

디지털시대의 음성언어자료 이해 및 활용

Leveraging the Potential of Spoken Data in the AI-Powered Digital Age

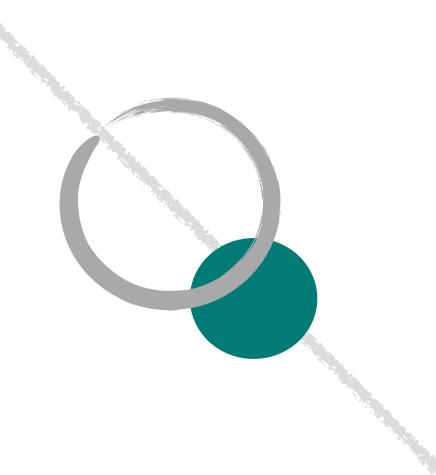
Miran Kim (mirankim@gnu.ac.kr)

Dept. of English Education, Gyeongsang National University



• 주최: 연세대학교 언어정보연구원

139회 학술발표회 2023. 1. 26 (목) 14:00 ~



- **Part 1: Understanding spoken language as digital data**

- Digital data
- Written vs. Spoken language
- Sound as physical entity
- Studies of sound
- Speech corpus

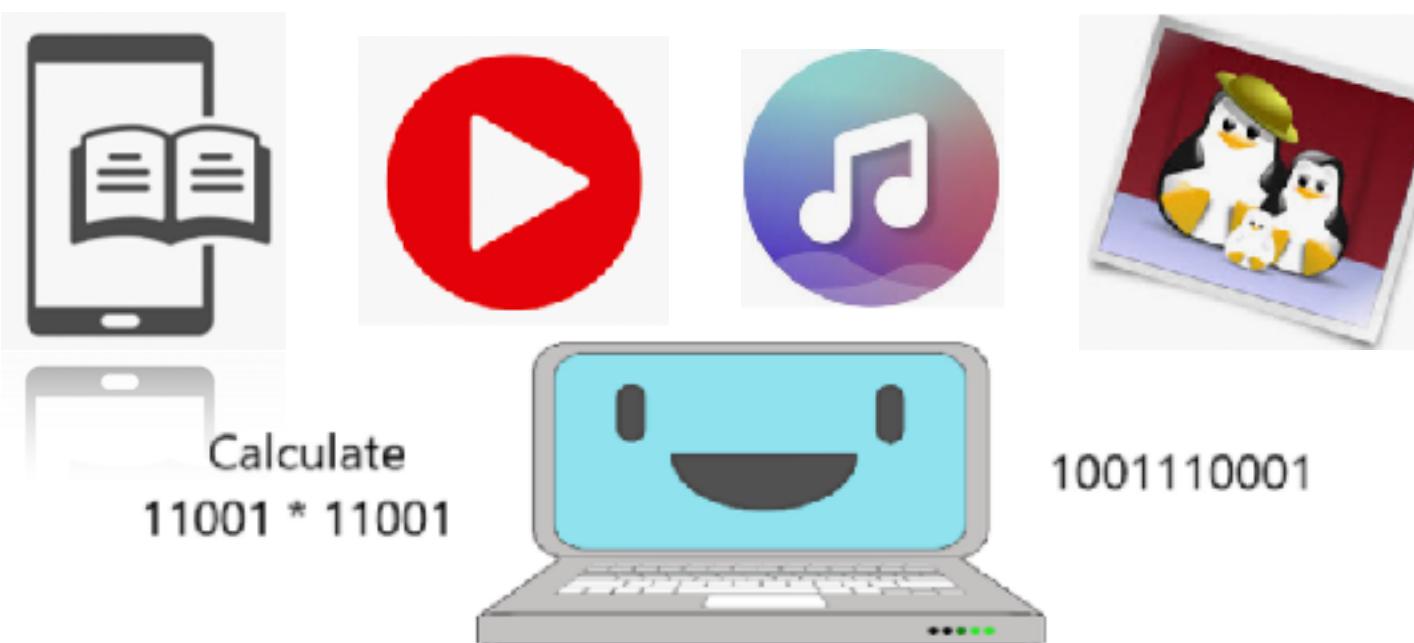
- **Part 2: Application of speech data**

- Text-to-Speech (TTS)
- Speech-to-Text (STT)
- Automated Speech Recognition (ASR)



