Sound and Hearing
“Sound” is the rhythmic compression and decompression of the air around us caused by a vibrating object.
Sound Wave: Amplitude and Frequency (Hz)

Sound Pressure is measured in units called Pascals
1 Pascal (Pa) = 1 Newton of force/m²
1 atmosphere = 100,000 Pa
Human absolute hearing threshold = 0.00002 Pa = 20 microPa
(i.e., 2 ten billionths of an atmosphere)

Frequency measured in cycles/sec = Hertz (Hz)
Nominal range of sensitivity: 20 – 20,000 Hz
The “decibel” (dB)

The decibel is a logarithmic unit used to describe a ratio (i.e., log \( (x/y) \))

In engineering analyses, it is used to normalize “power” measurements to a known reference and then compresses the resulting ratio using a \( \log_{10} \) operation.

This format is convenient for engineering analyses involving wide dynamic ranges (when very small and the very large magnitudes must be considered simultaneously).

\[
\text{dB} = 10 \log(\text{Observed Power} / \text{Reference})
\]
The transducers (microphones) on sound level meters measure sound pressure (i.e., N/m\(^2\) or Pascals).

Pressure needs to be converted to power prior to calculation of the decibel equivalent....i.e., \textit{acoustic power} = \textit{pressure}^2

Finally, we need to agree upon a Reference value. By convention, we use \textit{20 microPa} (i.e., the hearing threshold)

Thus:
\[ \text{dB} = 10 \log \left( \frac{\text{Observed Pressure}^2}{20 \text{ microPa}^2} \right) \]

However........
Prior to the advent of hand-held calculators and computers (circa 1970), performing a squaring operation was *computationally expensive* and prone to error.

To reduce computational demands, *hearing science* adopted a somewhat confusing convention in the specification of the $\text{dB}_{\text{SPL}}$ unit:

$$\text{dB}_{\text{SPL}} = 20 \log \left( \frac{\text{Observed Sound Pressure}}{20 \text{ microPa}} \right)$$

+6 $\text{dB}_{\text{SPL}}$ = doubling sound pressure  
+3 $\text{dB}_{\text{SIL}}$ = doubling acoustic power  
+20 $\text{dB}_{\text{SPL}}$ = 10x pressure  
+10 $\text{dB}_{\text{SIL}}$ = 10x acoustic power
## Some Typical Sound Amplitude Values

<table>
<thead>
<tr>
<th>Sound</th>
<th>Pressure Level (pascals)</th>
<th>Pressure Ratio $(P_s/P_r)$</th>
<th>Intensity Ratio $(I_s/I_r)$</th>
<th>dB_{SPL}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal audible sound</td>
<td>0.00002 $(P_r)$</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Soft whisper</td>
<td>0.0002</td>
<td>10</td>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td>Quiet office</td>
<td>0.002</td>
<td>$10^2$</td>
<td>$10^4$</td>
<td>40</td>
</tr>
<tr>
<td>Average conversation</td>
<td>0.02</td>
<td>$10^3$</td>
<td>$10^6$</td>
<td>60</td>
</tr>
<tr>
<td>Vacuum cleaner</td>
<td>0.2</td>
<td>$10^4$</td>
<td>$10^8$</td>
<td>80</td>
</tr>
<tr>
<td>Subway train</td>
<td>2</td>
<td>$10^5$</td>
<td>$10^{10}$</td>
<td>100</td>
</tr>
<tr>
<td>Loud thunder</td>
<td>20</td>
<td>$10^6$</td>
<td>$10^{12}$</td>
<td>120</td>
</tr>
<tr>
<td>Jet engine at takeoff (pain threshold)</td>
<td>200</td>
<td>$10^7$</td>
<td>$10^{14}$</td>
<td>140</td>
</tr>
<tr>
<td>Wind tunnel</td>
<td>2,000</td>
<td>$10^8$</td>
<td>$10^{16}$</td>
<td>160</td>
</tr>
<tr>
<td>Space shuttle</td>
<td>20</td>
<td>$10^9$</td>
<td>$10^{18}$</td>
<td>180</td>
</tr>
</tbody>
</table>
More about those pesky decibels

- **JND** for sound intensity is about 1 dB\textsubscript{SPL} for most of normal range of hearing
- **What does 0 dB\textsubscript{SPL} mean?**
  
  Hint: \(20 \log \left( \frac{20 \text{ microPa}}{20 \text{ microPa}} \right) = 0 \text{ dB}\textsubscript{SPL}\)
- If one machine emits 80 dB\textsubscript{SPL} then how much sound amplitude would be expected from two machines side-by-side?

\[
2 \times 80 = 160 \text{ dB}\textsubscript{SPL} \quad (That’s pretty intense)
\]

Convert from dB\textsubscript{SPL} back to raw pressure, sum the pressures, then convert sum to dB\textsubscript{SPL}

\[
\begin{align*}
80 \text{ dB}\textsubscript{SPL} & \rightarrow \text{antiLog}(80/20) \rightarrow 10,000 \\
20 \log (10,000+10,000) & = 86 \text{ dB}\textsubscript{SPL} \quad (approx.)
\end{align*}
\]
Inverse-Square Law

Area of sphere = $4\pi r^2$
<table>
<thead>
<tr>
<th>dB_{SPL}</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>130</td>
<td>Loud hand clapping at 1 m distance</td>
</tr>
<tr>
<td>110</td>
<td>Siren at 10 m distance</td>
</tr>
<tr>
<td>95</td>
<td>Hand (circular) power saw at 1 m</td>
</tr>
<tr>
<td>80</td>
<td>Very loud expressway traffic at 25 m</td>
</tr>
<tr>
<td>60</td>
<td>Lawn mower at 10 m</td>
</tr>
<tr>
<td>50</td>
<td>Refrigerator at 1 m</td>
</tr>
<tr>
<td>40</td>
<td>Talking; Talk radio level at 2 m</td>
</tr>
<tr>
<td>35</td>
<td>Very quiet room fan at low speed at 1 m</td>
</tr>
<tr>
<td>25</td>
<td>Normal breathing at 1 m</td>
</tr>
<tr>
<td>0</td>
<td>Absolute threshold</td>
</tr>
</tbody>
</table>
Most Sound Stimuli are Complex
Complex Sound = Sum of Sines
(Fourier Theorem Revisited)

J.B.J. Fourier
(1768-1830)

Fourier Sound Applet
Acoustic energy results from a traveling wave of rhythmic “compression” through a physical medium (e.g., air; water; steel).

It is the “compression” that travels not the medium, *per se*.

The characteristic speed of this travelling wave varies as a function of the medium (elasticity; density).

The speed of acoustic energy through the air (aka “sound”) is 331 m/sec (or **742 MPH**) at 0-deg C (Faster at higher temperatures).

<table>
<thead>
<tr>
<th>Medium</th>
<th>Sound Speed (metres/second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gas</td>
<td></td>
</tr>
<tr>
<td>Carbon dioxide</td>
<td>268</td>
</tr>
<tr>
<td>Air</td>
<td>331</td>
</tr>
<tr>
<td>Helium</td>
<td>972</td>
</tr>
<tr>
<td>Liquid</td>
<td></td>
</tr>
<tr>
<td>Ethanol</td>
<td>1,130</td>
</tr>
<tr>
<td>Fresh water</td>
<td>1,402</td>
</tr>
<tr>
<td>Sea water</td>
<td>1,522</td>
</tr>
<tr>
<td>Solid</td>
<td></td>
</tr>
<tr>
<td>Brass</td>
<td>4,700</td>
</tr>
<tr>
<td>Steel</td>
<td>5,790</td>
</tr>
<tr>
<td>Aluminum</td>
<td>6,420</td>
</tr>
</tbody>
</table>
Gross Anatomy of the Ear

- Flange
- Concha
- Semicircular canals
- Temporal bone
- Vestibulocochlear nerve
- Cochlea
- Oval window
- Round window
- Auditory ossicles
- Tympanic membrane
- External auditory canal
- Eustachian tube

Legend:
- Orange: Outer ear
- Purple: Middle ear
- Blue: Inner ear
Flow of Acoustic Energy
(The “Impedance Problem”)

- Malleus
- Incus
- Scala vestibuli
- Scala tympani
- Oval window
- Stapes
- Round window
- Basilar membrane
- Cochlea
- Tympanic membrane
- External auditory canal
- Sound waves

Legend:
- Yellow: Outer ear
- Purple: Middle ear
- Blue: Inner ear
99.9% of sound energy in the air is reflected at the air:water boundary \((10 \log(0.1/100)) = -30 \text{ dB loss}) \(1/1000x\)

How does the ear compensate for this loss as sound energy is transmitted from the air to the fluid that filled the cochlea?

2 dB gain via ossicular leverage \((1.6x)\)

25 dB gain via surface area condensation \((\text{eardrum} \rightarrow \text{stapes}) \ (316x)\)

~5 dB gain at mid-frequencies \((3x)\) due to pinna and auditory canal resonance
The Cochlea

- Vestibular organ
- Vestibulocochlear Nerve
- Stapes
- Cochlea
- Apex
- Helicotrema
- Base
- Reissner's membrane
- Scala vestibuli
- Cochlear duct
- Scala tympani
- Basilar membrane
- Stapes
- Oval window
- Round window
The Organ of Corti

3000-3500 Inner Hair Cells (IHC)
12,000 Outer Hair Cells (OHC)
Photomicrograph: Sensory Hair Cells

Three rows of Outer Hair Cells

One Row of Inner Hair Cells
Auditory Transduction
Basilar Membrane Modulation Effects upon Sensory Hair Cells

Note: $K^+$ ion concentration gradient across sensory hair cells (see pink cavities)
IHC Stereocilia “Tip Links”

“tip link” connects gate to adjacent cilia.

Shearing motion forces gate to open.

Mechanical open-and-close of gate modulates influx of potassium ions (much FASTER than slow chemical cascade in visual transduction).

$K^+$ depolarization of IHC triggers release of glutamate at cochlear nerve fiber synapse.
IHC Auditory Transduction
Innervation of 3000 IHCs versus 12,000 OHCs

30,000+ fibers in cochlear nerve. Nearly 10:1 fiber-to-IHC innervation ratio.

Sparse number of fibers carry info from OHC to brain.

Small number of fibers descend from brain to OHCs.

Role of OHC’s? Mechanical gain otoacoustic emission
Sound Amplitude Coding
(“Divide and Conquer”)

Multiple nerve fibers for each IHC.

Each nerve fiber tuned to a different 40 dB “range” of stimulus intensity.

Intensity-level multiplexing
Q: Why the broadening and asymmetry?  
A: Look to the Basilar membrane’s response
Ascending Pathways

- **Temporal lobe**: Auditory cortex
- **Thalamus**: Medial geniculate nucleus
- **Midbrain**: Inferior colliculus, Superior olive
- **Medulla**: Cochlear nucleus
- **Nerve VIII**: Cochlea

![Diagram of ascending pathways in the brain](https://example.com/diagram.png)
Tonotopic Organization of Primary Auditory Cortex (A1)

Also note:

Segregation of monaural versus binaural cells is maintained.

Binaural cells loosely organized according to spatial location of stimulus source.
Auditory Frequency Coding

(What is the neural code for “pitch”?)
Frequency Mechanism versus Place Mechanism

**Frequency Theory**
Ernest Rutherford (1871-1937)

**Place Theory**
Georg von Békésy (1899-1972)
Frequency Theory (Rutherford)

• Basilar membrane analogy to microphone diaphragm
• Each oscillation yields nerve pulse
• **Problem**: Max. neural response approx. 500 Hz
• **Solution**: Time division multiplexing (aka “Volley Principle”)
  Supported by “cochlear microphonic” (Wever & Bray; but consider Botox results)
von Békésy Place Theory:  
Focus on Basilar Membrane Dynamics
The Simple Beginnings for von Békésy’s Nobel Prize

- Low Frequency (Apex)
- Forearm as Organ or Corti
- High Frequency (Base)
- Basilar Membrane
- Eardrum
- Brass Tube Scala Tympani
- Malleus
Basilar Membrane Response to Pure Tone Stimulus

- Base: High frequency
- Apex: Low frequency

Dimensions:
- Base to Apex: 0.45 mm
- Base width: 0.15 mm
- Apex width: 34 mm
Von Békésy’s “Place Mechanism” as Biological Fourier Analyzer

Basilar Membrane Dynamic Simulation (animation)
Functional Aspects of Hearing
Human “Earscape”

- Pain threshold
- “High risk” threshold
- Audible range
- Music
- Speech
The average detection threshold for 18-yr-olds for a 1KHz tone at sea level is 20 microPascal (μPa).

Minimum occurs at approx. 3 KHz.

Binaural intensity thresholds are 3 dB lower than mono.
Clinical Audiogram (dB\textsubscript{HL})

dB-HL (Hearing Level) uses a different reference level for each test frequency.

That reference level represents the average threshold (18 yr-olds) demonstrated at that frequency.

Hence, a value of 0 dB-HL means “average” hearing level at the frequency under test.
Normal vs. Noise-Induced Hearing Loss

Note “notch” at 4 KHz.

Source: http://mustelid.physiol.ox.ac.uk/drupal/?q=acoustics/clinical_audiograms
Age-related Hearing Loss (Presbycusis)

Inevitable or preventable?
Loudness is non-linear
Stevens’ SONE SCALE of Loudness Perception

Perceptual Anchor:
1 sone = loudness of 1 KHz at 40 dB (40 phons)

Find the dB level that is twice as loud (2 sones) or half as loud (0.5 sones), etc. and construct a scale.
[i.e., Magnitude Estimation]

The psychological magnitude of sound (i.e., “Loudness”) grows at a slower rate than the physical magnitude of the sound stimulus.
Loudness

Using magnitude estimation techniques, S.S. Stevens has quantified this nonlinear relationship as:

\[
L = k \times P^{0.6} = k \times I^{0.3}
\]

$L=$loudness; $P=$sound pressure (µPa)
$I=$sound intensity (pW/m²)

**Stevens’ Power Law**; Linear in log-log plot; slope ≈ exponent

\[
\log(L)=\log(k)+0.3 \log(I) \quad \text{straight line}
\]

\[
\log(L)\approx0.3 \log(I)
\]

Hence, a log unit increase (10dB) of intensity yields 0.3 log (10^{0.3} or 2-fold) increase in loudness.

**Note**: Binaural presentation perceived as approx. 2x more loud than monaural equivalent.
# Sone Scale Landmarks

<table>
<thead>
<tr>
<th>Source Description</th>
<th>Sone Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal conversation</td>
<td>1-4</td>
</tr>
<tr>
<td>Automobile @ 10m</td>
<td>4-16</td>
</tr>
<tr>
<td>Vacuum cleaner</td>
<td>16</td>
</tr>
<tr>
<td>Major roadway @ 10 m</td>
<td>16-32</td>
</tr>
<tr>
<td>Long-term hearing damage dosage</td>
<td>32+</td>
</tr>
<tr>
<td>Jackhammer @ 1m</td>
<td>64</td>
</tr>
<tr>
<td>Brief-exposure hearing damage</td>
<td>256</td>
</tr>
<tr>
<td>Pain threshold</td>
<td>676</td>
</tr>
</tbody>
</table>
Pitch = f(Frequency)

MEL Scale

Reference unit of perceived PITCH:
1000 Hz = 1000 Mels

Perceived pitch increases “linearly” with stimulus frequency below 4KHz; but grows at a much slower rate at 4KHz and above.
Temporal Summation (< 200 msec) 
Complements Binaural (i.e., Spatial) Summation
Frequency differentiation is flattened at high amplitudes; Speech and music sounds “tinny” at high loudness levels; Remember change in cochlear nerve tuning at higher intensity levels.
Sound Localization
Localization Accuracy vs. Frequency

Signature of a dual-mechanism process?
Localization Accuracy vs. Frequency:
Low Freq – Interaural Time Difference
High Freq – Interaural Intensity Difference
Sound Shadowing
(Interaural Intensity Difference – IID)

High-frequency sound waves are “blocked” by the human head and cast a “shadow” at the far ear (Strong IID cue).

Low-frequency sound waves wrap easily around the head and cast little or no sound shadow (Weak IID Cue).
\[ \text{IID} = f(\text{Location}, \text{Frequency}) \]
ITD versus Location

ΔT

Interaural Time Difference (ms)

Azimuth Angle

0 30° 60° 90° 120° 150° 180°

0 0.1 0.2 0.3 0.4 0.5 0.6

- Straight Ahead
- Right Ear (Perpendicular)
- Straight Behind
Delay Line Theory
(How to Build a Cell tuned to delta-T Signals)

Delta-T = 200 microsec
“Active” Localization
(Continuous Sound Sources)
Relatively good localization performance despite same IID and ITD levels (i.e., zeros)

Differential sound distortion ("coloration") introduced by interaction with pinna

Modifying shape of pinna causes immediate reduction in localization accuracy (Hoffman, et al., 1998)

Listening through the ears of another yields “ahead” vs. “behind” confusion (chance performance)
Modifying the Pinna Transfer Function

(Hoffman, et al., 1998)

Earprints?
Cross-Section of a Head-Related Transfer Function
(Spectral Coloration by Head, Torso & Pinnae)
Auditory/Visual Integration

**Ventriloquism Effect**
Visual capture of sound localization

**McGurk Effect**
“Compromise” between conflicting sound and visual cues in speech understanding

*What you hear is what you see*