis and in dB on the vertical axis is used. The reference is the sensitivity value at 1 kHz.
- The graph always hits 0 dB at 1 kHz, by construction.
- Theoretically it tends toward $-\infty$ at 0 Hz and the behavior outside the 20-20000 Hz range is indifferent, because the ear no longer hears.

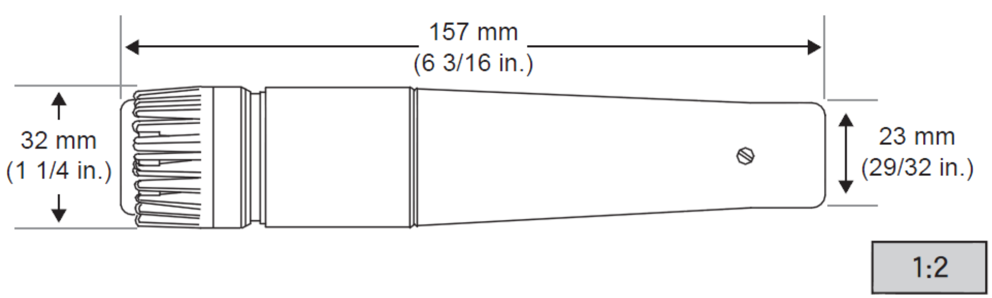


Frequency response

- Shure SM57 (dynamic, price around 110 euro):

Type	Dynamic
Frequency Response	40 to 15,000 Hz
Polar Pattern	Cardioid
Sensitivity (at 1,000 Hz Open circuit voltage)	Open circuit voltage: -56.0 dBV/Pa* (1.6 mV) *(1 Pa = 94 dB SPL)
Impedance	Rated impedance is 150 Ω (310 Ω actual) for connection to microphone inputs rated low impedance
Polarity	Positive pressure on diaphragm produces positive voltage on pin 2 with respect to pin 3
Case	Dark gray, enamel-painted, die-cast steel with a polycarbonate grille and a stainless steel screen
Connector	Three-pin professional audio connector (male XLR type)
Weight	284 g
Dimensions	157 mm L x 32 mm W

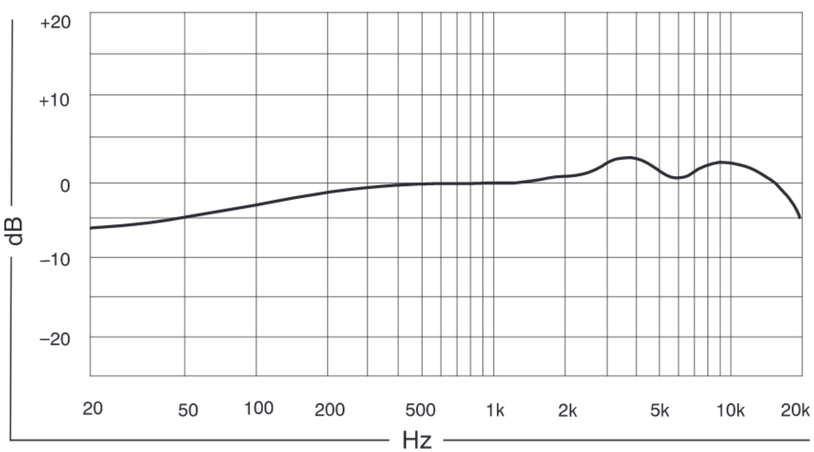


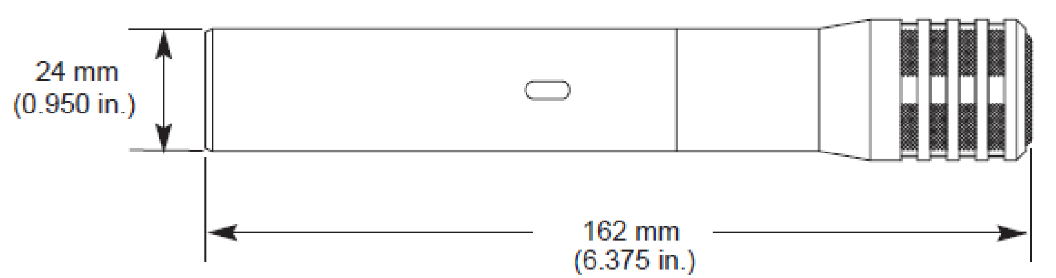


Frequency response

- Shure SM137 (condenser, price around 215 euro):

Type	Electret condenser
Frequency Response	20 to 20,000 Hz
Polar Pattern	Cardioid
Impedance	150 Ω (actual)
Sensitivity (at 1 kHz, Open circuit voltage)	–41 dBV/Pa 1 Pascal=94 dB SPL
Maximum SPL (1 kHz, < 1% THD)	1000 Ω load: 139 dB (154 dB, Pad on) 2500 Ω load: 144 dB (159 dB, Pad on)
Signal-to-Noise Ratio (referenced at 94 dB SPL at 1 kHz)	75 dB S/N ratio is difference between 94 dB SPL and equivalent SPL of self noise, A-weighted
Dynamic Range (at 1 kHz)	1000 Ω load: 122 dB 2500 Ω load: 128 dB
Common Mode Rejection (20 Hz to 20 kHz)	> 50 dB
Clipping Level (20Hz-20 kHz at 1% THD)	1000 Ω load: 3 dBV 2500 Ω load: 9 dBV THD of microphone preamplifier when applied input signal is equivalent to cartridge output at specified SPL
Self Noise (typical, equivalent SPL, A-weighted)	19 dB
Polarity	Positive pressure on diaphragm produces positive voltage on pin 2 with respect to pin 3
Weight	195 g
Switch	Attenuator: –15 dB
Dimensions	162 mm L x 24 mm
Connector	Three-pin professional audio (XLR), male
Power Requirements	48 Vdc phantom, 5.2 mA

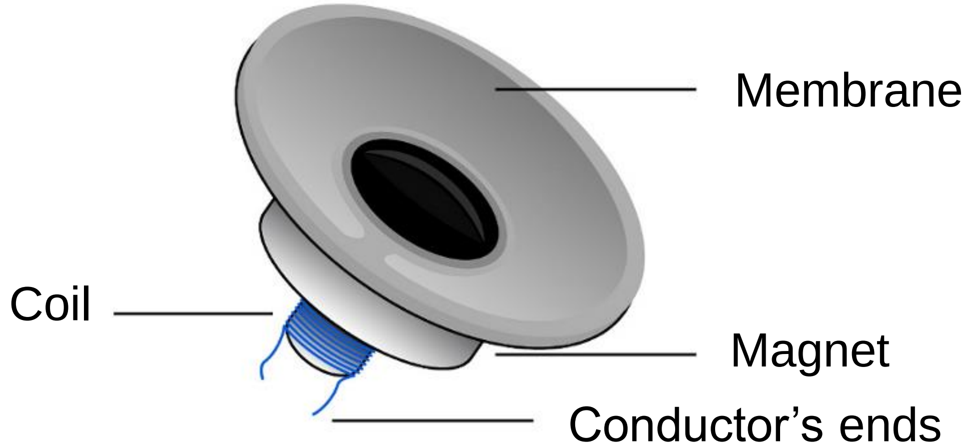




Sound diffusion systems

- Loudspeakers **transform electric signals** transporting sound information into **acoustic waves**.

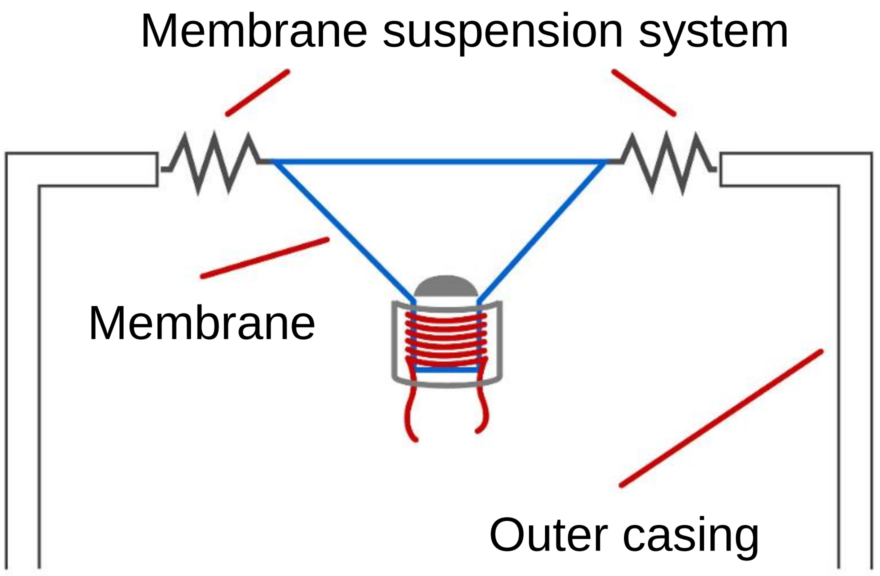
- **Functioning principals:** A coil, upon which a membrane has been set up, is placed within a circular magnet generating the acoustic wave from the electric signal applied to the coil.
- When an electric signal is applied at the far ends of a conductor, a current consisting of a flow of electrons passes through it.



Sound diffusion systems

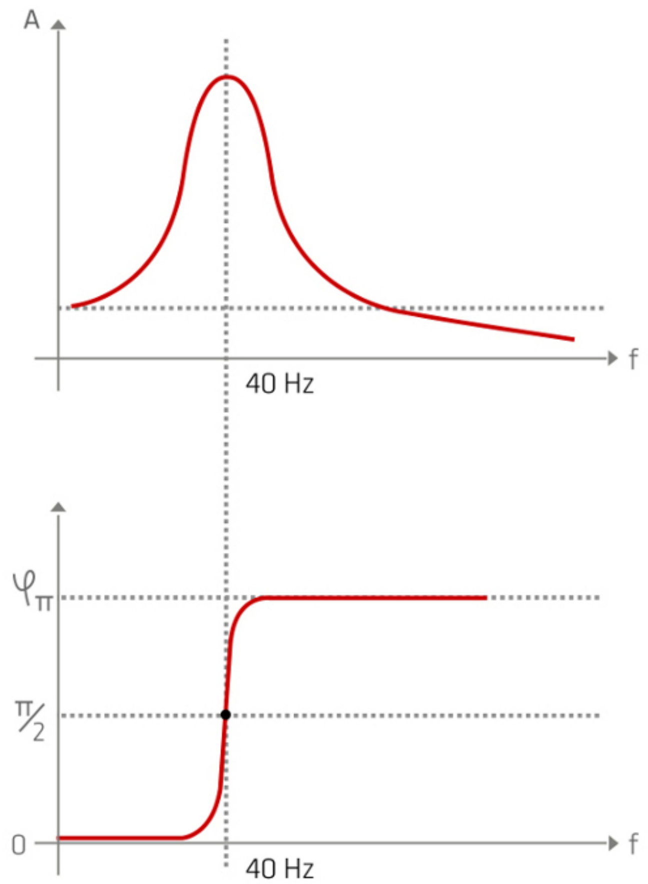
- This creates a magnetic field that can be concordant or discordant with the **magnetic field** generated by the magnet. This leads to a displacement of the coil.
- The entire coil moves up and down following the polarity applied at its far ends, in other words, depending on the applied electric signal.

- If we were to apply a **sinusoidal signal** with a certain amplitude, we'd see that the positive half wave would push the coil (and the membrane thereon) upwards, whereas during the negative half wave the coil (and the membrane) would be pushed downwards.



Resonant frequency of a loudspeaker

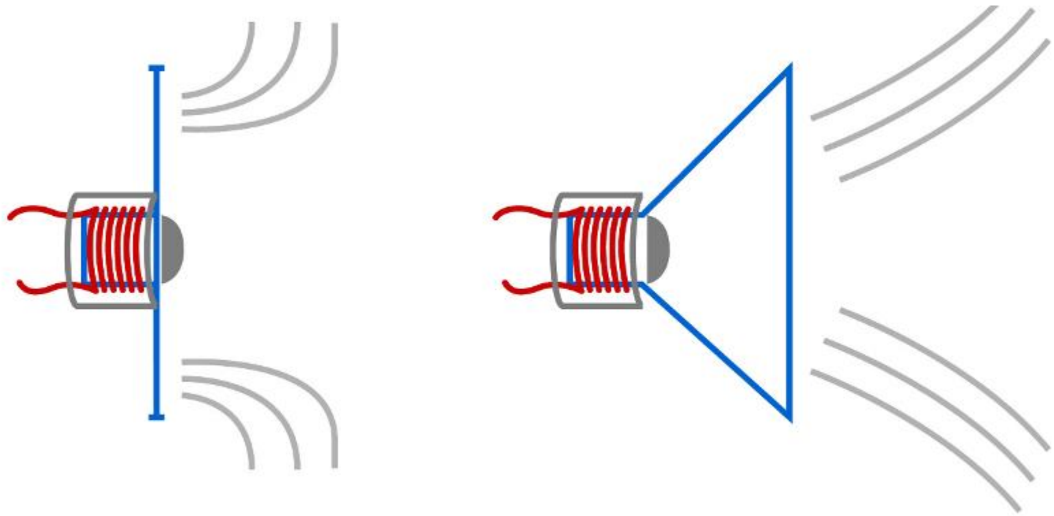
- An elastic system will begin to oscillate when the frequency of the stimulus **approximates the system's resonant frequency**.
- Example: Loudspeaker with resonant frequency of 40Hz.
- The diagrams plots the oscillation amplitude and phase versus frequency.



- 180 degrees phase displacement: undesirable situation because it alters the audio signal.
- No phase diagram of a real loudspeaker could have such a rate.

Efficiency of a loudspeaker

- This is the **effective measurement for the acoustic power** of a loudspeaker, i.e., its ability to transform electric energy into acoustic energy.
- The part of electric energy which doesn't get converted into acoustic energy is dissipated by the loudspeaker **as heat**.
- The **efficiency** of a loudspeaker is usually **very low**, in the order of 1-2% to a maximum of 8%.
- At the **lower frequencies**, cone shaped membranes are better than flat ones at gathering and moving the air in front of them.



Flat membrane Conic membrane

Sensitivity and Maximum applicable power

- Sensitivity is measured in dB_{spl}. It quantifies the intensity of sound pressure at one meter distance from the loudspeaker when an electric signal of 1 Watt is applied to it.
- For example, 93 dB_{spl}/m/W indicates that in the aforementioned conditions a sound pressure of 93 dB_{spl} was measured.

- Maximum applicable power is measured by two quantities.
  - **Peak power** is the maximum power the loudspeaker can handle without being damaged. If even once the signal goes beyond this value the loudspeaker will be damaged.
  - **RMS power** (Root Mean Square) is a unit that measures the average applicable power that can be applied for an amount of time before the heat begins to damage the loudspeaker.

Loudspeaker types

- The dimension of the membrane strongly conditions the functioning of a loudspeaker.
- The greater the dimension of the membrane and its mass, the smaller its resonance frequency (ideal for reproducing low frequencies).
- **Woofer:** kinds of loudspeakers used for the reproduction of low frequencies. The greater the quantity of air moved by it, the more power is needed to adequately feed the loudspeaker.
- **Subwoofer:** loudspeakers created for reproduction of very low frequencies (20-40 Hz).
- **Midrange:** loudspeakers used for the reproduction of middle frequencies (smaller than woofers).
- **Tweeter:** loudspeaker with very small membranes, for high frequencies.

