6.003: Signal Processing

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

- Quiz 2: November 2, 2-4pm, 50-340 (Walker)
  - Coverage up to and including all of week 7, including HW7.
  - Closed book except for two page of notes (four sides total)
  - No electronic devices. (No headphones, cellphones, calculators, ...)

- No HW8 – a practice quiz is posted.

October 28, 2021
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.
Controlled by complicated muscles, vocal cords are set in vibration by the passage of air from the lungs.

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.
Source/Filter Model of Speech Production

Vibrations of the vocal cords are “filtered” by the mouth and nasal cavities to generate speech.
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 natural way to describe vowel sounds.
Harmonic content is a natural way to describe vowel sounds.
Vibrations of the vocal cords are “filtered” by the mouth and nasal cavities to generate speech.
Demonstration

Physical model of the vocal tract.

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

Resonant frequencies of the vocal tract.

![Graph showing formants](http://www.sfu.ca/sonic-studio/handbook/Formant.html)

<table>
<thead>
<tr>
<th>Formant</th>
<th>heed</th>
<th>head</th>
<th>had</th>
<th>hod</th>
<th>haw’d</th>
<th>who’d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>270</td>
<td>530</td>
<td>660</td>
<td>730</td>
<td>570</td>
<td>300</td>
</tr>
<tr>
<td>F2</td>
<td>2290</td>
<td>1840</td>
<td>1720</td>
<td>1090</td>
<td>840</td>
<td>870</td>
</tr>
<tr>
<td>F3</td>
<td>3010</td>
<td>2480</td>
<td>2410</td>
<td>2440</td>
<td>2410</td>
<td>2240</td>
</tr>
<tr>
<td>Women</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>310</td>
<td>610</td>
<td>860</td>
<td>850</td>
<td>590</td>
<td>370</td>
</tr>
<tr>
<td>F2</td>
<td>2790</td>
<td>2330</td>
<td>2050</td>
<td>1220</td>
<td>920</td>
<td>950</td>
</tr>
<tr>
<td>F3</td>
<td>3310</td>
<td>2990</td>
<td>2850</td>
<td>2810</td>
<td>2710</td>
<td>2670</td>
</tr>
<tr>
<td>Children</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>370</td>
<td>690</td>
<td>1010</td>
<td>1030</td>
<td>680</td>
<td>430</td>
</tr>
<tr>
<td>F2</td>
<td>3200</td>
<td>2610</td>
<td>2320</td>
<td>1370</td>
<td>1060</td>
<td>1170</td>
</tr>
<tr>
<td>F3</td>
<td>3730</td>
<td>3570</td>
<td>3320</td>
<td>3170</td>
<td>3180</td>
<td>3260</td>
</tr>
</tbody>
</table>

http://www.sfu.ca/sonic-studio/handbook/Formant.html
Time plot & spectrogram of "flights from Denver to San Francisco."
Speech Production

Same glottis signal + different formants → different vowels.

We detect changes in the filter function to recognize vowels.
Singing

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”.

Speech Production

Same glottis signal + different formants $\rightarrow$ different vowels.

We detect changes in the filter function to recognize vowels.
Speech Production

Same glottis signal + different formants $\rightarrow$ different vowels.

We detect changes in the filter function to recognize vowels.
Speech Production

Same glottis signal + different formants $\rightarrow$ different vowels.

We detect changes in the filter function to recognize vowels.
Source/Filter Model

Model of speech production.

\[ N_p \]
\[ \text{Pulse Train} \]
\[ \text{Noise} \]
\[ H(\Omega) \]
\[ G \]
\[ s[n] \]

Acoustic sources:
- pulse train with period \( N_p \) for voiced utterances
- gaussian noise for unvoiced utterances

Gain: \( G \) controls loudness

Vocal tract: filter represented shapes of mouth, tongue, and lips
Time and Frequency Structure of Speech

Time plot & spectrogram of "flights from Denver to San Francisco."
Source/Filter Model

Model of speech production.

Acoustic sources:
- pulse train with period $N_p$ for voiced utterances
- gaussian noise for unvoiced utterances

Gain: $G$ controls loudness

Vocal tract: filter represented shapes of mouth, tongue, and lips
"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
"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
"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
"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
"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
"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
"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
"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
"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