1. Digital data

Video Tapes



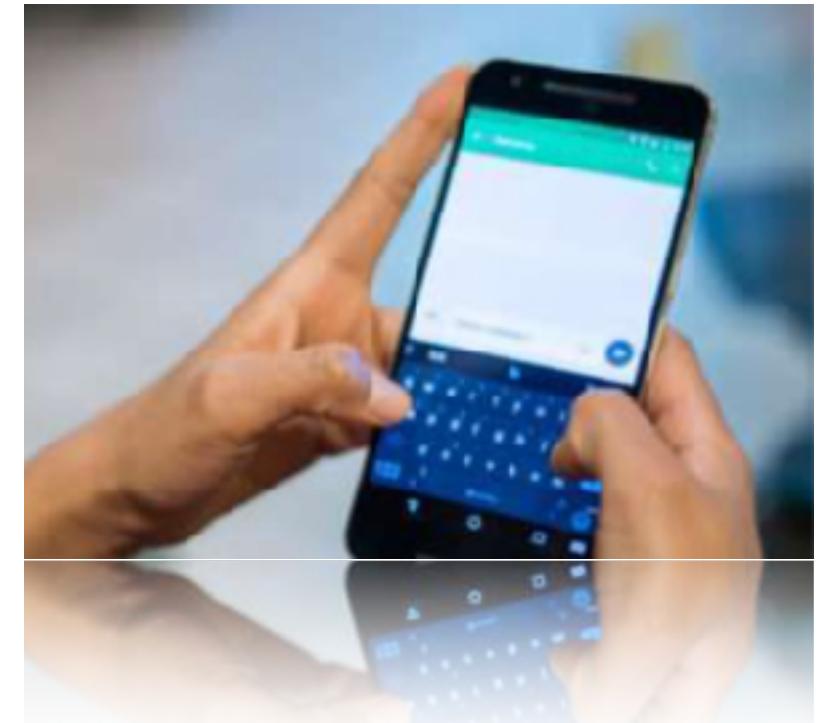
Binary representation of 25 is 11001



Written language in the digital age

- The use of written language has increased significantly:

e.g, instant messaging, email, text messages, and social media posts.



- More efficient and **effective communication** across long distances and time zones.
- New forms of written language: e.g., emoticons, acronyms and slang

DM

RT

별다줄

Spoken vs. Written language



Differences?

Transient	Permanent
Heavily rely on nonverbal cues	Words themselves
More informal / colloquial	More formal / polished
Cannot be easily edited / corrected	Edited / reviewed
In real-time communication	Delayed or asynchronous communication

e.g., Sumerian in ancient Mesopotamia around the 4th millennium BCE

Recording sound

- In the 4th century BCE, the Greek philosopher Aristotle wrote about **echoes** in his work "On the Heavens".



1 1859

Frenchman Édouard-Léon Scott de Martinville invents the first device capable of recording sound. Known as a **phonograph**, the machine captures sound by tracing the vibrations of a bristle onto a sheet of soot-covered paper.

These recordings, called **phonautograms**, were never intended to be played back – Martinville's purpose in making the device was to simply see what sound waves looked like.



- The earliest known recording was made by Scott on April 9, 1860, it was a **recording of himself singing "Au clair de la lune"** ("Under the Moon")

3 1938

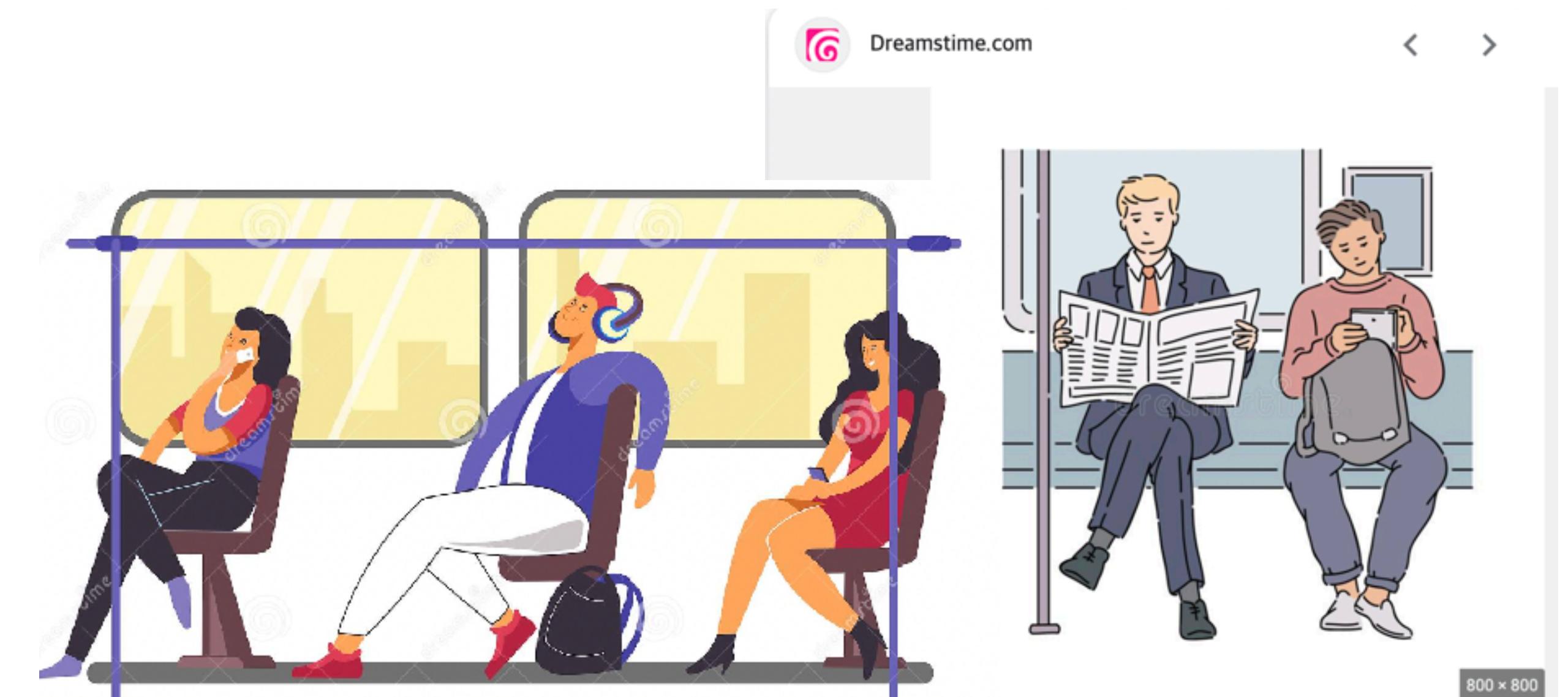
Alec Reeves, PCM and the Birth of Digital Communication – historictech

