Ultrasound

Articulograph

rtMRI
Articulation to acoustic

For vowels

after Fant
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>Formant 1 (Hz)</th>
<th>Formant 2 (Hz)</th>
<th>Formant 3 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>heed</td>
<td>i: 280</td>
<td>2620</td>
<td>3380</td>
</tr>
<tr>
<td>hid</td>
<td>i 360</td>
<td>2220</td>
<td>2960</td>
</tr>
<tr>
<td>head</td>
<td>e 600</td>
<td>2060</td>
<td>2840</td>
</tr>
<tr>
<td>had</td>
<td>æ 800</td>
<td>1760</td>
<td>2500</td>
</tr>
<tr>
<td>hudd</td>
<td>ñ 760</td>
<td>1320</td>
<td>2500</td>
</tr>
<tr>
<td>hard</td>
<td>a: 740</td>
<td>1180</td>
<td>2640</td>
</tr>
<tr>
<td>hod</td>
<td>u 560</td>
<td>920</td>
<td>2560</td>
</tr>
<tr>
<td>hoard</td>
<td>œi 480</td>
<td>760</td>
<td>2620</td>
</tr>
<tr>
<td>hood</td>
<td>œ 380</td>
<td>940</td>
<td>2300</td>
</tr>
<tr>
<td>Who’d</td>
<td>œi 320</td>
<td>920</td>
<td>2200</td>
</tr>
<tr>
<td>heard</td>
<td>ãi 560</td>
<td>1480</td>
<td>2520</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.
Formant

Frequency of second formant *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

Cochlea – organ of hearing
Tonotopic Mapping

- 60 Hz Sine Wave
- 300 Hz Sine Wave
- 2000 Hz Sine Wave
Central Auditory system
Audiogram
Theories of Speech Perception

• **Active theories** suggests that speech perception and production are closely related
  – Listener knowledge of how sounds are produced facilitates recognition of sounds

• **Passive theories** emphasizes the sensory aspects of speech perception
  – Listeners utilize internal filtering mechanisms
  – Knowledge of vocal tract characteristics plays a minor role, for example when listening in noise conditions
Bottom up Top Down

• Top-down processing works with knowledge a listener has about a language, context, experience, etc.
  – Listeners use stored information about language and the world to make sense of the speech

• Bottom-up processing works in the absence of a knowledge base providing top-down information
  – Listeners receive auditory information, convert it into a neural signal and process the phonetic feature information

Knowledge driven top-down approach are less resistant to Additive Noise in case of non-sense syllable, nonsense words, incoherent sentence, short utterance, ungrammatical sentence
Bottom-up approach results in error propagation upto top
Autonomous vs. Interactive

- **Autonomous theories** posit feed-forward processing with lexical influence restricted to post-perceptual decision processes (uni-directional)

- **Interactive theories** posit information and knowledge from many sources available to the listener at any or all stages of the processing of the signal (bi-directional)
Theories of Speech Perception

Marslen-Wilson’s Cohort Model

- Mental representations of words activated (in parallel) on the basis of bottom-up input (sounds)
- Can be de-activated by subsequent input
  - bottom-up (phonological)
  - top-down (contextual)

(Marslen-Wilson, 1980)
Theories of Speech Perception

TRACE Model

- Like the interactive-activation model of printed word recognition, TRACE has three sets of interconnected detectors
  - Feature detectors
  - Phoneme detectors
  - Word detectors
- These detectors span different stretches of the input (feature detector span small parts, word detectors span larger parts)
- The input is divided into “time slices” which are processed sequentially.

(McClelland & Elman, 1986)
Wilson et al., 2004

- Black areas are premotor and primary motor cortex activated when subjects produced the syllables.
- White arrows indicate central sulcus.
- Orange represents areas activated by listening to speech.
- Extensive activation in superior temporal gyrus.
- Activation in motor areas involved in speech production (!).
Theories of Speech Perception

Motor Theory

Motor theory postulates that speech is perceived by reference to how it is produced; that is, when perceiving speech, listeners access their own knowledge of how phonemes are articulated. Articulatory gestures such as rounding or pressing the lips together are units of perception that directly provide the listener with phonetic information.

(Liberman, et al., 1967; Liberman & Mattingly, 1985)
Theories of Speech Perception

Analysis by Synthesis (Stevens & Halle, 1960)
- In this model, speech perception is based on auditory matching mediated through speech production.

When a listener hears a speech signal, he or she analyzes it by mentally modeling the articulation (in other words, the listener tries to synthesize the speech his or herself). If the ‘auditory’ result of the mental synthesis matches the incoming acoustic signal, the hypothesized perception is interpreted as correct.
Theories of Speech Perception

Direct Realist Theory (Fowler, 1986)

- Direct realism postulates that speech perception is direct (i.e., happens through the perception of articulatory gestures), but it is not special. All perception involves direct recovery of the distal source of the event being perceived (Gibson).
  
  In vision, you perceive objects (e.g., trees, cars, etc.). Likewise with smell you perceive e.g., cookies, roses, etc. Why not in the auditory perception of speech?
- So, listeners perceive tongues and lips.

The articulatory gestures that are the objects of speech perception are not intended gestures (as in Motor Theory). Rather, they are the actual gestures.
Psycho-acoustic Experiments

Fletcher Munson Curve – loudness as a function of frequency and intensity