Pipeline Model

N-tube model
N-tube model

http://clas.mq.edu.au/speech/acoustics/frequency/vocal_tract_resonance.html

Labial
b, p, m

Labio-dental
f, v

Interdental
θ, δ

Alveolar
d, t, s, z, n
Palatal

š, ž

Velar

g, k, η
Four general classes of sounds in American English

- Vowels, diphthongs, semivowels, and consonants
- Each can be further divided according to articulators (manner, place)
Alternatively, phoneme classes can be divided into

- **Continuants**: produced by a fixed vocal tract configuration
  - Includes vowels, fricatives, and nasals
- **Non-continuants**: vocal tract configuration changes over time
  - Diphthongs, semivowels, stops and affricatives
Ultrasound

Articulograph

rtMRI
Articulation to acoustic

For vowels

[AH] as in "FATHER"

[EE] as in "HEED"

[OO] as in "POOL"
Formant

Formants are frequency peaks which have, in the spectrum, a high degree of energy. They are especially prominent in vowels. Each formant corresponds to a resonance in the vocal tract (roughly speaking, the spectrum has a formant every 1000 Hz). First three formant for few vowels (with example word and IPA symbol) are:

<table>
<thead>
<tr>
<th>Vowel</th>
<th>F1(Hz)</th>
<th>F2(Hz)</th>
<th>F3(Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>heed</td>
<td>iː</td>
<td>280</td>
<td>2620</td>
</tr>
<tr>
<td>hid</td>
<td>ɪ</td>
<td>360</td>
<td>2220</td>
</tr>
<tr>
<td>head</td>
<td>e</td>
<td>600</td>
<td>2060</td>
</tr>
<tr>
<td>had</td>
<td>æ</td>
<td>800</td>
<td>1760</td>
</tr>
<tr>
<td>hudd</td>
<td>æ</td>
<td>760</td>
<td>1320</td>
</tr>
<tr>
<td>hard</td>
<td>ɑː</td>
<td>740</td>
<td>1180</td>
</tr>
<tr>
<td>hod</td>
<td>ʌ</td>
<td>560</td>
<td>920</td>
</tr>
<tr>
<td>hoard</td>
<td>ɔː</td>
<td>480</td>
<td>760</td>
</tr>
<tr>
<td>hood</td>
<td>u</td>
<td>380</td>
<td>940</td>
</tr>
<tr>
<td>Who’d</td>
<td>ʊː</td>
<td>320</td>
<td>920</td>
</tr>
<tr>
<td>heard</td>
<td>ɔː</td>
<td>560</td>
<td>1480</td>
</tr>
</tbody>
</table>

Adult male formant frequencies in Hertz collected by J.C. Wells around 1960. Note how F1 and F2 vary more than F3.
Frequency of second formant \textit{versus} frequency of first formant for ten vowels by 76 speakers.
Speech production models

DIVA Model

Guenther, Ghosh, and Tourville (2006) *Brain and Language*
http://www.bu.edu/speechlab/research/the-diva-model/
Speech production models

Articulatory Model

Speech production models

TaDA Model


Forward Model

The human ear
Cochlea – organ of hearing
Tonotopic Mapping
Audiogram

- Frequency Hz:
  - Low pitch
  - High pitch

- Hearing level dB:
  - Quiet
  - Loud

- Normal hearing
- Mild
- Moderate
- Severe
- Profound