Edison Invents the Phonograph
Thomas Edison created many inventions, but his favorite was the phonograph. While working on improvements to the telegraph and the telephone, Edison figured out a way to record sound on tinfoil-coated cylinders. In 1877, he created a machine with two needles: one for recording and one for playback. When Edison spoke into the mouthpiece, the sound vibrations of his voice would be indented onto the cylinder by the recording needle. What do you think were the first words that Edison spoke into the phonograph?

Spoken language in the digital age

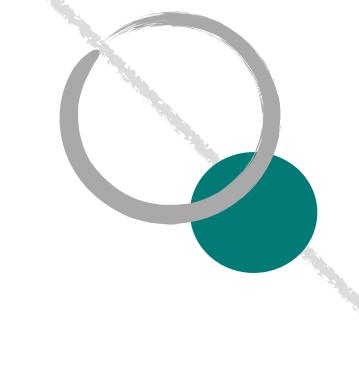
- Spoken language become increasingly important with the rise of **voice enabled technology**:

e.g, interaction with devices,
importance of audio contents
real-time interactions



Two Men in Subway Reading Newspaper and Using Tablet
Stock Vector - Illustration of device, city: 147073734
images from dreamtime.com

Spoken language characteristics in various contexts

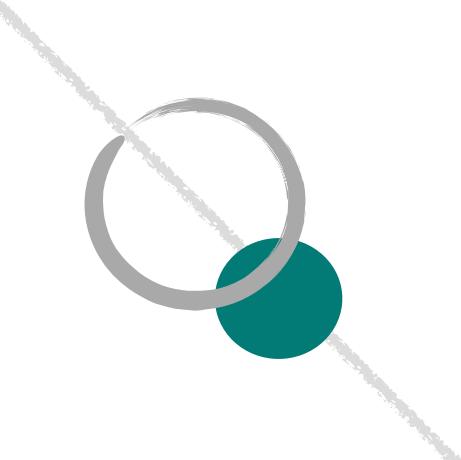


 The Washington Post < >



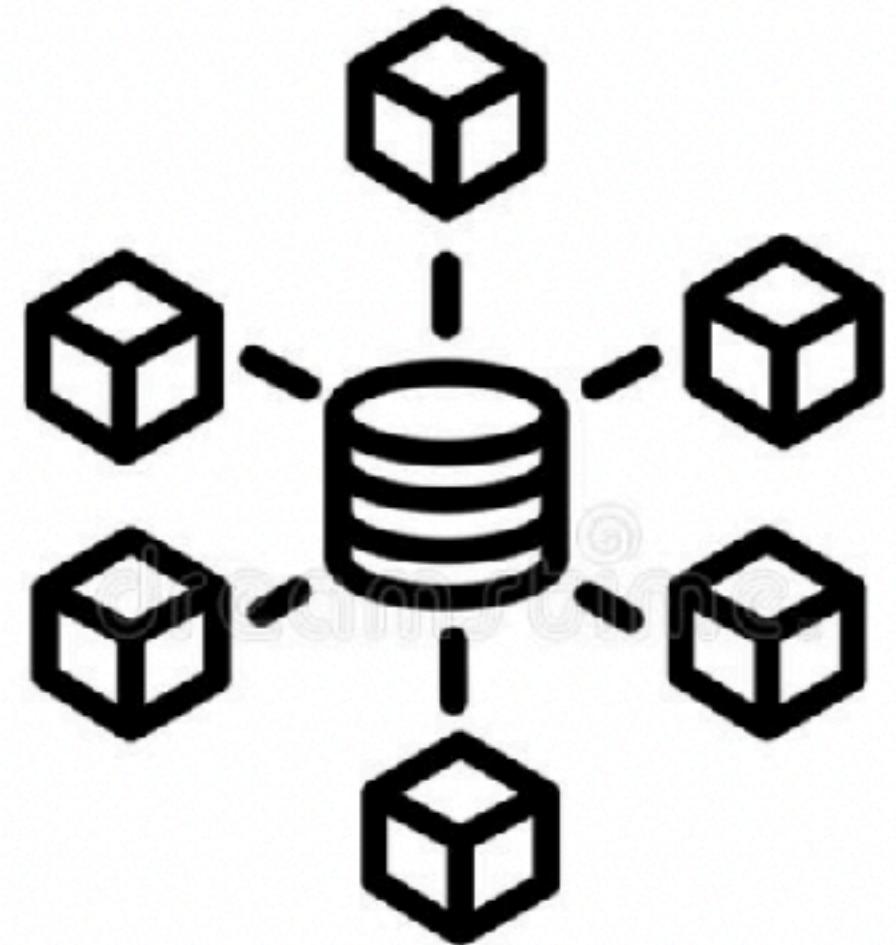
Children can be swayed by robot peer pressure, study says - The...

