Phonetics
- Anatomical, acoustic, neural properties
- Production and perception
- Focuses on continuous/physical properties
- Basic unit: phone (=speech sound)
- Narrow transcription: [pʰ]

Phonology
- Abstract representations mediate between the speech stream and its interpretation as meaningful words
- Discrete entities (phonemes, syllables, ...)
- Basic unit: phoneme (smallest contrastive unit of a language)
- Broad transcription: /p/

Phonology
- Phonemes: minimum meaning-differentiating units in a language
- Phonemic inventory: list of the phonemes of a given language
- American English has ~12 vowel phonemes, 3 diphthongs, ~24 consonants

Phonology
- Minimal pairs test: criterion for deciding whether two sounds belong to the same phonemic category.
  - If two words or phrases in a given language differ only in one phonological segment (e.g. English “pin” and “bin”) but are perceived as distinct words, then that sound contrast contains two distinct phonemes (/p/ and /b/).

Phonology
- Allophones: variants (pronunciations) of a given phoneme
  - In English, the “p” sound in “pot” is aspirated (pronounced [pʰɔt]) while the “p” in “spot” is unaspirated (pronounced [spɔt]).
  - [pʰ] and [p] are allophones of the phoneme /p/
Phonology

- Distinctive feature theory: universal set of (binary) features, combining articulation and perception
- Chomsky and Halle (1968). *The Sound Pattern of English*
  - Phonological representations as sequences of segments (phonemes) composed of **distinctive features.**
  - Two levels: **underlying** abstract representation and **surface** phonetic representation

Consonant Perception

- **What is a consonant?**
  - Consonants are sounds produced with a major constriction somewhere in the vocal tract.

English Consonants

<table>
<thead>
<tr>
<th>Place of articulation →</th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voicing →</td>
<td>v</td>
<td>s</td>
<td>z</td>
<td>s</td>
<td>s</td>
<td>s</td>
<td>s</td>
</tr>
<tr>
<td>Stop</td>
<td>p d b</td>
<td>t d b</td>
<td>h k</td>
<td>g</td>
<td>g</td>
<td>g</td>
<td>g</td>
</tr>
<tr>
<td>Nasal</td>
<td>m n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>Fricative</td>
<td>f v θ ð s</td>
<td>3 j รส</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
</tr>
<tr>
<td>Affricate</td>
<td><img src="image.png" alt="Image" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td><img src="image.png" alt="Image" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td><img src="image.png" alt="Image" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Consonant production

- American English has 24 consonant phonemes
- The standard phonetic classification system uses three main features to classify the consonants:
  - **Place of articulation**
  - **Manner of articulation**
  - **Voicing**

Consonant classification

- **Place of articulation**
  - bilabial, labiodental, dental, alveolar, palatal, velar, glottal
- **Manner of articulation**
  - stops, nasal, fricatives, affricates, approximant, lateral
- **Voicing**
  - voiced, voiceless

Stop consonants in English

<table>
<thead>
<tr>
<th>Place of articulation</th>
<th>Voicing</th>
<th>Manner of articulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>bilabial</td>
<td>voiced</td>
<td>stop consonants</td>
</tr>
<tr>
<td>/ba/</td>
<td></td>
<td>/pa/</td>
</tr>
<tr>
<td>alveolar</td>
<td>voiceless</td>
<td>/ta/</td>
</tr>
<tr>
<td>/da/</td>
<td></td>
<td>/ka/</td>
</tr>
<tr>
<td>velar</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Consonant production

- Speech production theory
  - Source-filter concept still applies
  - Relatively narrow constriction or closure
  - Secondary sound source in the upper vocal tract

Source-filter theory for consonants

- **Consonants**: sounds produced with a major constriction somewhere in the vocal tract.
- Source-filter theory still applies, but there are complications:
  1. secondary sound source within the vocal tract
  2. secondary source is aperiodic, turbulent noise (frication) rather than quasi-periodic like the glottal source
  3. secondary source may combine with the glottal source to produce mixed excitation (voiced fricatives)

Acoustic tube perturbation and vowel formants

Source: Stevens (1999, p. 286)

Filter properties of stop consonants

After Stevens (1997)

Filter properties of stop consonants

- **Obstruents** (fricatives, affricates, stops)
- **Sonorants** (nasals, liquids, glides)
- In obstruents, the constriction can divide the resonating vocal tract into two separate cavities.
  - coupled vs. uncoupled
  - interactions with the source
Place of articulation

- What are the *acoustic cues* for place of articulation (i.e., what acoustic properties do listeners use to distinguish /p/, /t/, /k/ (and /b/, /d/, /g/) from each other)?

