



SpeechDemo Activity Guide

Version of October 27, 2022

David S. Touretzky
Carnegie Mellon University

This work was funded by a grant from NEOM Company.



NEOM



This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>.

Speech Recognition Demo

Summary



Using your laptop's microphone, experiment with the Google Speech Recognition service to see how it generates alternative transcriptions of a sentence and applies multiple types of knowledge to deal with homophones and other types of ambiguity.

Grade Bands: 3-5, 6-8, 9-12

Materials/Resources Required

- 1.) Chrome or Chromium web browser (won't work in other browsers)
- 2.) Access to <https://www.cs.cmu.edu/~dst/SpeechDemo>

Vocabulary Terms

Flashcards for these vocabulary terms are located at the end of this guide.

- Ambiguity
- Homophone
- Phoneme
- Morpheme
- Syntax
- Semantics

Related Activities

- 1.) Waveforms and Spectrograms Activity
- 2.) Machine Translation Activity

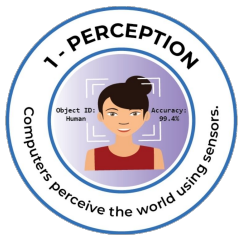
Learning Objectives

Students will be able to explain how the later stages of speech recognition work by:

- Explaining how ambiguity in the speech signal leads to multiple possible transcriptions
- Demonstrating how context is used to resolve ambiguity

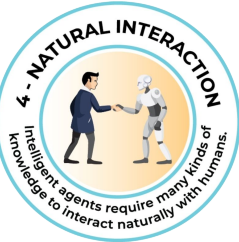
Guidelines and Standards

AI4K12 Guidelines



Big Idea 1: Perception

1-B-iii 3-5, 6-8, 9-12: Abstraction pipeline: language



Big Idea 4: Natural Interaction

4-A-i.K-2: Nature of language

4-A-ii.K-2: Ambiguity of language

4-A-iii.3-5: Reasoning about text

ELA Standards

Phonological Awareness:

CCSS.ELA-LITERACY.RF.K.2 - Demonstrate understanding of spoken words, syllables, and sounds (phonemes).

CCSS.ELA-LITERACY.RF.K.2.D

Isolate and pronounce the initial, medial vowel, and final sounds (phonemes) in three-phoneme (consonant-vowel-consonant, or CVC) words.¹ (This does not include CVCs ending with /l/, /r/, or /x/.)

CCSS.ELA-LITERACY.RF.K.2.E

Add or substitute individual sounds (phonemes) in simple, one-syllable words to make new words.

CCSS.ELA-LITERACY.RF.K.3.D

Distinguish between similarly spelled words by identifying the sounds of the letters that differ.

Preparation

Make sure the computers can access the SpeechDemo web page. Give the demo access to the microphone.

Operating the Demo

Try This! [Speech Demo](#)

1. Speak into the microphone and your voice will be transcribed to text
2. If there are several plausible transcriptions, all leading candidates will be displayed.
3. Click on the “*pause*” button to temporarily disable the microphone.
4. Click on “*resume*” to resume recording.
5. Click on the “*read back*” button to turn on computer readback of the recognized text.
6. Click on the pulldown menu to switch to a different language model (e.g., Spanish, Chinese, etc.)

Demo Investigations

1. Importance of context: longer sentences work better.

Accuracy for single words may not be that high, but if the words are strung together in a sentence, contextual information can be used to make better guesses about each word.

Try the following short words or phrases:

- “brat”
- “scat”
- “sat scat spat”
- “fat flat flap flag flab”

Now try these full sentences:

- “My little brother is a brat.”
- “I like scat singing.”
- “Lie flat and let the flag flap in the breeze.”

2. Importance of grammar: grammatical sentences work better than random words.

Try the following:

- “There is a sandwich heating up in the microwave.”
- “heating the in a up sandwich is microwave there”

- “I wish they made movies about flying goats.”
- “wish I movies made goats they about flying”

3. Careful enunciation helps.

Compare the following:

- “heating the in a up sandwich”
- “heating. the. in. a. up. Sandwich.” (pause between each word)

Demo Investigations Continued

4. Homophones can often be *disambiguated* by context.

Try these examples:

- “They’re claiming their dog is there.”
- “Which witch is which?”
- “If I teach from 12 to 2 can I eat lunch too?”
- “He led the study of lead poisoning.”

Many more examples at yourdictionary.com.

Can you get Google to recognize both of these sentences? They sound the same using normal enunciation.

- “How to recognize speech.”
- “How to wreck a nice beach.”

5. Common mispronunciations (or dialectal variants) may be corrected automatically to Standard American English.

Try these examples:

- “Let me ax you a question.”
- “I love fruit, expecially bananas.”
- “My hamster exscaped from his cage and is hiding under the bed.”

6. Non-words can sometimes be “corrected” based on context.

Try these examples where we replace “grapes” or “drapes” with the non-word “brapes”:

- “Brapes”
- “Start your fruit salad with brapes.”
- “Open the brapes to brighten the room.”

Here we replace “forks” or “Forbes” with the non-word “forps”.

- “Forps”
- “knives and forps”
- “forps magazine is edited by Steve Forps”

More examples of strong contextual bias:

- “He’s from Transylvoonia.”
- “Put the milk in the refroberator.”

In the absence of a strong contextual bias, non-words simply aren’t recognized. People can hear and perfectly repeat non-words or unfamiliar words, but Google’s speech recognition engine is not designed to do that.

- “You are a pain in the gabooch.”
- “We must visit Bwaduku.”
- “The fleeshmoddy is broken.”

Demo Investigations Continued

7. Common sayings are recognized as a whole.

This example is a famous palindrome:

- “Able was I ere I saw Elba.”

“Ere”, which means “before”, is archaic and not commonly used today, but it is recognized in this special context.

This next example also uses “ere”, but it’s not a famous saying. How well is “ere” recognized?

- “Able were you ere you saw Elba.”

Here is another famous quotation, from *The Hobbit*, that uses “ere.” Does Google get it? What if we change a few words?

- “We must away ere break of day.”
- “You must depart ere noon today.”

It’s not surprising that Google search knows famous phrases, and quotes from famous literature. But were you surprised to learn that Google’s speech recognition engine also makes use of this knowledge?

This famous quote is from John Donne (*Meditation XVII*). What if we change the last word?

- “No man is an island.”
- “No man is an eyelid.”
- “Now blink your eyelid.”

Google has no trouble recognizing “eyelid”, except when it occurs in a context where it strongly expects to hear “island”.

8. Language models make a big difference.

A language model determines not only the vocabulary that can be recognized, but also the set of phonemes (sounds) the recognizer will consider. Changing language models can have a dramatic effect on how English is “heard”.

Turn on the “read back” feature and try this sentence using the default “English (US)” language model:

- “I’m from Kalamazoo.”

Now change the language mode to “Spanish (Spain)” and try it again, using normal enunciation.

“I’m from Kalamazoo.” [Spanish (Spain)]

Now change the language once more, to “Chinese (Mandarin)”. Speak the sentence using normal enunciation. If you speak using very careful enunciation you can force the system to switch to an English language model and give you a better result

- “I’m from Kalamazoo.” [Chinese (Mandarin)]

Assessments

To be determined.

Other resources

Google's own speech demo page: <https://www.google.com/intl/en/chrome/demos/speech.html>

Copyright © 2022 AI4K12.org. Released under the [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/)

. This work was funded by a grant from NEOM Company.

Visual design by Pam Amendola.



Vocabulary Flash Cards



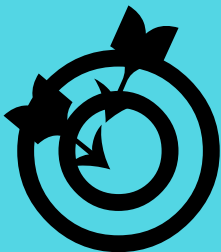
AMBIGUITY

The quality of being open to more than one interpretation; inexactness.



HOMOPHONE

Each of two or more words having the same pronunciation, but different meanings, origins, or spelling. For example: new and knew



MORPHEME

The smallest meaningful part of a word, e.g. “unbeatable” is composed of the morphemes “un,” “beat,” and “able.”

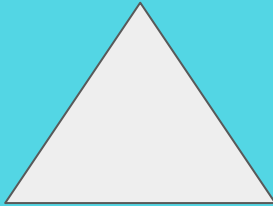


Vocabulary Flash Cards



PHONEME

Any of the perceptually distinct units of sound in a specified language that distinguish one word from another.



SEMANTICS

The meaning of a word, phrase, sentence or text



SYNTAX

The arrangement of words and phrases to create well-formed sentences in language.