- **Pitch:** typically higher pitched.
- **Speech rate:** tend to speak at a faster rate.
- **Articulation:** have difficulty producing some speech sounds correctly.
- **Vocabulary:** typically have a smaller vocabulary.
- **Grammar:** may have difficulty with grammar, resulting in incorrect word order, missing word forms, and overgeneralization of grammar rules.
- **Discourse:** tend to be less complex and less cohesive.
- **Prosody:** may use less intonation, stress, and rhythm, which can make their speech sound less expressive and less natural.



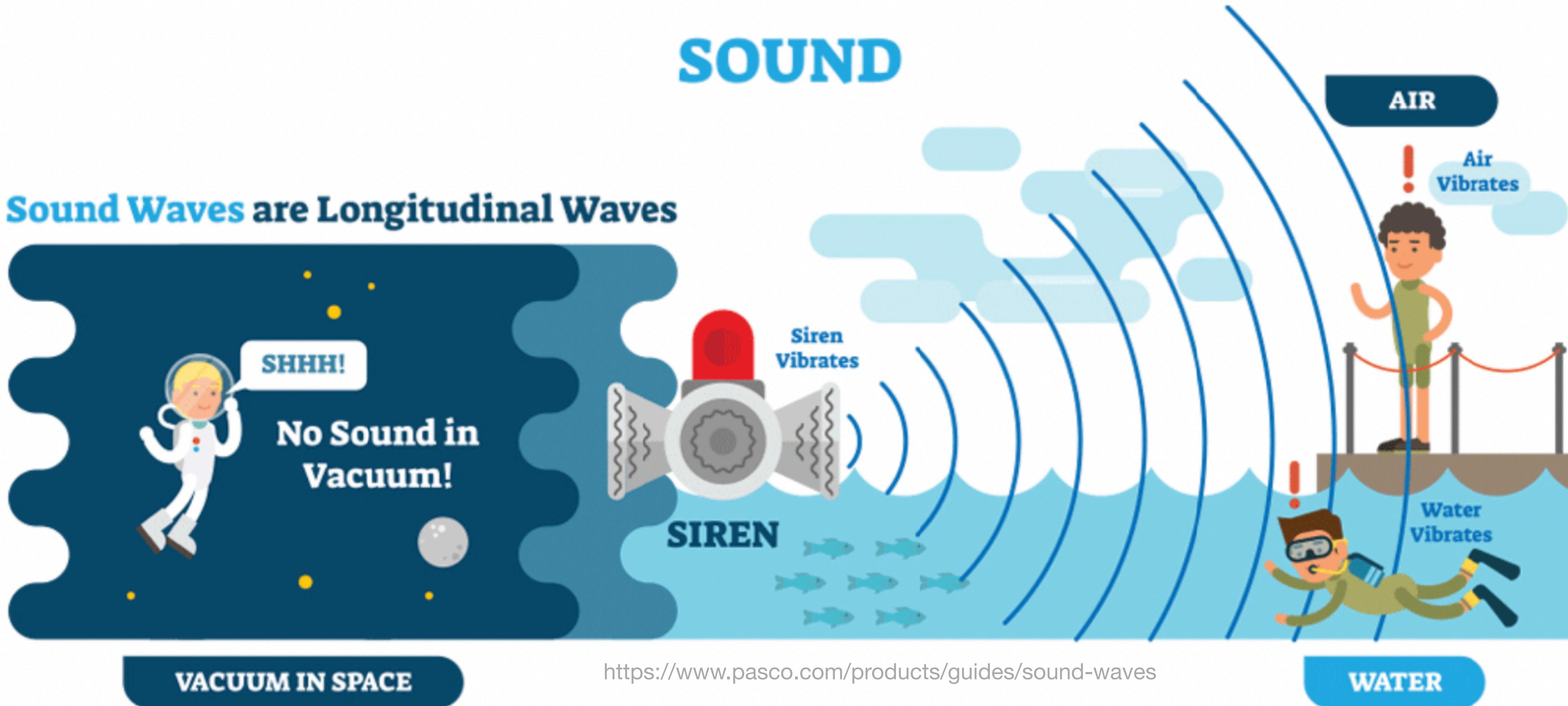
Language & Information

(written, spoken)



