Formant frequencies for vowels

- Close match between formant frequencies and traditional impressionistic phonetic descriptions of tongue height and advancement (Joos, 1948).
- Formant plots show a rotated/flipped version of the traditional (impressionistic) vowel space defined by height and advancement.

American English vowel space

- Peterson and Barney (1952) diagram showing formant frequencies for vowels.

Invariance problem

- Graphical interpretation of CLIH (sliding template) model.
Acoustic cues for vowels

- Formant frequencies (F1, F2, F3)
- Fundamental frequency (F0)
- Other aspects of spectral shape
- Duration differences
- Formant movement (VISC)

Perceptual role of formants

- Descriptive studies of vowel systems
- Correspondence with speech production
- Application in speech synthesis
- Correspondence with listener judgments
- Other (non-formant-related) aspects of vowel spectra are relatively unimportant perceptually (can be eliminated or removed, as in Pattern Playback)

Perceptual role of formants

- Formants can be hard to extract and measure (especially in children’s voices)
- Vowel classification using spectral shape features sometimes produce higher classification accuracy (but they may be relying on similar information)
- When formant peaks are removed from the spectrum listeners still hear the same vowel (Ito et al., 2001)

Consonant Perception

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

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

English Consonants

<table>
<thead>
<tr>
<th>Place of articulation</th>
<th>Bilabial</th>
<th>Labio-dental</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voicing</td>
<td>–</td>
<td>+</td>
<td></td>
<td>+</td>
<td>–</td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>Stop</td>
<td>p</td>
<td>b</td>
<td>t</td>
<td>d</td>
<td>k</td>
<td>g</td>
<td>–</td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td>s</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Fricative</td>
<td>f</td>
<td>s</td>
<td>h</td>
<td>th, f</td>
<td>sh, ssh</td>
<td>sh</td>
<td>–</td>
</tr>
<tr>
<td>Affricate</td>
<td>t</td>
<td>d</td>
<td>k</td>
<td>q</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Approximant</td>
<td>s, ss</td>
<td>r</td>
<td>j</td>
<td>w</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td>–</td>
<td></td>
<td>i</td>
</tr>
</tbody>
</table>

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>alveolar</td>
<td>voiced</td>
<td></td>
</tr>
<tr>
<td>velar</td>
<td>voiceless</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)

Filter properties of stop consonants

Source-filter theory for 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)?

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.

Formant transitions: /ba/ and /pa/
- Bilabials: low F2 and F3 onset frequencies
Formant transitions: /da/ and /ta/ 

- Alveolars: high F3 and relatively high F2 onset frequencies

Formant transitions: /ga/ and /ka/ 

- Velars: intermediate F2 and F3 frequencies; F2 and F3 appear to converge near the point of consonant release.

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: /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
**Bilabial stop**

- Time (ms)
- Frequency (kHz)
- 0
- 100
- 200
- 300
- 400
- 500
- 0
- 1
- 2
- 3
- 4

- Closure
- Release
- Formant transitions
- • low F2
- • low F3

**Velar stop**

- Time (ms)
- Frequency (kHz)
- 0
- 100
- 200
- 300
- 400
- 500
- 0
- 1
- 2
- 3
- 4

- Closure
- Release
- Formant transitions
- • high F2
- • low F3

**Burst cues**

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

**Continuum experiments**

- Generate a continuum of sounds that span two phonetic categories. For example:
  - **Place continuum**: vary formant transitions (F2 alone, or F2 and F3 combined) in small steps to see if this acoustic cue controls listeners' judgments of place of articulation, ranging from /ba/ vs. /da/.
  - **Voicing continuum**: vary the voice onset time (VOT) in small steps (e.g. 10 ms) between the category endpoints, /pa/ and /ba/.

**Phonetic identification task**

- Randomize the order of presentation and ask listeners to make a two-alternative forced choice between /pa/ and /ba/ (or in some cases a 3-way choice between /ba/, /da/, and /ga/).
- Plot the identification functions and measure the 50% crossover point (“phoneme boundary”).

**Continuum experiments**

- Typical results:
Formant transitions

- Speech analysis and synthesis experiments in the 1960s showed that F2 and F3 transitions provide important cues for place of articulation in stop consonants.

/b/ - /d/ - /g/ continuum

Formant transitions

- Unfortunately this simple picture becomes more complicated when the vowel changes.
- Formant patterns for /b,d,g/ in different vowels contexts.