- **Burst cues**
  - Release burst, frication noise, aspiration noise
  - Spectral shape
  - Static vs. dynamic

- **Formant transition** cues
  - F2, F3 onset vs. target (vowel) frequency

Formant transitions

- During the closure for a stop consonant the vocal tract is completely closed, and no sound is emitted. At the moment of release the resonances change rapidly. These changes are called *formant transitions*. 
Formant transitions

- The first formant (F1) always increases in frequency following the release of a stop closure. F2 and F3 increases or decreases, depending on the place of constriction and the flanking vowels.

Burst cues

- The burst is a brief acoustic transient that occurs at the moment of consonant release.
  - Burst intensity is higher for voiceless stops than for voiced stops
  - Burst frequency varies as a function of place

Burst cues

- Is consonant place information available when the burst is removed?

Stop consonants

(from Kent and Read, 1992)

- Syllable-initial prevocalic stop
- Release (noise burst)
- Transition (formant transitions)
- Closure (stop gap)

Alveolar stop

/ada/

Velar stop

/aga/

• high F2
• low F3
Bilabial stop

Stop consonants
(from Kent and Read, 1992)

Syllable-final postvocalic stop

Transition (formant transitions)
Closure (stop gap)
Release (noise burst)
No release (no burst)

Alveolar stop

Burst cues: /ba/ and /pa/

• Bilabials: Burst energy is concentrated at low frequencies

Burst cues: /da/ and /ta/

• Alveolars: Burst energy is concentrated at high frequencies
Burst cues: /ga/ and /ka/

• Velars: Burst energy is concentrated in the mid range

Place of articulation

What are the acoustic cues for place of articulation (i.e., what acoustic properties do listeners use to distinguish /p/, /t/, /k/ (and /b/, /d/, /g/) from each other)?

1. Frequency of burst
2. Formant transitions (F2 and F3) from the burst to the vowel

Acoustic cues for place of articulation in stops

- Burst cues
  - Release burst, frication noise, aspiration noise
  - Spectral shape
  - Static vs. dynamic
- Formant transition cues
  - F2, F3 onset vs. target (vowel) frequency

Formant transitions

- During the closure for a stop consonant the vocal tract is completely closed, and no sound is emitted. At the moment of release the resonances change rapidly. These changes are called formant transitions.

Formant transitions

- The first formant (F1) always increases in frequency following the release of a stop closure. F2 and F3 increases or decreases, depending on the place of constriction and the flanking vowels.
Burst cues

- The burst is a brief acoustic transient that occurs at the moment of consonant release.
- Burst intensity is higher for voiceless stops than for voiced stops.
- Burst frequency varies as a function of place.

Invariance problem for stop consonants

Cooper, Delattre, Liberman, Borst & Gerstman (1952) JASA 24, 597-606

Invariance problem for stop consonants

Liberman, Delattre, and Cooper (1952)

Transition + vowel without burst

Majority responses

Burst cues

- Is consonant place information available when the burst is removed?
Categorical Perception of stop consonants

Equal acoustic changes → unequal auditory percepts
place of articulation of stops: /b/ vs /d/ vs /g/

Liberman, Harris, Hoffman, and Griffith (1957)
Journal of Experimental Psychology 54, 358-368

Categorical Perception

1. Identification function shows steep slope (cross-over) between categories
2. Good between-category discrimination (near perfect) when the members of the pair belong to different categories
3. Poor within-category discrimination (near chance) when they are perceived to belong to the same category

Segmentation problem for stops

Where are the segments?

