Gridspace

GRIDSPACE IAP LECTURE 1
SOUND, ACOUSTICS, & VOICE

January 9, 2023

TODAY'S ROADMAP

- Course Overview
- Sound
- Acoustics & Psychoacoustics
- Microphones and Recording
- Voice

COURSE OVERVIEW

Introductions

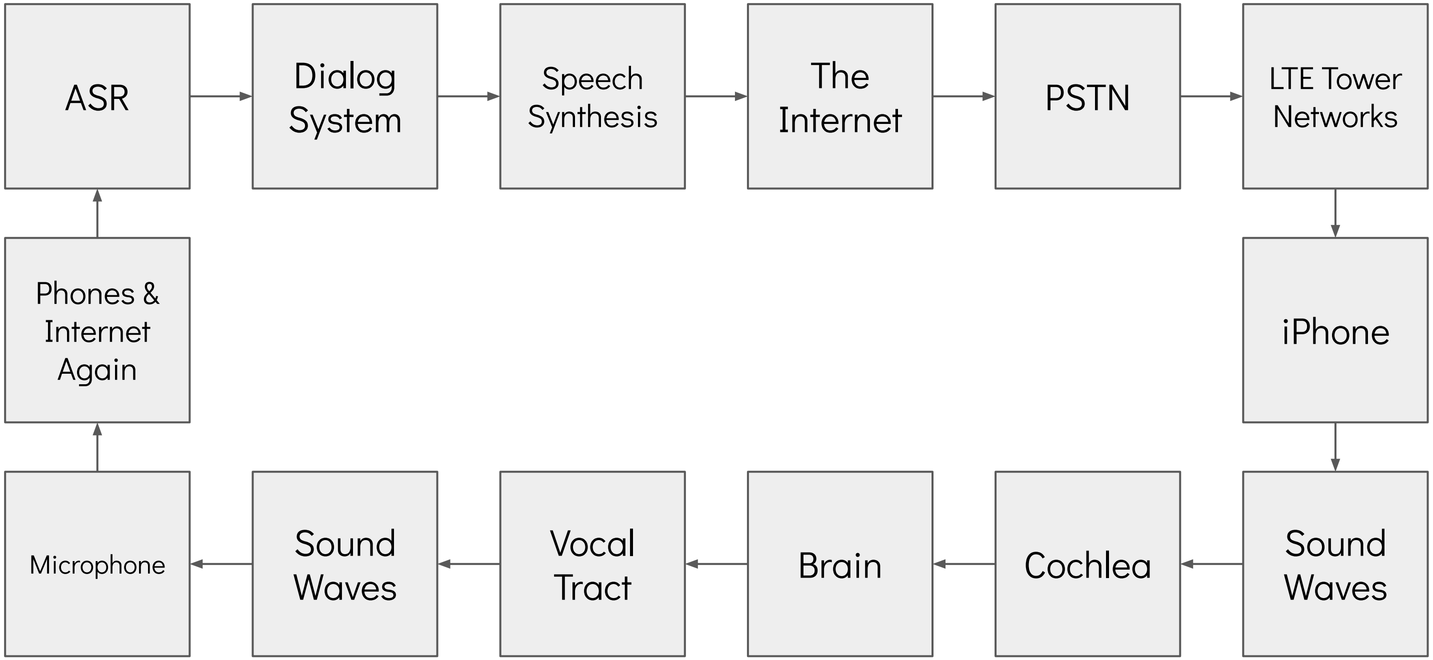
USE TRANSFORMERS FOR EVERYTHING
(THE END)

THANKS GUYS FOR JOINING OUR COURSE!!!

Where we're headed

- Integrated systems that deploy machine learning for
  - signal processing
  - speech recognition
  - language understanding
  - emotion recognition
  - language generation
  - speech synthesis
- All in the chaotic real world

Where we're headed


ASR
Dialog System
Speech Synthesis
The Internet
PSTN
LTE Tower Networks
iPhone
Sound Waves
Cochlea
Brain
Vocal Tract
Sound Waves
Microphone
Phones & Internet Again

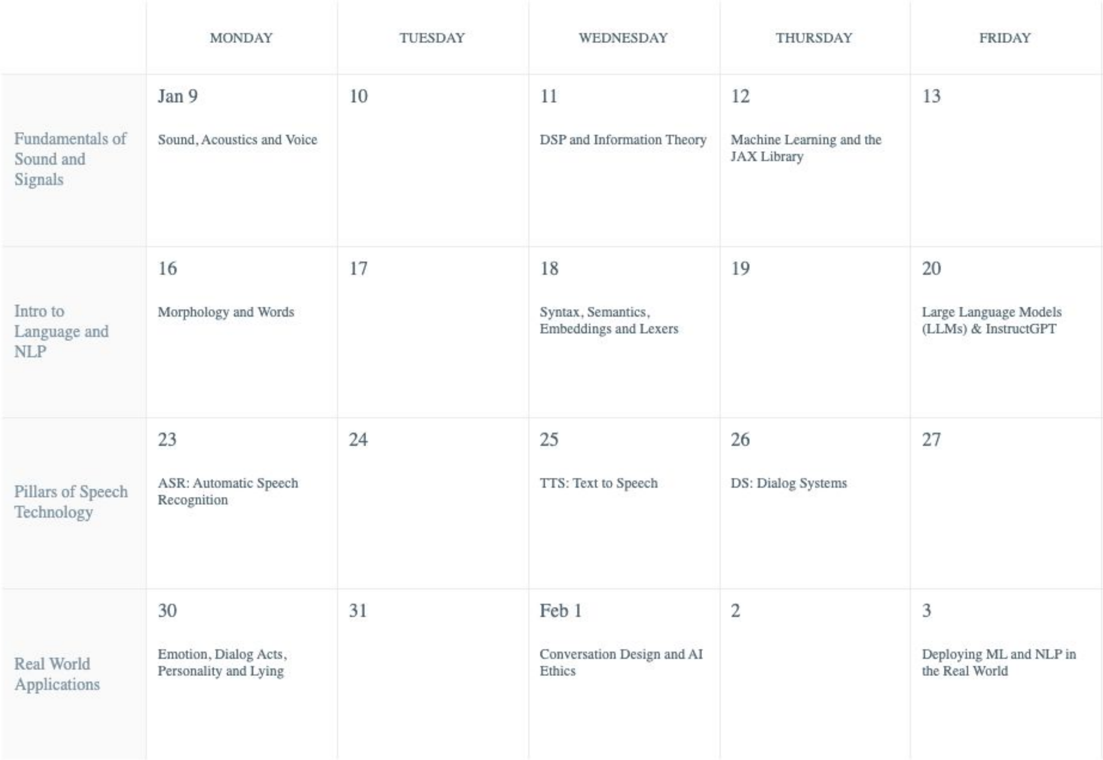
Objectives

- Cover breadth of spoken language topics beyond a typical introduction to machine learning or NLP
    - Sound
    - Signals
    - Linguistics
    - LLMs in messy real world deployments
    - Real world deployments
- Something for everyone; limited assumed knowledge
- Weekly projects and daily challenge questions

Audience & Prerequisites

- Excitement about ML & speech
- Some programming knowledge
- Some familiarity with the mathematical language of ML and common ML techniques

Course Schedule



	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY
Fundamentals of Sound and Signals	Jan 9 Sound, Acoustics and Voice	10	11 DSP and Information Theory	12 Machine Learning and the JAX Library	13
Intro to Language and NLP	16 Morphology and Words	17	18 Syntax, Semantics, Embeddings and Lexers	19	20 Large Language Models (LLMs) & InstructGPT
Pillars of Speech Technology	23 ASR: Automatic Speech Recognition	24	25 TTS: Text to Speech	26 DS: Dialog Systems	27
Real World Applications	30 Emotion, Dialog Acts, Personality and Lying	31	Feb 1 Conversation Design and AI Ethics	2	3 Deploying ML and NLP in the Real World

iap.gridspace.com

Staff & Admin

- Course lead: Phoebe Piercy (MIT '20 / '21)
- Course support: iap@gridspace.com
- Video Releases
- Meeting Invites
- YouTube Recordings
- Project Videos
- Remote versus in California

Structure

- Twelve lectures over four units
- Daily challenge questions
- Four weekly projects
- Opportunity to present your work
- Wide span of topics

But, first, the basics…

SOUND

Kinetic Theory of Gases

- The Rules:
  - Gas is made of tiny particles.
  - There are many, many particles. Behaves like a continuum.
  - Constantly colliding with walls of the container and each other.
  - No forces between particles at a distance.
  - Collisions are elastic (no energy is lost).
- The Result:
  - The ideal gas law
- https://ciechanow.ski/sound/

Ideal Gases

$$PV = nRT$$

- Pressure of the gas
- Volume of the gas
- number of moles of the gas
- R is the ideal gas constant
- Temperature of the gas

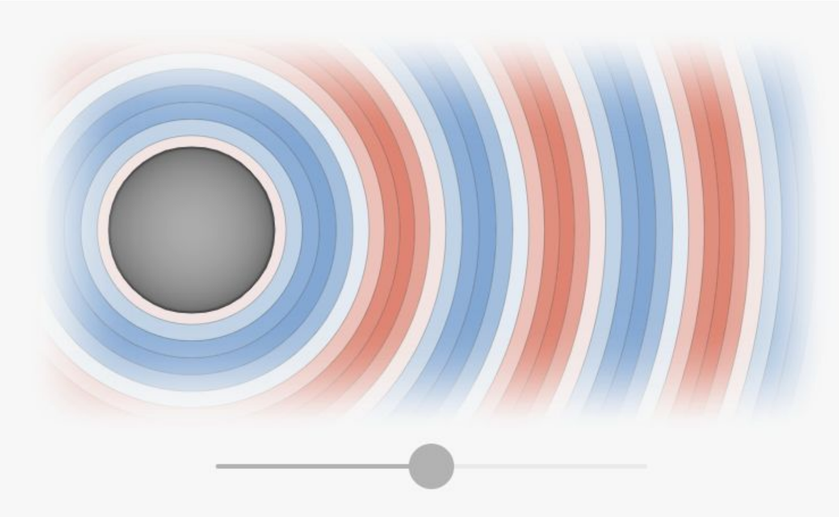
The Wave Equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$$

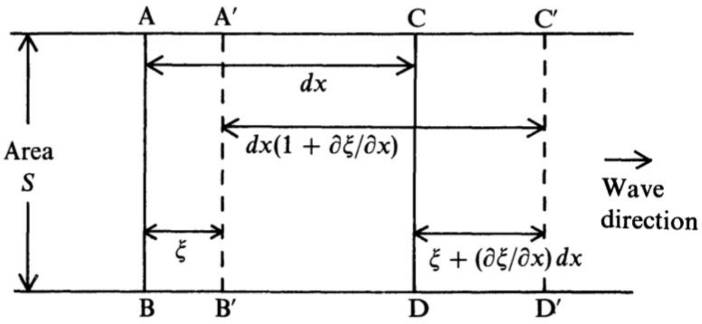
Sound Waves

- Most mathy part of this lecture I promise, but this is where sound comes from.

