Speech Sounds

Introduction to Linguistics for Computational Linguists
Speech Sounds

• Phonetics - Physical basis of speech sounds
  – Physiology of pronunciation, perception
  – Acoustics of speech sounds

• Phonology - Patterns of combination of speech sounds
  – Which sequences are allowed (phonotactics)
  – Effects of context on speech
Design Problem

• $10^4 - 10^5$ words in vocabulary
  – Compare to animal systems $<< 100$
• “Open” system -- new items added easily
• Rapid learning: 10 year-old has 40K items
  – 11 new words daily
  – Some learned after a single experience
• Problem: how can all this be learned?
Large, Learnable Symbol Set

- 40K independent symbols would be unlearnable.
- Solution:
  - No direct pairing of pronunciation-meaning
  - Symbols are combinations of small set of discrete elements
  - Rules of combination are general -- independent of the meaning of the sign
  - Combination rules learned through entire experience
Miller’s “Exponential Principle”

- Small set of discrete elements combine into large numbers of strings
  - consider (very simple!) language with 8 cons., 5 vowels, syllables only in form CV, words with 4 syllables
    
    \[
    \text{CV. CV.CV.CV} \\
    8 \times 5 \times \ldots \times 8 \times 5 = 40^4 = 2,560,000 \text{ words}
    \]

- Symbols created through combination
Phonology

- Elements (8 consonants, 5 vowels) are phonemes.
- Syllables (only in form CV) are one form of organizing principle.
- Organizing (phonological) principles are neuropsychology
- Physical (phonetic) encoding/decoding is physical/physiological
Speech Chain

SPEAKER

Brain

Vocal muscles

Motor nerves

Sensory nerves

Ear

Feedback link

Sound waves

LISTENER

Brain

Sensory nerves

Ear

Linguistic level

Physiological level

Acoustic level

Physiological level

Linguistic level
Speech Organs

- “Second-Hand Use”

<table>
<thead>
<tr>
<th>Organ</th>
<th>Primary</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lungs</td>
<td>Breath</td>
<td>Power</td>
</tr>
<tr>
<td>Vocal Folds</td>
<td>Protection, Rigidity</td>
<td>Fluid -&gt; Acoustic Conversion</td>
</tr>
<tr>
<td>Tongue</td>
<td>Digestion</td>
<td>Fine Resonance</td>
</tr>
<tr>
<td>Lips, Teeth</td>
<td>Chewing</td>
<td>Resonance</td>
</tr>
</tbody>
</table>
“Buzz, Hiss, Pop”

• Speech sounds mostly can be described as
  – buzz of air through larynx
  – hiss of air through tight passageway ([s])
  – pop of air after pressure build-up ([p])

• Often in combination
Speech Breathing

• Every syllable powered by muscle contraction (intracostal muscles between ribs)

• 80/20 exhalation/inhalation (vs. 60/40 during quiet breathing)

• Flow of air through larynx may result in voicing (buzz of vibration)
  – compare [s] vs. [z], [f] vs. [v]
  – listen / fingers on larynx / fingers in ears
Laryngeal “Buzz”

- “Adam’s Apple” at top of trachea (windpipe)
- Contain vocal folds (cords) that vibrate if close
- Bernoulli effect causes fast opening & shutting (compare “raspberry”).
- Regular vibration results in a perceptible tone.
Shaping the Buzz

- Tongue tip, front/blade, back, root
- Lips, teeth, alveolar ridge, palate, velum, uvula
- Velum controls passage to nasal cavities
Bell’s Visible Speech

- Melville Bell invented a system for transcribing speech “Visible Speech”
- Tool for teaching deaf children
  - But first system for recording sounds exactly
- 1860’s lecture tour of Bell’s boys -- Melville, Edward & Alexander.
  - One brother leaves room (out of earshot)
  - Volunteers solicited for interesting speech, transcribed
  - Brother returns and repeats on basis of transcription
IPA

- One of Bell’s boys went on to invent the telephone
- Visible speech evolved into the *International Phonetic Alphabet*, standard for phonetic transcription.
- Enough here to transcribe standard German
- Transcriptions in brackets [ˈbræ.kɛts]
Hiss of Fricatives

- Turbulent flow past a narrow constriction produces a *hiss of frication*
- Present initially in *Fuß* [f], *Wut* [v], *Hut* [h], medially in *lassen* [s], *lasen* [z], *laschen* [ʃ], *lachen* [χ]
- In foreign words a voiced version of [ʃ]: *Journal* [ʒ]
“Pop” of Pressure Release

• When air flow is stopped and then suddenly released, we hear a “pop”
• These are stop consonants aka *plosives*
• Initially in *Pass* [p], *Bass* [b], *Tasse* [t], *dass* [d], *Kasse* [k], *Gasse* [g]
• Also (in German) before initial vowels *As* [ʔ]
Place of Maximal Constriction

- Stops differ in *where* they block flow
  - Velum closed (nonnasal); raised would be [m, n, ɳ]
- Fricatives likewise [f, v], [s, z], [χ], and [h]
(Too Many) Consonants

- Standard German /r/ is uvular [R]; Bavarian alveolar [r]
- Affricates are stop + fricative (in one phoneme)
  - Pfennig [pf], Zeit [ts]
- Los [l] is alveolar lateral
Vowels

- Vowels involve no constriction of air flow
- More fluid, variable than consonants
- But still a limited set of discrete elements
- Good way to see this is to collect the set of vowels
Lexical Set

- [i] Siehe, Biene, Glied
- [y] Bühne, Blüte, Kühe
- [I] bitte, Schimmel, Blitz
- [y] müssen, füttern, Lücke
- [e] Lehne, nehmen, geben
- [ø] Stöhnen, Föhn, Söhne
- [ε] Wetter, besser, kess
- [œ] Löcher, können, Töpfer
- [æ] nähme, gäbe, täte
  - if distinctive from [e:]
More on Vowels

- [Ə] unstressed neutral vowel
  - Liebe, genau, allemal

- Lots of dialect variation
  - e.g., Bavarian [iɛ] lieb

- *Pure* vowels can be extended in pronunciation, while *diphthongs* involved a change
  - [i,I,u,y,..] vs. [au,ai,ɔI]
Relative Properties

Front

- [i] Siehe, Biene, Glied
- [y] Bühne, Blüte, Kühe
- [I] bitte, Schimmel, Blitz
- [y] müssen, füttern, Lücke
- [e] Lehne, nehmen, geben
- [ø] Stöhnen, Föhn, Söhne
- [ɛ] Wetter, besser, kess
- [œ] Löcher, können, Töpfer
- [æ] nähme, gäbe, täte
  - if distinctive from [e:]

Back

- [u] Huhn, Mut, Kuh
- [ʊ] Mutter, plus, Kuss
- [o] Sohn, Pfote, los
- [ɔ] Motte, Topf, Sonne

Central

- [a] kahl, Aachen, Raten
- [ɑ] Ball, lachen, Ratten
- [au] Haus, Trauben, Couch
- [ai] heiser, mein, leicht
- [ɔI] Häuser, Reue, Meute
Relations

- **Front/Back**
  - \([i, I, y, y, e, \varnothing, e, \varepsilon, \varepsilon, \ddot{a}]\) Front
  - \([u, u, o, o]\) Back
  - \([a, a]\) Central

- **Close/Open (High/Low)**
  - \([i, I, y, y, u, u]\) Close
  - \([\varepsilon, \varnothing, o, \Theta]\) Close-Mid
  - \([\varepsilon, \varepsilon, o, o]\) Open-Mid
  - \([\ddot{a}, a]\) Open

- **Round**
  - \([y, y, \varnothing, \varepsilon]\) and all back vowels

Where symbols appear in pairs, the one to the right represents a rounded vowel.
The Sounds Produced

- **Source** -- Buzz, Hiss, Pop
- **Filter** -- Resonance of Oral Cavity
- **Source/Filter** -- Guitar String/Body

- **Example:** Vocal folds vibrating at 100 Hz produce harmonics (overtones) at 200, 300, 400, etc. (but with decreasing strength). Mouth and nose strengthen some frequencies, dampen others.
Müller’s ‘Source-Filter Theory of Voice Production’
--From G. Miller’s
*The Science of Words*
Role of Articulators

- *Articulators* -- organs that shape sound
- Tongue, lips, palate, velum
- Role -- filter that strengthens some frequencies, dampen others

- Formants - characteristic resonant frequencies of vowels, caused by configuration of articulators
3 filters on spectra
--Miller
Voicing in Consonants

- \([p/b, t/d, k/g]\) distinguished by voicing
  - stops build pressure above larynx
- Voice Onset Time (VOT) is start of vocal cord vibration relative to stop release

When does voicing start?

```
release

stop

When does voicing start?

---

time
```
Stop + Vowel

- Prevoicing (French, Russian [b,d,g])

- Unaspirated (French, Russian [p,t,k], German, English [b,d,g])

- Aspirated (German, English [p,t,k]) -- note puff!
Voicing Across Languages

• Few languages distinguish three levels of voicing, but, e.g., Thai does

• Most distinguish two levels
  – voiced/unvoiced
  – always adjacent VOT types
    • either prevoiced vs. unaspirated
      – Dutch, Yiddish among Germanic languages
    • of unaspirated vs. aspirated
Phonetics - Summary

• Phonetics - physical basis of speech sounds
• Design problem: provide extendible, learnable symbol set of size $10^5$
  – Soln: small set of phonemes in different orders
• Written via visible speech, IPA
• Production has source and filter
  – Example: vowels’ source - vocal fold vibration, filter - resonance of oral cavity