Human Auditory System

- Anatomy and physiology of human ears
  - Anatomy
  - Physiology
- Psychoacoustics
  - Frequency masking
  - Temporal masking
- Demos
  - Anatomy, masking, emotion, deception
A Demo of Human Auditory System

An Auditory Tour
Physiology of Human Auditory System

- Outer ear: collects sound waves
  - amplifies sound waves in some frequencies
  - vibrations of air are translated to vibrations of the tympanic membrane

- Middle ear: vibrations of tympanic membrane are translated to oscillations of liquid in inner ear
  - performed by the ossicles
  - amplification 15:1

- Inner ear:
  - Cochlea transforms mechanical vibrations into nerve impulses
    - Movement of basilar membrane causes the hairs to bend (3K hairs in cochlea)
    - What information is in the nerve impulses? Not well understood
  - nerve endings (30K nerve fibers).
Anatomy of Human Auditory System

- Outer ear: it channels sound waves through the ear canal to the eardrum.
- Middle ear: vibrations caused air pressure changes in the ear canal are transmitted to three small bones called "ossicles".
- Inner ear: it houses the "cochlea", a spiral-shaped structure that contains the organ of "Corti" - the most important component of hearing. The Corti sits in an extremely sensitive membrane called the "basilar membrane". Whenever the basilar membrane vibrates, small sensory hair cells inside the Corti are bent, which stimulates the sending of nerve impulses to the brain.
Physiology of Human Auditory System

- Frequency discrimination
  - Early work suggested that specific places in the cochlea may be responsible for hearing specific frequencies of sound.
  - Place principle: each hair cell and neuron in the cochlea is tuned to respond to a specific frequency and that that frequency can be determined based on its position in the cochlea.
  - Volley principle: frequency discrimination is not done based on basilar membrane resonance, but on timing information.
Physiology of HAS (Con’t)

- **Amplitude discrimination**
  - Sound Pressure Level (SPL) \(L_{SPL} = 20 \log_{10} \frac{p}{p_0} (dB)\)
  - \(p\) – sound pressure of stimulus in Pascals, \(p_0\) - 20\(\mu\)Pa
  - Human hearing has a dynamic range of approximately 110dB

- **Temporal information**
  - Phase locking: neurons seem to prefer to fire at only certain times in their center frequency's waveform
  - Used to tell the difference in the position of a sound source
Psychoacoustics Overview

- **Absolute threshold of hearing**
  - Characterize the amount of energy needed in a pure tone such that it can be heard

- **Critical bands**
  - Nonlinear warping of frequency bands to better match the frequency-dependent sensitivity of human ears

- **Simultaneous masking (frequency masking)**

- **Non-simultaneous masking (temporal masking)**
Sensitivity of HAS in Quiet

Characterize the amount of energy needed in a pure tone such that it can be heard by a listener in a noiseless environment.
Critical Bands

- Human auditory system has a limited, frequency-dependent resolution
  - Cochlea can be viewed as bank of highly overlapping bandpass filters
- The perceptually uniform measure of frequency can be expressed in terms of the width of the *Critical Bands.*
  - It is less than 100 Hz at the lowest audible frequencies, and more than 4 kHz at the high end. Altogether, the audio frequency range can be partitioned into 25 critical bands
Bark (a new unit of frequency)

- 1 Bark = width of one critical band
- For frequency < 500 Hz, it converts to \( \frac{\text{freq}}{100} \) Bark
- For frequency > 500 Hz, it is \( 9 + 4 \log_2(\frac{\text{freq}}{1000}) \) Bark.

\[
z(f) = 13 \tan^{-1}(0.00076f) + 3.5 \tan^{-1}\left(\frac{f}{7500}\right)^2 \text{(Bark)}
\]
Critical Bands Illustration

MPEG/AUDIO FILTER BANK BANDS

CRITICAL BAND BOUNDARIES

<table>
<thead>
<tr>
<th>Band No.</th>
<th>Center Freq. (Hz)</th>
<th>Bandwidth (Hz)</th>
<th>Band No.</th>
<th>Center Freq. (Hz)</th>
<th>Bandwidth (Hz)</th>
<th>Band No.</th>
<th>Center Freq. (Hz)</th>
<th>Bandwidth (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>-100</td>
<td>10</td>
<td>1175</td>
<td>1080-1270</td>
<td>19</td>
<td>4800</td>
<td>4400-5300</td>
</tr>
<tr>
<td>2</td>
<td>150</td>
<td>100-200</td>
<td>11</td>
<td>1370</td>
<td>1270-1480</td>
<td>20</td>
<td>5800</td>
<td>5300-6400</td>
</tr>
<tr>
<td>3</td>
<td>250</td>
<td>200-300</td>
<td>12</td>
<td>1600</td>
<td>1480-1720</td>
<td>21</td>
<td>7000</td>
<td>6400-7700</td>
</tr>
<tr>
<td>4</td>
<td>350</td>
<td>300-400</td>
<td>13</td>
<td>1850</td>
<td>1720-2000</td>
<td>22</td>
<td>8500</td>
<td>7700-9500</td>
</tr>
<tr>
<td>5</td>
<td>450</td>
<td>400-510</td>
<td>14</td>
<td>2150</td>
<td>2000-2320</td>
<td>23</td>
<td>10,500</td>
<td>9500-12000</td>
</tr>
<tr>
<td>6</td>
<td>570</td>
<td>510-630</td>
<td>15</td>
<td>2500</td>
<td>2320-2700</td>
<td>24</td>
<td>13,500</td>
<td>12000-15500</td>
</tr>
<tr>
<td>7</td>
<td>700</td>
<td>630-770</td>
<td>16</td>
<td>2900</td>
<td>2700-3150</td>
<td>25</td>
<td>19,500</td>
<td>15500-</td>
</tr>
</tbody>
</table>
CB vs. Frequency

CB (Bark)

Freq. (Hz)
Masking

- Masking refers to a process where one sound is rendered inaudible because of the presence of another sound
  - Simultaneous masking (frequency masking)
  - Non-simultaneous masking (temporal masking)
- Masker: primary tone (masking tone)
- Maskee: secondary tone (test tone)
- Spread of masking: a masker at some CB also has predictable effect on detection thresholds in other CBs
Frequency Masking

- Question: do receptors interfere with each other?

  - Experiment:
    - Play 1 kHz tone (masking tone) at fixed level (60 dB). Play test tone at a different level (e.g., 1.1 kHz), and raise level until just distinguishable.
    - Vary the frequency of the test tone and plot the threshold when it becomes audible.
    - Repeat for various frequencies of masking tones

  - The result will be a collection of curves showing the frequency masking effect
Frequency Masking (Con’d)
Three Types of Frequency Masking

- **Noise-Masking-Tone (NMT):** $\text{SMR}=4\text{dB}$
- **Tone-Masking-Noise (TMN):** $\text{SMR}=24\text{dB}$
- **Noise-Masking-Noise (NMN):** $\text{SMR}=26\text{dB}$
Temporal Masking

- If we hear a loud sound, then it stops, it takes a little while until we can hear a soft tone nearby.
- Experiment: Play 1 kHz *masking tone* at 60 dB, plus a *test tone* at 1.1 kHz at 40 dB. Test tone can't be heard (it's masked). Stop masking tone, then stop test tone after a short delay.
- Adjust delay time to the shortest time when test tone can be heard (e.g., 5 ms).
- Repeat with different level of the test tone and plot
Temporal Masking (Con’d)
Pre-masking vs. Post-masking
Total Effect of Both Frequency and Temporal Masking
Summary of Psychoacoustics

- Frequency and temporal masking
  - Two sounds could become indistinguishable if their frequencies are close
  - A test tone could be masked by a masking one if it is too close and relatively weaker

- Implication into audio compression
  - Time: we do not need to represent all tones faithfully if some are masked by others
  - Frequency: we do not need to represent all frequencies if some are masked by others