Speech Processing 11-492/18-492

Human Speech Processing
Phonetics and Phonology
The vocal tract

- alveolar ridge
- tongue tip
- tongue blade
- tongue body
- epiglottis
- larynx
- hard palate
- soft palate
- uvula
- pharynx
- tongue root
From meat to voice

- **Blow air through lungs**
  - Vibrate larynx
  - Vocal tract shape defines resonance
  - Obstructions modify sound
    - *Tongue, teeth, lips, velum (nasal passage)*
The ear

- Outer Ear
- Middle Ear
- Inner Ear
- Malleus
- Incus
- Facial Nerve
- Auditory Nerve
- Cochlea
- Stapes
- Ear Drum
- Ear Canal
From sound to brain waves

Sound waves

- Vibrate ear drum
- Cause fluid in cochlear to vibrate
- Spiral cochlear
  - Vibrate hairs inside cochlear
  - Different frequencies vibrate different hairs
  - Converts time domain to frequency domain
From grunts to meaning

- Grunts and vocalization
  - Lots of variation available
    - (continuous systems – not discrete)
  - Noises become distinct, recognizable
- Grow into languages, dialects and idiolects
- What are the fundamental units?
Articulatory Movements
Electromagnetic Articulograph
**Defined as fundamental units of speech**

- If you change it, it (can) change the meaning

  “pat” to “bat”
  “pat” to “pam”
• One or two banded frequencies (formants)
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Pronunciation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>wAshington</td>
<td>AE</td>
</tr>
<tr>
<td>AH</td>
<td>bUt, hUsh</td>
<td>AO</td>
</tr>
<tr>
<td>AW</td>
<td>hOW, sOUth</td>
<td>AX</td>
</tr>
<tr>
<td>AY</td>
<td>hlde, bUY</td>
<td>EH</td>
</tr>
<tr>
<td>ER</td>
<td>makER, sEARch</td>
<td>EY</td>
</tr>
<tr>
<td>IH</td>
<td>blt, shlp</td>
<td>IY</td>
</tr>
<tr>
<td>OW</td>
<td>lOne, nOse</td>
<td>OY</td>
</tr>
<tr>
<td>UH</td>
<td>fUll</td>
<td>UW</td>
</tr>
</tbody>
</table>
English Consonants

- **Stops:** P, B, T, D, K, G
- **Fricatives:** F, V, HH, S, Z, SH, ZH
- **Affricatives:** CH, JH
- **Nasals:** N, M, NG
- **Glides:** L, R, Y, W

**Note:** voiced vs unvoiced:
- P vs B, F vs V
Number of Phonemes in Language

- **US English**: 43
- **UK English**: 44
- **Japanese**: 25
- **Hindi**: 81

**Numbers aren’t definite though**

- Depends on who you ask,
- And what you want it for
Not all variation is Phonetic

- **Phonology: linguistically discrete units**
  - May be a number of different ways to say them
  - /r/ trill (Scottish or Spanish) vs US way

- **Phonetics vs Phonemics**
  - Phonetics: discrete units
  - Phonemics: all sounds

- /t/ in US English: becomes “flap”
  - “water” /w ao t er /
  - “water” /w ao dx er /
Dialect and Idiolect

- **Variation within language (and speakers)**
- **Phonetic**
  - “Don” vs “Dawn”, “Cot” vs “Caught”
  - R deletion (Haavaad vs Harvard)
- **Word choice:**
  - Y’all, Yins
  - Politeness levels
Not all languages use the same set

- **Asperated stops (Korean, Hindi)**
  - P vs PH
  - English uses both, but doesn’t care
  - Pot vs sPot (place hand over mouth)

- **L-R in Japanese not phonological**

- **US English dialects:**
  - Mary, Merry, Marry

- **Scottish English vs US English**
  - No distinction between “pull” and “pool”
  - Distinction between: “for” and “four”
Different language dimensions

- **Vowel length**
  - Bit vs beat
  - Japanese: shujin (husband) vs shuujin (prisoner)

- **Tones**
  - F0 (tune) used phonetically
  - Chinese, Thai, Burmese

- **Clicks**
  - Xhosa
Co-articulation

- Voicing actually doesn’t always stop
  - “have honey”, “impossible”

- Nasalized voices, lip rounding
  - “min” vs “bit”, “sow” vs “see”

- Lexical stress:
  - EMphasis, emPHAsis
  - PROject, proJECT

- Reduction, contraction
  - “A boy is riding a bike”
  - “I want to go to Disneyland.”
  - “I will go tomorrow”
Prosody

- **Intonation**
  - Tune

- **Duration**
  - How long/short of each phoneme

- **Phrasing**
  - Where the breaks are
Rate of vibration during voiced speech
- Males: 80-140 times a second
- Females: 130-220 times a second
- Children: 180-320 times a second

Used for:
- Emphasis
- Style: questions, statements, confidence etc
Intonation Information

- Large pitch range (female)
- Authoritive since goes down at the end
  - News reader
- Emphasis for Finance $H^*$
- Final has a raise – more information to come

- Female American newsreader from WBUR
- (Boston University Radio)
Intonation Examples

- Fixed durations, flat F0.
- Decline F0
- “hat” accents on stressed syllables
- Accents and end tones
- Statistically trained
Words

- The things with space around them (sort of)
- Chinese, Thai, Japanese doesn’t use spaces
- Speech doesn’t use spaces
  - Blackboard vs Black Board
- English
  - Morphology: walk, walks, walking, walked
- Japanese
  - Morphology: aruku, arukimasu, arukimashita, aruite, aruikitai, aruikitakatta, arukemasu, ....
Speech Acts

- **Words aren’t always what they seem**
  - Can you pass the salt?
  - Boston. Boston! Boston?
  - Yeah, right

- **Multiple ways to say the same thing:**
  - I want to go to Boston.
  - Yes
Human Speech

- Human production and perception
  - Quite different from computers
- Phonology
  - Defining the alphabet of speech
  - Different languages make different distinctions
- Intonation
  - How it's said