EE 6424

Digital Audio Signal Processing

Bi Guoan

Email:	egbi@ntu.edu.sg
Office:	S1-B1a-27
Tel:	6790 4823

Welcome to EE6424 and very happy to see you again!

TEXT BOOKS

1. M. Bosi and R. Goldberg, *Introduction to Digital Audio Coding and Standards*, Kluwer Academic Publishers, 2003
2. K. C. Pohlmann, “*Principles of Digital Audio*,” McGraw-Hill, 6th Ed., 2010.
3. A. Spanias et al., *Audio Signal Processing and Coding,* Wiley , 2007.

REFERENCE BOOKS

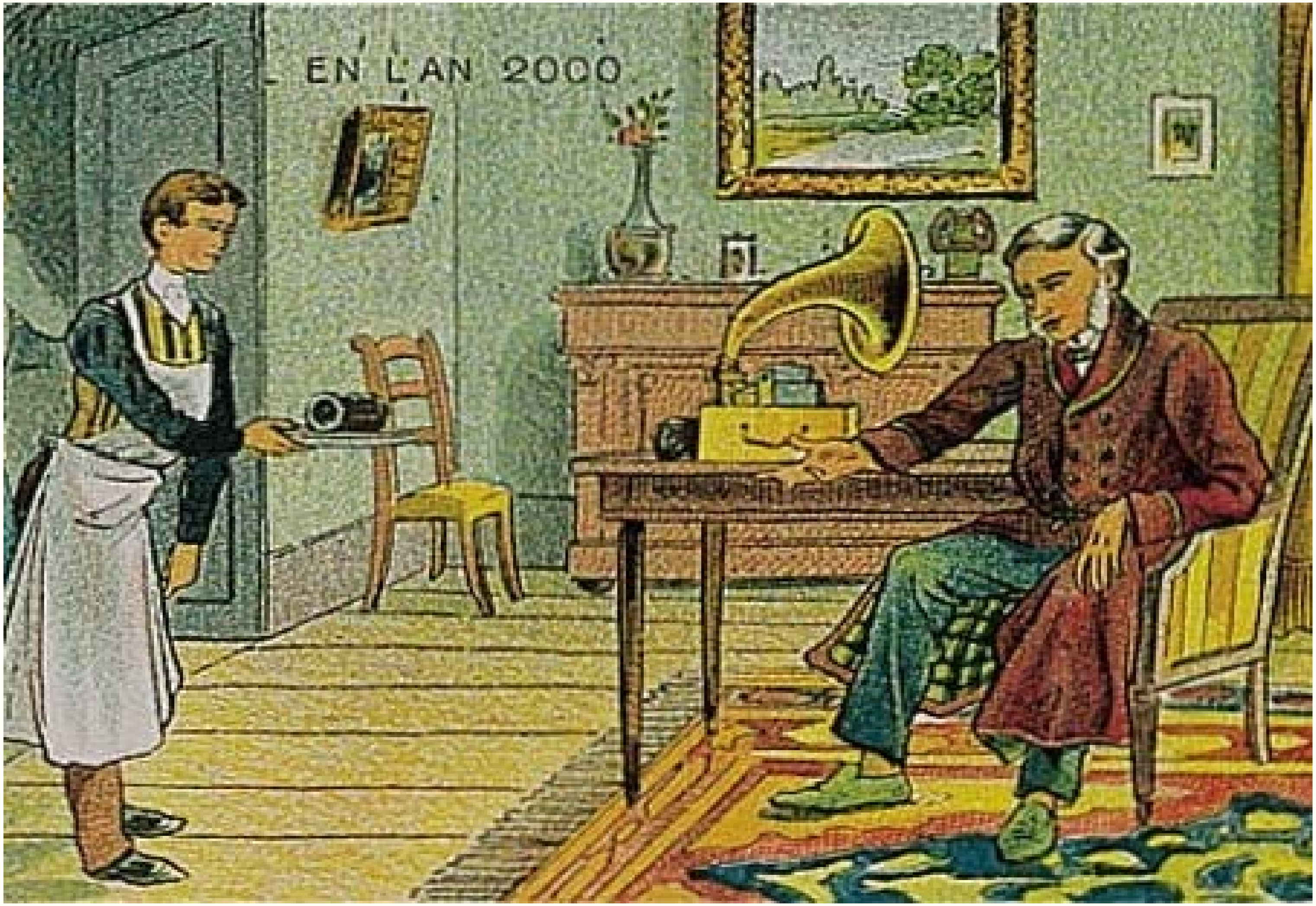
1. B. C. J. Moore, “*An Introduction to the Psychology of Hearing*,” Academic Press, 4th Edition, 1997.

Better Understanding

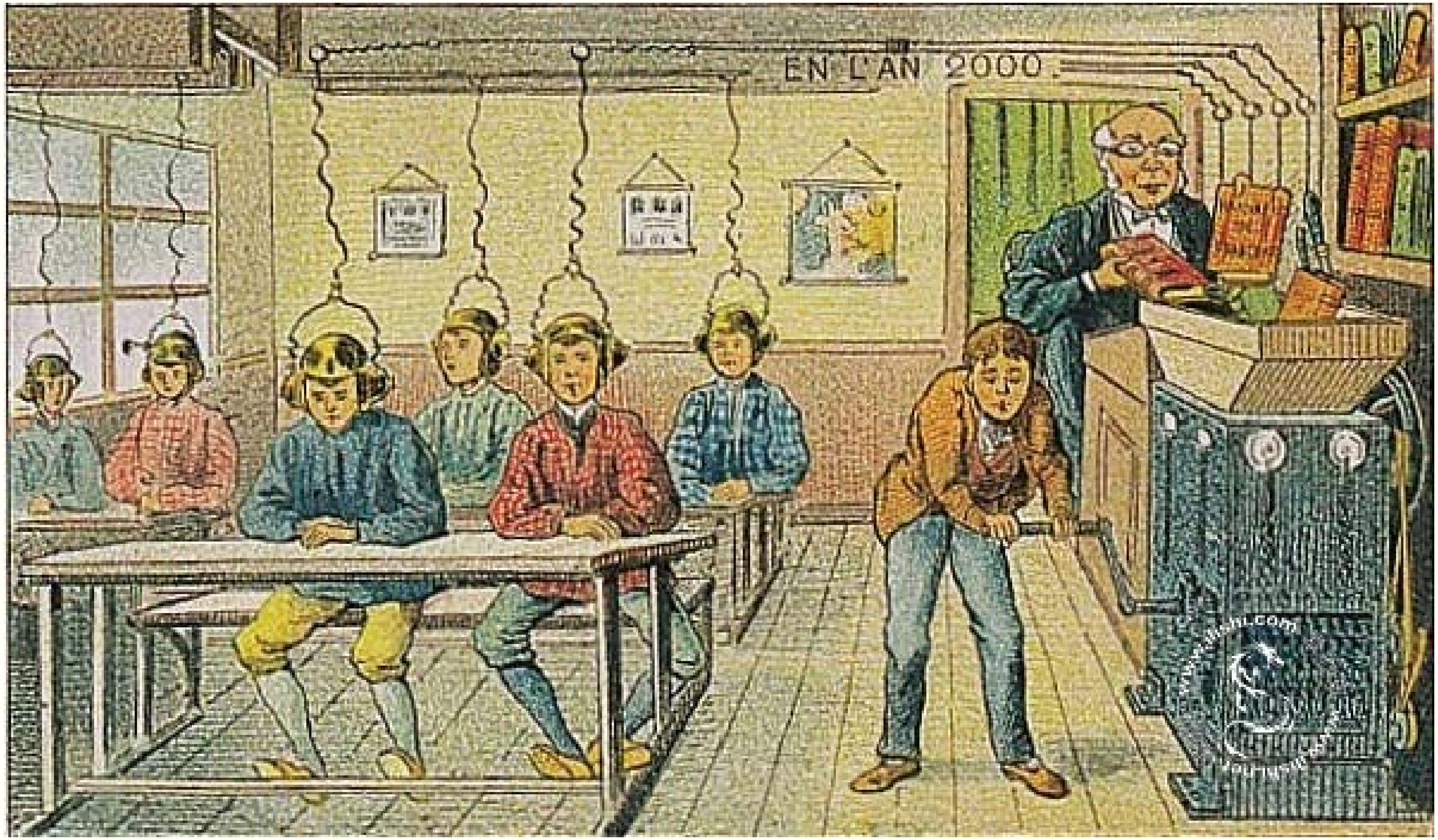
- This subject has involved too many innovations accumulated during the last thirty years in various scientific research fields, such as
  - bio-engineering for modelling the sound generation and receiving process of human beings – *understanding regeneration and reception*
  - information presentation and processing – *needing extensive computation*
  - coding for storage and communications – *leading to compression*
- By extensive signal analysis and processing, the amount of data for storage or transmission is significantly reduced without sacrificing the perception quality (signal band width limited applications)
- Due to limited time, we can touch only the basic concepts of audio signals, audio coding and compression
- Because students may not have enough DSP background, we will provide introduction to the basic DSP concepts used in this course without providing details of these operations.
- For better understanding of the subject, the following learning practices are necessary.
  - Attending lectures
  - Do revision regularly to understand the concepts TIMELY
  - The best and most direct way is to ask the lecturer
  - Do tutorial questions TIMELY

Lecture Notes Outlines

- Basics of Sound and Hearing
- DSP Basics
  - Sampling and quantization
  - Spectrum analysis
- Spectrum Analysis
  - Subband coding
  - Transform coding
- Audio Coding
  - MPEG 1



- The voice mail service at the year of 2000 predicted in 1900
- We have achieved much more than this prediction



The learning process predicted at 1900

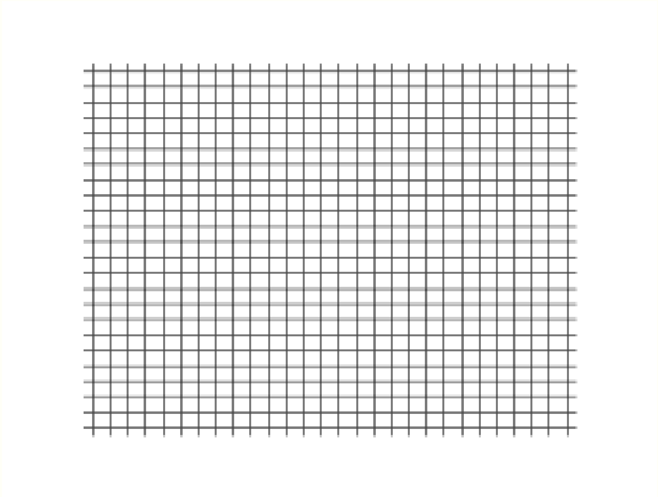
Basics of Sound and Hearing

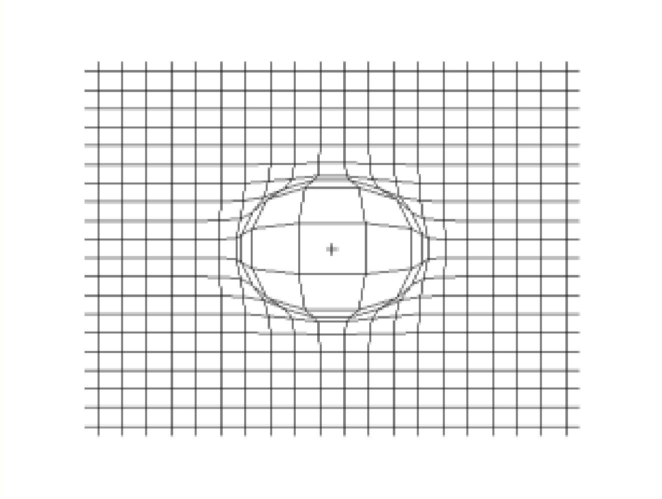
Sound Generation

- Sound waves are generated by vibrations.
- There are two main kinds of sound generators (thus, two main kinds of musical instruments).
  - Vibrating or oscillating piston: Examples include the soundboard of a piano, the surfaces of drums and cymbals, the diaphragm of loudspeakers, etc.
  - Valve quickly open and close (or vibrate): Examples include trumpets, sirens, organs, saxophones, and trombones.

Sound Propagation

- The sound energy travels through a medium (air, liquid or solid) as a fluctuation in pressure





- When the air pressure reaches the ear, humans' brain *perceives it as sound*.

Properties of Sound

- The speed of sound in air, in meters per second, is given by

$$c = \sqrt{\frac{\gamma R T}{M}}$$

where	$\gamma \approx 1.4$	adiabatic constant,
	$R = 8.3145 \text{ J mol}^{-1} \text{ K}^{-1}$	Molar gas constant
	T	temperature (Kelvin)
	$M = 0.0289645$	Molar mass (kg per mol)

- Speed of sound in air is independent of the frequency and is approx. 340 meters/second at 14 ^{o}C .
- The speed of sound is proportional to the square-root of temperature.

- Frequency, in Hz, is defined by the *rate of vibration* of the sound source. If the source is vibrating at 100 Hz, then it will vibrate once per one hundredth of a second.
- T is known as the *period of the wave*, i.e.,

$$T = \frac{1}{f} = \frac{1}{100} = 0.01 \text{ s},$$

where *T* is in second and *f* is the frequency of the wave in Hz.

- The wavelength, λ, defined in meters, of the sound is then related to the speed of the sound by

$$c = f\lambda.$$

Inverse square law

- The intensity of a sound source is defined as the *sound power per unit area* given by

$$I = P/A$$

where *P* is the sound power (Watts) and *A* is the area of sphere (m^2)

- Defining *r* as the radius of a sphere, the area is given by

$$A = 4\pi r^2,$$

and the intensity of the spherical wave varies *inversely* with the square radius, i.e.,

$$I \propto \frac{1}{r^2}.$$

- The above relationship is known as the *inverse square law* for acoustic propagation.

Example: Find the drop in acoustic intensity when the distance to the acoustic source is increased by twice the distance.

- Let I_1 be the intensity of the sound at distance r_1 . Since we have $r_2 = 2r_1 \quad \Rightarrow \quad r_2^2 = 4r_1^2,$

$$\frac{I_2}{I_1} = \frac{r_1^2}{r_2^2} = 0.25.$$

- Taking logarithm on both sides and rearranging, we have

$$\begin{aligned} I_2 \text{ (dB)} &= I_1 \text{ (dB)} + 10\log_{10}(0.25) \\ &= I_1 \text{ (dB)} - 6.02 \text{ dB.} \end{aligned}$$

- Therefore, there is a drop of 6.02 dB in intensity for every twice the distance one moves away from the source.

Pure tones

- The simplest sound is the sine wave having a line spectrum consisting of only one frequency.
- which is known as a *pure tone*.

Square wave

- Multi-tone waveforms contain *multiple frequency components*, i.e., they are frequency-rich.
- An example is the square wave which can be expressed in terms of odd harmonics (Fourier series) given by:

$$
\begin{aligned}
s(t) &= \sum_{n\in Y}^{N} \frac{\sin(2n\pi ft)}{n}, \qquad Y \in \mathbb{Z}^{+} : \mathrm{mod}(Y,2) = 1 \\
&= \sin(2\pi ft) + \frac{\sin(2\times\pi 3ft)}{3} + \frac{\sin(2\times\pi 5ft)}{5} + \ldots + \frac{\sin(2\times\pi Nft)}{N}.
\end{aligned}
$$

Demo: Observe and listen

Triangular wave

- Triangular (sawtooth) waveforms are another type of multi-tone wave which can be expressed as

$$
\begin{aligned}
s(t) &= -\frac{2}{\pi}\sum_{n=1}^{N}\frac{\sin(2\pi n f t)}{n} \\
&= -\frac{2}{\pi}\left[\sin(2\pi f t) + \frac{\sin(2\pi \times 2ft)}{2} + \frac{\sin(2\pi \times 3ft)}{3} + \ldots + \frac{\sin(2\pi \times Nft)}{N}\right].
\end{aligned}
$$

Random noise

- Not possible to predict the future value from its past samples.
- A random noise is made up of a *random combination* of an *infinite number* of sine wave components.
- White noise, in theory, has all frequencies from zero to infinity with equal energy (wide-band), therefore does not contain any information
- Often used to analyze the behavior of the ear and to quantify the performance of algorithms.