big data

3. Sound as physical entity

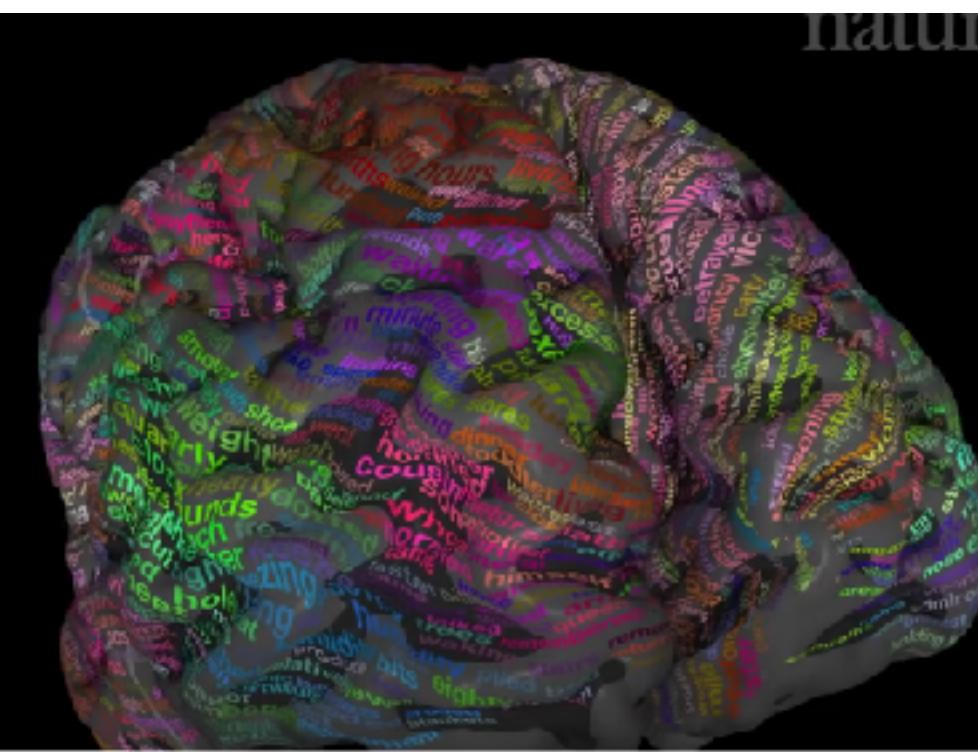


4. Studies of sound



Phonetics