Invariance problem for place of articulation in stop consonants
Release burst

- Additional information for identifying the place of articulation of stop consonants comes from the burst.
- Varying the frequency of the burst alters the identity of the consonant. High frequency bursts tend to be heard as alveolar /t,d/, mid-frequency bursts as velar /k,g/ and low-frequency bursts are heard as bilabial /p,b/.

Invariance problem for stop consonants

Majority responses

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

Invariance problem for stop consonants

Liberman, Delattre, and Cooper (1952)

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

Categorical Perception demo

- Labeling experiment (BA-DA-GA continuum)
- Discrimination experiment (BA-DA-GA continuum)

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/

http://www.ling.gu.se/~anders/KatPer/Applet/index.eng.html
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

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
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.

Voicing cues in stop consonants

Role of pitch in speech

- Pitch is the dimension of auditory perception that makes it possible to rank sounds on a scale from low to high.

Pitch perception

- Sounds that evoke a clear sensation of pitch are periodic; models of pitch perception describe how the auditory system estimates the periodicity (repetition rate, $T$) or its inverse, the fundamental frequency, $F_0$.

Harmonicity and Periodicity

- Period: regularly repeating pattern in the waveform

Period duration, $T_0 = 6$ ms
Harmonicity and Periodicity

- **Harmonic**: regularly repeating peak in the amplitude spectrum

\[ F_0 = \frac{1000}{6} = 166 \text{ Hz} \]

\[ F_0 = \frac{1}{T_0} \]

Pitch in speech is determined by the rate of vocal fold oscillation

Pitch variation in speech

- \( F_0 \) – repetition rate of vocal fold vibration
- Pitch – perceived aspect
  - log scale:
    - 1 octave = 12 semitones (1 semitone ≈ 6%)
    - The semitone difference, \( D \), between two frequencies \( f_1 \) and \( f_2 \) is calculated as:
      \[
      D = 12 \log_2 \left( \frac{f_1}{f_2} \right) = 12 \log_{10} \left( \frac{f_1}{f_2} \right)
      \]

Log scale

ERB-rate scale

- Moore and Glasberg (1989) auditory model
- ERB: Equivalent Rectangular Bandwidth
- ERB units provide approximately equal distances along the basilar membrane
  \[
  E = 16.7 \log_{10} (1 + f / 165.4)
  \]
  and
  \[
  f = 165.4(10^{0.06E} - 1)
  \]
ERB-rate scale

F₀ range in speech
- 80-200 Hz for adult males
- 180-400 Hz for adult females
- 200-600 Hz for young children
- Even-tempered Scale for the Octave Above Middle C

F₀ measurement

F₀ as a function of age and sex

F₀ distribution – child talkers

F₀ distribution – males (blue)
Intrinsic pitch (Intrinsic f₀)

- On average, high vowels like /i/ and /u/ are produced on a higher intrinsic pitch than low vowels like /a/.
- Involuntary side-effect or deliberate strategy?
  - Infants show similar pattern (Whalen)
  - Katz & Assmann (2001) neutralized intrinsic pitch differences in vowels but found no change in identification accuracy.

Tone languages

- In tone languages such as Mandarin, a difference in the level and/or movement of pitch associated with a given phoneme can serve to differentiate pairs of words.
- Tone languages may have several tones (e.g. Mandarin: high level, high rising, low falling, high falling)

Role of Pitch Information

- Pitch contour is the primary cue for tone recognition
  - Tonal languages rely on pitch level and differences to convey lexical meanings within syllables
- Pitch helps to segregate auditory components from different sound sources

Tone contrasts in Mandarin

Lexical tones in Mandarin

Demo:
http://hctv.humnet.ucla.edu/departments/Linguistics/VowelsAndConsonants/vowels/chapter2/index/recording2-1.html
Pitch and vowel identification

- There is a systematic relationship between F0 and formant frequencies across voices (low-pitched voices tend to have lower formants than high-pitched voices and vice versa).

Pitch and consonant voicing

- Voice pitch is higher following a voiceless consonant compared to a voiced consonant.
- Listeners perceive these small changes; voicing judgments are influenced by F0 of the following vowel.

Perceptual units in speech perception

- words
- syllables
- phonemes, segments
- context-sensitive allophones
- diphones