Sound Waves



Sound Waves



ξ is a tiny displacement of air in a passing pressure wave
S is the area of the volume of air the wave passes through
x is along the direction of wave travel

Sound Waves

Differential change in volume as wave passes:

$$V + dV = S\,dx\left(1 + \frac{\partial \xi}{\partial x}\right). \qquad (1)$$

Sound Waves

The Bulk Modulus K acts like the string constant in Hooke's law, a linear relationship between volume changes and pressure

$$dp_a = -K\frac{dV}{V}. \qquad (2)$$

Sound Waves

Combining (1) and (2). (2) let's us express changes in volume instead as a linear change in pressure. The more springly the fluid, the more the pressure changes with displacement.

$$p = -K \frac{\partial \xi}{\partial x}. \qquad (3)$$

Sound Waves

- Newton's 3rd Law: F=ma
- F = P*S (Force is pressure times area)
- Differentially, for a pressure wave of dx size, we can rewrite this as:

$$-S\left(\frac{\partial p}{\partial x}dx\right)$$

Sound Waves

- acceleration *a* can be written like this:

$$\frac{\partial^2 \xi}{\partial t^2}$$

Sound Waves

- mass *m* can be written like this in terms of density ρ times the volume (S*dx)

$$\rho S dx$$

Sound Waves

- Put together in the rewritten Newton's 3rd law:

$$-S\left(\frac{\partial p}{\partial x}dx\right) = \rho S dx \frac{\partial^2 \xi}{\partial t^2}$$

Sound Waves

- Cancelling the dx and S

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial^2 \xi}{\partial t^2} \qquad (4)$$

Sound Waves

- From (3) ($p = -K\frac{\partial \xi}{\partial x}$.) and (4) ($-\frac{\partial p}{\partial x} = \rho\frac{\partial^2 \xi}{\partial t^2}$) :

$$\frac{\partial^2 \xi}{\partial t^2} = \frac{K}{\rho}\frac{\partial^2 \xi}{\partial x^2} \quad (5)$$

(Displacement wave equation)

Sound Waves

- We want to convert this into the pressure form
- Differentiate (3) with respect to x and differentiate (5) with respect to t twice and we get:

$$\frac{\partial^2 p}{\partial t^2} = \frac{K}{\rho} \frac{\partial^2 p}{\partial x^2}$$

(Pressure wave equation)

Sound Waves

- By comparison $c^2 = K/\rho$

$$\frac{\partial^2 p}{\partial t^2} = \frac{K}{\rho} \frac{\partial^2 p}{\partial x^2}$$

$$\boldsymbol{\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}}$$

Sound Waves

- We now need to put K (dP/dV) in terms of stuff we know (pressure, temperature, density), so we can get the speed of sound
- If things were adiabatic (quick with no thermal transfer) we could use the ideal gas law as the "equation of state", but it's not adiabatic so we need to use this formula (γ is the specific heat ratio for the gas at constant pressure and volume, for air it's 1.4):

$$p_{\mathrm{a}} V^{\gamma} = \text{constant},$$

Sound Waves

- Taking the log and differentiating p with respect to x (to get K), we finally get the speed of sound of a gas is

$$c^2 = \frac{K}{\rho} = \frac{\gamma p_{\mathrm{a}}}{\rho},$$

Why did I just do all this?

- Nearly all properties of sound can be derived from the wave equation:
    - Vibration of mechanical systems like drums or strings
    - Propagation through air
    - Refraction, reflection, diffraction, etc…
- Everything the rest of the lecture comes from solving that one PDE

The Speed of Sound

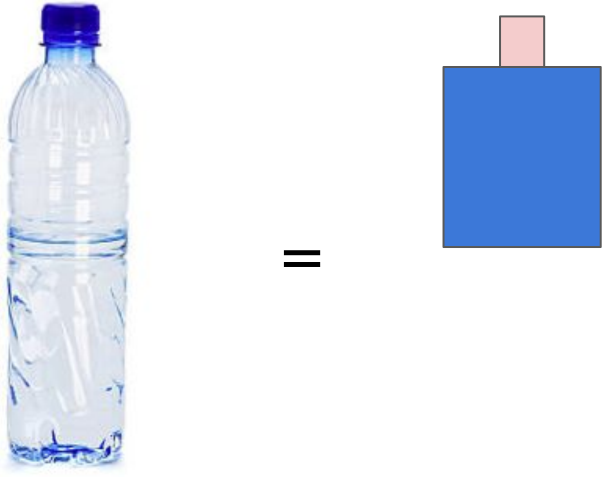
- 43 metres per second
- 1,235 km/h
- 1,125 ft/s
- 767 mph

The Helmholtz Oscillator

A simple acoustic model that shows one common way mechanical vibration becomes sound.

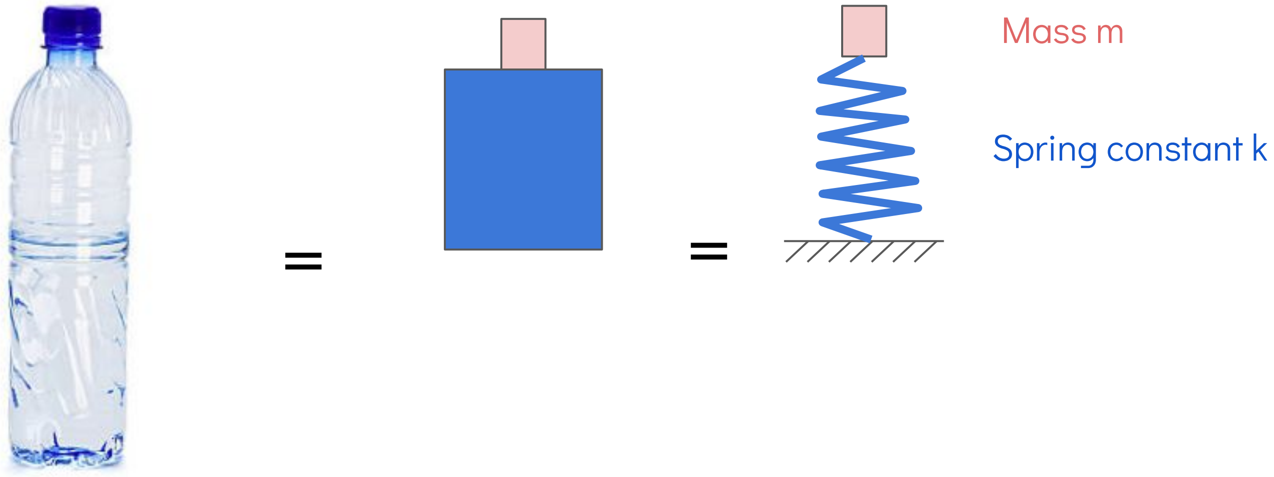
The Helmholtz Oscillator

A simple acoustic model that shows one common way mechanical vibration becomes sound.



The Helmholtz Oscillator

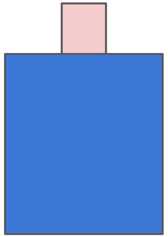
A simple acoustic model that shows one common way mechanical vibration becomes sound.



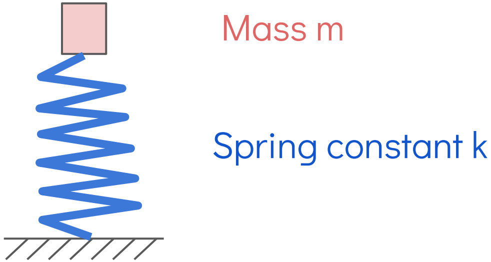
The Helmholtz Oscillator

A simple acoustic model that shows one common way mechanical vibration becomes sound.

=

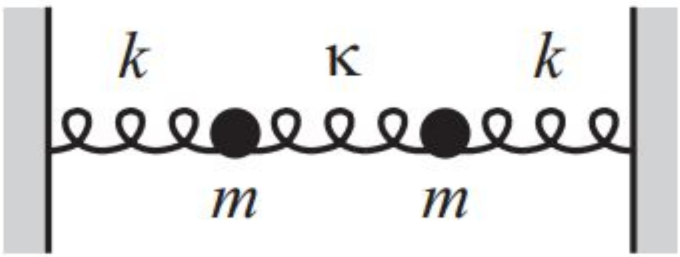


=



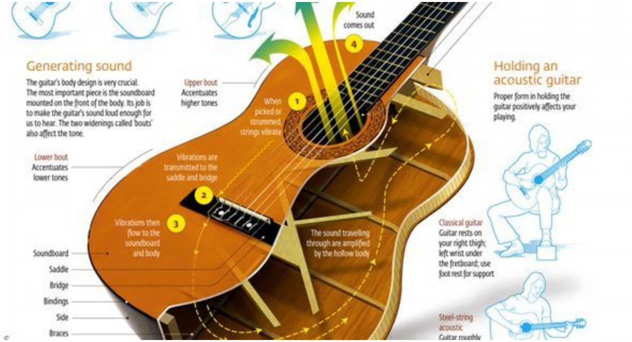
$$f_H = \frac{v}{2\pi}\sqrt{\frac{A}{V_0 L_{eq}}}.$$

Coupled Oscillators



Coupled Oscillators

- <u>The Transfer Function</u>. The relationship between two oscillating systems.
- Most sound is born from the transfer between multiple mechanical oscillations and the air



Generating sound

The guitar's body design is very crucial. The most important piece is the soundboard mounted on the front of the body. Its job is to make the guitar's sound loud enough for us to hear. The two widenings called 'bouts' also affect the tone.