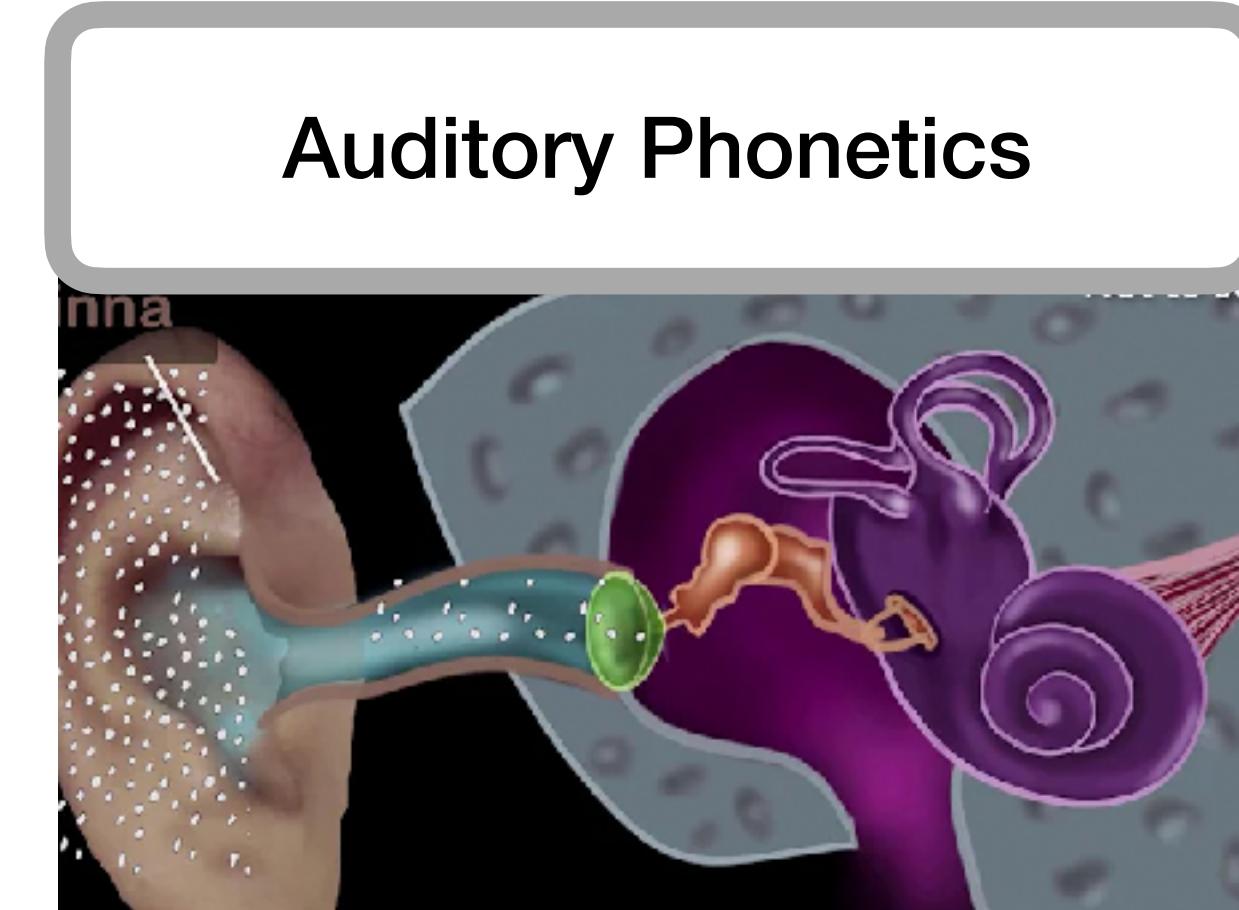
Phonology



Articulatory Phonetics

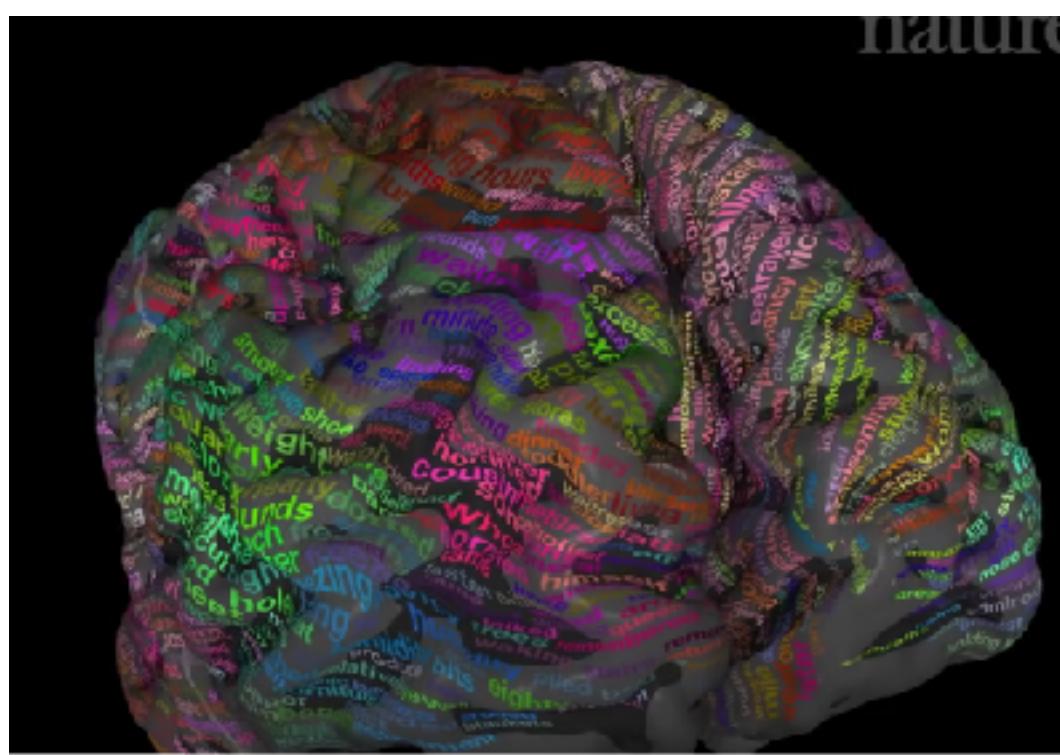
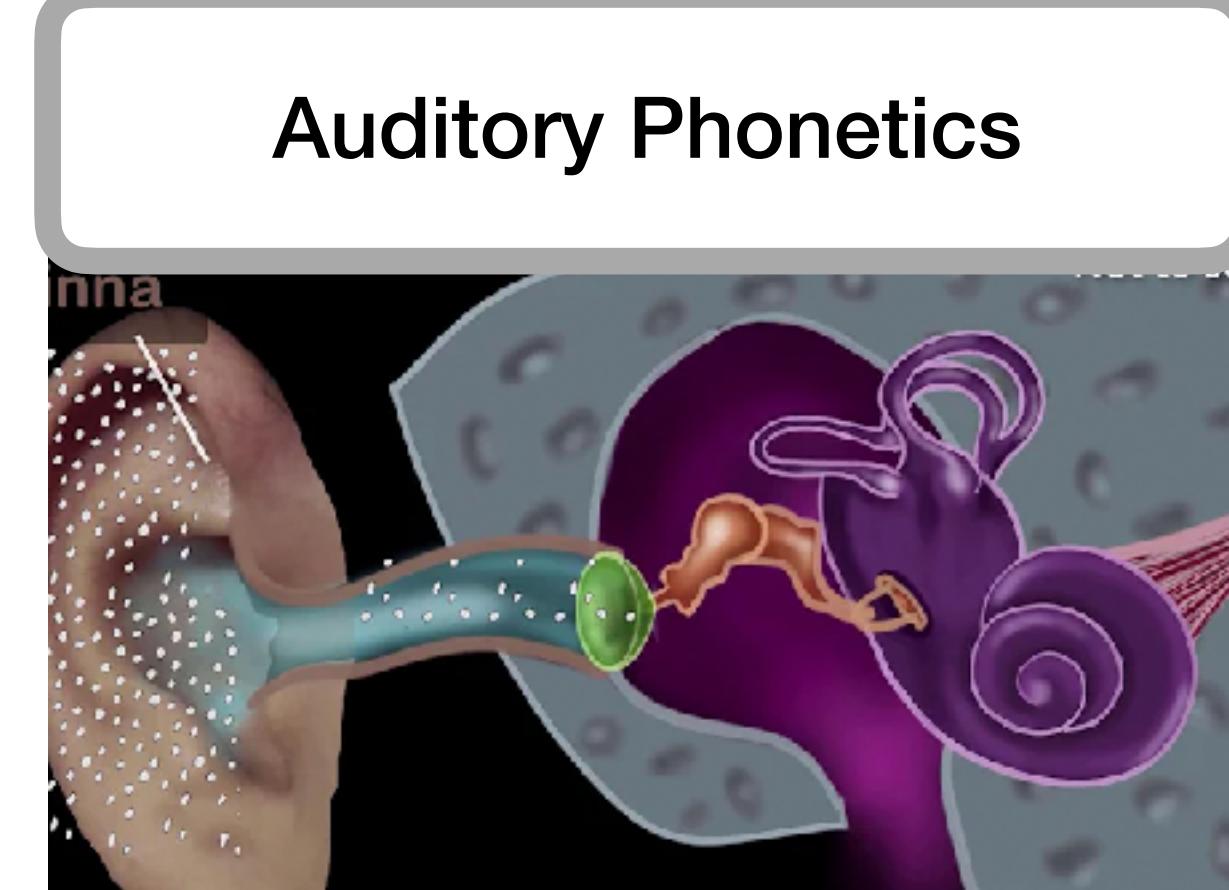
