6.003: Signal Processing

Speech

- source/filter model of speech production
- speech analysis
- speech synthesis

Quiz 2: April 12, 2-4pm
- Coverage up to and including April 5 and HW 8.
- Closed book except for two pages of notes (four sides total)
- No electronic devices. (No headphones, cellphones, calculators, ...)
- No HW 9 – a practice quiz is posted (along with solutions)
- Quiz Practice Session: Friday 3-5pm in 36-144
  (in addition to regular office hours on Friday 1-3pm in 36-144)

April 7, 2022

Source/Filter Model of Speech Production

Speech is generated by the passage of air from the lungs, through the vocal cords, mouth, and nasal cavity.

Vibrations of the vocal cords are “filtered” by the mouth and nasal cavities to generate speech.

During voiced speech, the glottis generates puffs of air that are a few ms in duration. The frequency of puffs ranges from 100–300 Hz.

Speech Production

X-ray movie showing speech in production.

Vowels sound different because mouth and lip positions are different.
**Source/Filter Model of Speech Production**

Harmonic content is a natural way to describe vowel sounds.

<table>
<thead>
<tr>
<th>bat</th>
<th>bat</th>
<th>bet</th>
<th>bet</th>
</tr>
</thead>
<tbody>
<tr>
<td>bait</td>
<td>bait</td>
<td>bought</td>
<td>bought</td>
</tr>
<tr>
<td>bet</td>
<td>beet</td>
<td>beet</td>
<td>beet</td>
</tr>
<tr>
<td>bit</td>
<td>bit</td>
<td>boot</td>
<td>boot</td>
</tr>
<tr>
<td>bite</td>
<td>bite</td>
<td>bought</td>
<td>bought</td>
</tr>
<tr>
<td>bought</td>
<td>bought</td>
<td>boat</td>
<td>boat</td>
</tr>
<tr>
<td>but</td>
<td>but</td>
<td>boot</td>
<td>boot</td>
</tr>
</tbody>
</table>

**Demonstration**

Physical model of the vocal tract.

Buzzer represents sound from glottis.
Machined cavities represent vocal tract.

**Formants**

Resonant frequencies of the vocal tract.*

<table>
<thead>
<tr>
<th>Formant</th>
<th>heed</th>
<th>heed</th>
<th>had</th>
<th>hod</th>
<th>haw’d</th>
<th>who’d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>F1</td>
<td>270</td>
<td>530</td>
<td>660</td>
<td>730</td>
<td>570</td>
</tr>
<tr>
<td></td>
<td>F2</td>
<td>2290</td>
<td>1840</td>
<td>1720</td>
<td>1090</td>
<td>840</td>
</tr>
<tr>
<td></td>
<td>F3</td>
<td>3010</td>
<td>2480</td>
<td>2410</td>
<td>2440</td>
<td>2410</td>
</tr>
<tr>
<td>Women</td>
<td>F1</td>
<td>310</td>
<td>610</td>
<td>850</td>
<td>590</td>
<td>370</td>
</tr>
<tr>
<td></td>
<td>F2</td>
<td>2790</td>
<td>2330</td>
<td>2050</td>
<td>1220</td>
<td>920</td>
</tr>
<tr>
<td></td>
<td>F3</td>
<td>3310</td>
<td>2990</td>
<td>2810</td>
<td>2850</td>
<td>2710</td>
</tr>
<tr>
<td>Children</td>
<td>F1</td>
<td>370</td>
<td>690</td>
<td>1030</td>
<td>680</td>
<td>430</td>
</tr>
<tr>
<td></td>
<td>F2</td>
<td>3200</td>
<td>2610</td>
<td>2320</td>
<td>1370</td>
<td>1060</td>
</tr>
<tr>
<td></td>
<td>F3</td>
<td>3730</td>
<td>3570</td>
<td>3320</td>
<td>3170</td>
<td>3180</td>
</tr>
</tbody>
</table>

* [http://www.sfu.ca/sonic-studio/handbook/Formant.html](http://www.sfu.ca/sonic-studio/handbook/Formant.html)

**Speech Production**

Same glottis signal + different formants → different vowels.

We detect changes in the filter function to recognize vowels.

* [https://linguistics.ucla.edu/people/hayes/103/Charts/VChart](https://linguistics.ucla.edu/people/hayes/103/Charts/VChart)
We detect changes in the filter function to recognize vowels ... at least sometimes.

Demonstration.

“la” scale.
“lore” scale.
“loo” scale.
“ler” scale.
“lee” scale.
Low Frequency: “la” “lore” “loo” “ler” “lee”.
High Frequency: “la” “lore” “loo” “ler” “lee”.

“Flights from Denver ...” was analyzed with the source/filter model and a new sound was produced using a modified model.

What part of the model was changed?

1. Original
2. Modification #1
3. Modification #2
4. Modification #3

Summary
Introduction to speech processing
- source/filter model of speech production
- speech analysis
- speech synthesis