Blackboard and other Secrets

• But, ummm, professor, where is the syllabus?
• [http://tinyurl.com/3tyq3zr](http://tinyurl.com/3tyq3zr)

• Non-textbook readings and lecture slides are available under course materials on BB
Last time

What is psycholinguistics?
The study of language processing, production, and learning.

The nature of linguistic knowledge
Rich & subtle, but implicit and hard to articulate. This is the type of knowledge access during language production and comprehension

(Traxler: Design features of language)
Questions?
Articulation, acoustics, and phonological features

Language & Psycholinguistics
BCS 152 / LIN 217 / ASL 260 / PSY 152
(Articulation) ➔

Energy distribution over frequencies over time ➔

Acoustic features / Phonetic features ➔

Phonological features ➔

Allophones (speech sounds) ➔

Phone recognition
Today

• **Speech production and perception**
  – Physiology and articulation
  – Physics of speech sounds
  – Key acoustic properties of speech
  – Phonological systems
(Articulation)

↓

Energy distribution over frequencies over time

↓

Acoustic features / Phonetic features

↓

Phonological features

↓

Allophones (speech sounds)

↓

Phone recognition
Speech articulation

- Our vocal tract is specialized for speech
  - Creates choking hazard.
  - Breathing affected.
Source-Filter Model

• Sound source: larynx (coll. ‘voice box’)

• Changeable resonators:
  – pharynx (throat)
  – mouth
  – lips
  – nose

http://en.wikipedia.org/wiki/Larynx
Glottis - Vocal Chords

• See also this recording of Mel Blanc, the voice of many cartoon characters from the outside:
  https://youtu.be/95eXCS0Qy2Y

  the glottis in action:
  https://youtu.be/ejVQEFbIfmI

• Can’t get enough?
  https://youtu.be/y2okeYVclQo for a video of speech pathologist
  https://youtu.be/-XGds2GAvGQ for a video of singers
X-ray view of speech

Why did Ken set the soggy net on top of the deck?
On youtube: https://youtu.be/DcNMCB-Gsn8
Today

- Physiology and articulation
- **Physics of speech sounds**
- Key acoustic properties of speech
- Phonological systems
Sound waves

- **Simple sounds waves** are described by their **frequency** (Hz) and **amplitude** (dB).

- **Complex sound signals** consist of layers of different sound waves.
  - We can average or sum the signal over time and describe it by a power spectrum: **energy** and **frequency**.
  - Each time point in a signal can also be understood as a distribution of energy across the different frequencies.
Spectrum of a Pure Tone vs. Noise

(a) Pure tone (continuous)

(e) White noise (continuous)
The speech signal occupies a middle ground between pure tones and unstructured noise. It clearly exhibits complex structure.
The speech signal and what we perceive

The signal \(\quad\) via several intermediate steps of auditory processing \(\quad\) Perceived linguistic units

• Continuous
• Multi-dimensional
• Noisy (we’ll see what that means)

• Categorical
• Lower dimensionality

Can you think of why?
Question

• How do these different dimensions get mapped to categorically perceived linguistic units? For example, how do we extract vowels and consonants from the speech signal? How do speakers map linguistic units onto the desired speech signal properties?
(Articulation) ➡️

Energy distribution over frequencies over time ➡️

Acoustic features / Phonetic features ➡️

Phonological features ➡️

Allophones (speech sounds) ➡️

Phone recognition
Today

- Physiology and articulation
- Physics of speech sounds
- **Key acoustic properties of speech**
- Phonological systems
Energy distribution over frequencies over time

↓

Acoustic features / Phonetic features

↓

Phonological features

↓

Allophones (speech sounds)

↓

Phone recognition
Key properties

• The linguistic signal is a physical signal and its acoustic properties map onto phonological features (though not deterministically!)