Decibels

- The pressure of sound, with a unit of *Pascal (Pa)*, is given by

$$p = F/A$$

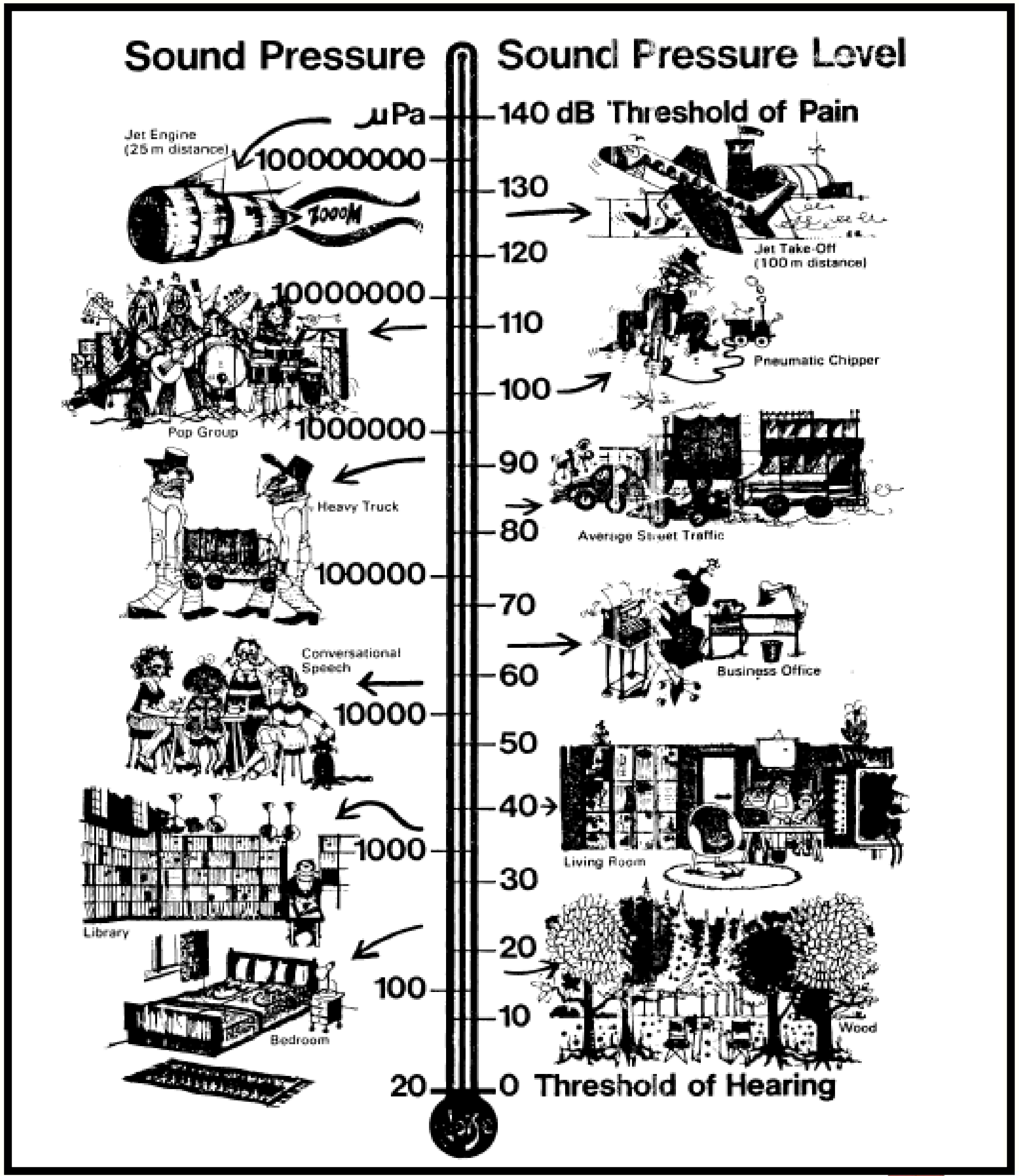
where *F* is force (Newton) and *A* is the area (*m*^{2}).

- The range of pressure to which the ear responds is between 20μPa (1μ=10^{-6}) to 120 Pa. Because of this wide range, a logarithmic unit in decibel has been introduced.
- The decibel unit is a unit of *relative level*. Hence, **sound pressure level (SPL)** is defined in relation to a reference level, normally w.r.t 20 μPa, i.e.,

$$\mathrm{dB} = 20\log_{10}\frac{p}{20\ \mu\mathrm{Pa}}.$$

Example:

- SPL is a ratio between the absolute sound pressure and the lowest intensity sound that can be heard by most people.
- SPL is measured in decibels (dB), because of the broad range of intensities we can hear.
- SPL can be measured with a Sound Pressure Level Meter, available from some electronics stores for fairly reasonable prices.
- Figure shows the events with different SPLs.



Loudness of tones – The response of listeners

- A magnitude estimation experiment for loudness presents listeners with a series of stimuli having same relative spectrum but different intensities.
- The listener then assigns numbers on any scale he chooses.
- These numerical estimates are proportional to the power of the sound and can be regarded as a measure of psychological magnitude given by

$$\psi = kI^p,$$

where *I* is intensity (W/m^2) and *k* and *p* are constants

- Taking the logarithm of both sides, we arrive at

$$\log \psi = \log k + p \log I.$$

- We note that the sound intensity level can be expressed as

$$L = 10\log\left(\frac{I}{I_{\text{ref}}}\right),$$

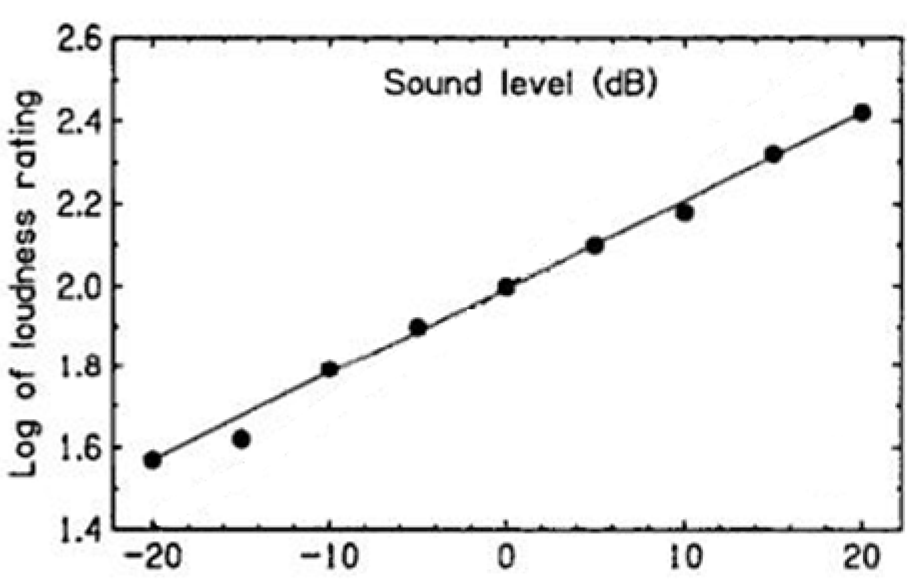
where $I_{\text{ref}} = 1 \times 10^{-12}\ \text{Wm}^{-2}$.

- Consequently, the relationship between loudness estimate *ψ* and sound intensity level *L* is given by

$$\log\psi = \log\left(kI_{\text{ref}}^{p}\right) + (p/10)L.$$

- This equation states that
  - the log of the loudness estimate is a linear function of the sound level *L* in dB, and
  - the slope of the line is given by *p*/10 and is a factor of the exponent in the power law.

- Figure below illustrates magnitude estimates obtained by using
  - 64 subjects and
  - 9 different broadband noise levels over a range of 40 dB sound level.



  - From this result (solid line), we see that the slope is approximately 0.022 giving a value of $p = 0.22$.

Example: Find the sound intensity level change required such that the loudness is doubled.

Solution: If loudness is doubled, we have the following relationship:

$$\frac{\psi_2}{\psi_1} = \left(\frac{I_2}{I_1}\right)^p = 2.$$

Taking logarithm of both sides we obtain

$$\begin{aligned} \log 2 &= p \log\left(\frac{I_2}{I_1}\right) \\ 0.3 &= \frac{p}{10}\triangle L \\ \triangle L &= 3/p. \end{aligned}$$

Hence, if *p*=1/3 then we need to increase sound intensity by 9 dB for doubling the loudness.

Sone scale

- The Sone scale is the basis for the current international loudness scale.
- The reference for this absolute scale is that a 1 kHz sine tone with a level of 40 dB SPL will correspond to one **Sone**.
- Hence, one Sone is equivalent to 40 **phons**.
- The definition of Sone is given by

$$\begin{aligned}\psi\ (\text{sones}) &= \frac{1}{15.849}\left(\frac{I}{I_{\text{ref}}}\right)^{0.3}\\ \log\psi\ (\text{sones}) &= -\log(15.849) + 0.03\times 10\log\left(\frac{I}{I_{\text{ref}}}\right).\end{aligned}$$

- The relationship between Sone and phons can be derived via:

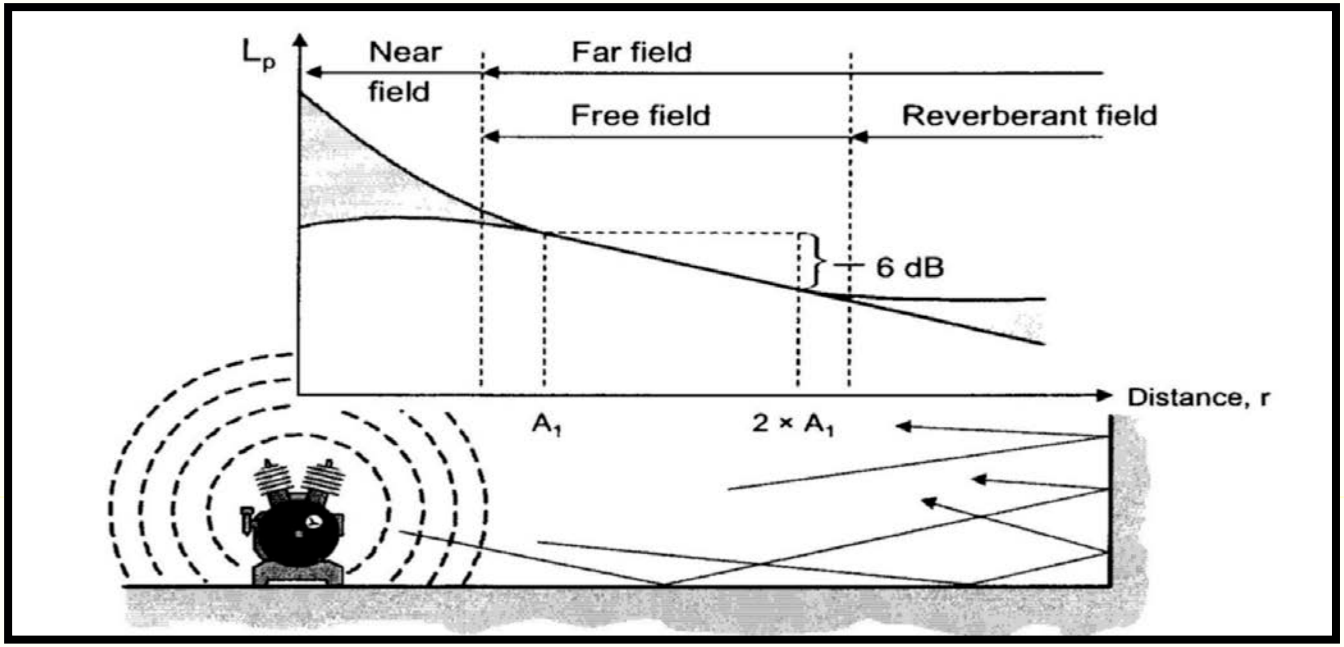
$$\log \psi \ (\text{sones}) = -1.2 + 0.03 L_{\phi}$$

where L_{ϕ} is the loudness level in phons.

sone	**1**	**2**	**4**	**8**	**16**	**32**	**64**	**128**	**256**	**512**	**1024**
phon	40	50	60	70	80	90	100	110	120	130	140

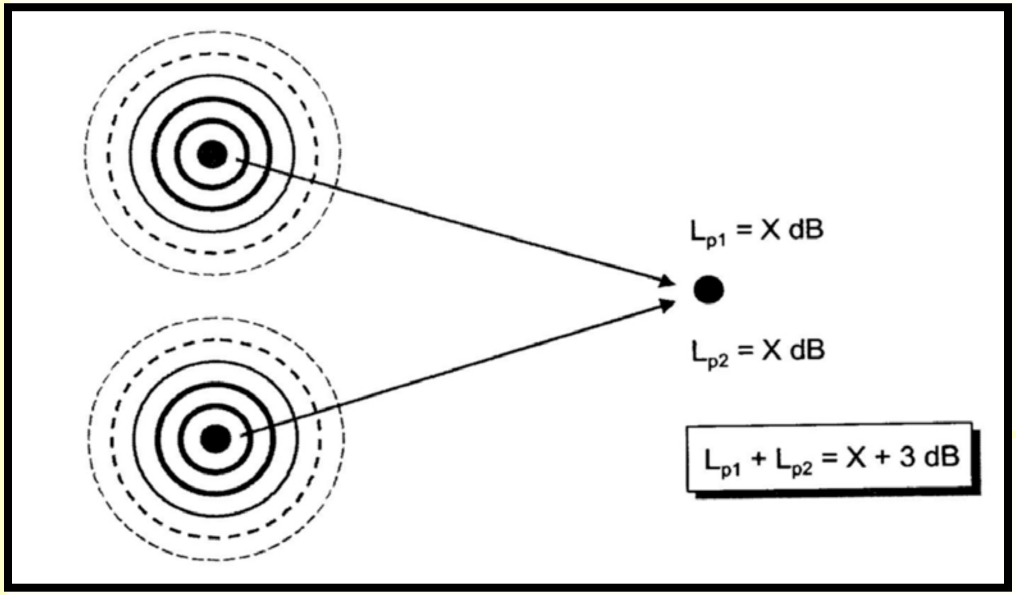
Sound Fields

- **Near field** : very close to the sound source, usually being away less than the wavelength of the lowest frequency emitted from source. Avoid sound pressure measurement in this area.
- **Far field** : divide into free field and reverberant field. In free field, sound drops 6 dB for a doubling in distance from source.
- In reverberant field, reflection from wall may be just as strong as direct sound.