1 When picked or strummed, strings vibrate

2 Vibrations are transmitted to the saddle and bridge

3 Vibrations then flow to the soundboard and body

4 Sound comes out

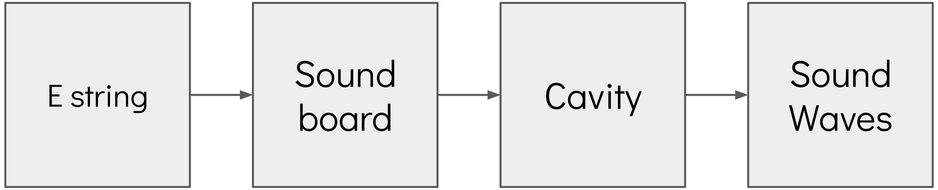
The sound travelling through are amplified by the hollow body

Holding an acoustic guitar

Proper form in holding the guitar positively affects your playing.

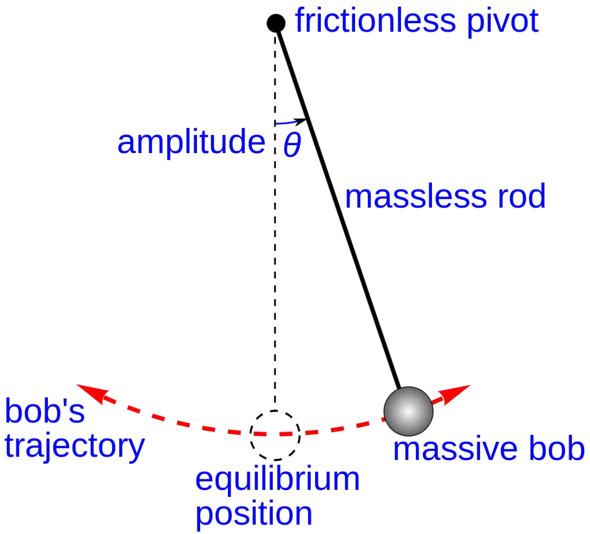
Classical guitar Guitar rests on your right thigh; left wrist under the fretboard; use foot rest for support

Steel-string acoustic Guitar roughly

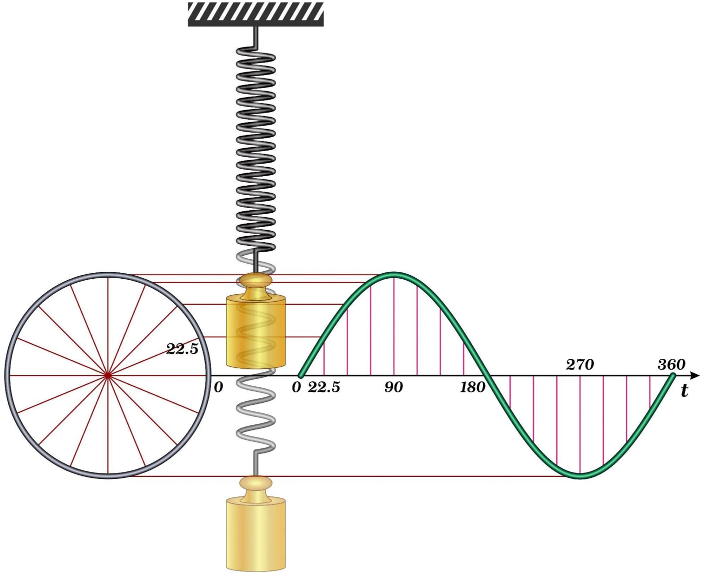


E string → Sound board → Cavity → Sound Waves

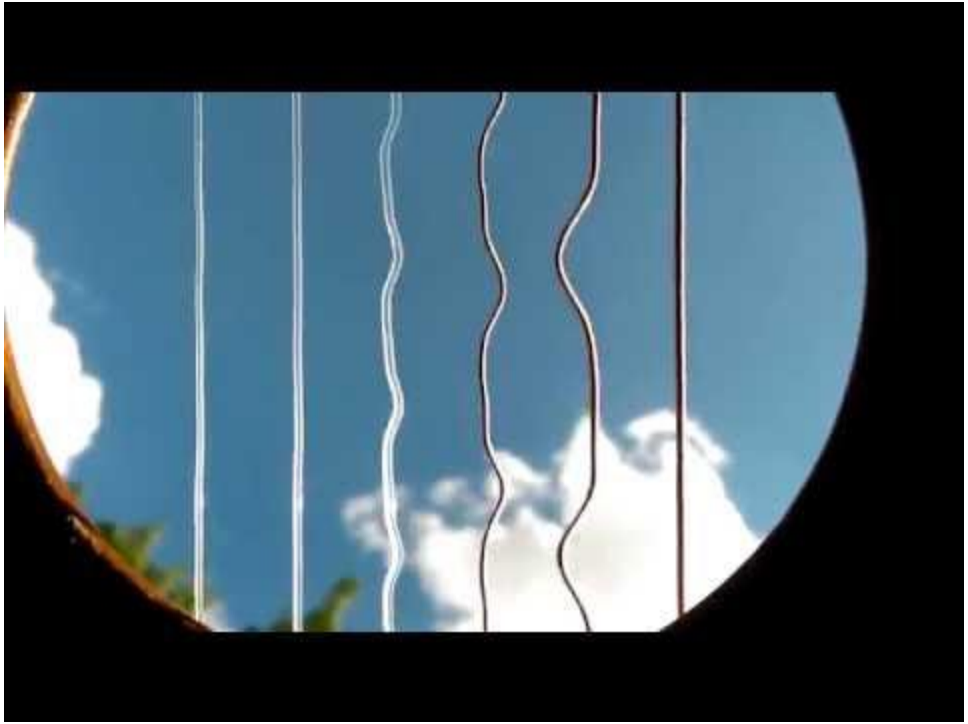
Types of Mechanical Oscillators: Simple Harmonic



Types of Mechanical Oscillators: Elastic



Types of Mechanical Oscillators: String

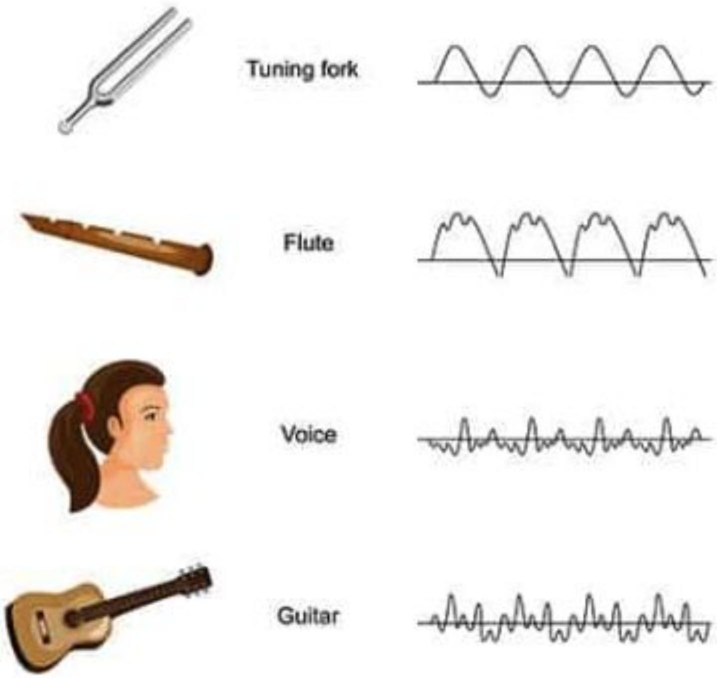


Types of Mechanical Oscillators: Membrane

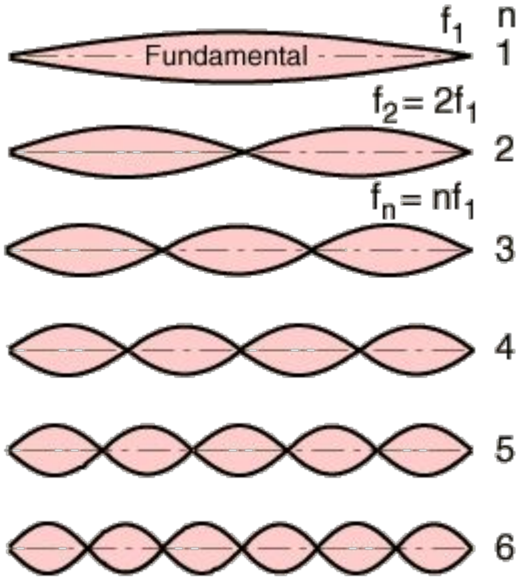
Types of Mechanical Oscillators: Bar

Types of Mechanical Oscillators: Pipes

Vibration Spectra and Timbre

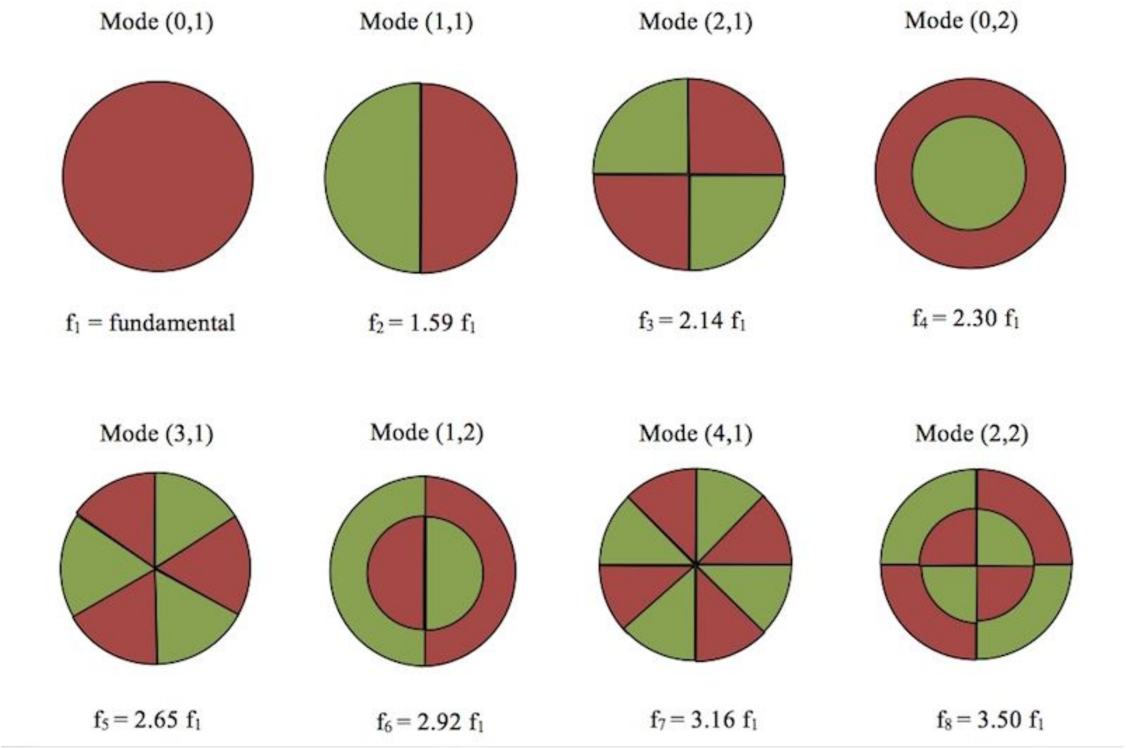


Strings and Modes



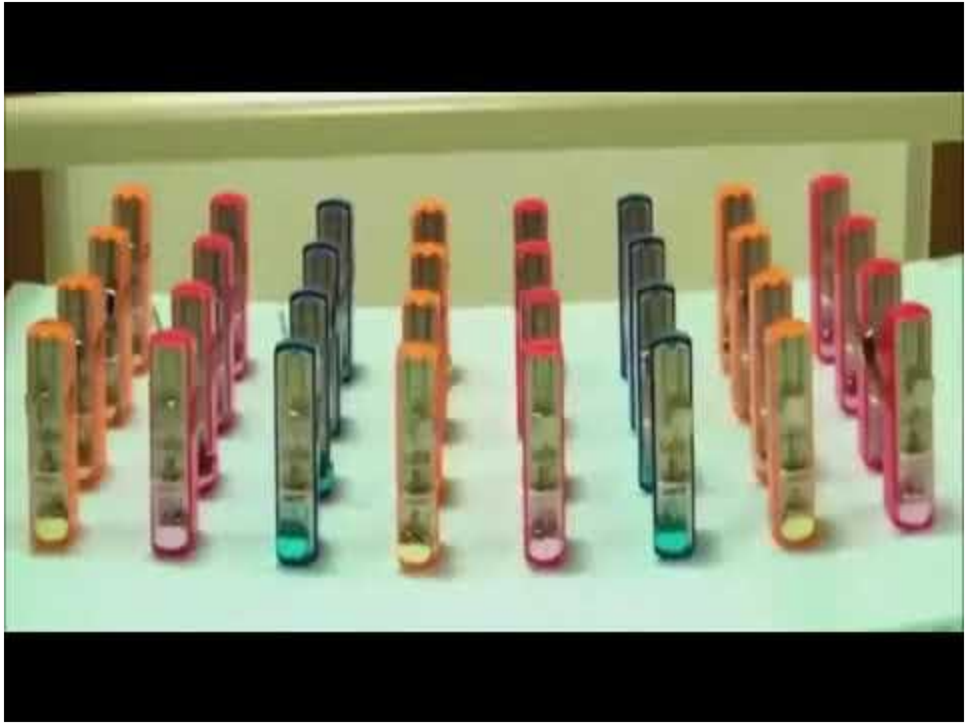
Chladni Patterns

Modes of a Drumhead



ACOUSTICS
&
PSYCHOACOUSTICS

Resonance

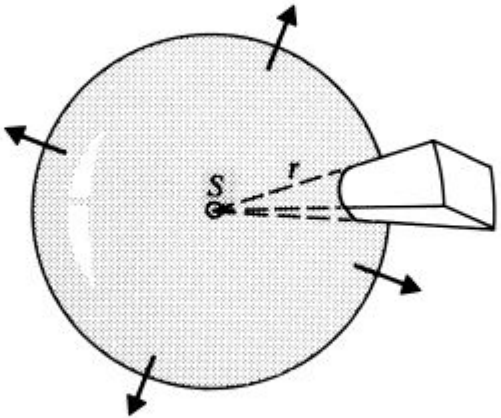


Pressure + Power

- Intensity is the power per unit area of a sound wave
    - $I \sim p/A$
- Intensity and power both grow as pressure squared
    - $W \sim I \sim p^2$

Inverse Square Law

- In free space, acoustic energy (and power) die off with the inverse of the distance squared.



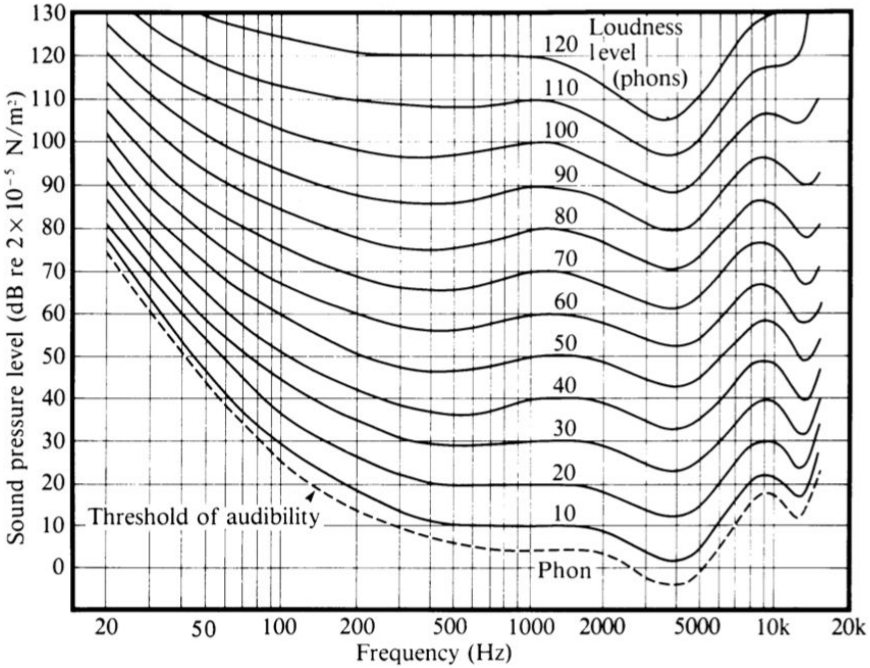
Loudness + Decibels

- Our ears perceive intensity logarithmically
- We often describe sound with "decibels" (dB) relative to some reference intensity, power, or pressure
- A "bel" is the log of relative values, decibel then is 10 bels
- Often, the reference for sound power level is $W_0 = 10^{-12}$ W
- So, $L_W = 10 \log(W/W_0)$

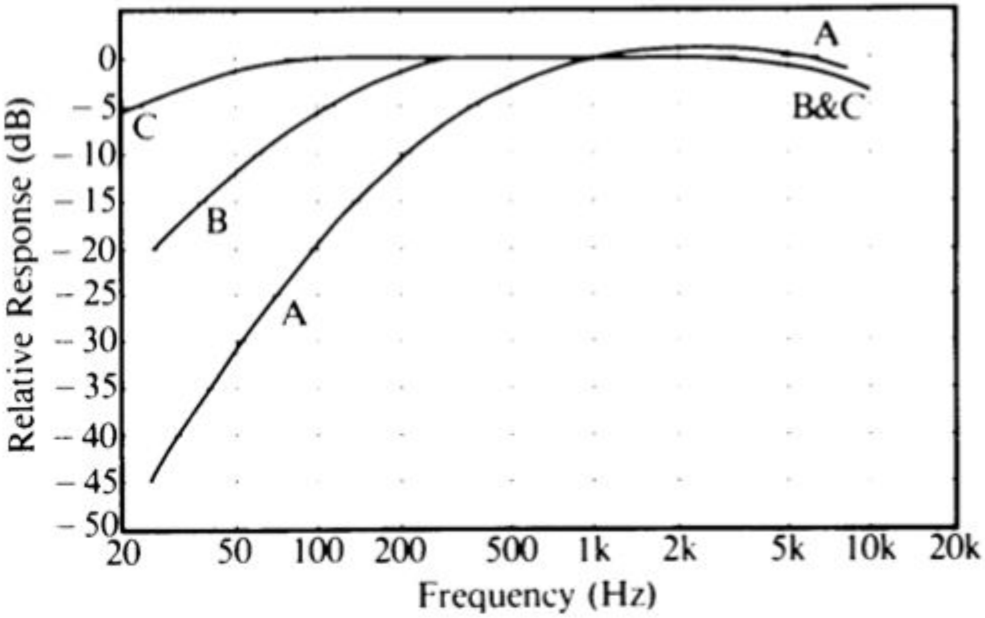
Multiple Sources

- What happens when the number of sources double? What's the change in decibels and, thus, perceived intensity?