• The next few slides summarize some key acoustic properties – an exhaustive list is beyond the scope of this class

• I’ll skip over some of them without going into detail, but download the slides and re-read those slides later.
• **Fundamental frequency** \( (F_0) \)
  – Men: 80-240Hz
  – Women: 140-500Hz
  – Children: 170-600Hz

(determined by length and thickness of vocal chords)
Pitch (F0)

When I was little, my parents had one.
Pitch and Sentence Meaning

• Changes in F0 can be used to mark semantic distinctions using **pitch accents**, e.g. in English:
  – Contrast (*That wasn’t ME. It was AL PACINO*)
  – Emphasis (*and it was SOOOO adorable*)
  – Focus (*David only threw a chicken at PETER vs. David only threw a CHICKEN at Peter*)
Pitch and Sentence Structure

• Pitch information is also used to convey **syntactic phrase boundaries** (from Price et al, 1991):

  A) *My experience with slow learners had shown one thing. When you learn* **gradually**, *you worry more.**

  B) *As you begin to study about nuclear war it becomes frightening. When you learn,** **gradually you** worry more.**
• **Pitch range** is co-determined by language/culture
Pitch (or tone) and lexical meaning

- Other languages use F0 (a.k.a. Tone) for lexical distinctions (e.g. Mandarin Chinese)

http://www.wku.edu/~shizhen.gao/Chinese101/pinyin/tones.htm

妈妈骂马的麻吗

1st tone mā means “mother”
2nd tone mǎ means “hemp”
3rd tone má means “horse”
4th tone mà means “to scold”
Key Properties of Speech: Amplitude

- Speech signals also differ in terms of their **amplitude** or **intensity** (intensity over time can be thought of as energy) → one of the primary determinant of **perceived loudness**.
- Not the same as pitch
Intensity, duration and lexical meaning

• Intensity and duration can mark *lexical stress*, which can distinguish different words
  – *desert* (Noun), as in that hot dry place that one shouldn’t get lost in
  – *desert* (Verb), as in to desert someone

• Sometimes together with pitch, intensity and duration can also be associate with *prosodic boundaries*, including (phonological or prosodic) word boundaries and phrase boundaries, e.g.
  – *When you learn gradually, you get better.*
  – *When you learn, gradually you get better.*
When I was little

Word lengthened

Pitch accent
• **Durational** information can be part of phonological contrasts (see later part of this lecture).

• For example, many languages have words with different meanings that are only contrasted by the duration of one of the segments (e.g. short and long vowels in Finnish).
Key Properties of Speech: Formants

• Resonators – cavities amplify certain frequencies and dampen others
  – Bigger cavities = low formant frequencies (think male speakers, who are on average bigger)
  – Smaller cavities = high formant frequencies

• Formants (F₁, F₂, etc.) – Strongest frequencies
  (result from the size and shape of the resonating cavities)
Key Properties of Speech

• Resonance properties modulated by manipulating the articulators
  → More energy on some frequencies
  → Less energy on other frequencies
  → Changes formants
  → Changes sound
When I was little, my parents had one.
When I was little, my parents had one.
• There are many more features that have been identified to be involved in speech recognition.

• Voice onset time
• Spectral tilt
• Etc.
Today

- Physiology and articulation
- Physics of speech sounds
- Key acoustic properties of speech
- Phonological systems
Energy distribution over frequencies over time

\[\downarrow\]

Acoustic features / Phonetic features

\[\downarrow\]

Phonological features

\[\downarrow\]

Allophones (speech sounds)

\[\downarrow\]

Phone recognition
Background: Phonology

- **Phoneme** – a speech sound that can be used to distinguish between two words.
  - **Minimal pairs** define phonemes: *bin* vs. *pin*
  - /b/ and /p/ are phonemes
What is a minimal pair?

- A **minimal pair** is a pair of words with different meanings (i.e. not just two different instances of pronouncing the same word) that have only one sound difference between them.

  Minimal pairs: “tab” and “tag”

  **Not** minimal pairs: “three” and “these” (more than one sound is different)

→ **Conclusion:** /b/ and /g/ are different phonemes!

- **Phonemes are part of language users implicit knowledge of language and hence are cognitively real**
Questions

• How could you prove that /l/, /k/, and /w/ are a phoneme in English?

• Are phonemes language-specific?
  – The set of phonemes differs between languages (follows from definition of minimal pairs)
    • Japanese: no /r/ vs. /l/
    • Hindi: aspirated vs un-aspirated plosives
Energy distribution over frequencies over time

Acoustic features / Phonetic features

Phonological features

Allophones (speech sounds)

Phone recognition
Background: Phonology

- **Phoneme** – a speech sound that can be used to distinguish between two words.
  - *Minimal pairs* define phonemes: *bin* vs. *pin* → */b/* and */p/* are phonemes

- **Allophones** – variants of the same phoneme
  - Compare the */p/* sound in “*pit*” and “*spit*”
  - Both are */p/*, but they sound a little different

- **Phone** – a speech sound (phoneme or allophone).
Putting it all together

(articulation)
\[\downarrow\]
Energy distribution over frequencies over time
\[\downarrow\]
Acoustic features / Phonetic features
\[\downarrow\]
Phonological features
\[\downarrow\]
Allophones (speech sounds)
\[\downarrow\]
Phone recognition
Let’s go through this for the English vowel system

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Sound</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>ee</td>
<td>as in</td>
</tr>
<tr>
<td>ɪ</td>
<td>ɪ</td>
<td>as in</td>
</tr>
<tr>
<td>e</td>
<td>e</td>
<td>as in</td>
</tr>
<tr>
<td>ɛ</td>
<td>ɛ</td>
<td>as in</td>
</tr>
<tr>
<td>æ</td>
<td>ae</td>
<td>as in</td>
</tr>
<tr>
<td>ø</td>
<td>a</td>
<td>as in</td>
</tr>
<tr>
<td>o</td>
<td>o</td>
<td>as in</td>
</tr>
<tr>
<td>ɔ</td>
<td>u</td>
<td>as in</td>
</tr>
<tr>
<td>u</td>
<td>oo</td>
<td>as in</td>
</tr>
<tr>
<td>ə</td>
<td>uh</td>
<td>as in</td>
</tr>
<tr>
<td>ɛr</td>
<td>er</td>
<td>as in</td>
</tr>
<tr>
<td>ai</td>
<td>ai</td>
<td>as in</td>
</tr>
<tr>
<td>øu</td>
<td>au</td>
<td>as in</td>
</tr>
<tr>
<td>ɔɪ</td>
<td>oi</td>
<td>as in</td>
</tr>
</tbody>
</table>

Examples:
- *beat*
- *bit*
- *bait*
- *bet*
- *bat*
- *bought*
- *boat*
- *book*
- *boot*
- *cut*
- *bird*
- *bite*
- *bout*
- *boil*
Speech sounds: Distinctive features

- **Vowels**: Unrestricted vocal tract, lips open, vocal cords vibrating

1. part of tongue (front vs. back)
   - *bet* vs. *butt*

2. position of tongue (high, middle, low)
   - *beet* vs. *bat*
Distinctive features: Vowels

Front

High

Mid

Low

Back
Vowel articulation: Tongue position
Videos taken from the Vocal Tract Visualization Laboratory, University of Maryland
http://speech.umd.edu/3dtongues.html
Vowels distinguished by formants

- Vowels differ in terms of their formants.
- Mostly F1 and F2 are sufficient to distinguish between vowels.

/hid/  /hid/  /hed/  /heed/  /hod/  /hood/  /hud/
heed  hid  head  had  hawed  hood  who'd.
Refresh of spectra

(a) Pure tone (continuous)

(e) White noise (continuous)
Spectra of Vowel Sounds

(h) Vowel ‘eee . . .’ (continuous)

(i) Vowel ‘ah . . .’ (continuous)
## Consonants of English

<table>
<thead>
<tr>
<th>Consonant</th>
<th>Example</th>
<th>Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>as in pea</td>
<td>v</td>
</tr>
<tr>
<td>t</td>
<td>as in tea</td>
<td>ð</td>
</tr>
<tr>
<td>k</td>
<td>as in key</td>
<td>z</td>
</tr>
<tr>
<td>b</td>
<td>as in bee</td>
<td>ʒ</td>
</tr>
<tr>
<td>d</td>
<td>as in do</td>
<td>j</td>
</tr>
<tr>
<td>g</td>
<td>as in go</td>
<td>m</td>
</tr>
<tr>
<td>f</td>
<td>as in fin</td>
<td>n</td>
</tr>
<tr>
<td>θ</td>
<td>as in thin</td>
<td>ŋ</td>
</tr>
<tr>
<td>s</td>
<td>as in sin</td>
<td>l</td>
</tr>
<tr>
<td>š</td>
<td>as in shin</td>
<td>r</td>
</tr>
<tr>
<td>č</td>
<td>as in chin</td>
<td>w</td>
</tr>
<tr>
<td>h</td>
<td>as in honk</td>
<td>j</td>
</tr>
</tbody>
</table>
Distinctive features: consonants

• Restricted vocal tract

• Ways consonants can vary:

  1. **place of articulation**
     
     - bot
     - dot
     - got
     - bilabial dental velar

  2. **manner of articulation**
     
     - e.g. bot shot jot
     - stop fricative affricate

  3. **voicing**
     
     - got - caught
     - dot - tot
     - voiced unvoiced
Distinctive features: consonants

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Postalveolar</th>
<th>Retroflex</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Pharyngeal</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>p b</td>
<td></td>
<td>t d</td>
<td>t q</td>
<td>c j</td>
<td>k g</td>
<td>q g</td>
<td>N</td>
<td></td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Nasal</td>
<td>m mɲ</td>
<td></td>
<td>n</td>
<td>η</td>
<td>η</td>
<td>η N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td>B</td>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tap or Flap</td>
<td>ʋ v ř ŏ ō s z ʃ ʒ ʒ̊</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>φ β f v θ δ s z ʃ ʒ ʒ̊</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral fricative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>ʋ j</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral approximant</td>
<td>ǀ ǀ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Examples: stops

• Each consonant can be specified by a combination of phonological features

• For example, different types of stops in English:
  – /g/: voiced, velar, stop
  – /k/: unvoiced, velar, stop
  – /b/: voiced, labial, stop
  – /p/: unvoiced, labial, stop
  – /d/: voiced, dental, stop
  – /t/: unvoiced, dental, stop
The tongue

Videos taken from the Vocal Tract Visualization Laboratory, University of Maryland
http://speech.umd.edu/3dmovies.html
Examples: fricatives and nasals

• Fricatives:
  – /z/: voiced, dental, fricative
  – /s/: unvoiced, dental, fricative

• Nasals:
  – /m/: voiced, labial, nasal
  – /n/: voiced, dental, nasal
Spectra of fricatives

(compare to vowels)

(k) Consonant ‘ssss . . . ’ (continuous)

(l) Consonant ‘ffff . . . ’ (continuous)
Compare: spectra of vowels

(h) Vowel ‘eee . . . ’ (continuous)

(i) Vowel ‘ah . . . ’ (continuous)
How many phonemes does a language have?

- English has about 30-45 phonemes
- Vowels:
  - Some languages have as few as 2 distinctive vowels
  - Ngwe (Bantu language family) has 38 vowels
- Consonants:
  - Some languages have as few as 6 consonants
  - Ubyx (Northwestern Caucasian) supposedly has eighty-one distinct consonants (but only very few vowels).
- Overall:
  - as few as eleven phonemes to as many as 112 or even 141 (Khoi-San “click”-language)
Cross-linguistic commonalities in phonological systems

• Some phonemes are very common across all languages:
  – Vowels: /i/, /e/, /a/, /o/, /u/
  – Consonants: /p/, /t/, /k/, /m/, /n/

• Can you think of a reason why all or almost all languages have the schwa sound (roughly like in “I got an A”)?
Question

• Why don’t all languages have all the phonemes?

• What are some advantages and disadvantages of having more phonemes?
Today

- Speech production
  - Physiology and articulation
  - Physics of speech sounds
  - Key acoustic properties of speech
  - Phonological system of English

- From theory to reality
  - Coarticulation
Coarticulation

- Sounds are not produced out of context: sounds surrounding a phoneme influence its realization \( \leftarrow \text{co-articulation} \)

→ One reason why synthesized speech often sounds unnatural
Continuous speech

• Recall that spontaneous speech is really fast

• Co-articulation: adjust pronunciation of current sound to take into account preceding and following sounds
  – *kill* vs. *cool*

• Can you describe the **difference in the position of your articulators** during the /k/ production of these two words?
Vowel context alters consonant cues

- Schematic F1 and F2 for the realization of the phoneme /d/ in different contexts
What does this mean?

• We mapping linguistic units (vowels, consonants) onto a multi-dimensional set of articulatory features

• The features are not just spatially defined, but temporally.

• The time course of the realization of a feature depends on surrounding phonological material, including effects of both recently produced and upcoming articulatory movements.
Co-articulation makes sense

• Not only is co-articulation necessary or at least efficient from a speaker’s point of view, but it is also rational from a comprehender’s point of view: Information for a segments overlap so we can get out more in a shorter amount of time
Next time

Speech perception

Co-instructor
Professor Jaeger
Prepare:

- Categorical perception example:
  [http://psych.rice.edu/mmtbtn/language/sPerception/dag_aSound/catDis.html](http://psych.rice.edu/mmtbtn/language/sPerception/dag_aSound/catDis.html)

- McGurk:

- Gating task:
For the ultrasound and 3d recordings of the tongue, see the following publications:

