Topic 3

Human Auditory Sensation

(Some slides are adapted from Bryan Pardo’s course slides on Machine Perception of Music)
The Ear

Pinna

Concha

Ear canal

Ossicles (Middle ear Bones)

Cochlea (inner ear)

Eardrum ( tympanum)

Eustachian tube

Outer Ear

Middle Ear

Inner Ear
Function of the Ear

• Outer ear: shape the sound spectrum
  – Torso, head, pinnae: head-related transfer function (HRTF). Interaural difference.
  – Concha, canal: increase sound level of about 10-15dB between 1.5k-7kHz, due to resonances

• Middle ear: effective and efficient transfer
  – Eardrum: effective area about 55 mm^2 (where the oval window is about 3 mm^2 size.
  – Three ossicles: a lever system
  – The last ossicle is called stapes, the smallest bone in the human body
Function of the Ear

• Inner ear:
  – Vestibular system: sense of balance
  – Eustachian tube: provides ventilation to middle ear
  – Cochlea: the primary auditory organ of the inner ear

• Vestibular system starts to develop (around 8 weeks) much earlier than the auditory system (around 20 weeks)!
The Cochlea

- Each point on the Basilar membrane resonates to a particular frequency
- At the resonance point, the membrane moves

Apex (wide and flexible)
Base (narrow and stiff)
Gammatone Filterbank

- Impulse response (n=4)

\[ g(t) = at^{n-1}e^{-2\pi bt} \cos(2\pi ft + \phi)u(t) \]

- Gamma
- Tone

Diagram showing impulse responses for different frequencies and a graph of magnitude gain vs. frequency.
• When the membrane moves, it moves hairs.
• When hairs move, they fire nerve impulses.
Hair Cells

• Inner hair cell
  – The actual transducer

• Outer hair cell
  – Makes small but very fast movements (maybe ~100kHz)
    – Feedforward amplifiers, nonlinear

• They are damaged by age and hard to regrow

• Let’s look at a dance by an outer hair cell!
  – http://www.youtube.com/watch?v=Xo9bwQuYrRo
Auditory Nerve Fibers

- Each IHC -> 10-20 type I fibers
- ~10 OHCs -> ~6 type II fibers

- This is the evidence that IHCs are the actual transducers.

- Fibers are arranged to maintain the tonotopy of basilar membrane.
Frequency Sensitivity

- single nerve measurements
- (roughly) symmetric in log of frequency
Encoding Loudness

More spikes/second = louder
More fibers firing = louder

One fiber:
~ 25 dB dynamic range

Hearing dynamic range > 100 dB

Intensity / dB SPL

Spikes/sec
Phase Locking

For high frequency tones, the fibers phase lock to low frequency modulations.

(from Chris Darwin)
Poststimulus Time Histogram (PSTH)

- Single fiber firing pattern to a click sound

A half-wave rectified gammatone function!

Histogram calculated by superimposing neural responses of repeated experiments

1/central frequency
PSTH to a Tone Burst

Onset response

Adaptation

Recovery

Tone duration

(from Gelfand 1998)
Coding Frequency Information

• Frequencies under 5 kHz
  – Individual harmonics are resolved by the cochlea
  – Coded by place (which nerve bundles along the cochlea are firing)
  – Coded by time (nerves fire in synchrony to harmonics)

• Frequencies over 5 kHz
  – Individual harmonics can’t be resolved by the inner ear and the frequency is revealed by temporal modulations of the waveform amplitude (resulting in synched neuron activity)
Loudness

Loudness is a subjective measure of sound pressure (or intensity).

- How does the intensity of a sound relate to its perceived loudness?
  - Does frequency matter?
  - Is broadband noise different from narrow band?
  - How can we find out?
Measuring Frequency’s Effect

• Pick a reference frequency (like 1000Hz)

• Play a sine wave of a defined intensity at that frequency (say, 30 dB-SPL)

• Pick another frequency (any one)

• Play a sine wave at the new frequency, $f$

• Adjust the intensity of the sine at $f$ until its loudness equals the reference
Equal Loudness Contours

[Graph showing equal loudness contours with sound pressure level (dB re 2 × 10^{-5} N/m^2) on the y-axis and frequency (Hz) on the x-axis. The contours represent different loudness levels (in phon) with lines indicating thresholds of audibility.]
Phons

The phon is a unit of perceived loudness for pure tones. The purpose of the phon scale is to compensate for the effect of frequency on the perceived loudness of tones. By definition, 1 phon is equal to 1 db-SPL at a frequency of 1000 Hz.
Sensitivity to Loudness

ECE 477 - Computer Audition, Zhiyao Duan 2017
JND

The **just noticeable difference** (JND) is the smallest difference in sensory input that is detectable by a human being. It is also known as the **difference limen** (DL).
Weber’s Law (named after Ernst Heinrich Weber, 1795-1878) attempts to describe the relationship between the physical magnitudes of stimuli and the perceived intensity of the stimuli.