Adding Sound Sources

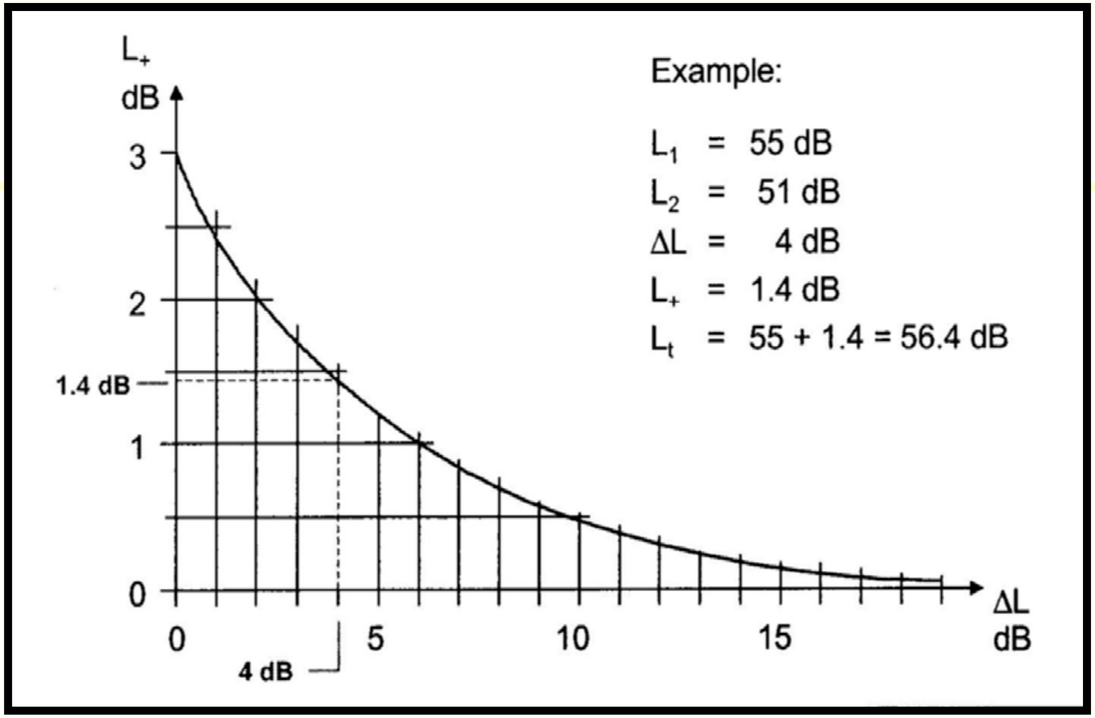
- o If 2 sound sources radiate the same amount of energy and a measuring device is position equidistant from both sources, sound intensity is twice as high as only one source.
- o This is equivalent to increase in sound pressure of √2 or 3dB since intensity is proportional to the square of pressure.
- o What is the total sound pressure level when both sources operate in 50 dB ?
- o The total energy of N similar sound sources is L+10log$_{10}$N



Assume sound sources are incoherent so that they do not have fixed relative phase.

What happen when 2 sound sources are different in intensity ?

- Use the curve to derive the additional dB SPL value.
- What happen when the $\Delta L = 0$?
- What happen when $\Delta L > 10\text{dB}$?



Example:

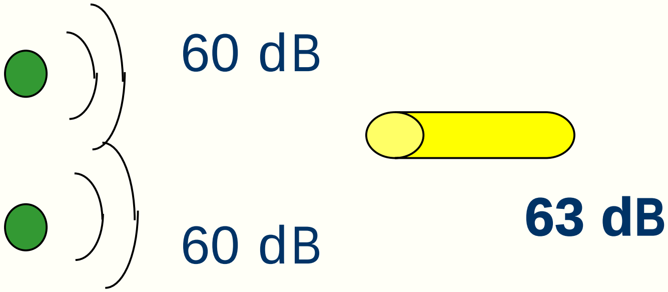
L_1	=	55 dB
L_2	=	51 dB
ΔL	=	4 dB
L_+	=	1.4 dB
L_t	=	55 + 1.4 = 56.4 dB

Add and subtract sound levels

- **Adding sound levels**

  With more sources being measured separately, the combined sound level is:

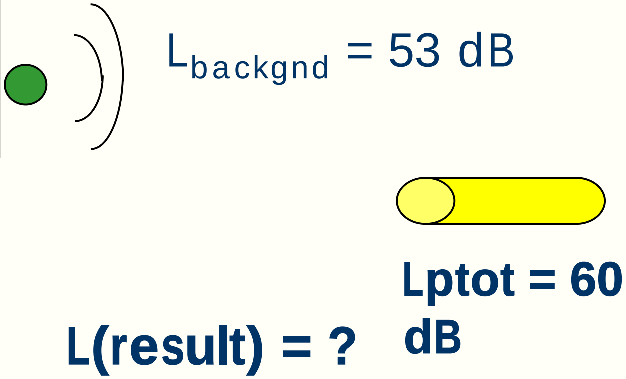
$$\boxed{L(tot) = 10 * \log(10^{\frac{Lp1}{10}} + 10^{\frac{Lp2}{10}} + 10^{\frac{Lp3}{10}} + ...)}$$



- **Subtracting sound levels**

  Subtract background noise from the total sound level:

$$\boxed{L(result) = 10 * \log(10^{\frac{Lptot}{10}} - 10^{\frac{Lpbackground}{10}})}$$

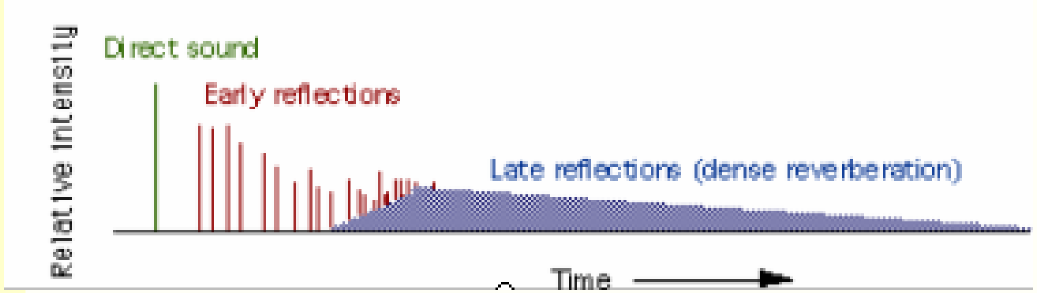


- If ΔL < 3 dB, background noise is too high and correct sound level cannot be found
- If ΔL > 10 dB, ignore background noise.

Sound Wave in Room

Room Acoustics : important consideration in concert hall, cinema, living room.

- **(a) Direct sound:** sound radiated in a straight line between the source and the receiver. It follows the inverse square law.
- **(b-c) Discrete reflection:** sound wave hit off the surface or objects in the room. Can be joined by 2nd-order reflection and so on. Exponential decay which is densed within 100-200 msec of the 1st direct wave
- **(d) Reverberation:** Occurs after enough discrete reflections, that are not distinct. No apparent direction



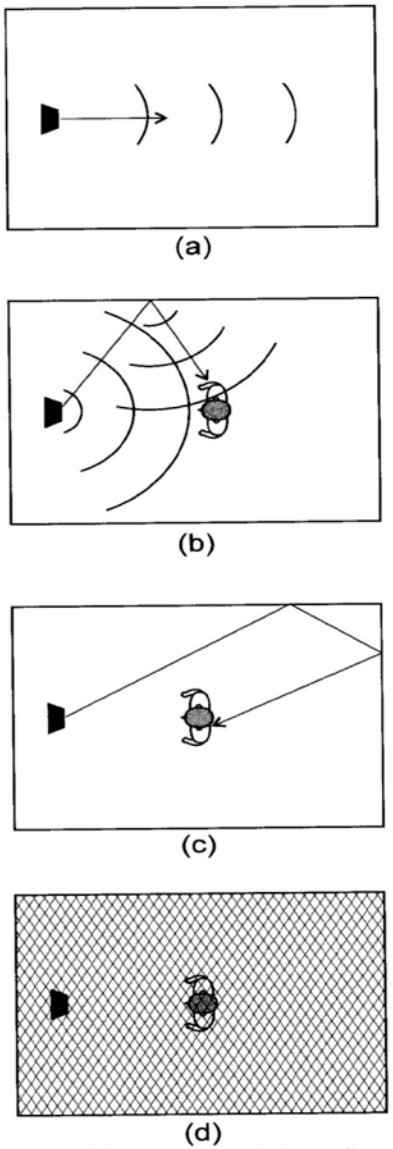
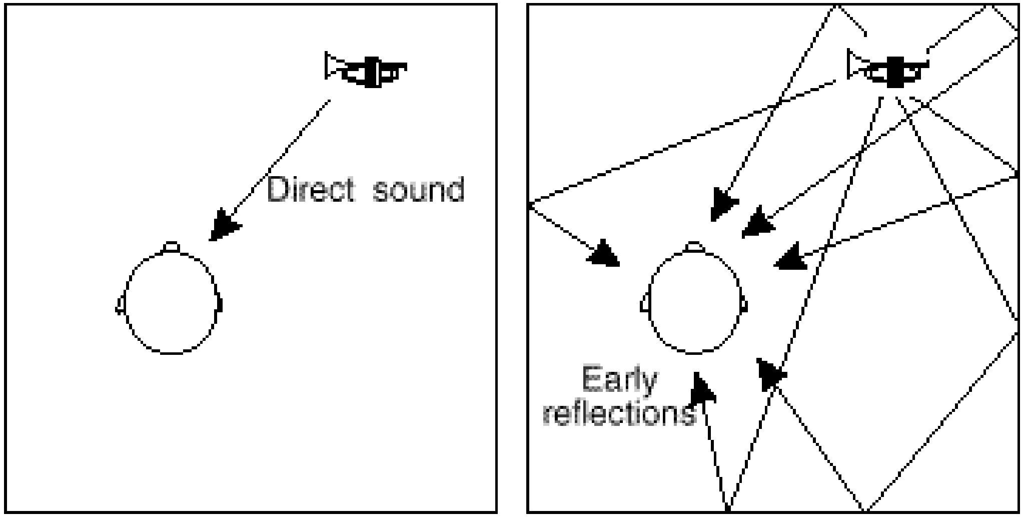


Fig. 1.16 Sound fields in a room: (a) shows direct sound, (b) 1st-order reflected sound, (c) higher-order reflected sound, and (d) a reverberant field where there are so many reflections per unit of time that each becomes indistinguishable from its peers.

Sound in rooms

- When a sound is confined in a room, surrounding walls can cause sounds to be reflected and/or absorbed.
- Each reflected sound behaves as a new sound source. This new source is known as the image of the original source and further images are formed from each reflection.
- Due to these multiple reflections, a sound field at a particular point in the room will be produced by superposition.



- For a rectangular room of dimensions *L* x *W* x *H* (*m*) the resonance frequencies are given by

$$f = \frac{c}{2}\sqrt{\left(\frac{p}{L}\right)^2 + \left(\frac{q}{W}\right)^2 + \left(\frac{r}{H}\right)^2}$$

  where *c* is the speed of sound and *p*, *q* and *r* are integers, determine the mode of resonance.

- **Example:** Assume *c* = 340 ms^{-1}, a room of length supports resonance frequencies along the length given by *L*=6.8 m. With *q* and *r* being 0

$$\begin{aligned}
f_1 &= \frac{340}{2}\sqrt{\frac{1^2}{6.8^2}} = 25 \text{ Hz},\\
f_2 &= \frac{340}{2}\sqrt{\frac{2^2}{6.8^2}} = 50 \text{ Hz},\\
f_3 &= \frac{340}{2}\sqrt{\frac{3^2}{6.8^2}} = 75 \text{ Hz},\\
f_4 &= \frac{340}{2}\sqrt{\frac{4^2}{6.8^2}} = 100 \text{ Hz}, \qquad \text{etc.}
\end{aligned}$$

- As the resonance frequencies increases, the number of modes supported by the room increases, i.e., we have larger values of *p*, *q* and *r*.

- Since the resonance frequency is inversely proportional to the size of the room, one needs to have a good loudspeaker and an adequately large room in order to achieve good reproduction of bass frequencies.

Acoustic Reverberation

- For a mid-frequency sound wave, the room acoustic effect can be viewed as a linear time-invariant sum of attenuated, filtered and delayed original signal.
- Room audio character is determined by the geometry of room, position of source and receiver, absorption properties of boundaries.
- When wave front hit a boundary, a part of the energy get reflected, and other part is absorbed.
- The absorption coefficient states the fraction of the energy that is being absorbed and is also frequency dependent.

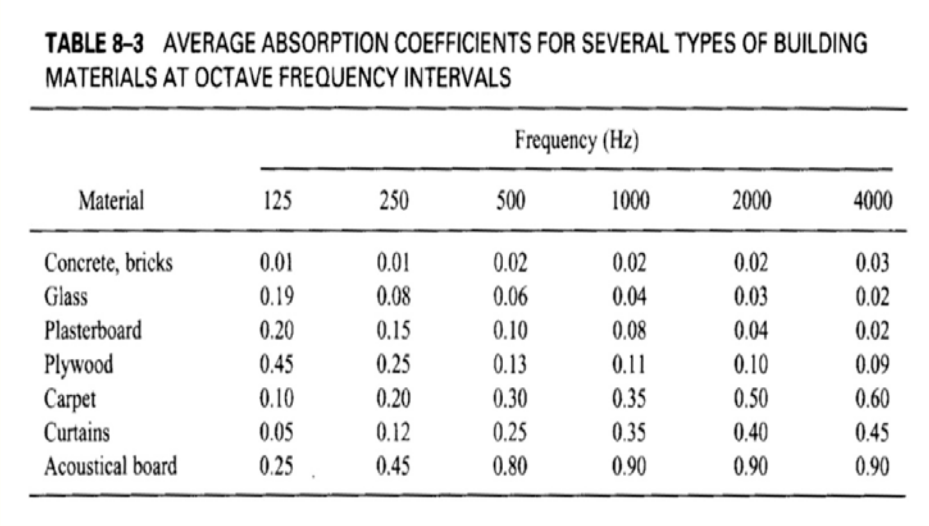


TABLE 8-3 AVERAGE ABSORPTION COEFFICIENTS FOR SEVERAL TYPES OF BUILDING MATERIALS AT OCTAVE FREQUENCY INTERVALS

	Frequency (Hz)					
Material	125	250	500	1000	2000	4000
Concrete, bricks	0.01	0.01	0.02	0.02	0.02	0.03
Glass	0.19	0.08	0.06	0.04	0.03	0.02
Plasterboard	0.20	0.15	0.10	0.08	0.04	0.02
Plywood	0.45	0.25	0.13	0.11	0.10	0.09
Carpet	0.10	0.20	0.30	0.35	0.50	0.60
Curtains	0.05	0.12	0.25	0.35	0.40	0.45
Acoustical board	0.25	0.45	0.80	0.90	0.90	0.90

(1) Glass, plywood and plasterboard are more absorptive at low frequency.

(2) Curtain, carpet and acoustical board are less absorptive at low frequency.

(3) Concrete has little change of absorption at all frequency. Generally, Low absorption.

Reverberation Time

- The reverberation time is normally characterized by the time it takes a steady-state noise to decay by 60 dB (which is commonly known as RT60).
- There are many formulae for RT60. A simple approximation of RT60 of an enclosed room was proposed by Sabine:

$$RT_{60} = \frac{0.16 \times V}{A},$$

where V (m^3) is the volume of the room and A(m^2) is the total room absorption in Sabins

$$\begin{aligned} A &= \sum_{n=1}^{N} S_n \alpha_n \\ &= S_1\alpha_1 + S_2\alpha_2 + \ldots + S_N\alpha_N \end{aligned}$$

where S_n is the area of the surface (m^2) and α_n is its absorption coefficient.

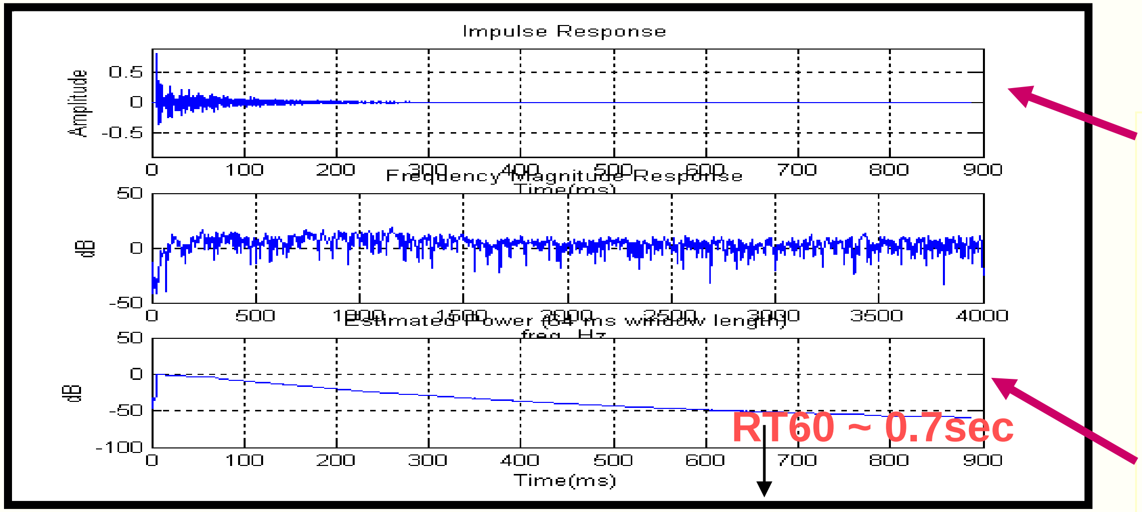
Sabine’s Formula

- A drop of 60 dB corresponds to how big a change in pressure amplitude? intensity?
- What does a long RT60 mean?
- Frequency dependent : higher frequencies have faster decay rate.

ritical Distance:

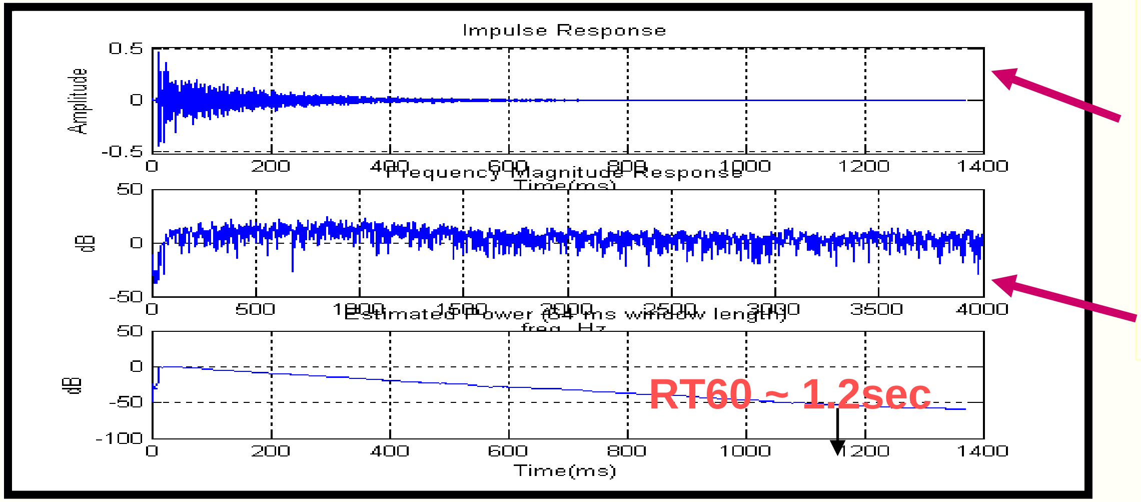
- Further away from source, direct sound level drops
- In contrast, reverberant sound level stays constant everywhere in the room

Sound demonstration on reverb



Original signal

Small room

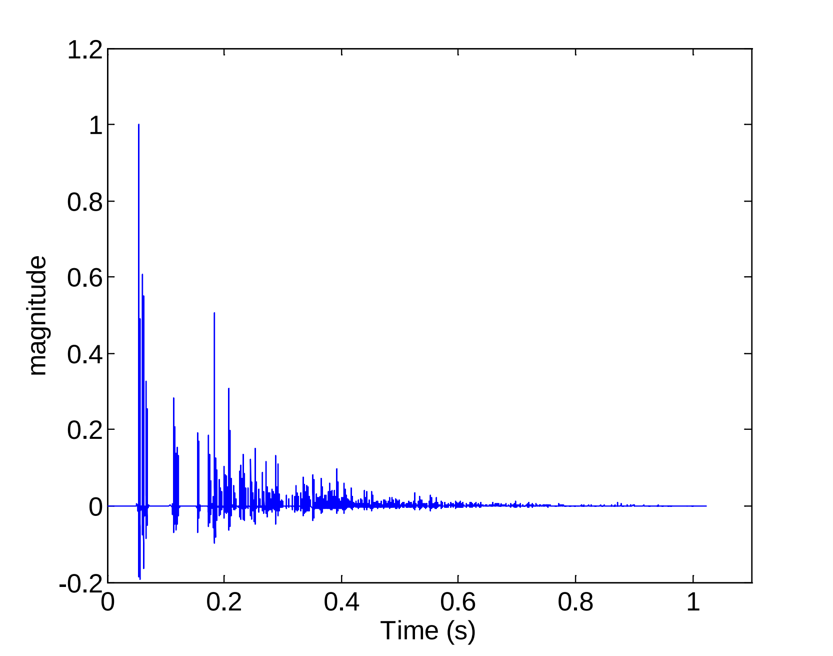


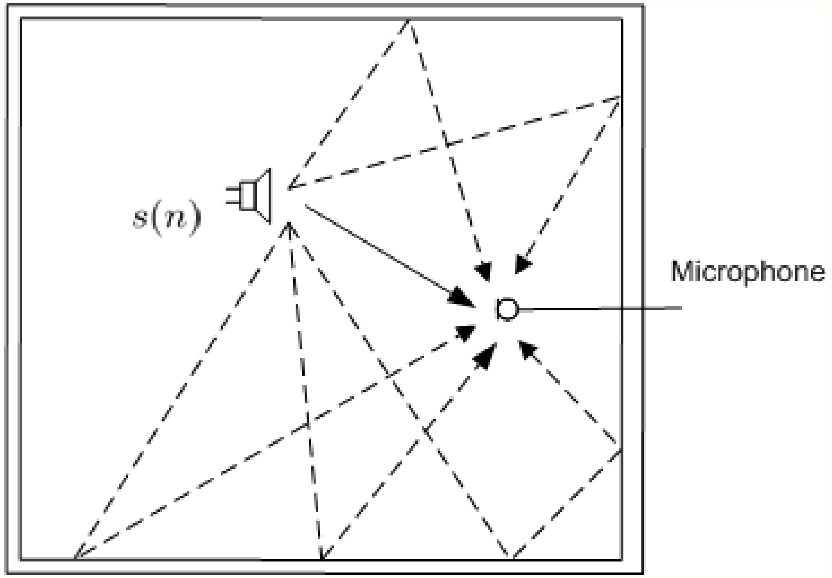
Original signal

Large room

Example:

reverberation time	$RT_{60} = 1.024$ s,
room dimension	$\{30 \times 30 \times 30\}$ m,
source position	$\{1,\ 5,\ 1.6\}$ m,
microphone position	$\{20,\ 5,\ 1.6\}$ m,





Reverberation, intelligibility and music

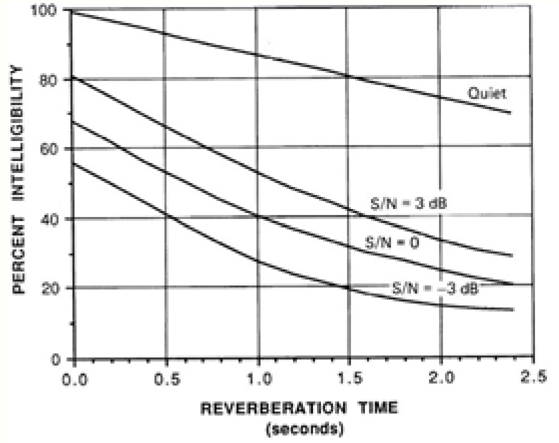
Reverberation affects the quality of the sound.

- Without any reverberation, difficult to have good intelligibility
- For speech

$RT_{60} > 1\ \mathrm{s}$ intelligibility drops

$RT_{60} > 2\ \mathrm{s}$ ability to perceive every syllable drops significantly

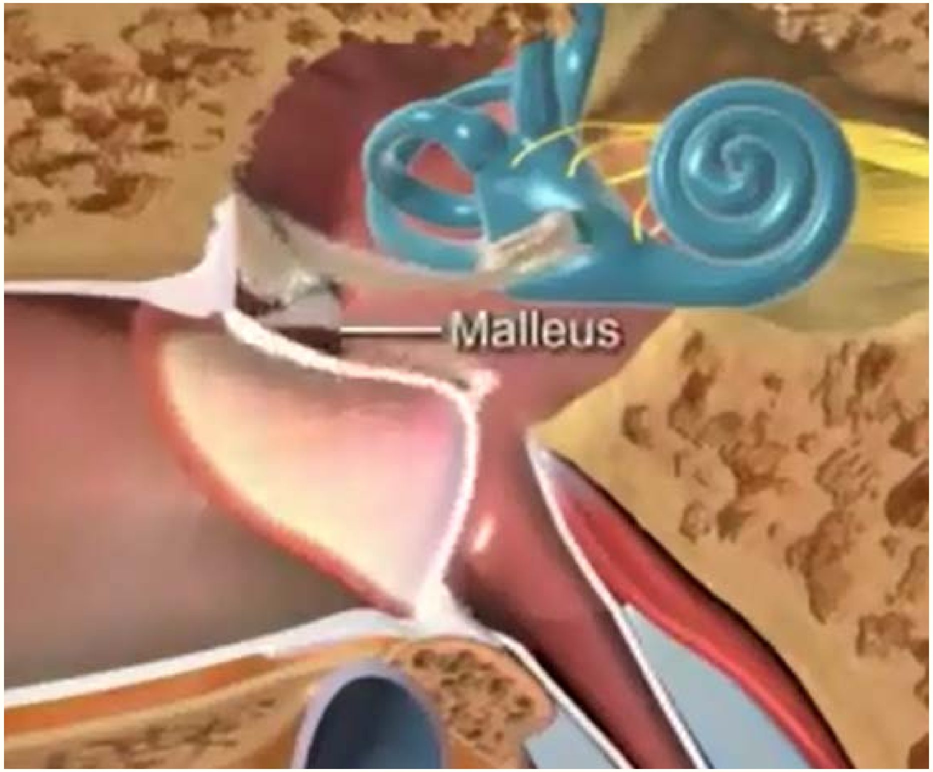
CleanSpeech $RT_{60} = 1\ \mathrm{s}$ $RT_{60} = 4\ \mathrm{s}$



- For concert halls, more reverberation increases the “richness” of the music so that they do not sound too simple.

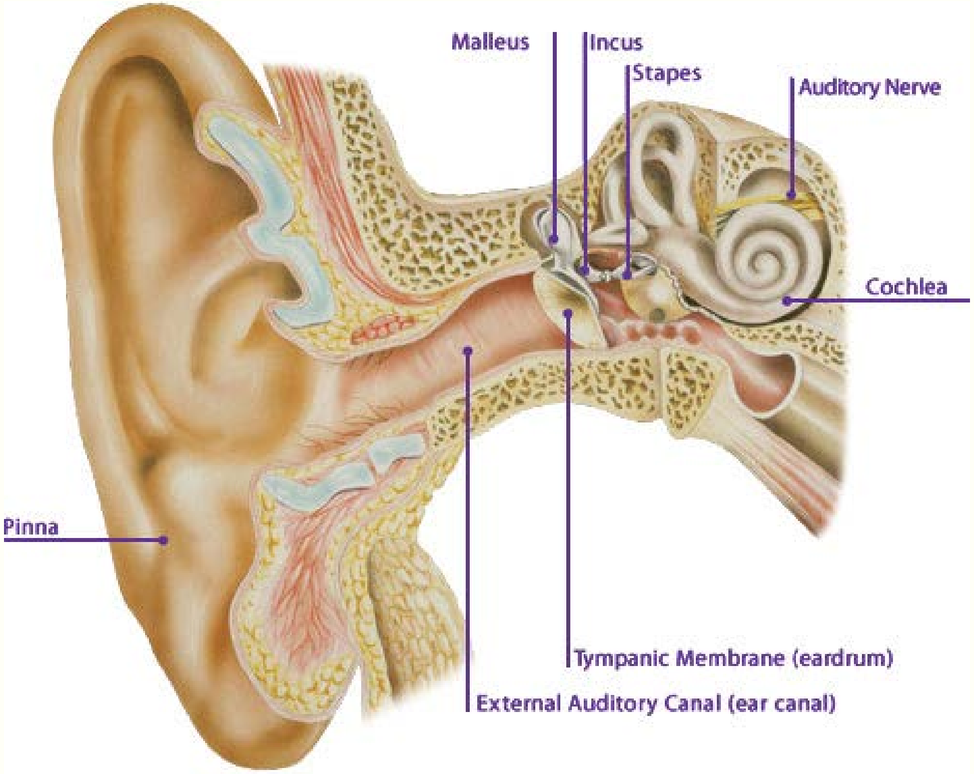
Human Hearing

Demo: How sound travels in human ears

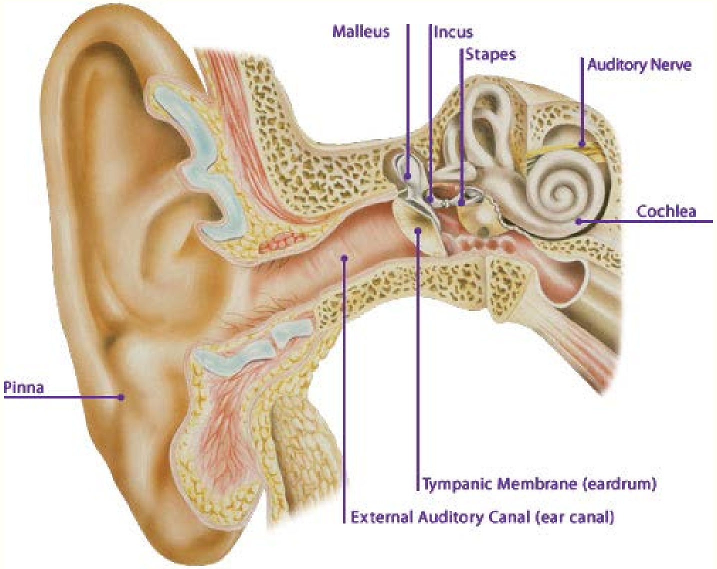


Physiology of the human ear

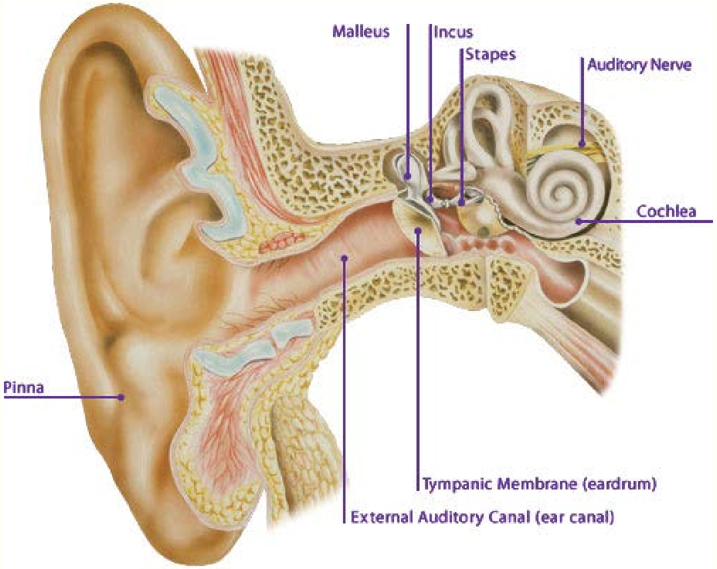
- The human auditory system is divided into
  - the pinna
  - the auditory canal
  - the middle ear
  - the inner ear


Malleus
Incus
Stapes
Auditory Nerve
Cochlea
Pinna
Tympanic Membrane (eardrum)
External Auditory Canal (ear canal)

- The pinna (outer ear)
  - Limited function as it is not big enough to act as a horn to collect much sound energy.
  - Alter human's perception and to determine if the sound is in-front-of or behind the head.
  - Humans need to train their listening according to their pinna.