Decibel Scales



Decibel Scales

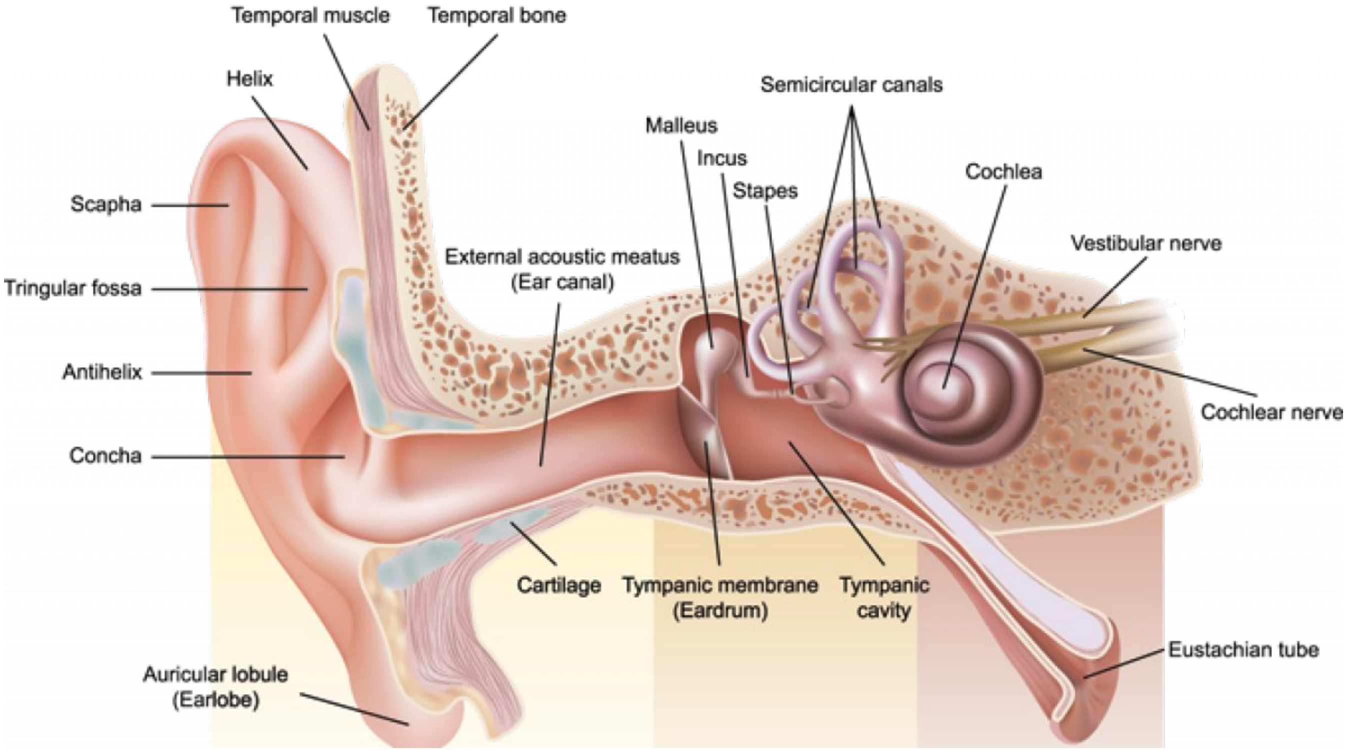


(b)

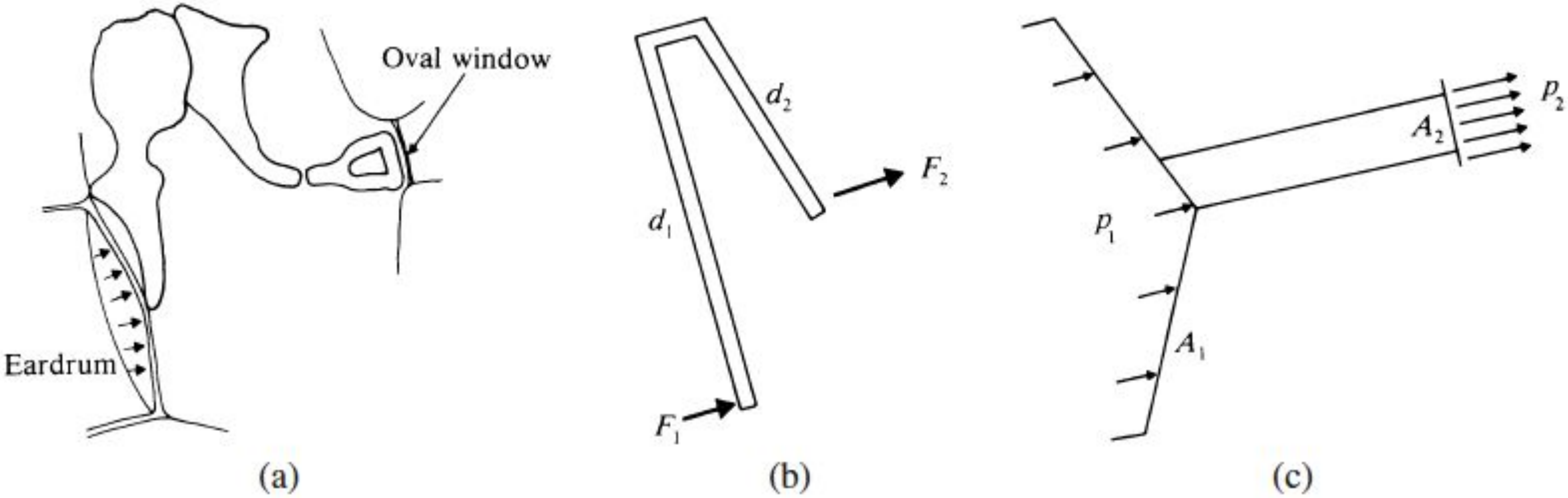
Signal to Noise Ratio (SNR)

- SNR = P_{signal} / P_{noise}
- Typically also presented in dB
  - SNR (dB) = $\log(P_{signal}) - \log(P_{noise})$
- Often matters more for intelligibility than intensity
- Critical to speech recognition accuracy (for both ASR and humans)

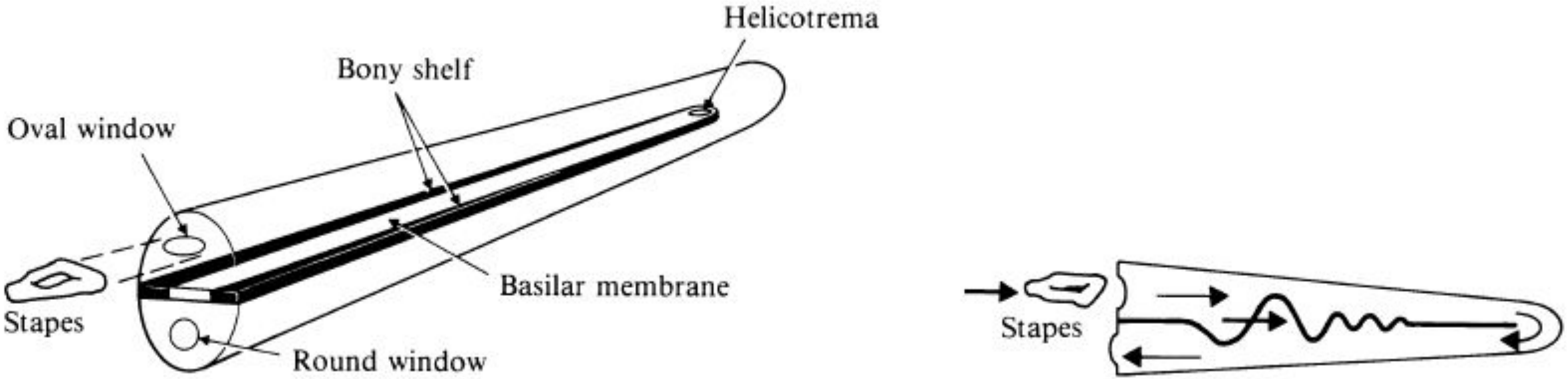
The Ear


Temporal muscle
Temporal bone
Helix
Semicircular canals
Malleus
Incus
Stapes
Cochlea
Scapha
External acoustic meatus
(Ear canal)
Vestibular nerve
Tringular fossa
Antihelix
Cochlear nerve
Concha
Cartilage
Tympanic membrane
(Eardrum)
Tympanic
cavity
Eustachian tube
Auricular lobule
(Earlobe)