“bag”

F3

F2

F1

time

Same noise - different consonant

Different transition - same consonant

1400 Hz
Motor Theory

- Invariance problem
- Segmentation problem
- Categorical perception
- Mapping from continuous to discrete
- Encodedness of speech
- Units of perception (segments, syllables..)
- Linguistic units, acoustics and articulation

K.N. Stevens

- Quantal theory
  - Certain sounds are favored in the languages of the world because their acoustic properties can be produced with a wide range of articulations.

K.N. Stevens

- Distinctive feature theory
  - syllables /bi/, /du/, ...
  - segments /b/, /d/, ...
  - features [+voiced], [-rounded], ...

K.N. Stevens JASA 2002

- Acoustic Landmarks
  - An early stage in the processing of speech identifies acoustic landmarks in the signal such as syllable nuclei and acoustic discontinuities corresponding to consonantal closures and releases.
  - Landmarks point to regions of the signal where a more detailed phonetic analysis is carried out.

K.N. Stevens JASA 2002

- Lexical access from features
  - The acoustic signal in the vicinity of these landmarks is processed by a set of modules, each of which identifies a phonetic feature.

K.N. Stevens JASA 2002

- Lexical access from features
  - From these landmarks and features, lexical (word) hypotheses are evaluated.
  - This is done using analysis-by-synthesis, in which a word sequence is hypothesized, a possible pattern of features from this sequence is internally synthesized, and the synthesized pattern is tested for a match against an acoustically derived pattern.
Burst spectrum

- Blumstein and Stevens (1979, 1980) invariance hypothesis (static spectral shape model)
  - The shape of the spectrum — sampled at the time of burst release — provides invariant cues specifying the place of articulation for the English stop consonants.

Temporal window

Step 2: Position a 25.6 ms half-Hamming window

Acoustic landmarks

Step 3: Compute spectrum of windowed segment

Static spectral shape model

- Predictions:
  1. spectral shape for a given place of articulation is the same for different talkers and in different phonetic contexts.
  2. acoustic modifications that distort other aspects of the syllable but preserve the spectral shape near the burst do not affect consonant identity.
**Dynamic properties**

- Kewley-Port (1983) hypothesized that the identification of stop consonants is based on *time-varying changes* in the spectrum from the onset of the burst into the transition region.

**Locus Equations**

- Delattre et al. (1954) described the locus as the frequency location of F2 extrapolated back in time to the consonant release.

![Diagram of locus equations](image)

---

**Problems with the locus concept**

- No physical energy exists at the time/frequency position of the locus; actual extrapolation of the formant transition may lead to a change in consonant identity.

**Locus equations**

- Sussman et al. (1991, 1993) proposed that locus equations provide *invariant relational cues* for the perception of place of articulation in stop consonants.

- Locus equations describe the relationship between the frequency of F2 at burst onset and F2 in the vowel.

![Diagram of locus equations](image)
Smits, ten Bosch, and Collier (1996)

- **Gross cues**
  - spectral features that are distributed across frequency or time
    - overall spectral shape or its relative change over time
- **Detailed cues**
  - features that are narrowly localized in frequency or time
    - center frequency of prominent spectral peaks and their dynamic transitions.

Cues for voicing

- **VOT** – voice onset time
  - Short lag (0-20 ms) = voiced sounds /b,d,g/
  - Long lag (80-100 ms) = unvoiced sounds /p,t,k/
Voicing

- **Lisker and Abramson (1970)**
  - Studied voice onset time (VOT) in several languages, some with 2 voicing categories (e.g., English, Spanish, Cantonese) and others with 3 or 4 voicing categories (e.g., Thai, Hindi, Korean)
  - Prevoiced
  - Voiced
  - Voiceless aspirated
  - Voiceless unaspirated

VOT in Dutch

Prevoiced and unaspirated stops

VOT in an English-Dutch bilingual child

Voicing

- Identification functions for English VOT continuum.
- Frequency histograms of measured VOT in English stops.

Voicing

- Identification functions for Thai VOT continuum.
- Frequency histograms of measured VOT in Thai stops.

Identification and discrimination of English VOT continuum

Identification function

Discrimination function

Hypothetical data
Categorical Perception

- Relationship between VOT and identification functions is nonlinear, with steep transitions and a clearly defined boundary.
- Discrimination functions show a peak near the phoneme boundary; near chance levels elsewhere.