- DL in intensity is proportional to the intensity itself, i.e. $\frac{\Delta I}{I}$ is constant.
The Sone

- The sone is a unit of perceived loudness, proposed by S. Stevens in 1936.
  - At 1kHz, 1 sone = 40 phon = 40 dB-SPL
  - A stimulus that is $n$ sones loud is judged to be $n$ times as loud as 1 sone.

\[ P_2 - P_1 = 10(\log_2 S_2 - \log_2 S_1) \]

Twice loudness
10 phon, 10 dB
Intensity $\Leftrightarrow$ Sone

- Tone at 1kHz with intensity $> 40$ dB SPL
- To make the tone $n$ times as loud, how many times should we increase the intensity?
  - We want to have $\frac{S_{new}}{S} = n$.
  - Therefore, we need $P_{new} - P = 10 \log_2 n$.
  - That is, we need $10 \log_{10} \frac{I_{new}}{I} = 10 \log_2 n$.
  - So $\frac{I_{new}}{I} = 10^{\log_{10} n} = 10^{\log_{10} 2} = n^{\log_2 10} \approx n^{3.32}$.
  - Roughly 8 people make as twice loud as 1 person.
- That’s why some papers use $\sqrt[3]{I}$ to describe loudness.
Masking

- A loud tone masks perception of tones at nearby frequencies
**Critical Band**

- **Critical band** – the frequency range over which a pure tone interferes with perception of other pure tones.

- Think about the masked threshold as a bandblock filter.

- How to measure the bandwidth?
Equivalent Rectangular Bandwidth (ERB)

- \( \text{ERB}(\text{Hz}) = 24.7(4.37 \, f(\text{kHz}) + 1) \)
- The bandwidth increases with frequency.
Frequency Difference Limen

- The smallest difference between the frequencies of two sine tones that can be discriminated correctly 75% of the time.
Pitch (ANSI 1994 Definition)

- That attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch depends mainly on the frequency content of the sound stimulus, but also depends on the sound pressure and waveform of the stimulus.
Pitch (Operational)

• A sound has a certain pitch if it can be reliably matched to a sine tone of a given frequency at 40 dB SPL.
Pitch and Intensity

• Stevens Rule
  – The pitch of low frequency (below 1000Hz) sine tones decreases with increasing intensity (low loud sounds go flat).
  – The pitch of high frequency tones (over 3000 Hz) increases with intensity (high loud sounds go sharp)

220Hz 7040Hz
Mel Scale

- A perceptual scale of pitches judged by listeners to be equal in distance from one another. The reference point between this scale and normal frequency measurement is defined by equating a 1000 Hz tone, 40 dB SPL, with a pitch of 1000 mels.
Mel Scale

$$\text{Mel} = 2595 \log_{10}(1 + \frac{f}{700})$$
Mel Scale

• Above about 500 Hz, larger and larger intervals are judged by listeners to produce equal pitch increments.

• The name *mel* comes from the word *melody* to indicate that the scale is based on pitch comparisons.

• Proposed by Stevens, Volkman and Newman (Journal of the Acoustic Society of America 8(3), pp 185-190, 1937)
Ear Craziness

- **Binaural Diplacusis**
  - Left ear hears a different pitch from the right.
  - Can be up to 4% difference in perceived pitch

- **Otoacoustic Emissions**
  - Healthy ears can make noise.
  - Thought to be a by-product of the sound amplification system in the inner ear.
  - Caused by activity of the outer hair cells in the cochlea.
Harmonic Sound

• A sound with strong sinusoid components at integer multiples of a fundamental frequency. These components are called **harmonics** or **overtones**.

• Harmonic sounds are the sounds that may give a perception of “pitch”.

ECE 477 - Computer Audition, Zhiyao Duan 2017
Classify Sounds by Harmonicity

- Sine wave
- Strongly harmonic
Classify Sounds by Harmonicity

- Somewhat harmonic (quasi-harmonic)

![Waveform and Spectrum of Marimba](image1)

![Waveform and Spectrum of Human Voice](image2)
Classify Sounds by Harmonicity

- **Inharmonic**

![Amplitude vs Time](image1.png)
![Magnitude vs Frequency](image2.png)

### Sounds

<table>
<thead>
<tr>
<th>Sounds</th>
<th>Instrument family</th>
<th>Instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harmonic</td>
<td>Woodwind</td>
<td>Piccolo, flute, oboe, clarinet, bassoon, saxophone</td>
</tr>
<tr>
<td></td>
<td>Brass</td>
<td>Trumpet, horn, euphonium, trombone, tuba</td>
</tr>
<tr>
<td></td>
<td>Arco string</td>
<td>Violin, viola, cello, double bass</td>
</tr>
<tr>
<td></td>
<td>Pluck string</td>
<td>Piano, guitar, harp, celesta</td>
</tr>
<tr>
<td></td>
<td>Vocal</td>
<td>Voiced phonemes</td>
</tr>
<tr>
<td>Quasi-harmonic</td>
<td>Pitched percussive</td>
<td>Timpani, marimba, vibraphone, xylophone</td>
</tr>
<tr>
<td>Inharmonic</td>
<td>Non-pitched percussive</td>
<td>Drums, cymbal, gong, tambourine</td>
</tr>
</tbody>
</table>

ECE 477 - Computer Audition, Zhiyao Duan 2017
What determines pitch?

• Complex tones
  – Strongest frequency?
  – Lowest frequency?
  – Something else?

• Let’s listen and explore...
Hypothesis

- Pitch is determined by the lowest strong frequency component in a complex tone.
The Missing Fundamental
Hypothesis

- Pitch is determined by the lowest strong frequency component in a complex tone.

- The case of the missing fundamental proves that ain’t always so.
Hypothesis – “It’s complicated”

- by the loudest frequency
- by the common frequency that divides other frequencies
- by the space between regularly spaced frequencies
Pitch and Music

• How do we tune pitch in music?

• How do we represent pitch in music?

• How do we represent the relation of pitches in music?
Equal Temperament

- Octave is a relationship by power of 2.
- There are 12 half-steps in an octave

\[ f(n) = 2^{\frac{n}{12}} f_{\text{ref}} \]

- frequency of desired pitch
- number of half-steps from the reference pitch
- frequency of the reference pitch
• 100 Cents in a half step
• 2 half steps in a whole step
• 12 half steps in an octave

Number of cents

\[ c = 1200 \log_2 \left( \frac{f}{f_{\text{ref}}} \right) \]

Virtual keyboard
A=440 Equal tempered tuning
Musical Intervals (from C)
Interval Names

- minor 2nd (half step)
- major 2nd (whole step)
- minor 3rd
- major 3rd
- perfect 4th
- tritone (augmented 4th)
- diminished 5th
- perfect 5th
- minor 6th (augmented 5th)
- major 6th
- minor 7th (augmented 6th)
- major 7th
- octave
Half-steps: 0 1 2 3 4 5 6 7 8 9 10 12

- C → C: 12 half-steps, $2^{\frac{12}{12}} = \frac{2}{1}$
- C → G: 7 half-steps, $2^{\frac{7}{12}} = 1.4983 \approx \frac{3}{2}$
- C → F: 5 half-steps, $2^{\frac{5}{12}} = 1.3348 \approx \frac{4}{3}$
- C → E: 4 half-steps, $2^{\frac{4}{12}} = 1.2599 \approx \frac{5}{4}$

Are these just coincidence?
• How about defining pitches this way, so that they sound more harmonic?
Pythagorean Tuning

- Frequency ratios of all intervals are based on the ratio 3:2, i.e. perfect fifth (P5), which is 7 half-steps.
Circle of Fifths
Problem with Pythagorean Tuning

- One octave = 2f
- A perfect 5th = (3/2)f

- What happens if you go around the circle of 5ths to get back to your original pitch class?
  
  - \((3/2)^{12} = 129.75\)
  - Nearest octave is \(2^7 = 128\)
  - 128 != 129.75

- Not convenient for key changes
Overtone Series

- Approximate notated pitch for the harmonics (overtones) of a frequency