The Ear

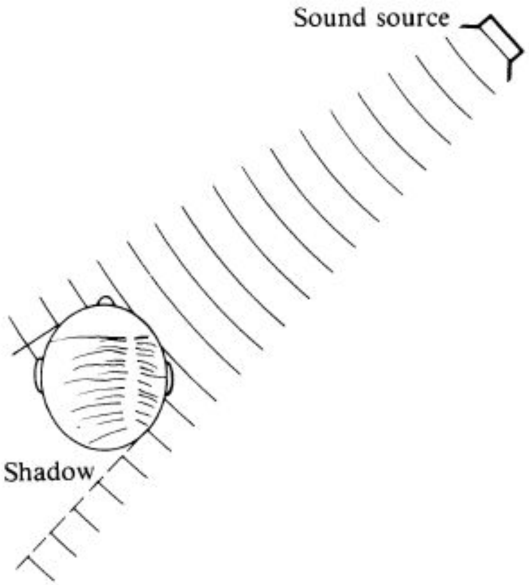


The Ear

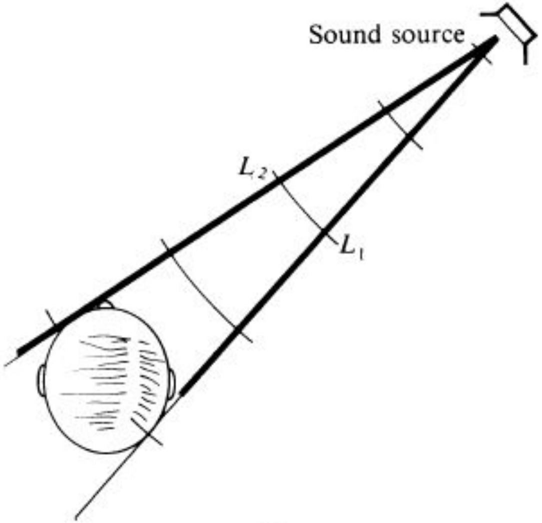

Helicotrema
Bony shelf
Oval window
Basilar membrane
Stapes
Round window
Stapes

(a) (b)

Binaural Hearing



(a)



(b)

Psychoacoustics

- Perception of intensity is logarithmic
- Perception of frequency is logarithmic (and cultural)
- Perception of intensity if frequency dependent
- And weakly on other factors

Fechner's Law

- Perception of most stimulus is logarithmic

Sones

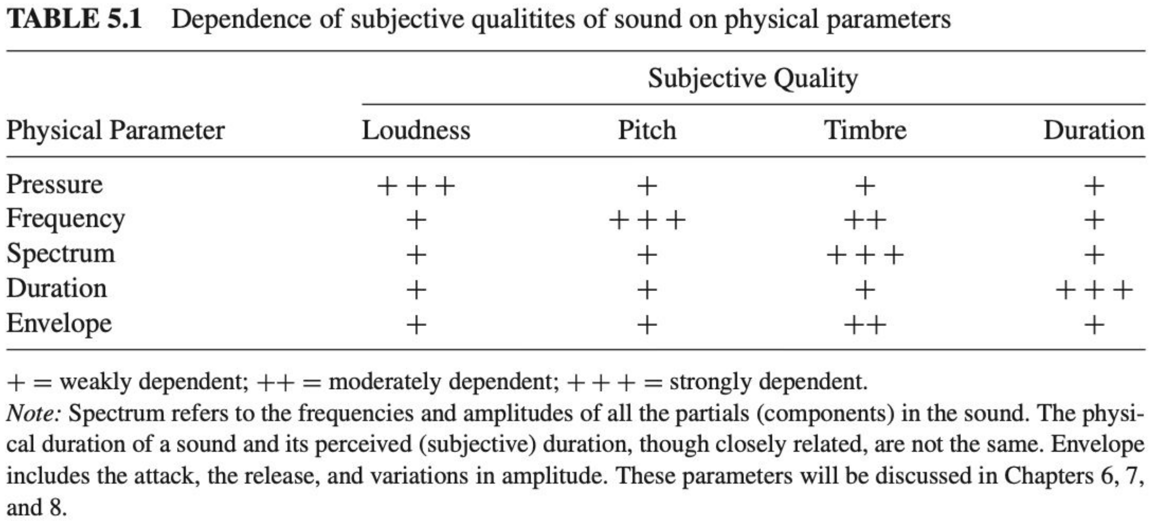


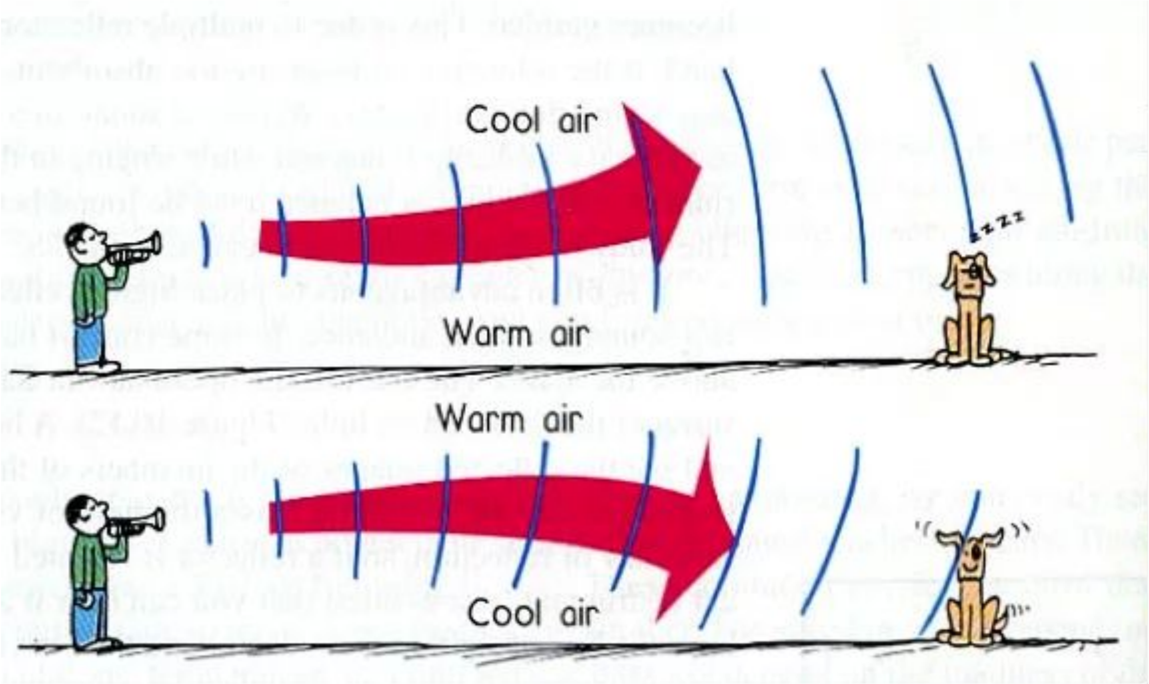
TABLE 5.1 Dependence of subjective qualitites of sound on physical parameters

	Subjective Quality			
Physical Parameter	Loudness	Pitch	Timbre	Duration
Pressure	+ + +	+	+	+
Frequency	+	+ + +	++	+
Spectrum	+	+	+ + +	+
Duration	+	+	+	+ + +
Envelope	+	+	++	+

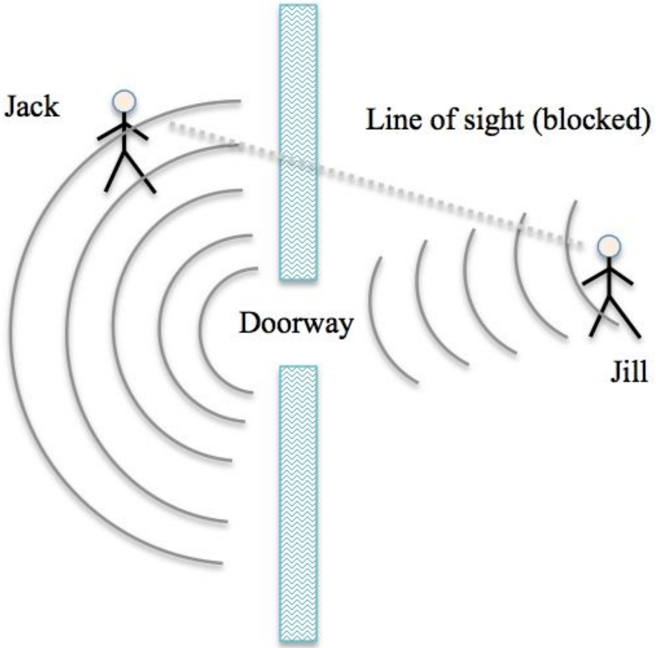
+ = weakly dependent; ++ = moderately dependent; + + + = strongly dependent.
Note: Spectrum refers to the frequencies and amplitudes of all the partials (components) in the sound. The physical duration of a sound and its perceived (subjective) duration, though closely related, are not the same. Envelope includes the attack, the release, and variations in amplitude. These parameters will be discussed in Chapters 6, 7, and 8.

Reflection

Refraction



Diffraction



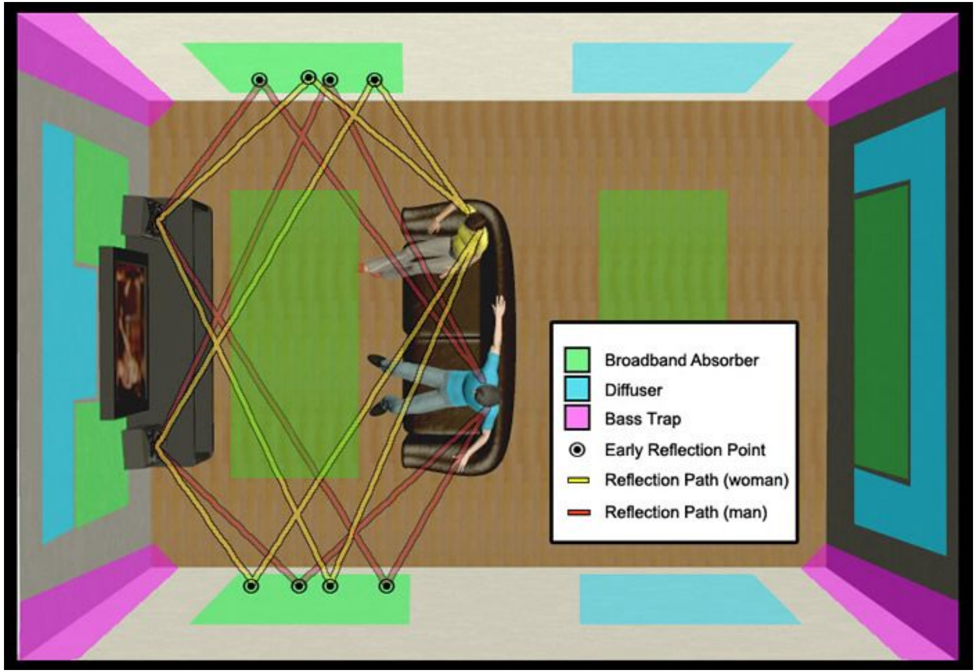
Interference

Dispersion

Doppler Effect

Reverberation

Room Acoustics



MICROPHONES
&
RECORDING

Sensors

- Chains of transducers changing the form of energy, typically ending with an electric signal.

Pressure Measurement

- Gauge pressure is a measurement of pressure made relative to the ambient pressure. I.e. tire pressure reads 35 psi.

Pressure Measurement

- Absolute pressure is a measurement made relative to a pure vacuum condition, such as the vacuum of space. I.e. air pressure.

Pressure Measurement

- Differential pressure is a measurement of the pressure difference between two pressure values. I.e. pitot tubes

Pressure Measurement

- Vacuum pressure is the pressure measurement of values that are in a negative direction with respect to atmospheric pressure. I.e. on a vacuum pump

Pressure Sensors

- Potentiometric pressure sensors use a Bourdon tube, capsule, or bellows which drives a wiper arm, providing relatively course pressure measurements.
- Inductive pressure sensors use a linear variable differential transformer (LVDT) to vary the degree of inductive coupling that occurs between the primary and secondary coils of the transformer.

Pressure Sensors

- <u>Capacitive pressure sensors</u> use a diaphragm that is deflected by the applied pressure which results in a change in the capacitance value, which can then be calibrated to provide a pressure reading.
- <u>Piezoelectric pressure sensors</u> rely on the ability of materials such as ceramic or metalized quartz to generate an electrical potential when the material is subjected to mechanical stress.

Pressure Sensors

- Strain gauge pressure sensors rely on a measurement of the change in resistance that occurs in a material such as silicon when it is subjected to mechanical stress, known as the piezoresistive effect.

Pressure Sensors

- Variable reluctance pressure sensors make use of a diaphragm that is contained in a magnetic circuit. When pressure is applied to the sensor, the diaphragm deflection causes a change in the reluctance of the circuit, and that change can be measured and used as an indicator of the applied pressure.

Types of Microphones:
Crystal

- Vintage broadcasting
- Cheap

Types of Microphones:
Dynamic (Coil)

- Robust
- Low Sensitivity
- Low Distortion

Types of Microphones:
Condenser Microphone

- Powered
- High frequency response

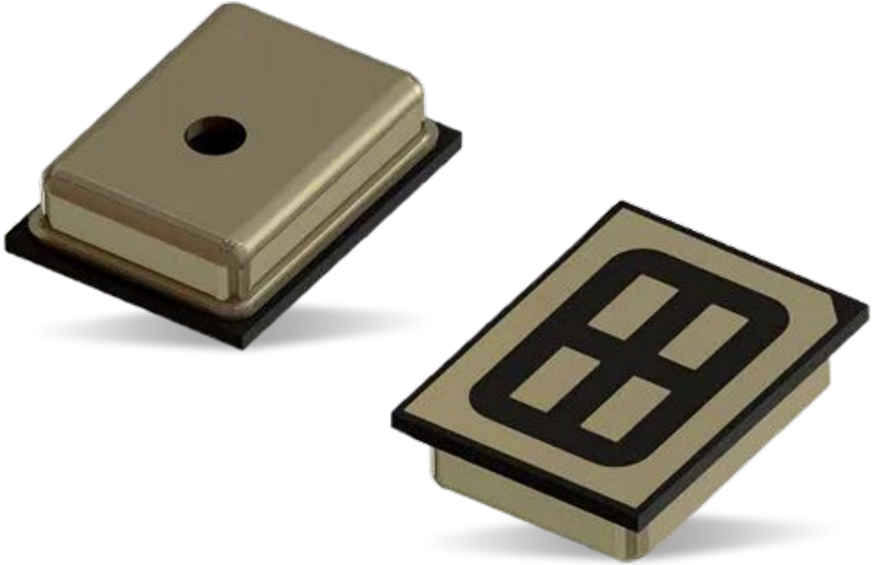
Types of Microphones:
Electret-Condenser Microphones

- Like Condenser but lower voltage requirements

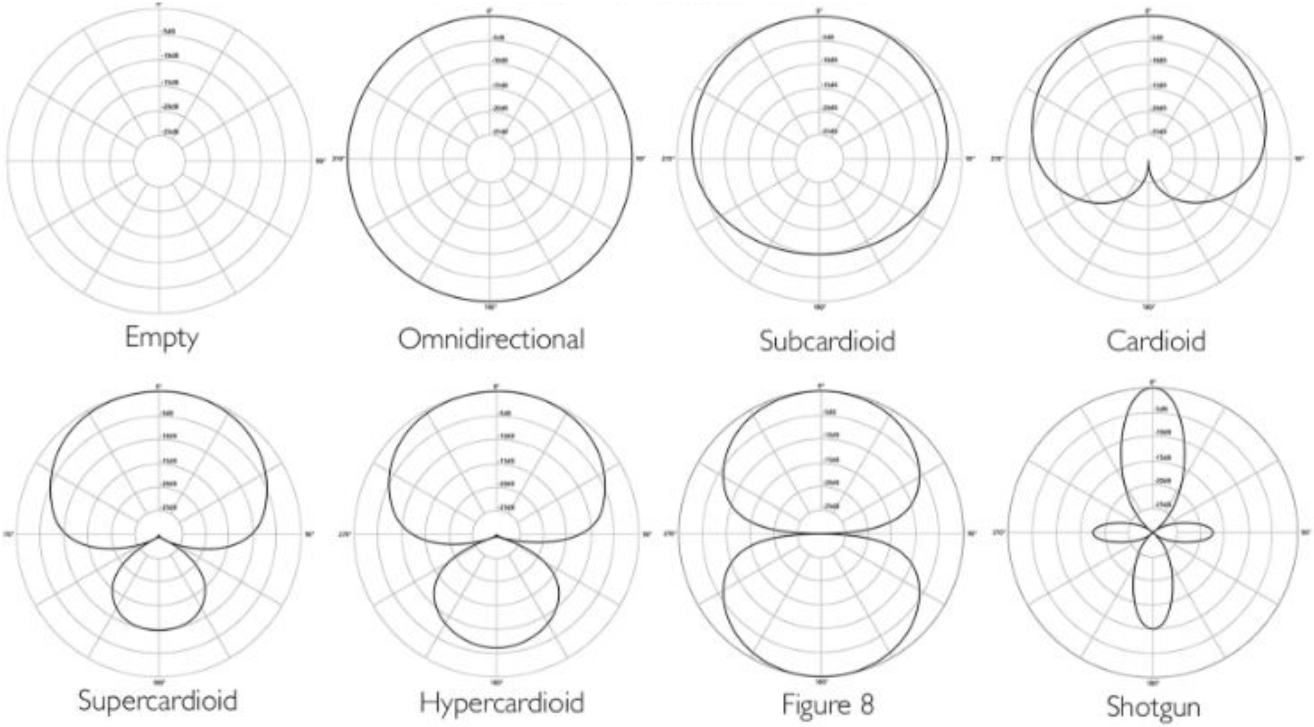
Types of Microphones: Ribbon Microphone

- Responds to velocity not pressure
- Early radio
- Specialized (very close singer)

Types of Microphones:
MEMS Microphone



Pickup Patterns



Other Microphone Properties

- Microphone impedance
- Microphone sensitivity
    - Voltage sensitivity
    - Power sensitivity

Amplifiers

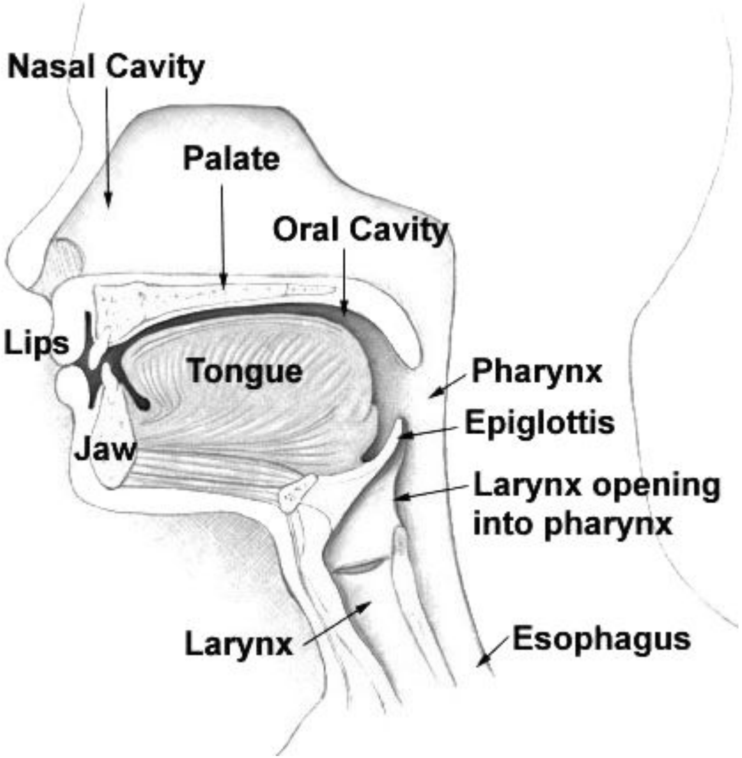
- Often required by the microphone type before digitization or playback
- Gain expressed in dB
- Can be a major source of noise

Distortion

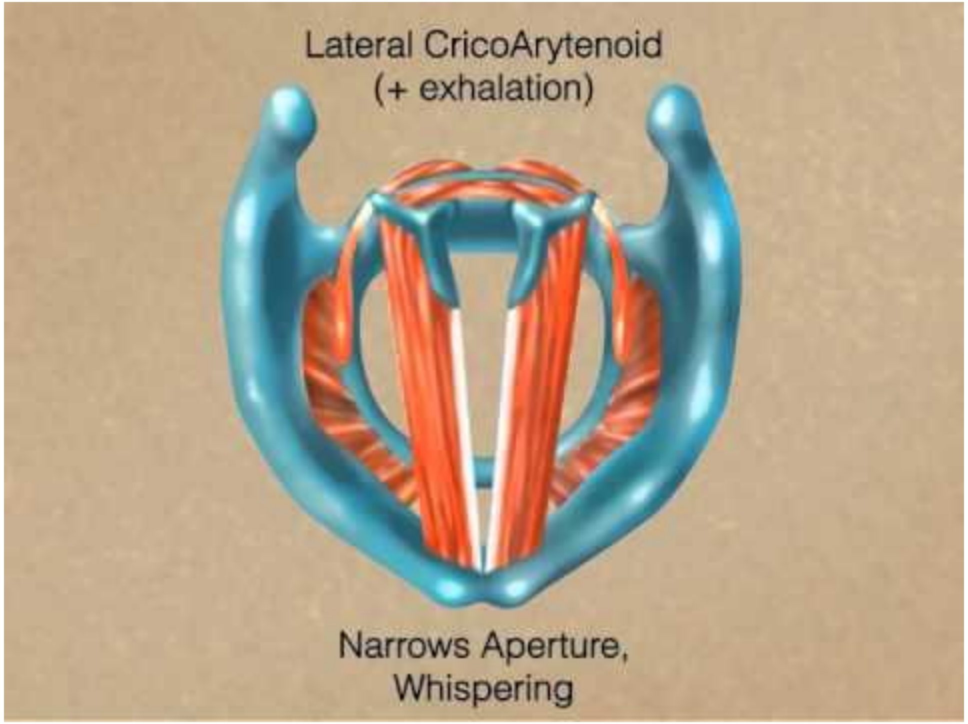
- Harmonic Distortion
  - Clipping
- Intermodulation Distortion
  - Tones of two different frequencies
- Transient distortion
  - Rapidly changing signal
- Many other types!
  - Envelope, digital clipping, aliasing, saturation, phase…

HUMAN VOICE

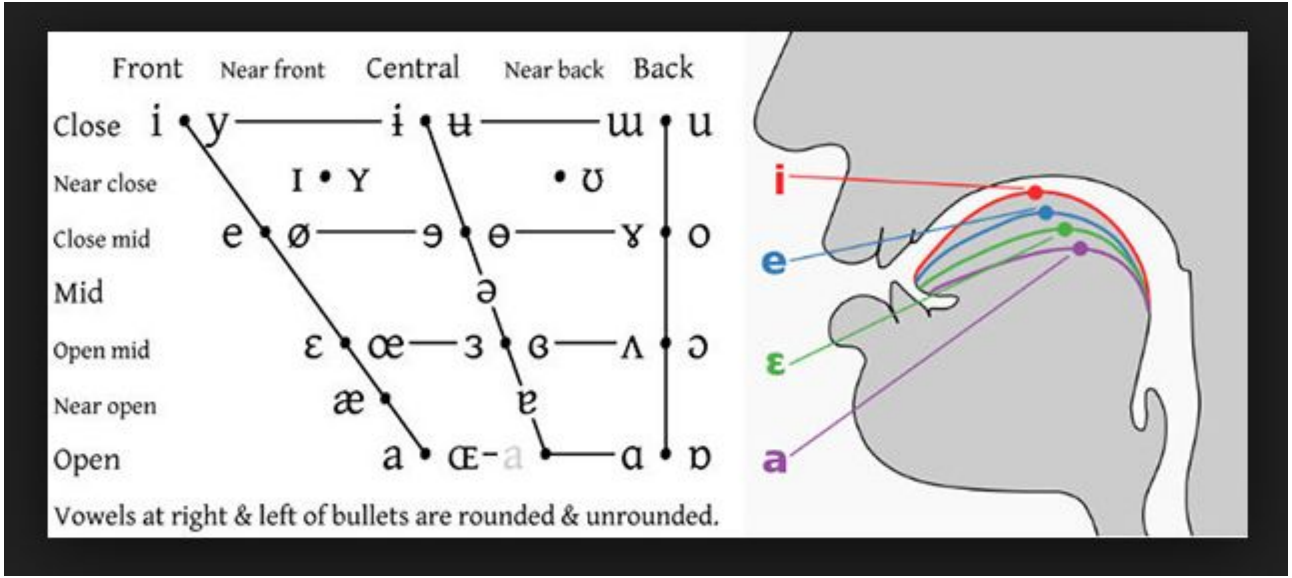
Human vocal apparatus



Human vocal apparatus

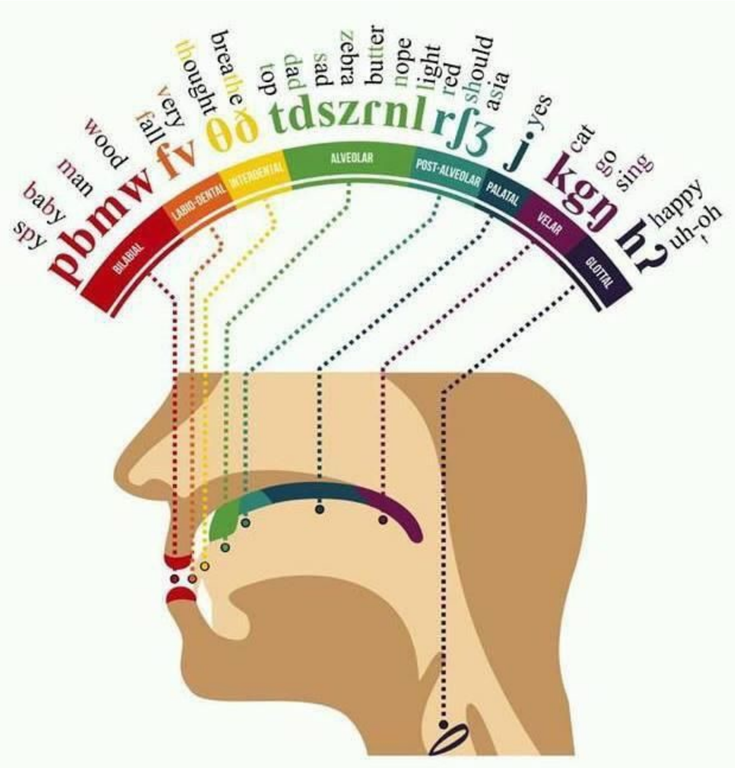


Human vocal apparatus



Vowels at right & left of bullets are rounded & unrounded.

Human vocal apparatus



Exercises for Next Time

- A vuvuzela produces 116 dB at 1m. How loud is it a soccer field away in dB? How loud would it be if there were 20,000 people in a stadium playing vuvuzelas at that distance?

- In the trenches of WW1, on September 28, 1915, German artillery in Belgium could be heard more than sixty miles away, however not between thirty and sixty miles away. Why not?

- A voiced consonants uses the vocal cords. You can tell if a sound is voiced by touching your throat when you make the sound. /z/ (as in "<u>z</u>inc") is the voiced version of the /s/ (as in "<u>s</u>ink") alveolar fricative. What is the voiced version of the palato-alveolar fricative /ʃ/ (as in "<u>sh</u>ip")?

Exercises from Last Time

- A vuvuzela produces 116 dB at 1m. How loud is it a soccer field away in dB? How loud would it be if there were 20,000 people in a stadium playing vuvuzelas at that distance?

Soccer field is ~100m. By $1/r^2$, $116dB - 10*\log_{10}((1m/100m)^2) = 116dB - 40dB = 76dB$
Now, if there are 20,000 of them: $76dB + 10*\log_{10}(20{,}000) = 76dB + 43dB = 119dB$
Hearing loss occurs at 120dB!!

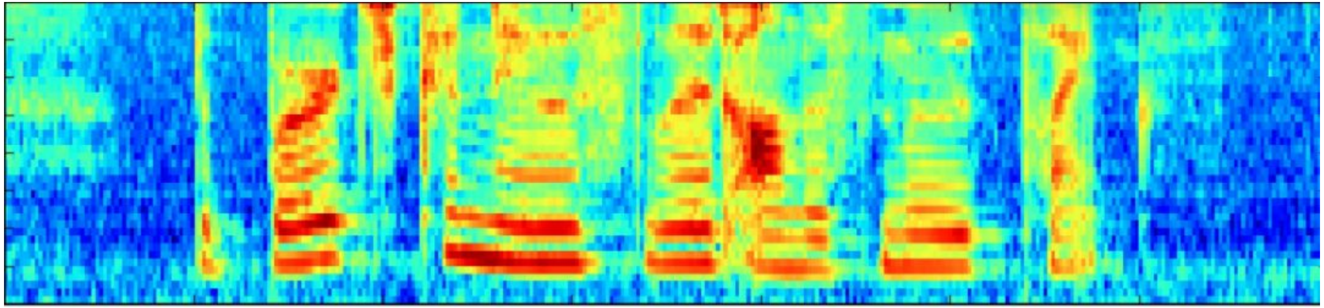
- In the trenches of WW1, on September 28, 1915, German artillery in Belgium could be heard more than sixty miles away, however not between thirty and sixty miles away. Why not?

A thermal gradient caused distant sound to refract over a dead zone.

- A voiced consonants uses the vocal cords. You can tell if a sound is voiced by touching your throat when you make the sound. /z/ (as in "zinc") is the voiced version of the /s/ (as in "sink") alveolar fricative. What is the voiced version of the palato-alveolar fricative /ʃ/ (as in "ship")?

The voiced palato-alveolar fricative is /ʒ/ as in "pleasure" and "vision".

Stay tuned on Wednesday for how to make things that look like this:



and how to interpret signals like this: