Speech Perception

- Dynamics of Speech
- Perception of English speech sounds.
- Cues for manner-place-voicing
- Suprasegmentals
- Context dependence
- Categorical perception
- Hearing Loss and speech perception

Dynamics of Speech

- intensity
- frequency
- suprasegmentals

Dynamics of Speech

- intensity
  - whisper - 20 dB HL
  - normal conversational speech - 50 to 60 dB HL
  - loud speech - 70 dB HL
  - shouting - 90 dB HL
Dynamics of Speech

- frequency
  - about 250 to 8000 Hz
  - most important speech sounds - 500 through 6000 Hz

<table>
<thead>
<tr>
<th>Class</th>
<th>Intensity</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowels</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Nasals</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Stops</td>
<td>Varies</td>
<td>Mid-high</td>
</tr>
<tr>
<td>Fricatives</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>F2 transitions</td>
<td>Varies</td>
<td>Mid-high</td>
</tr>
</tbody>
</table>
Perception of English Speech Sounds

- Vowels
- Diphthongs
- Semivowels
- Nasal consonants
- Stop/plosives
- Fricatives and affricates

Vowels

- Most important cue is relationship between the F1 & F2 formant frequencies.

Vowels

- Other cues
  - For high-front vowels: F3 formant
  - For low-back vowels: A strong F1 frequency
  - Contextual cues
Vowels

Diphthongs

- Most important cue is gliding formants

\[
\begin{array}{c}
\text{FREQUENCY} \\
\hline
\text{TIME} \\
\end{array}
\]

- Specifically …
- Rate of frequency change. The faster the rate of change the easier it is to perceive
- Brief steady state formant pattern at the beginning and the end of the diphthong also improves perception.
  - i.e., tense monothongs are more difficult to perceive because they lack the long steady state formant.
Semivowels

- Glides -----> /j/ and /w/
- Liquids -----> /r/ and /l/
- As with diphthongs, perception relies on vowel formant transitions.
- Most important transition is F2
- For /r/ and /l/, the F3 formant is also important
- Semivowels differ from diphthongs since they have more rapid F2 transitions which makes them consonant like.
**Nasals**

- Most important decisions when perceiving nasals.
  - Is the speech segment nasal or non-nasal?
  - If it is a nasal, which nasal is it?
- Important Cues
  - Transition between nasal and adjacent vowel
  - Weakening of upper formants
  - Lengthening of nasal tract causes primary resonant frequency to drop from 500 to about 250 Hz.
  - Lowering of intensity when compared to vowels.

---

**Nasals (continued)**

- In perceiving specific nasals the most important cues are…
- Transition to and from adjacent vowels are similar to different stops…
  - /m/ same as /p-b/
  - /n/ same as /t-d/
  - "ng" same as /k-g/
Nasals (continued)

- Frequency cues in F2 transition from adjoining vowels
  - /m/ lowest in frequency and shortest in duration
  - /n/ higher in frequency and a bit longer in duration
  - /ng/ highest in frequency and shortest in duration.

Nasals (continued)

- Resonance and anti-resonant patterns are different for different nasals. This is important for discriminating between /m/ and /n/, but not important for perceiving /ng/.
- Nasal murmur

Stops

- Manner-place of articulation-voicing are important cues.
Speech Perception

22

Stops (manner)

• Cues for manner
  – Oral occlusion is heard either as a …
    • Silence in the voiceless stop /p-t-k/
    • Brief attenuation of sound in voiced stops /b-d-g/.
  – Stopped air is often released and heard as a plosive (i.e., and aperiodic sound source)
  – Stops have much shorter transition to and from adjacent vowels than semi-vowels.

Speech Perception

23

Stops (manner)

• To review…
  • /p-b/ …. Bilabial
  • /t-d/ …. Alveolar
  • /k-g/ …. Velar (or palatal)
Stops (place of articulation)

- Cues for place include …
- Frequency position of the burst in relation to F1 and F2 formants.
  - When burst is above 3000 Hz, the sound is perceived as a /t/.
  - When burst is below F2 formant the vowel was perceived as a /p/, with the exception of high-back vowels.
  - When burst was just above F2 formant, phoneme perceived as a /k/.

F2 transition. Depending upon direction of transition, different stop plosives will be perceived. Refer to previous slide.
Stops (voicing)

- Voicing cues include
- Low frequency voice bar.

Stops (voicing)

- Presence of aspiration
- Onset of F1 Formant for a following vowel (F1 cutback)

Stops (voicing)
Stops (voicing)

- Voice onset time (VOT)
  - Longer the VOT prior to a vowel and after a stop… the more likely you will perceive it as voiceless.

- Vowel duration preceding stop.
  - Longer duration of vowel, the more likely it will be perceived as voiced.