Malleus
Incus
Stapes
Auditory Nerve
Cochlea
Pinna
Tympanic Membrane (eardrum)
External Auditory Canal (ear canal)

- The auditory canal
  - About 20-35 mm long and serves as a passage for sound.
  - Sounds resonate at about 3-4 kHz.
  - This resonance increases the transmission of sound energy substantially in this frequency.
  - Thus humans are most sensitive to these frequencies (3-4 kHz.)


Malleus
Incus
Stapes
Auditory Nerve
Cochlea
Pinna
Tympanic Membrane (eardrum)
External Auditory Canal (ear canal)

human	64-23,000
dog	67-45,000
cat	45-64,000
cow	23-35,000
horse	55-33,500
sheep	100-30,000
rabbit	360-42,000
rat	200-76,000
mouse	1,000-91,000
gerbil	100-60,000
guinea pig	54-50,000
hedgehog	250-45,000
raccoon	100-40,000
ferret	16-44,000
opossum	500-64,000
chinchilla	90-22,800
bat	2,000-110,000
beluga whale	1,000-123,000
elephant	16-12,000
porpoise	75-150,000
goldfish	20-3,000
catfish	50-4,000
tuna	50-1,100
bullfrog	100-3,000
tree frog	50-4,000
canary	250-8,000
parakeet	200-8,500
cockatiel	250-8,000
owl	200-12,000
chicken	125-2,000

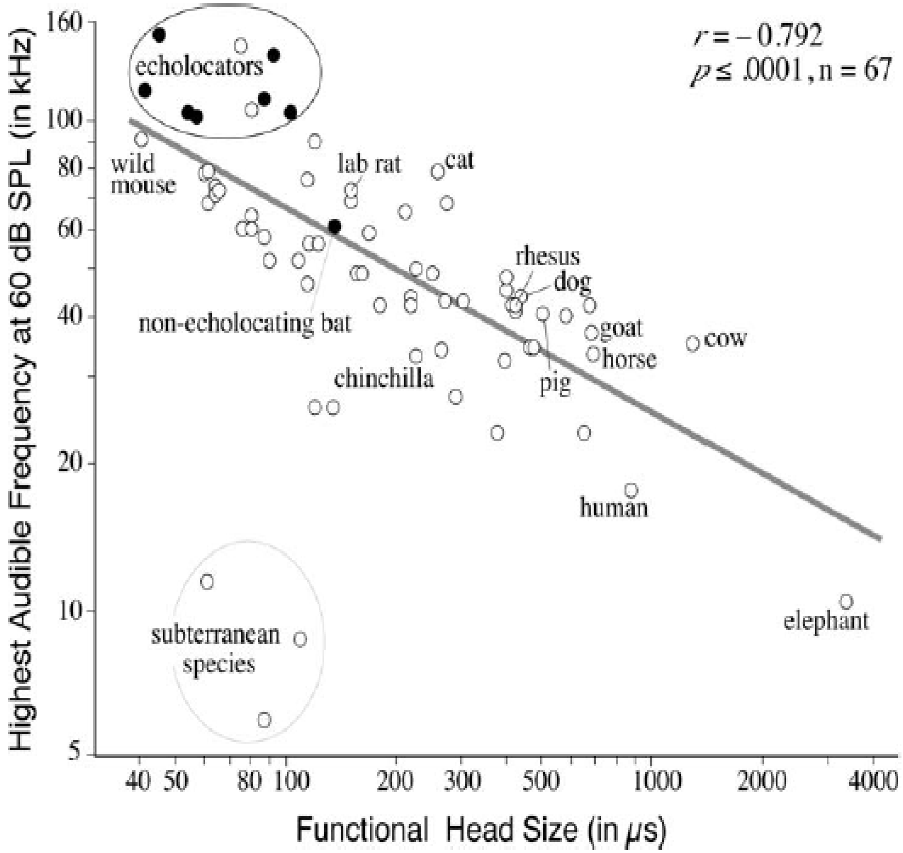
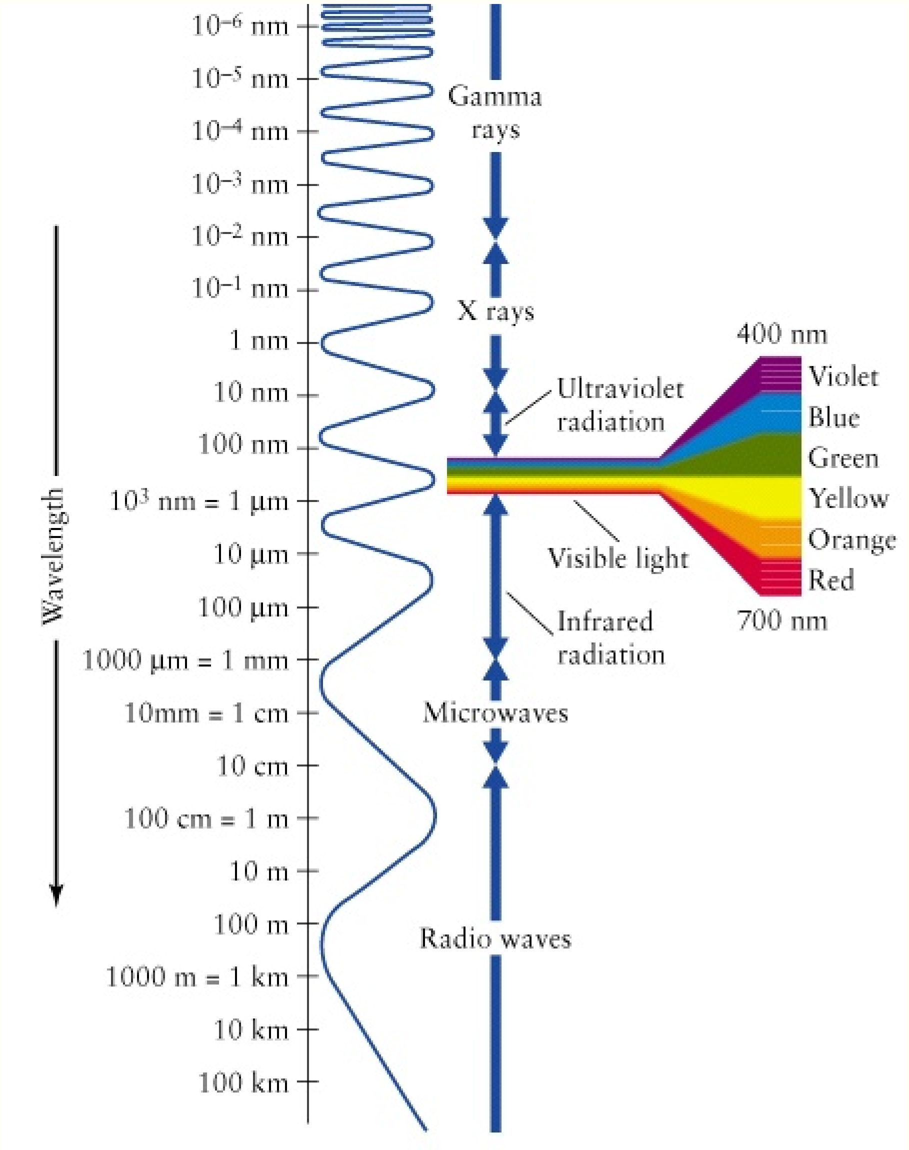


Figure 1. Relation between functional head size and high-frequency hearing (highest frequency audible at 60 dB sound pressure level) for mammals. This relationship is explained by the need of small mammals need to hear higher frequencies than larger mammals in order to use the binaural spectral-difference cue and/or pinna cues to localize sound. Note that the subterranean species (naked mole rat, blind mole rat, and gopher), which do not localize sound, have lost the ability to hear high frequencies. Echolocating bats hear slightly higher than predicted based on their functional head size. Filled circles indicate bats, open circles indicate all other mammals. (The open circles among the echolocators are two species of cetacea.) For references to individual audiograms, see Koay, G. *et al*., 1998a, and Heffner, R. S. *et al*., 2003; For tables of the absolute thresholds of mammals, go to the website at http://psychology.utoledo.edu/lch

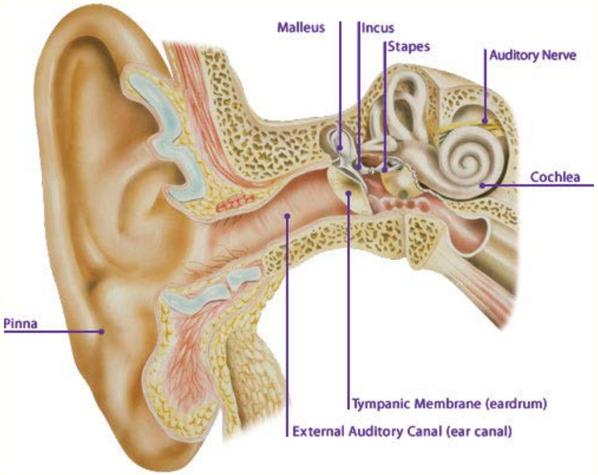
- For sound waves in air, the speed of sound is 343 m/s (at room temperature and atmospheric pressure).
- The wavelengths of sound frequencies audible to the human ear (20 Hz – 20 kHz) are thus between approximately 17 m and 17 mm, respectively.
- The wavelengths in audible sound are much longer than those in visible light.



- The middle ear and eardrum
  - Vibration of the eardrum is transferred by the bones of the middle ear to inner ear where sound reaches the cochlea.
  - Air is a medium of low density and therefore has a low acoustic impedance Z where

$$Z = v \text{ x } \rho$$

    such that v is the velocity of sound and ρ is the density of medium.
  - The function of the middle ear is to "match" the low impedance of the air (in the eardrum) to the high impedance of the cochlea fluid by a system of levers.

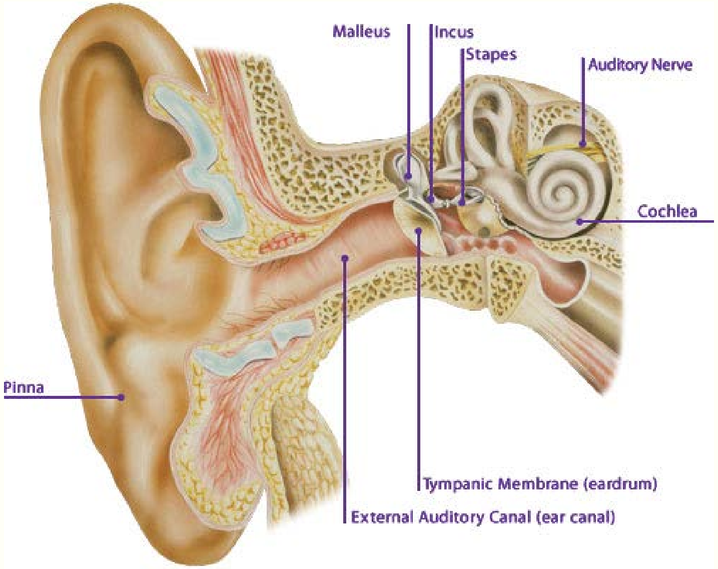

Malleus
Incus
Stapes
Auditory Nerve
Cochlea
Pinna
Tympanic Membrane (eardrum)
External Auditory Canal (ear canal)

- The middle ear resonates at about 1.5 kHz and the ear becomes less sensitive to lower frequencies.

- This reduction in sensitivity makes us insensitive to manmade and natural noise (wind, traffic)

- At high frequencies, the bones of the middle ear, and the tissues joining them, form a filter which prevent transmission of sound at frequencies above 20 kHz.

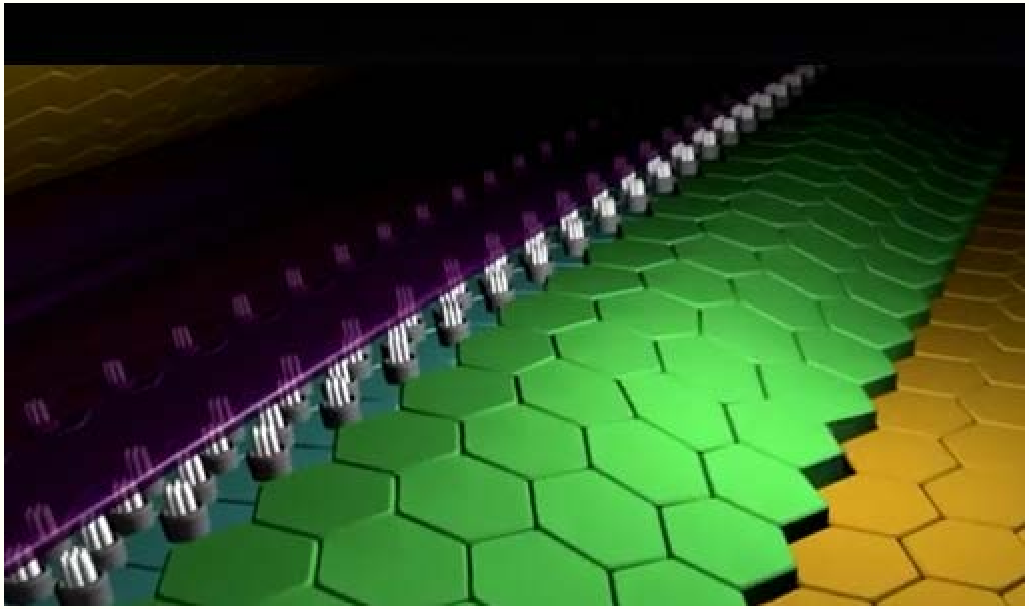
- The inner ear
  - Consist of a spirally coiled tube (with the shape of a snail shell) known as the *cochlea.*
  - The cochlea is divided along its length by the basilar membrane containing hair cells which couples vibration of the cochlear to the auditory nerves.


Malleus
Incus
Stapes
Auditory Nerve
Cochlea
Pinna
Tympanic Membrane (eardrum)
External Auditory Canal (ear canal)

  - For any particular sound frequency, the fluid in the cochlea vibrates with a peak at one point on the basilar membrane.
  - Hence, frequency of a sound is converted to a point of maximum stimulation on the membrane which in turn provides the basis of the perception of pitch.

Demo:

The cochlea responds to different amplitude and frequencies



Audio perception

- At the simplest level,

  intensity ⇒ loudness,

  frequency ⇒ pitch.

- In practice, frequency and intensity interact and the loudness of the sound depends on both.
- Loudness is a **subjective** quantity and **cannot be measured** directly. It is however useful to assign numerical values to the experience of loudness.

Demo 1: Audio perception of different frequencies

Demo 2: https://www.animations.physics.unsw.edu.au/jw/hearing.html

Open by Explorer????