Digital audio

- The **transformation** of the signal from analogue (continuous) to digital (discreet) is called ***sampling***.
- The etymology of this word comes from the fact that samples are picked up from the original signal, undergoing this process **at regular time intervals** *(**sampling frequency**).*
- **Nyquist**'s theorem states that, when sampling is carried out at **a frequency equal to at least double the band of the signal** being sampled, the transition from analogue to digital takes place without any loss of information.
- **Unfortunately**, in the chain of operations carried out to recover the analogue signal starting from stored samples, we'll nevertheless have some loss of information if we compare it to the original signal. **(quantization)**
- **Aliasing frequency**

Sampling frequency

- As previously mentioned, the sampling frequency must comply with Nyquist's theorem (at least twice the band of the signal to be sampled), to avoid the occurrence of aliasing frequencies.

- The lower the sampling rate than the signal band, the more new aliasing frequencies emerge.
- Example: guitar arpeggio progressively undersampled, with decreasing sampling frequency:

- Undersampled 440Hz sinusoid:

Quantization

- Sampling a signal at frequency f means extracting f samples per second from the signal.
- In the case of an audio signal sampled with CD standard, 44100 samples are extracted each second. Every sample is represented by a binary 16 bit number. So, a stereo signal produces the following number of samples per second:

  2 (stereo) x 16 (bit) x 44100 (samples) = 1411200 bit/s

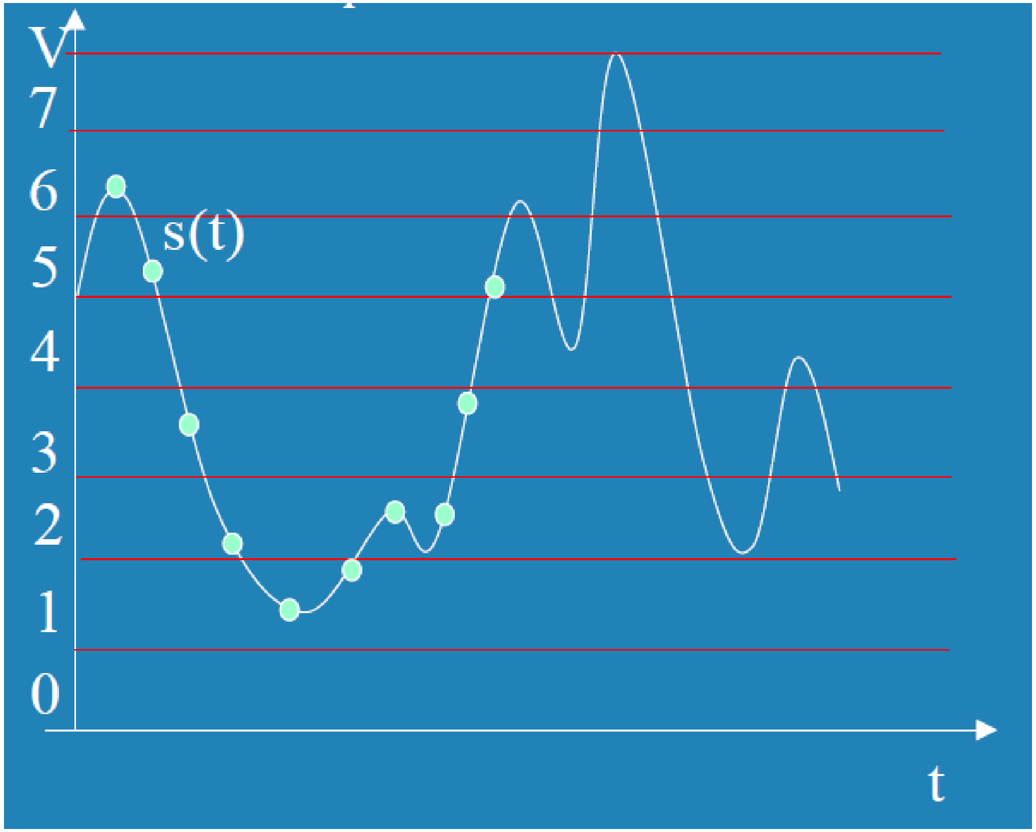
- if we wanted to express this result in bytes we'd have:

  1411200/8 = 176400bytes = 172.26Kb

- A normal CD can record 74 minutes, therefore it has:

  172.26 (Kb/s) x 60 (sec) x 74 min = 764Mb

Quantization



Each level can be represented with a sequence of 3 bits.

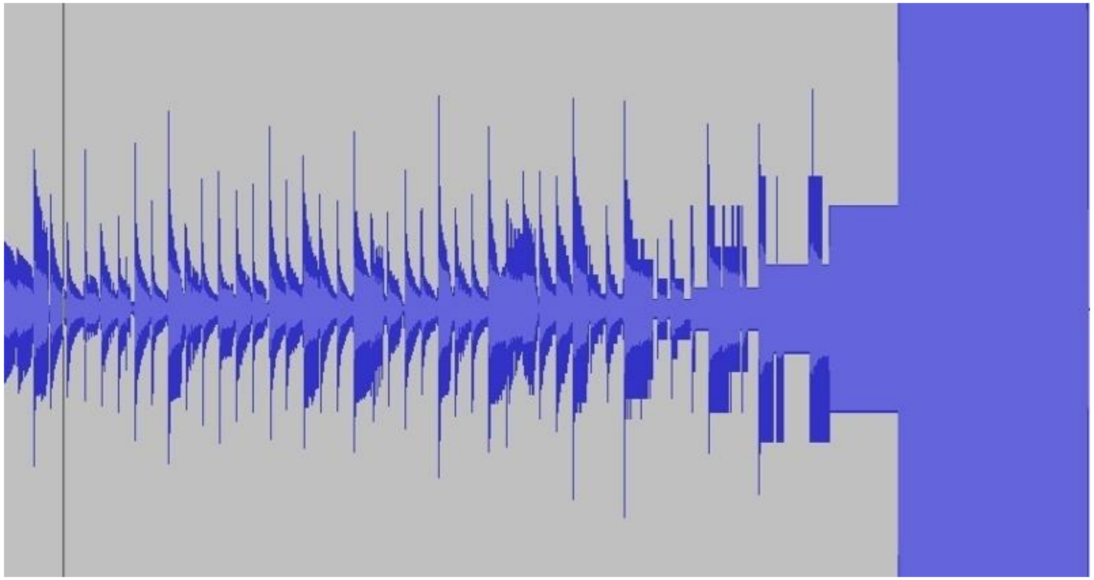
The sequence of quantized samples: 6 5 3 2 1 1 2 2 3 5 becomes

6	5	3	2	1	1	2	2	3	5
110	101	011	010	001	001	010	010	011	101

Quantization bit

- Our samples are quantized using 16 bits. Let's listen to what happens if the quantization bits are progressively decreased, up to 1 (only two possible levels):

- The following figure shows the waveform as it is progressively under quantized.



- Under quantized 440Hz sinusoid:

Audio media parameters

Here is a table with the indicative quality parameters specific to some audio media. Note in particular the case of the telephone, whose bandwidth is just sufficient to transmit the voice with reasonable intelligibility, but completely inadequate for the transmission of music. In fact, the voice remains intelligible, even if distorted, if the region of the spectrum in which the formants fall, which is within 5 kHz, is preserved.

medium	sampling frequency (kHz)	bit per sample (bit)	speed of information transmission (kB/s)	information size in 1 minute of music
telephone	8	8 (mono)	8	480 kB
AM radio	11.025	8 (mono)	11	660 kB
FM radio	22.050	16 (stereo)	88.2	5.3 MB
CD	44.1	16 (stereo)	176	10.6 MB