An English speaker only knows a finite vocabulary, yet can form an infinite number of unique sentences. What is the basic building block of language, and how do these building blocks fit together to create words and sentences? How can we use this information to create a program capable of speech synthesis?

In this session, we will go over some basic linguistic concepts and learn how speech is produced by both humans and computers. For our activity, we will try to guess all the sounds in the English language, make recordings of these sounds, and use these recordings to create a voice for a speech synthesis program. Once this is done we will be able to type any English sentence we want into the computer and the computer will vocalize it. Then, we will try to "trick" the program by finding sentences that don't sound quite right, and use this information to create more recordings and improve upon our system. During our wrap-up, we will reflect on the difficulties of speech synthesis and what we could add to make an even better speech synthesizer.
Spoken Language

● How many unique sentences in English?
  ○ Infinite!
  ○ Example: add “said John”

● How many unique words in English?
  ○ Google says: 1,025,109.8
  ○ Always new words
Spoken Language

- Smallest part of a word?
  - Phonemes!

- How many phonemes in English?
  - ~44 unique phonemes
  - Consonant sounds
  - Vowel sounds
English Phonemes

- Consonant sounds
  - What are all the consonant sounds in English?

- Vowel sounds
  - What are all the vowel sounds in English?
How can computers speak?

● Given a string such as “Hello World”, must turn it into speech
  ○ Ability to produce all the sounds (phonemes)
  ○ Able to determine how letters correspond to phonemes
How can computers speak?

- Electrical currents (1939)
  - [https://www.youtube.com/watch?v=0rAyrmr7vv0](https://www.youtube.com/watch?v=0rAyrmr7vv0)

- Computer-generated (1961)
  - [https://www.youtube.com/watch?v=ebK4wX76RZ4](https://www.youtube.com/watch?v=ebK4wX76RZ4)

- Computer text-to-speech, Siri (2011)
  - [https://www.youtube.com/watch?v=c2tNN-9VLeQ](https://www.youtube.com/watch?v=c2tNN-9VLeQ)
C A T → /kæt/ →
Activity

Talking Computers
Getting set up

- open Terminal (use spotlight search)
- type in “setup”
Let’s record some sentences!

What sounds should appear in our sentences?

To build the voice:
ctrl-D to close festival
Type in “make-voice”
Make it talk!

ctrl D to close festival, make-voice to update voice
- Type in “run”
- Your cursor should be next to “festival>”
- Type “(voice_cmu_us_tn_cg)” to specify our voice
- Put on your headphones
- (SayText “sentence you want to say”)

What sentences sound good? Are there sounds that don’t work?
Wrap-Up
Reflections

How did our speech synthesizer work?
Consonant sounds

Which ones did we get?

- p, b, t, d, k, g (stops)
- f, v, s, z, th, dh, sh, zh (fricatives)
- j, ch (affricates)
- w, y, r (approximants)
- l (lateral)
Vowel sounds

Which ones did we get?

- bead
- bid
- bayed
- bed
- bad
- bard
- body
- bawd
- buddhist
- bode
- booed
- bud
- bird
- bide
- bowed
- Boyd
- beer
- bare
- byre
- boor
Why is speech synthesis so hard?

- Phonemes
- Transitions
- Intonation
- Context
The voice behind Siri

- [https://www.youtube.com/watch?v=z2bTymnb1uE](https://www.youtube.com/watch?v=z2bTymnb1uE)
  - 4 hours a day for a month
  - Lots of random sentences