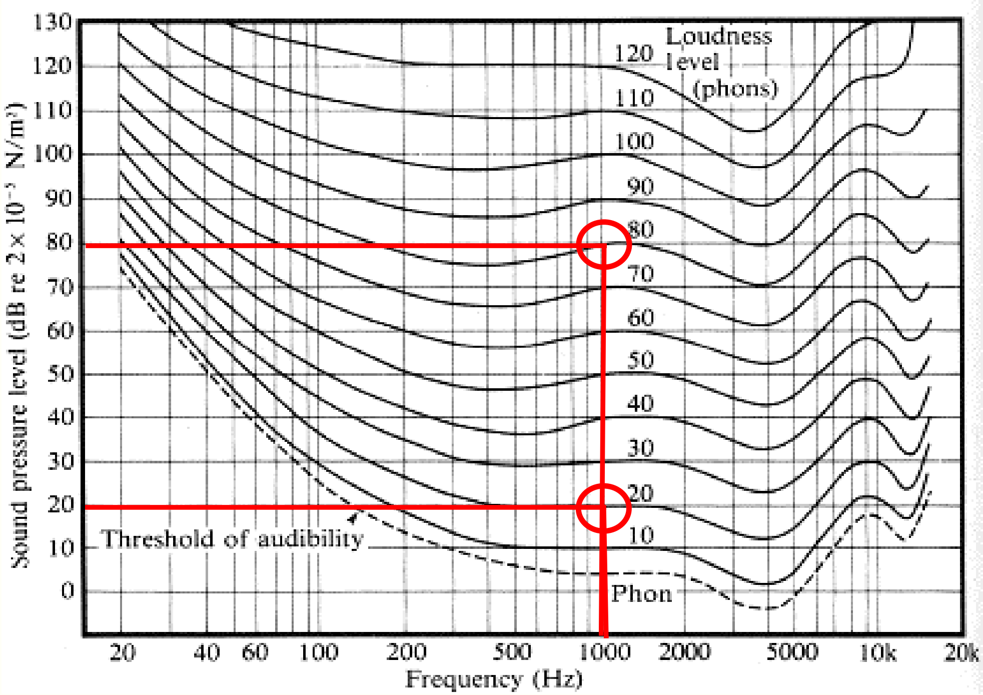
Reading Materials:
http://www.animations.physics.unsw.edu.au/waves-sound/human-sound/

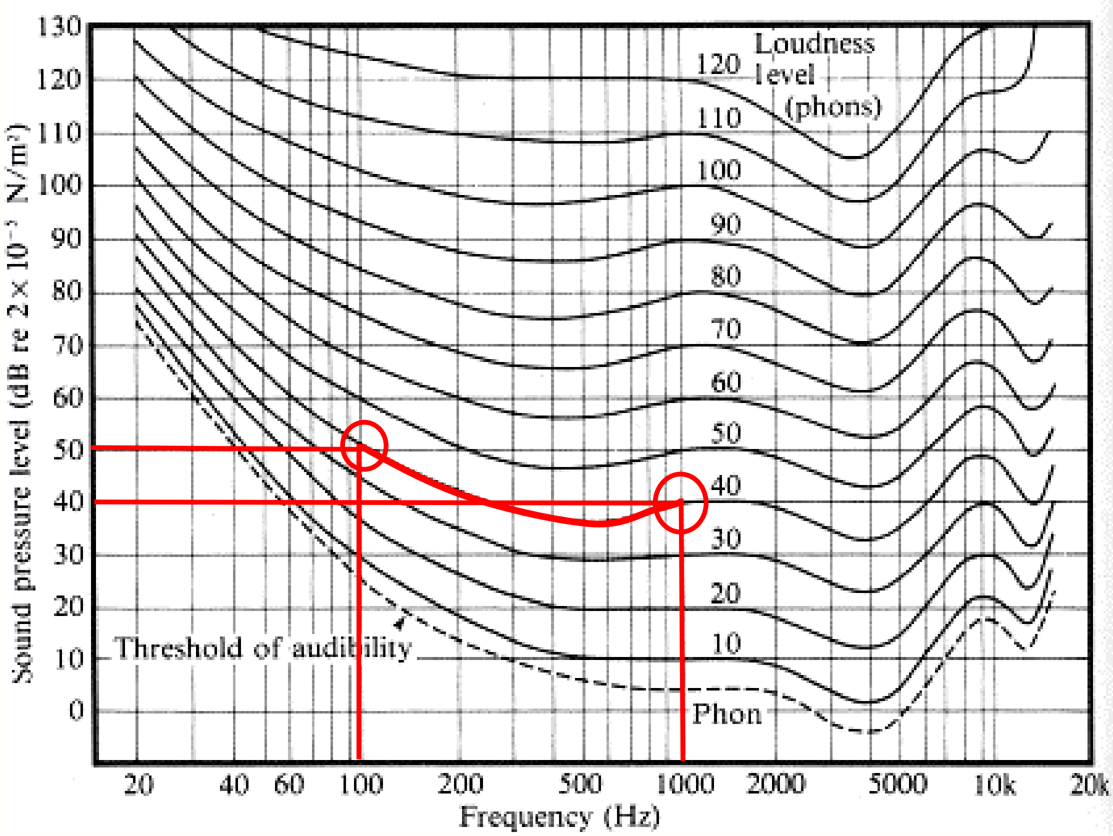
SPL, Phons and Sones

- Loudness level is defined as the sound pressure level (SPL) in dB of a 1000 Hz tone to be as loud as the sound under examination.
- By definition, 1 phon is defined as 1 dB SPL at frequency of 1 kHz. Hence, 30 phons means "as loud as 30 dB, at 1000 Hz frequency."
- The phon level found as described above only makes comparisons between different sounds. It does tell us nothing
- The use of the phon as a unit of loudness is an improvement over just quoting the level in decibels, but it is still not a measurement which is directly proportional to loudness.
- The **sone** scale provides a linear scale of loudness.

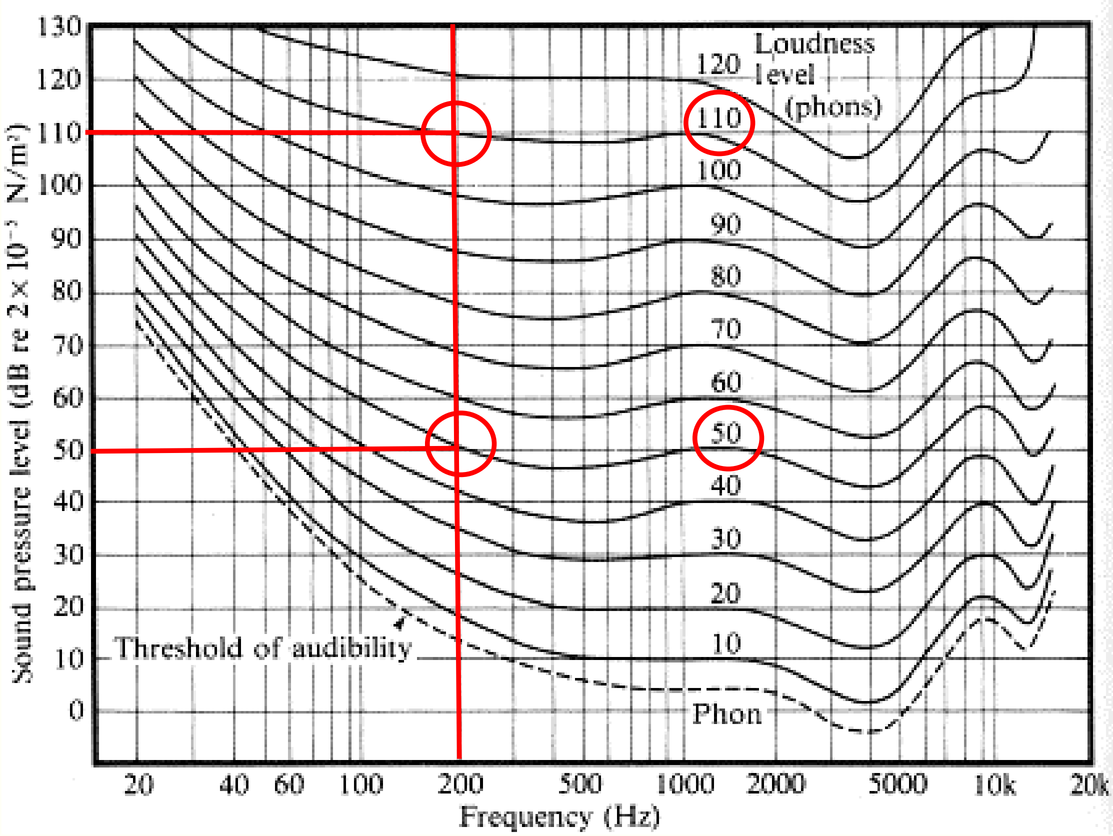
Phons	40	50	60	70	80	90	100
Sones	1	2	4	8	16	32	64

- Loudness level comparisons have been made between 20 Hz to about 15000 Hz and at various sound pressure levels.
- This leads to the production of “**equal loudness curves**.”
- **Interpretations:**
  1) 20 phons is 20 dB SPL at 1 kHz
  2) 80 phons is 80 dB SPL at 1 kHz etc



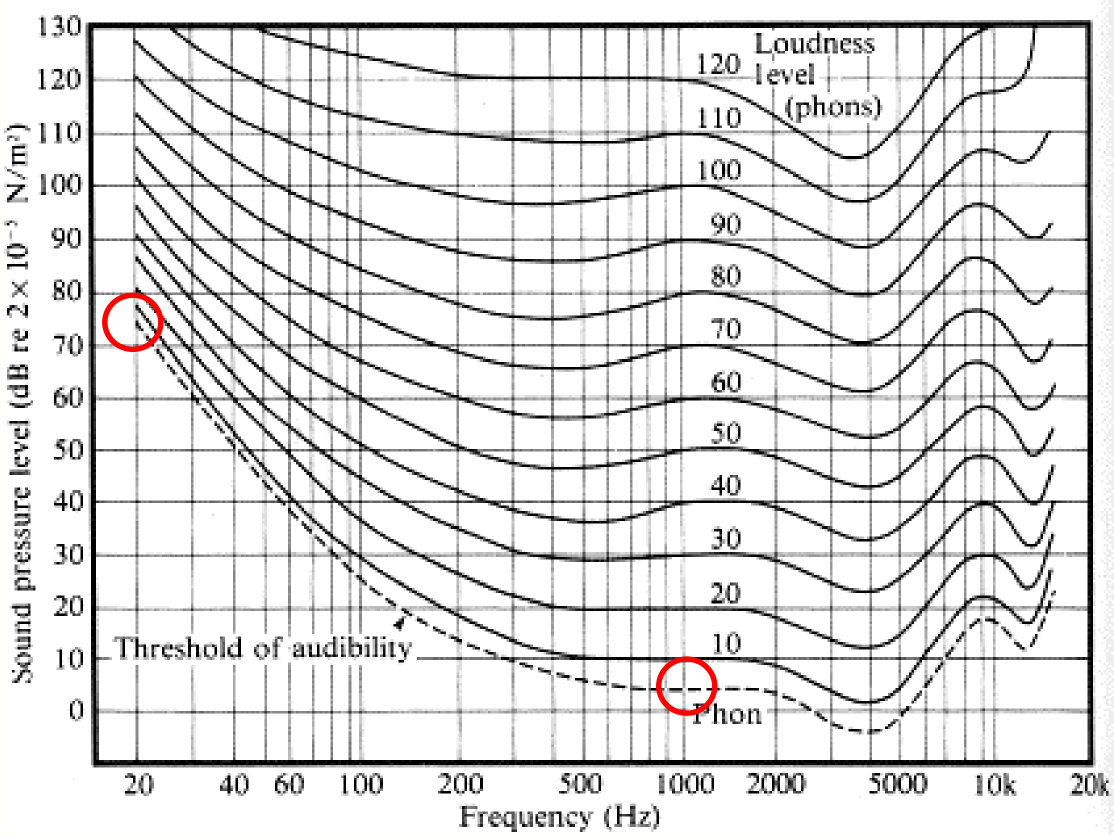


- **Interpretations:** All points on a given contour have equal loudness.
  - A tone of 100 Hz at 50 dB SPL sounds as loud as 1000 Hz at 40 dB SPL.

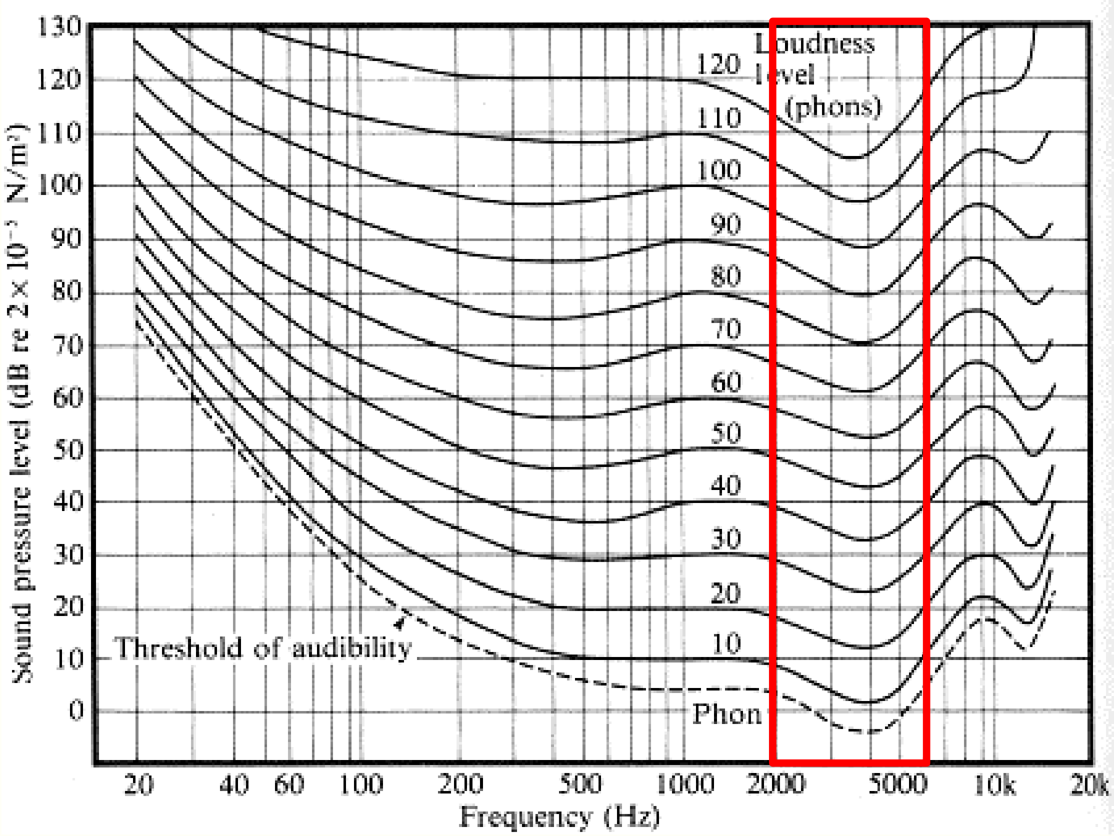


- **Interpretations:** Increasing the SPL will increase the level of loudness for each frequency.

  For 200 Hz, increasing from 50 dB SPL to 110 dB SPL will increase the perceive loudness from 50 phons to 110 phons.



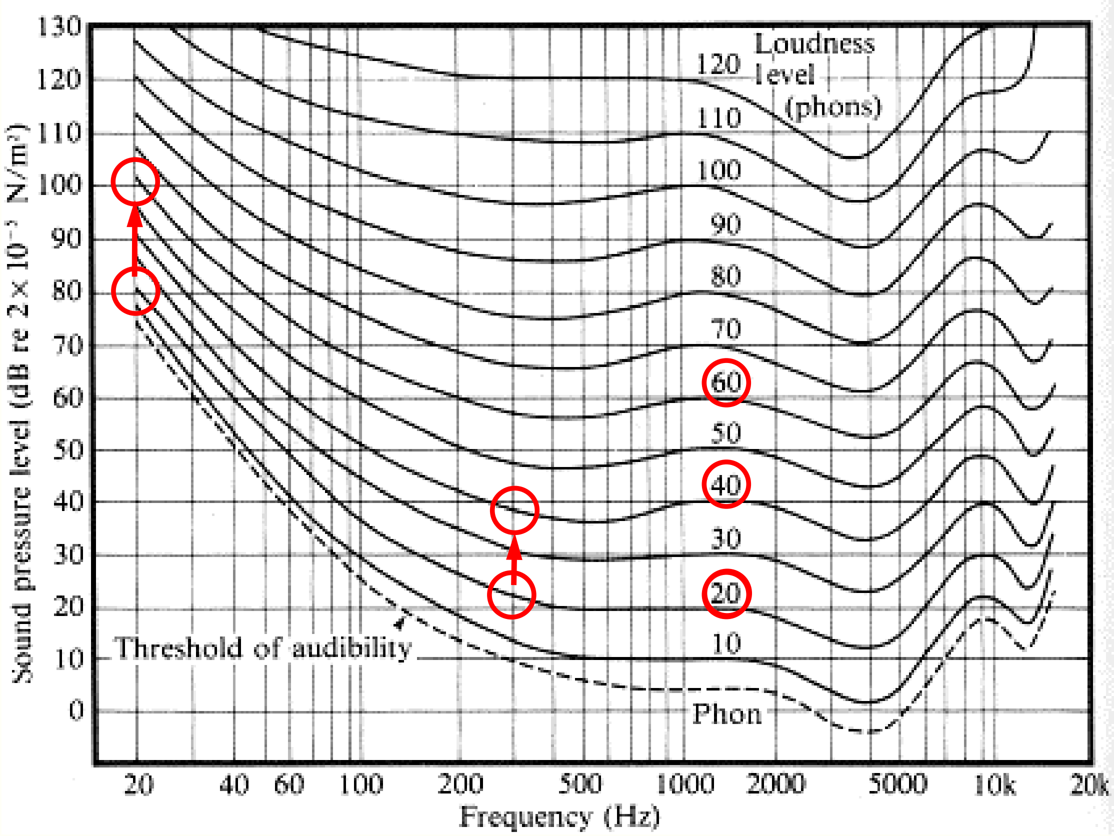
- **Interpretations:** Higher SPL is required for the lower frequencies to sound as loud as that for mid-frequencies.
  - We can only hear 20 Hz at 75 dB but can detect 1 kHz even at 5 dB.



- **Interpretations:** Humans are most sensitive to frequency range 2 kHz to 6 kHz.

20 dB SPL increase at 20 Hz (20 to 60 phons)

20 dB SPL increase at 300 Hz (20 to 40 phons)



- **Interpretations:** Steep gradient at lower frequencies and gentler at higher frequencies.
  - If the sound is reproduced at a higher level than it was recorded, then low frequencies will sound much louder compared to mid/high frequency.

Pitch and Frequency

- The pitch is our perception of a sound, and frequency is the physical oscillations involved in producing the sound
- Higher/low pitches correspond to higher/low frequencies.
- Pitch is relative and frequency is absolute
- Frequency describes a physical phenomenon while pitch describes a perceptual phenomenon
- Therefore, pitch is the perceived frequency.

- The ear is not good at distinguishing between two tones with slightly different frequencies.
- When the two frequencies in the same critical band, they hardly sound differently.

Demo: Human ear frequency sensitivity

Masking implications/Applications:

- When talking in a shopping mall, we tend to raise voice for clear hearing, which means the high volume voice **masks** the background noise.
- People feel comfortable at beach because the sound from the sea water masks other sounds.

Demo: Office privacy and interference reduction

- Transmission volume can be reduced if masked sound is not transmitted (since these masked signals cannot be heard by other party). In general, reducing
  - Channel bandwidth for transmission;
  - Storage (hard disk, CD)
  - Computational complexity

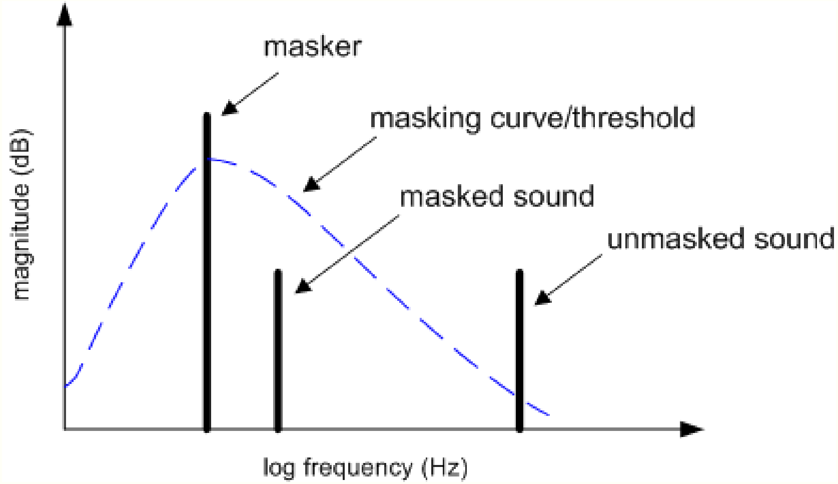
Masking

- For the humans, perception of sound is important:
  - We do not perceive *frequency* but *pitch*;
  - We do not perceive *pressure level* but *loudness*;
  - We do not perceive spectral shape, modulation depth, modulation frequency but sharpness, fluctuation strength, roughness;
  - We do not perceive time directly but subjective duration.

- Masking plays an important role in both frequency as well as time domain.
- Masking can be classified into two main categories:
  - Simultaneous masking (frequency masking)
  - Non-simultaneous masking (temporal masking)

Simultaneous masking

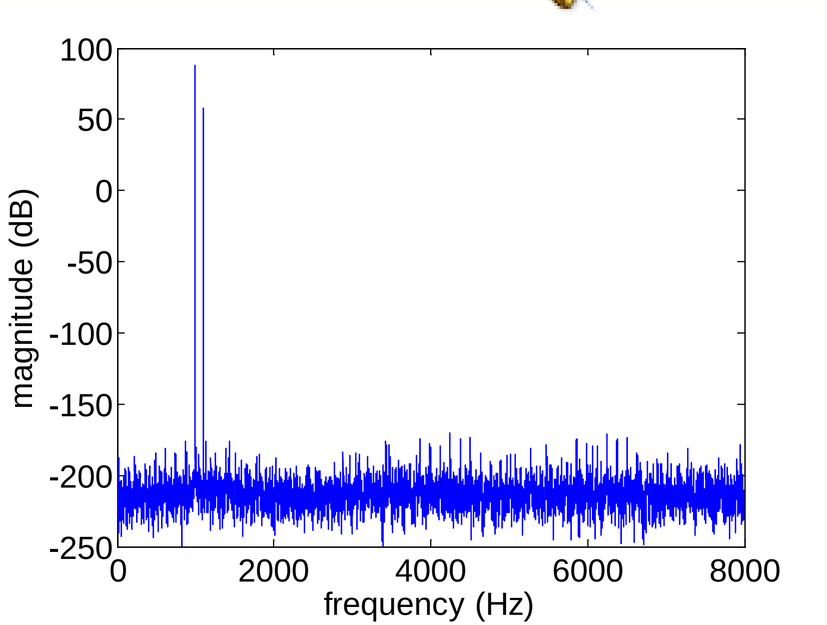
- Simultaneous masking is a frequency domain phenomenon where a low-level frequency component can be made inaudible by a simultaneously occurring stronger frequency component if both are close enough to each other in frequency domain.



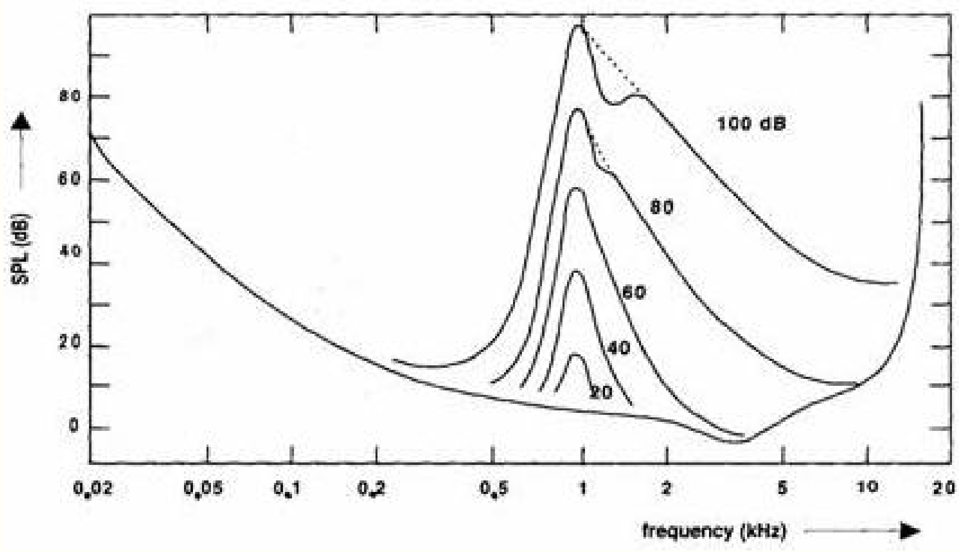
- Masking is usually described in terms of minimum SPL of a test sound (tone) that is audible in the presence of a masker. This is how the masking curve is generated.

Listening demo : Masker (1 kHz) is 30 dB higher than masked tone

1 kHz + 1.1 kHz

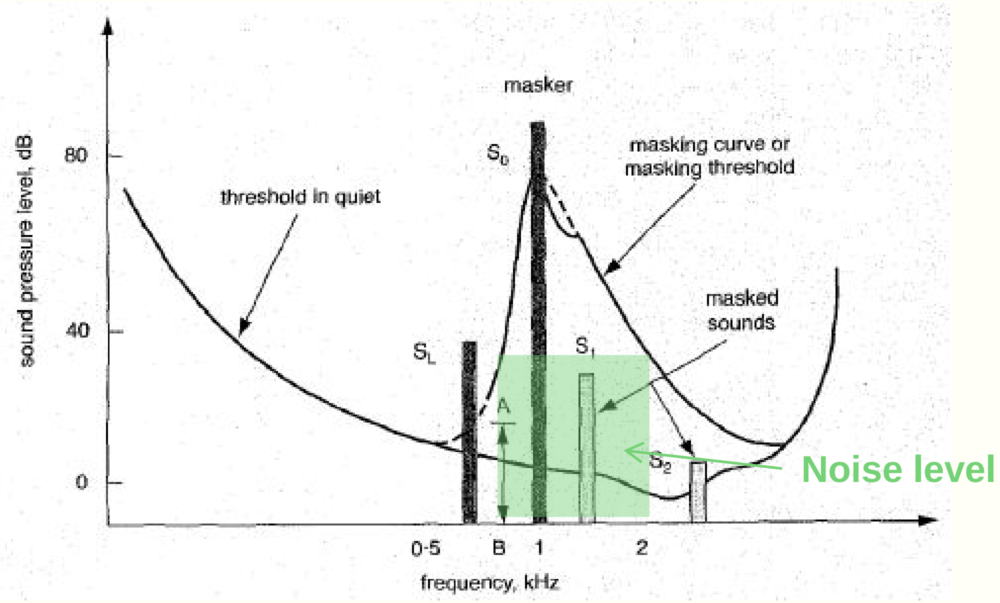


- Variation of masking curve with SPL



- Masking thresholds of narrow band signal with bandwidth of 90 Hz centered at 1 kHz at various SPL.
- Double peaks at 80 and 100 dB SPL are due to nonlinear phenomena in hearing.
- Signals with frequencies higher than the masker frequency are masked more effectively than signals with frequencies lower than masker frequency.

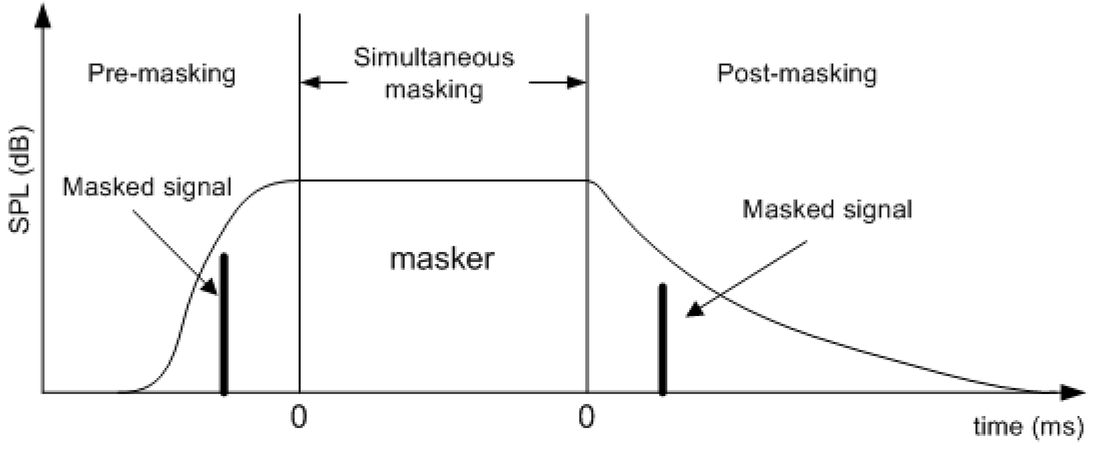
Example



- Signals S_1 and S_2 are masked by S_0 completely.
- Signal S_L cannot be entirely masked by S_0.
- It is possible to increase the quantization noise in the subband containing S_L up to the level AB. This means *fewer bits are needed* to represent the signal at this level.
- Masking threshold in this context is known as threshold of just noticeable distortion (JND).
- Distance between the level of masker and masking threshold is called signal-to-mask ratio (SMR).

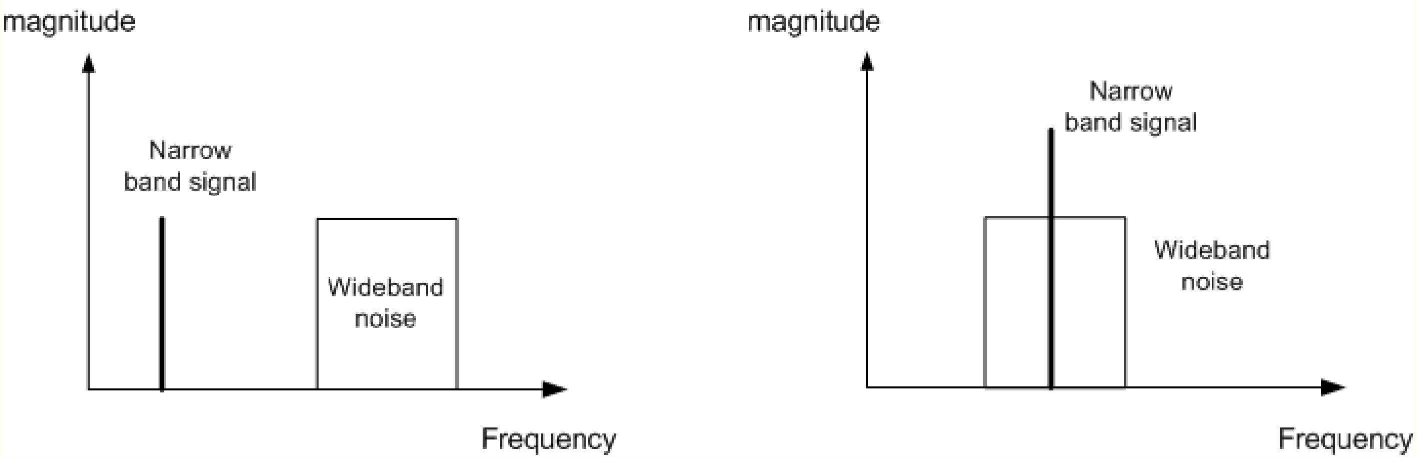
Non-Simultaneous masking

- Non-simultaneous masking is also known as temporal masking.
- It can occur when two sounds appear within a small interval.
- A strong signal can mask a weaker signal that occurs after and before it.
- There are two types of temporal masking
  - Pre-masking
  - Post-masking
- Post-masking last longer than pre-masking.

