

6.3000: Signal Processing

Speech

Quiz 2: April 16, 2-4pm, 50-340 (Walker).

- Closed book except for two pages of notes (four sides total)
- No electronic devices. (No headphones, cellphones, calculators, ...)
- Coverage up to and including classes on April 9 and HW 8.

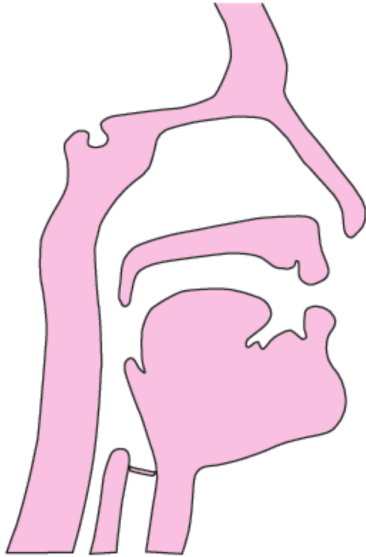
More information under the **Quiz 2 Information** tab of the 6.300 website.
There is no HW 9.

If you have personal or medical difficulties, please contact S³ and/or 6.3000-instructors@mit.edu for accommodations.

April 14, 2026

Speech Production

Motions of lips and chin are essential to speech production.
But how does it work?



Cross-section of human head showing forehead, nose, lips, chin, and neck.

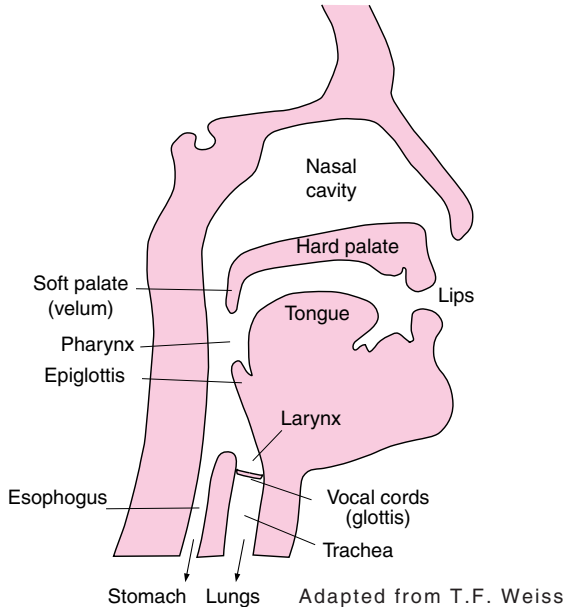
Speech Production

X-ray movie showing speech in production.



Source/Filter Model of Speech Production

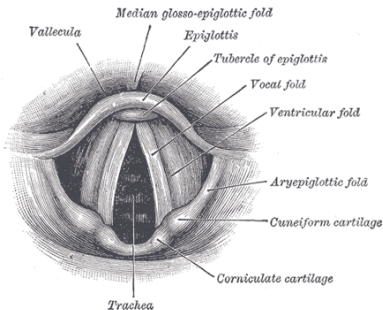
Two parts of speech production: the **source** and the **filter**.



Source/Filter Model of Speech Production

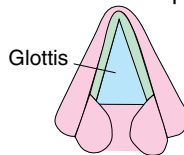
Controlled by complicated muscles, vocal cords are set in vibration by the passage of air from the lungs.

Looking down the throat:

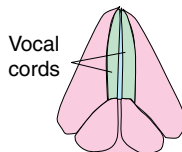


Gray's Anatomy

Vocal cords open



Vocal cords closed

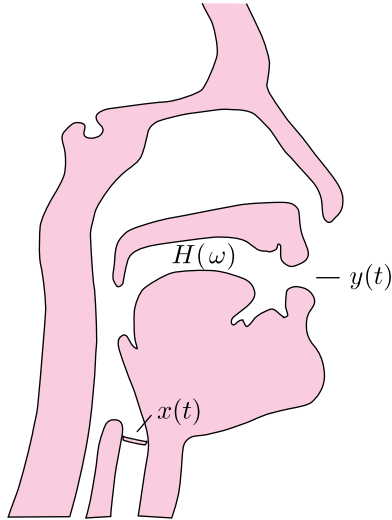


Adapted from T.F. Weiss

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.



buzz from
vocal cords



mouth, lips and
nasal cavities



speech

Demonstration

Physical model of the vocal tract.



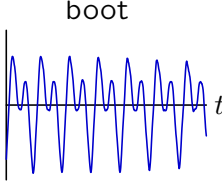
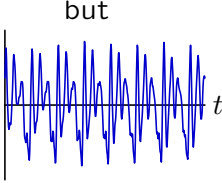
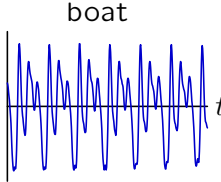
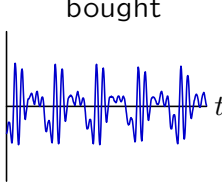
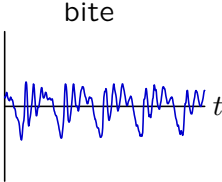
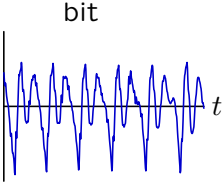
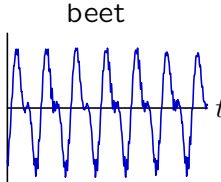
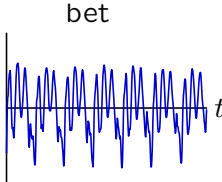
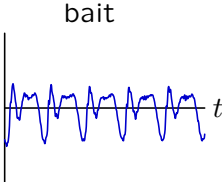
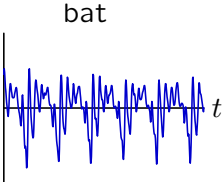
Buzzer represents sound from glottis.

Machined cavities represent vocal tract.

Chiba and Kajiyama Model replicated by Takayuki Arai.

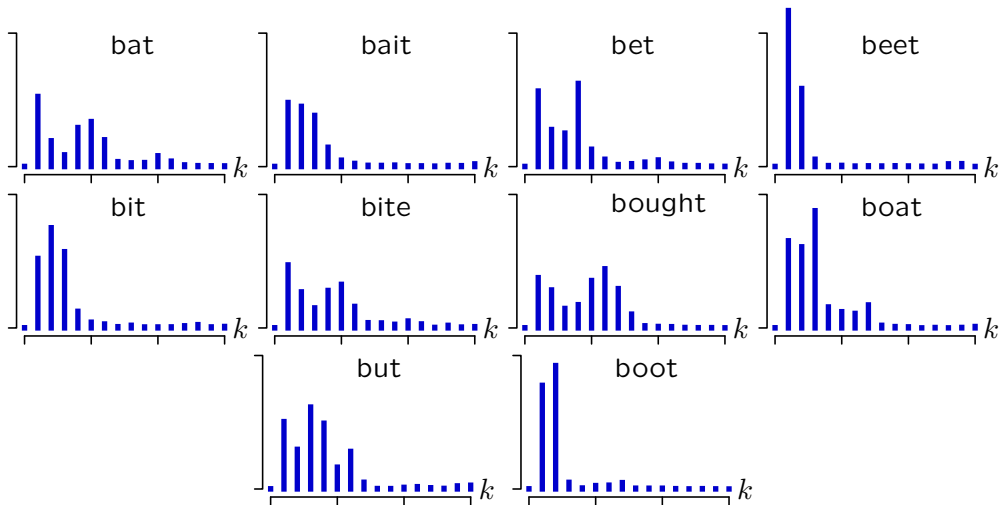
Source/Filter Model of Speech Production

Vowels sound different because mouth and lip positions are different.



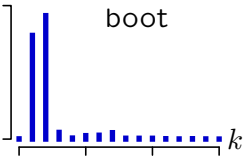
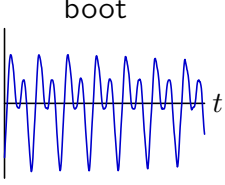
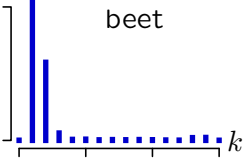
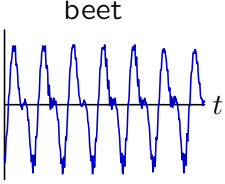
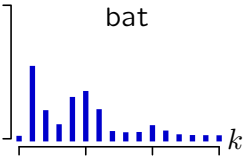
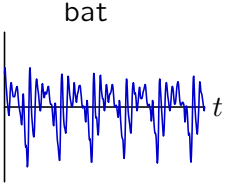
Source/Filter Model of Speech Production

Harmonic content is natural way to describe vowel sounds.



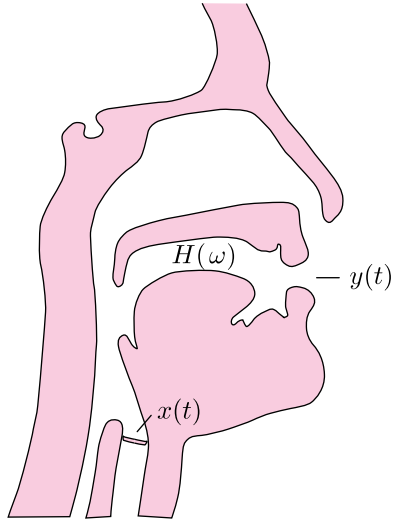
Source/Filter Model of Speech Production

Harmonic content is natural way to describe vowel sounds.



Source/Filter Model of Speech Production

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



buzz from
vocal cords



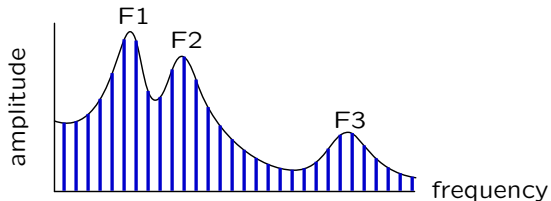
throat and
nasal cavities



speech

Formants

Resonant frequencies of the vocal tract.*

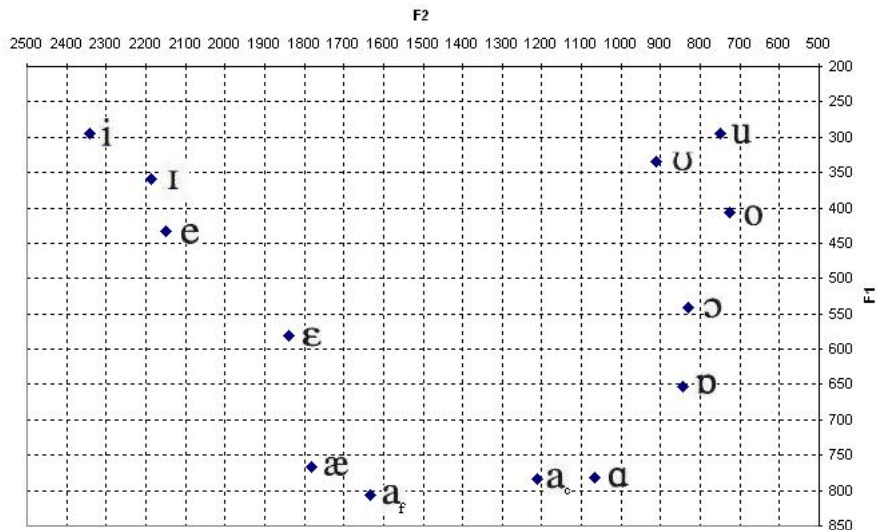


| | Formant | heed | head | had | hod | haw'd | who'd |
|----------|---------|------|------|------|------|-------|-------|
| Men | F1 | 270 | 530 | 660 | 730 | 570 | 300 |
| | F2 | 2290 | 1840 | 1720 | 1090 | 840 | 870 |
| | F3 | 3010 | 2480 | 2410 | 2440 | 2410 | 2240 |
| Women | F1 | 310 | 610 | 860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

* <http://www.sfu.ca/sonic-studio/handbook/Formant.html>

Formants

Formant frequencies for common vowels.*

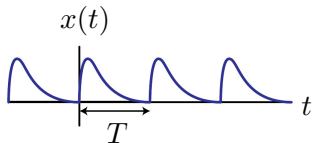


* <https://linguistics.ucla.edu/people/hayes/103/Charts/VChart>

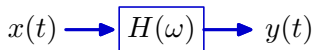
Speech Production

Same glottis signal + different formants \rightarrow different vowels.

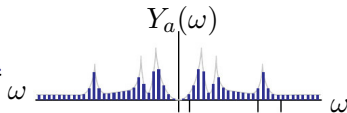
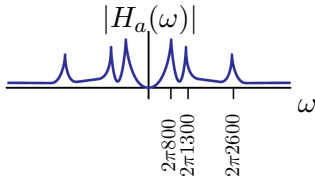
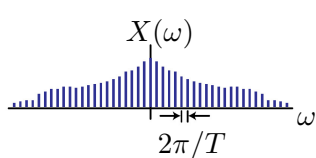
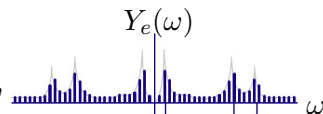
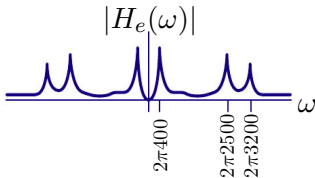
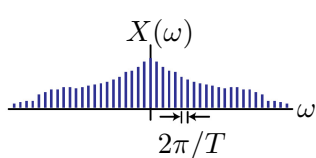
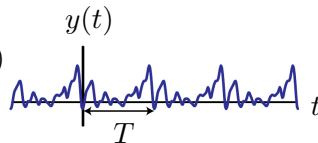
glottis signal



vocal tract filter



vowel sound



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.

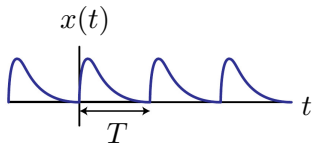
Low Frequency: “la” “lore” “loo” “ler” “lee” .

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

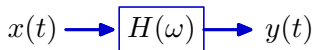
Speech Production

Same glottis signal + different formants \rightarrow different vowels.

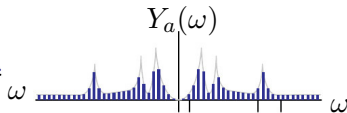
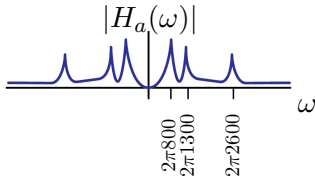
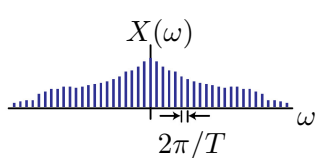
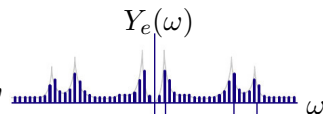
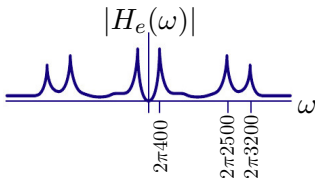
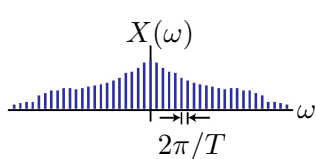
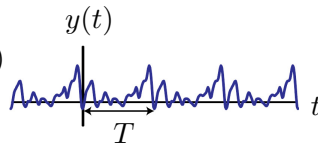
glottis signal



vocal tract filter



vowel sound

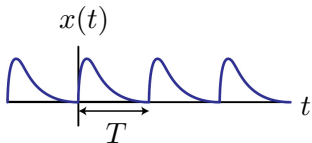


We detect changes in the filter function to recognize vowels.

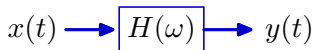
Speech Production

Same glottis signal + different formants \rightarrow different vowels.

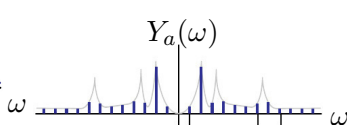
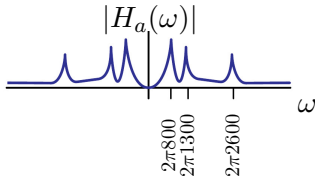
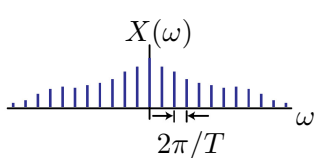
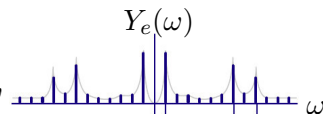
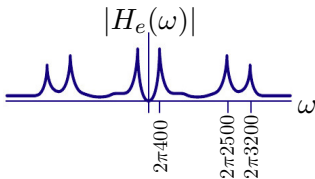
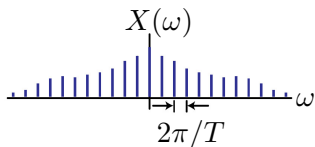
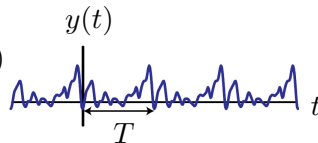
glottis signal



vocal tract filter



vowel sound

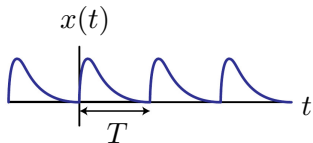


We detect changes in the filter function to recognize vowels.

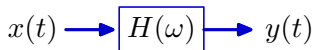
Speech Production

Same glottis signal + different formants \rightarrow different vowels.

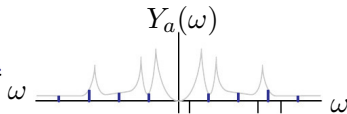
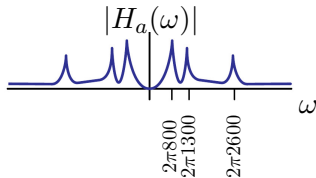
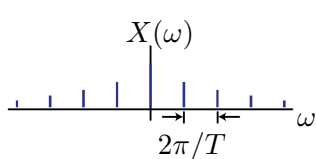
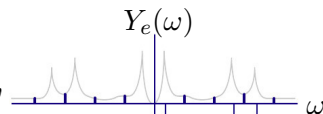
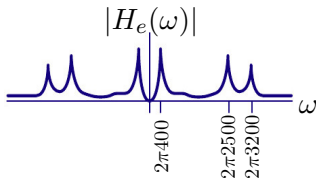
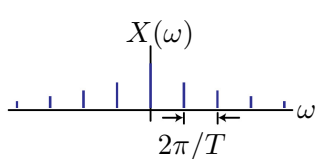
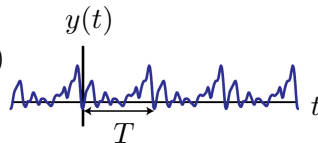
glottis signal



vocal tract filter



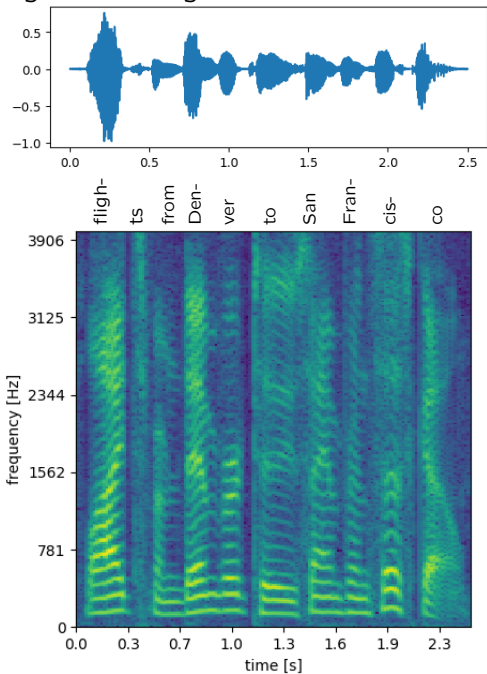
vowel sound



We detect changes in the filter function to recognize vowels.

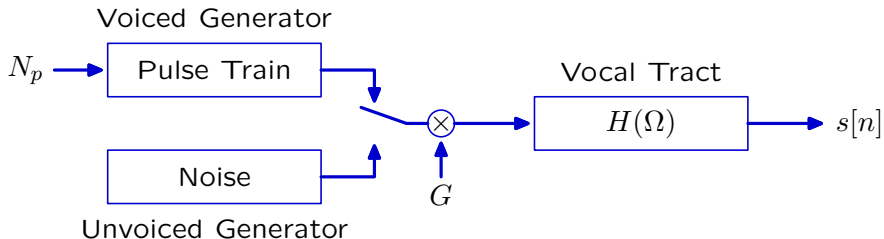
Time and Frequency Structure of Speech

Time plot & spectrogram of "flights from Denver to San Francisco."



Model of Running Speech

Model of speech production.



Acoustic sources:

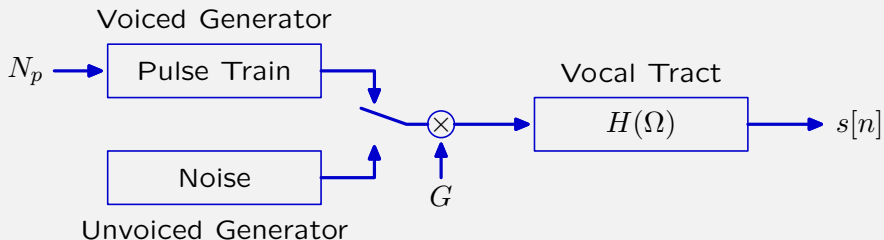
- pulse train with period N_p for voiced utterances
- gaussian noise for unvoiced utterances

Gain: G controls loudness

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

Model of Running Speech

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

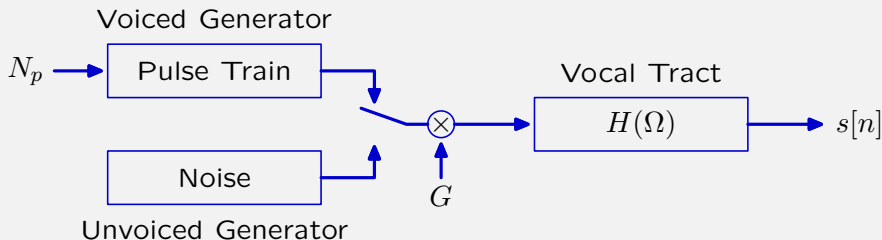


What part of the model was changed?

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

Source/Filter Model

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

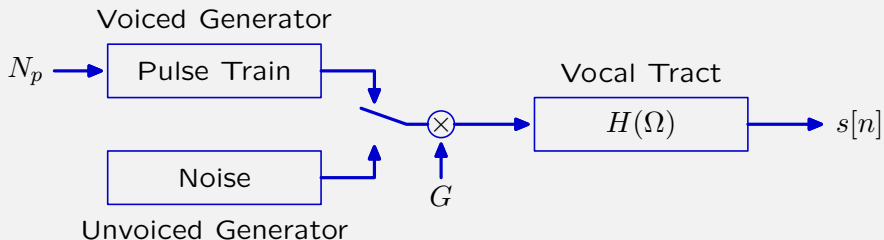


What part of the model was changed?

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

Model of Running Speech

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

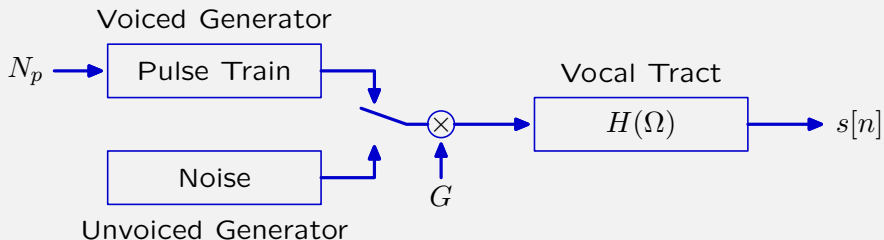


What part of the model was changed?

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

Source/Filter Model

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

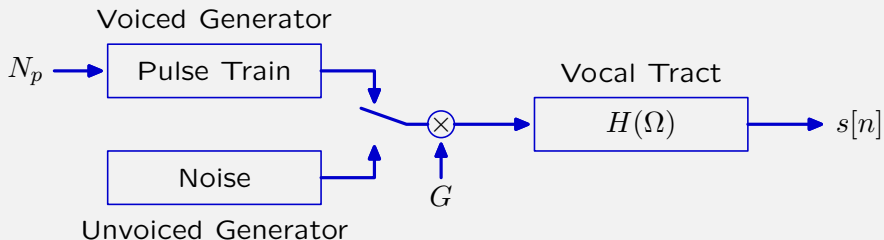


What part of the model was changed?

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

Model of Running Speech

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

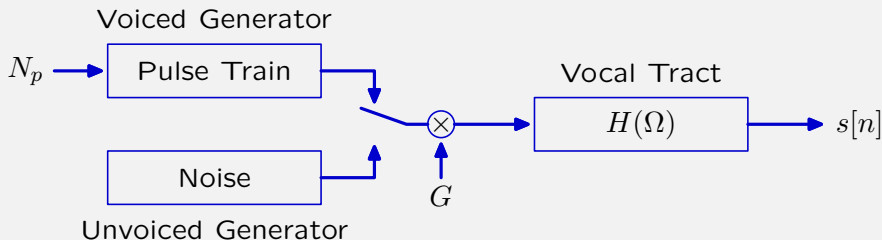


What part of the model was changed?

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

Source/Filter Model

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

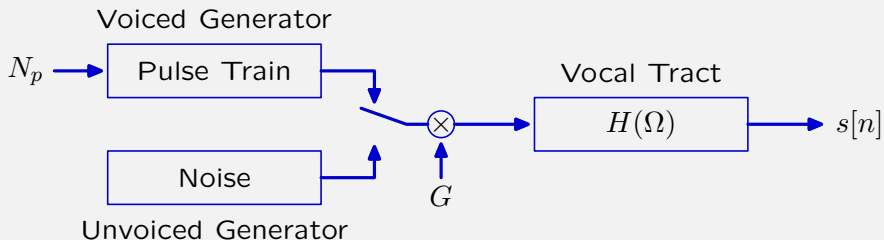


What part of the model was changed?

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

Model of Running Speech

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

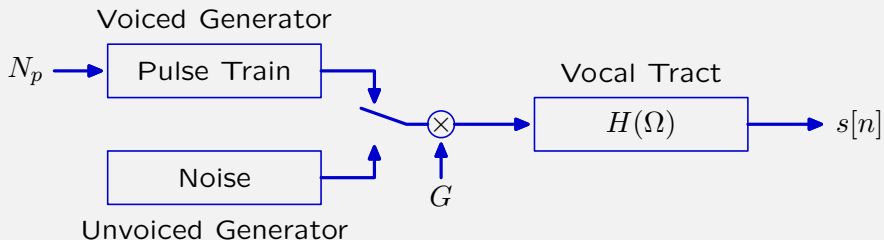


What part of the model was changed?

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

Source/Filter Model

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

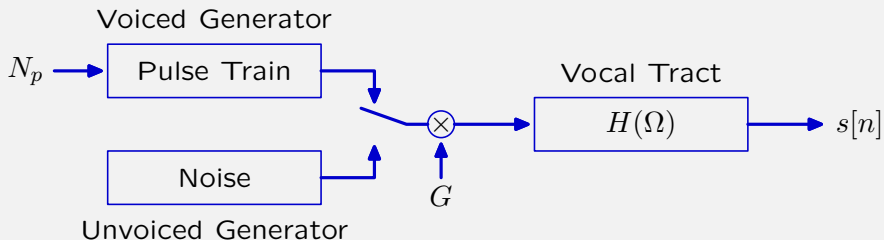


What part of the model was changed?

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

Model of Running Speech

"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

Question of the Day

The “filter” in the source filter model of speech production can be described by F1, F2, and F3.

Part 1. Describe what these numbers mean.

Part 2. Are these numbers important in whispered speech?

Please enter your responses at the following url:



<http://bit.ly/4qehFmF>