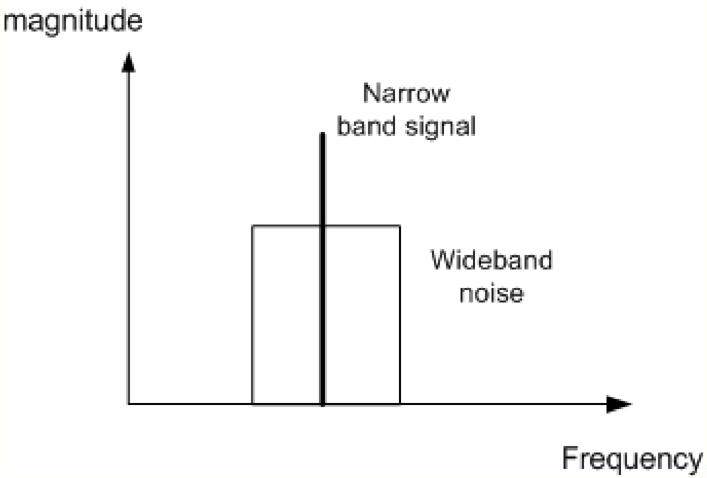


Critical band

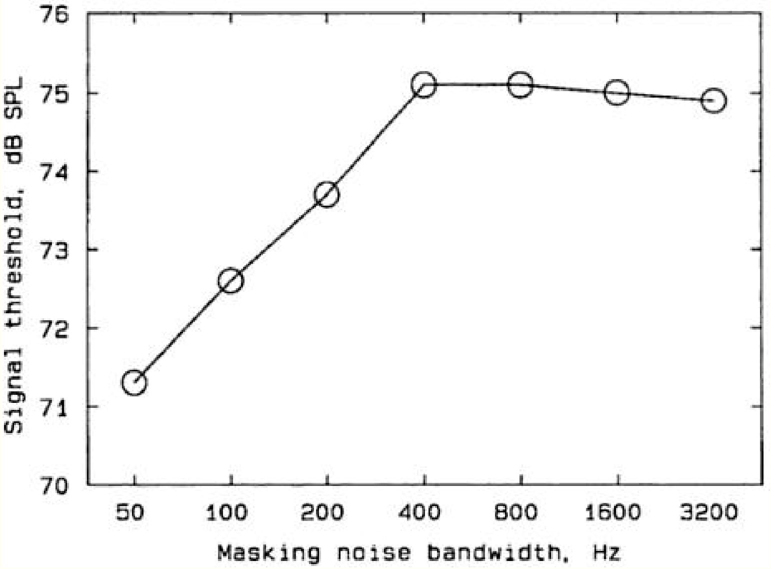
- The concept of critical band is often used for masking.
- Basically, the critical band is a frequency selective "channel" of psychoacoustic process. Only noise falling within the critical bandwidth can contribute to the masking of a narrow band signal.
- The auditory system consist of a whole series of critical bands each filtering out a specific portion of the audio spectrum. These filters are called "auditory filters."

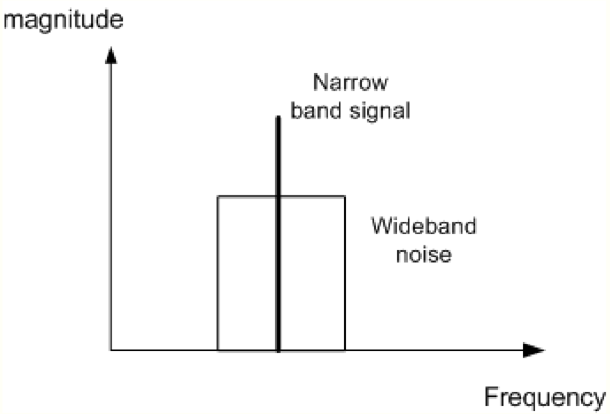


- When trying to detect a signal in a noisy environment, the listener is assumed to make use of a filter with a center frequency close to that of the signal.
- Only components in the noise which pass through the filter have any effect in masking the signal.
- To determine the critical bandwidth of the auditory system
  - Play a tone at a particular frequency.
  - Play a bandpass noise centered at this frequency.
  - Adjust tone so that it is just audible (this is the threshold of this tone).
  - Measure the threshold of this tone as a function of the bandwidth of the bandpass noise masker.



- The threshold of a 2 kHz sinusoidal signal plotted as a function of the bandwidth of a noise masker centered at 2 kHz.
- Note that the threshold increases with increasing masker bandwidth and then remains fairly constant.
- The bandwidth at which the signal threshold ceased to increase is known as the *critical bandwidth*.

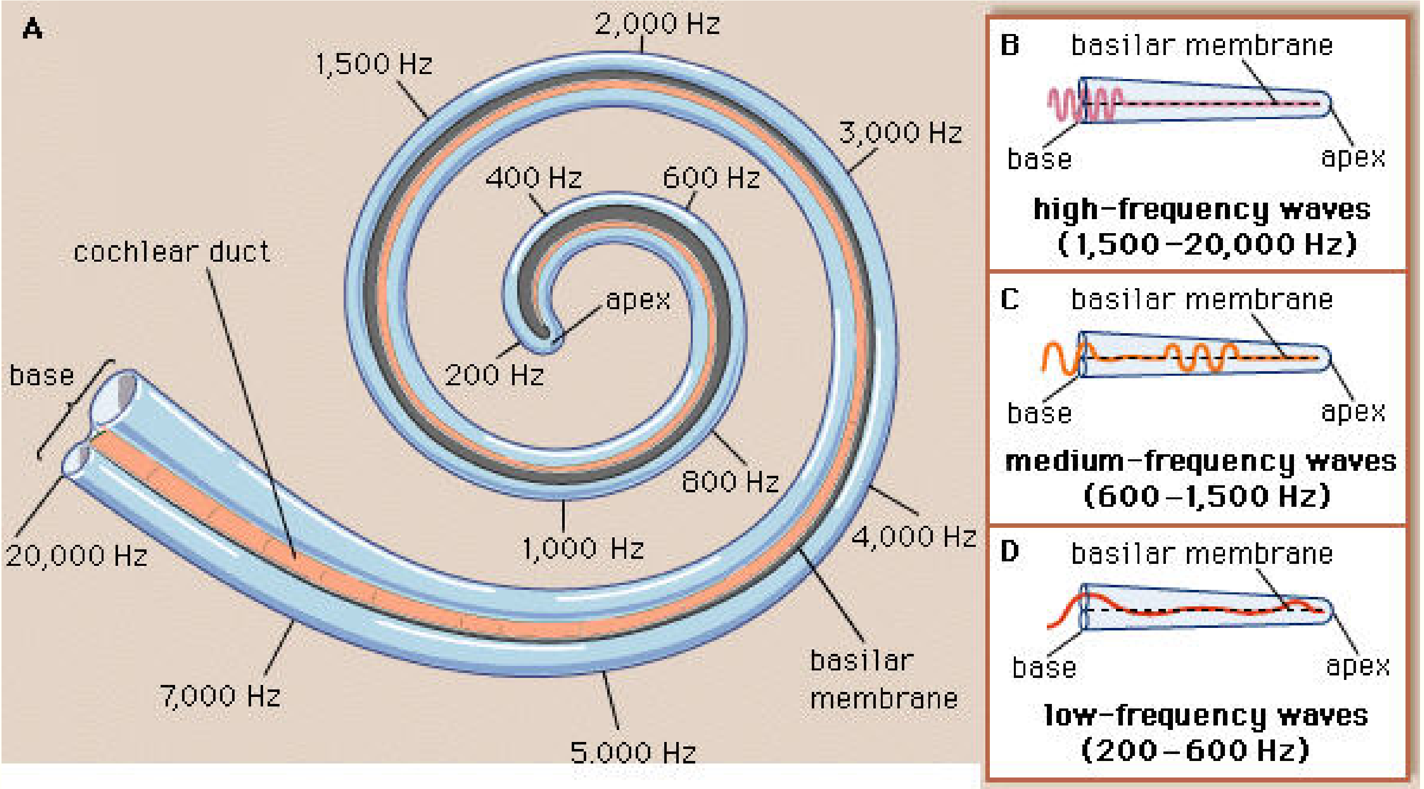




- Hence, all spectral energy that falls into one critical band is summed up and masks (or disables) the detection of a sinusoidal tone centred within that critical band as long as its level is below this masked threshold.
- The critical band is roughly constant at about 100 Hz for low center frequencies less than 500 Hz. For higher frequencies, critical bandwidth increases with center frequency.
- Twenty five critical bands are required to cover frequencies up to 20 kHz.

Bark scale

- The bark scale is named after German physicist Heinrich Barkhausen.
- It illustrates the concept of critical bands. There is a strong correlation between the critical bands and the excitation patterns of the basilar membrane (BM).
- If the basilar membrane is of 32 mm long and is divided into uniform intervals, each segment is equivalent to a critical band.
- This uniform subdivision of the basilar membrane is widely known as the Bark scale.
- Hence the Bark scale provides a mapping between frequencies and locations on the basilar membrane of the cochlear.



http://www.hearingcentral.com/howtheearworks.asp

- Formulas to convert Bark Scale to frequency

  - We first define the 25 lower band edge frequencies in Hertz:

$$\begin{aligned} f_L \quad = \quad & [0, 100, 200, 300, 400, 510, 630, 770, 920, \\ & 1080, 1270, 1480, 1720, 2000, 2320, 2700, \\ & 3150, 3700, 4400, 5300, 6400, 7700, 9500, 12000, 15500]. \end{aligned}$$

  - The Bark index can be computed using two methods

$$\text{Bark index} = \left[13 \times \tan^{-1}(0.00076 \times f_L)\right] + \left[3.5 \times \tan^{-1}\left(\frac{f_L^2}{7500^2}\right)\right];$$

or

$$\begin{aligned} \mathcal{B} \quad &= \quad \left[\frac{26.81 f_L}{1960 + f_L}\right] - 0.53, \\ \text{Bark index} \quad &= \quad \begin{cases} 0.15(2 - \mathcal{B}) & \text{if } \mathcal{B} < 2, \\ 0.22(\mathcal{B} - 20.1) & \text{if } \mathcal{B} > 20.1. \end{cases} \end{aligned}$$

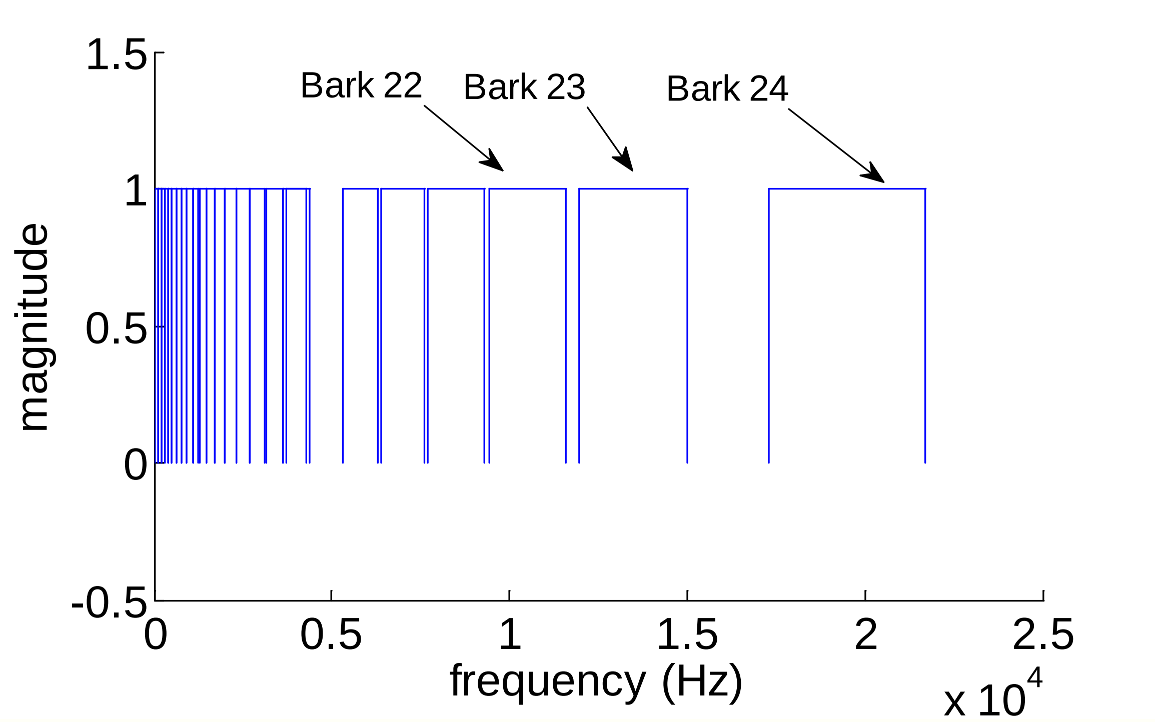
  - The bandwidth for each Bark index can be computed as

$$\text{BW} = 94 + 71 \times \left[\frac{f_L}{1000}\right]^{3/2}.$$

- Bark scale and their frequencies:

Bark	f_L (Hz)	f_c (Hz)	BW
0	0	50	100
1	100	150	100
2	200	250	100
3	300	350	100
4	400	450	110
5	510	570	120
6	630	700	140
7	770	840	150
8	920	1000	160
9	1080	1170	190
10	1270	1370	210
11	1480	1600	240
12	1720	1850	280
13	2000	2150	320
14	2320	2500	380
15	2700	2900	450
16	3150	3400	550
17	3700	4000	700
18	4400	4800	900
19	5300	5800	1100
20	6400	7000	1300
21	7700	8500	1800
22	9500	10500	2500
23	12000	13500	3500
24	15500		

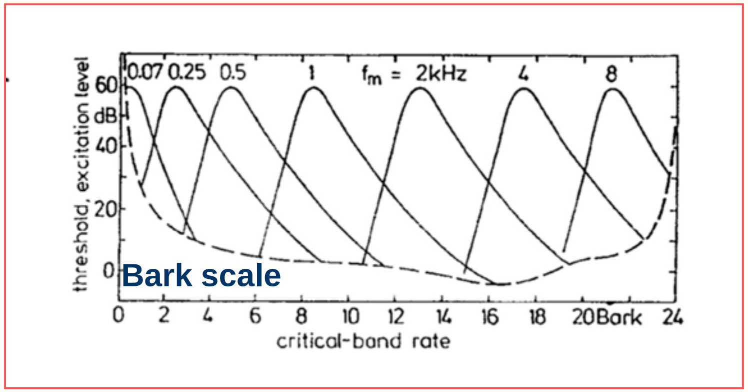
- Illustration of Bark index against frequencies

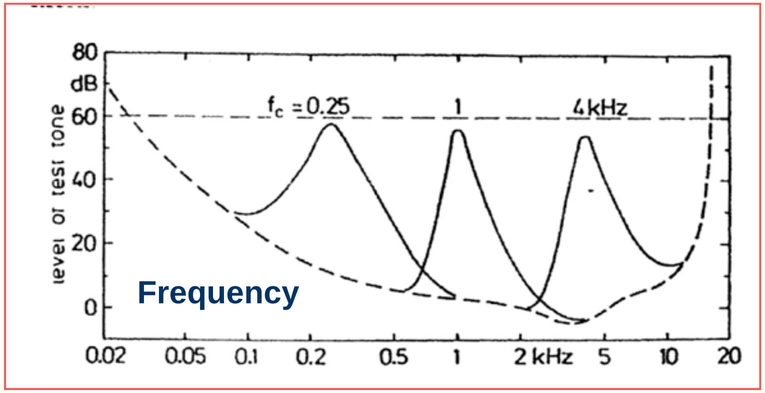


- The critical band is roughly constant at about 100 Hz for low center frequencies less than 500 Hz. For higher frequencies, critical bandwidth increases with center frequency.

Advantages of Bark scale

- The masking curve expressed in bark scale has a fairly constant shape, that is independent of masker frequency (see figures below).





  - The slope is more asymmetric when represented in Bark scale.
  - Because more overlapping between masker at lower freq., low-freq. masker is a better masker.