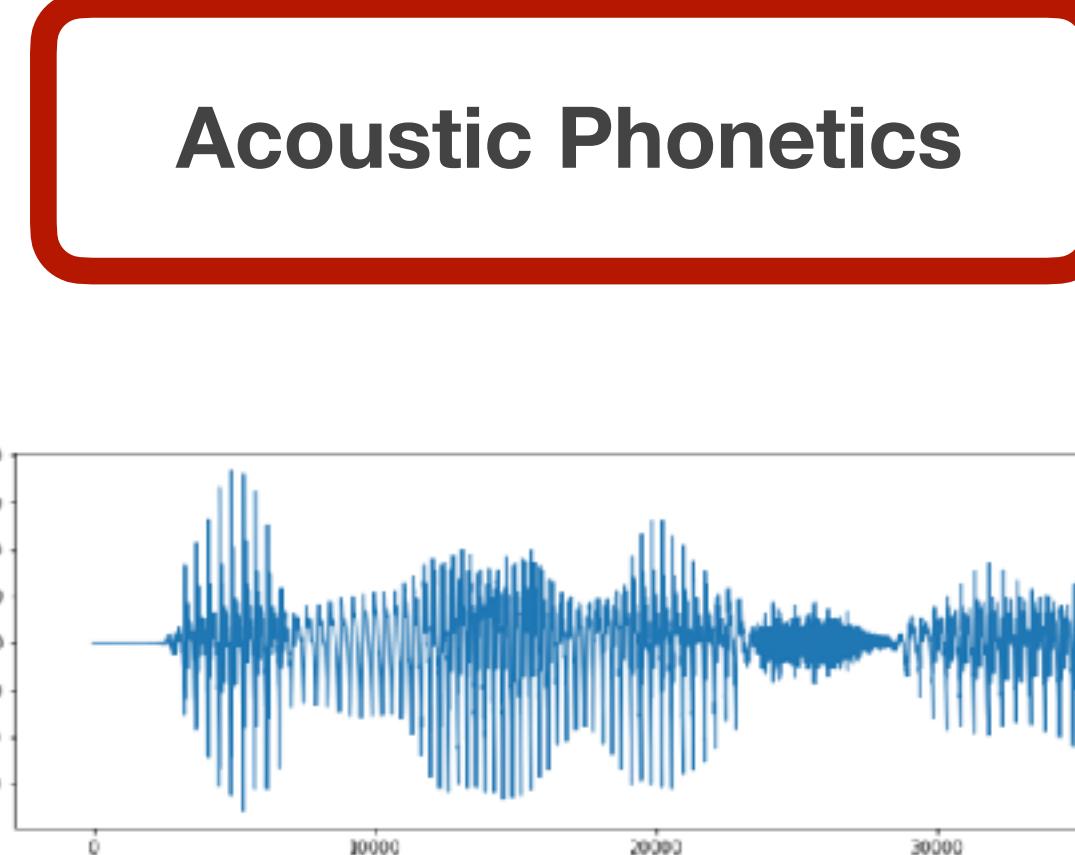
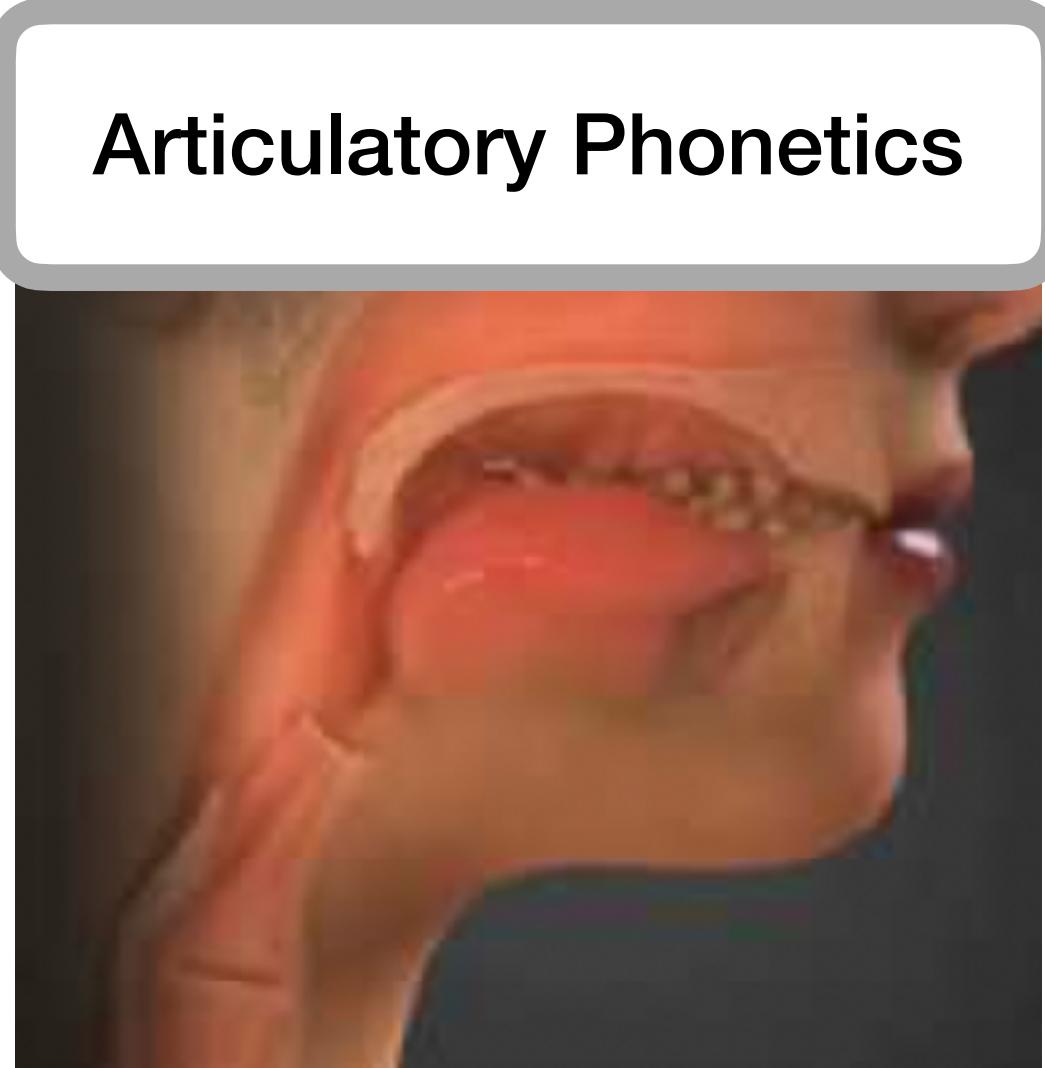
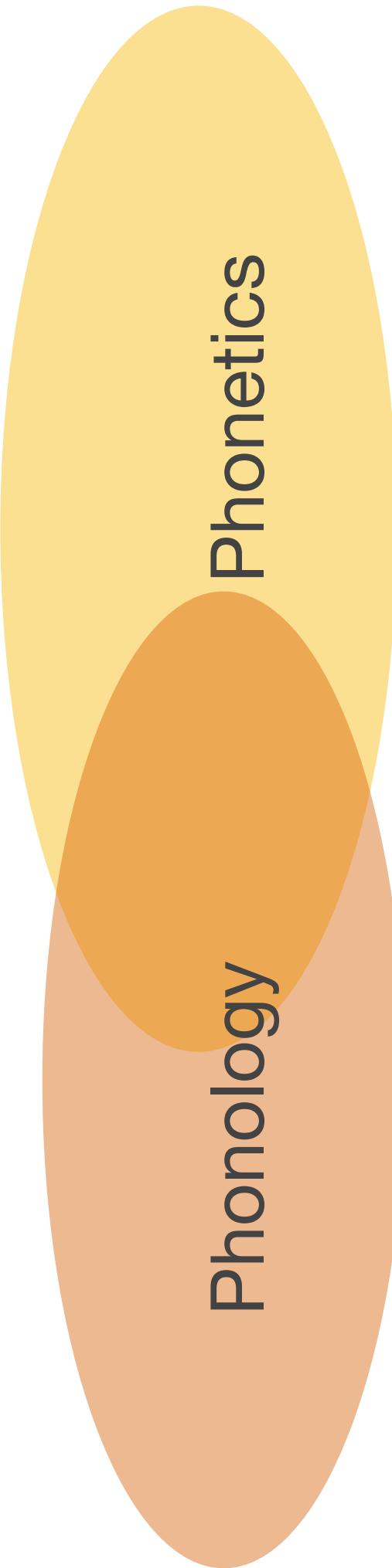


Acoustic Phonetics



Auditory Phonetics

4. Studies of sound

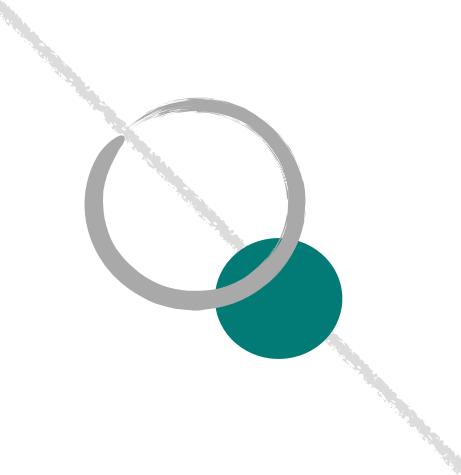


Speech (audio signal) processing

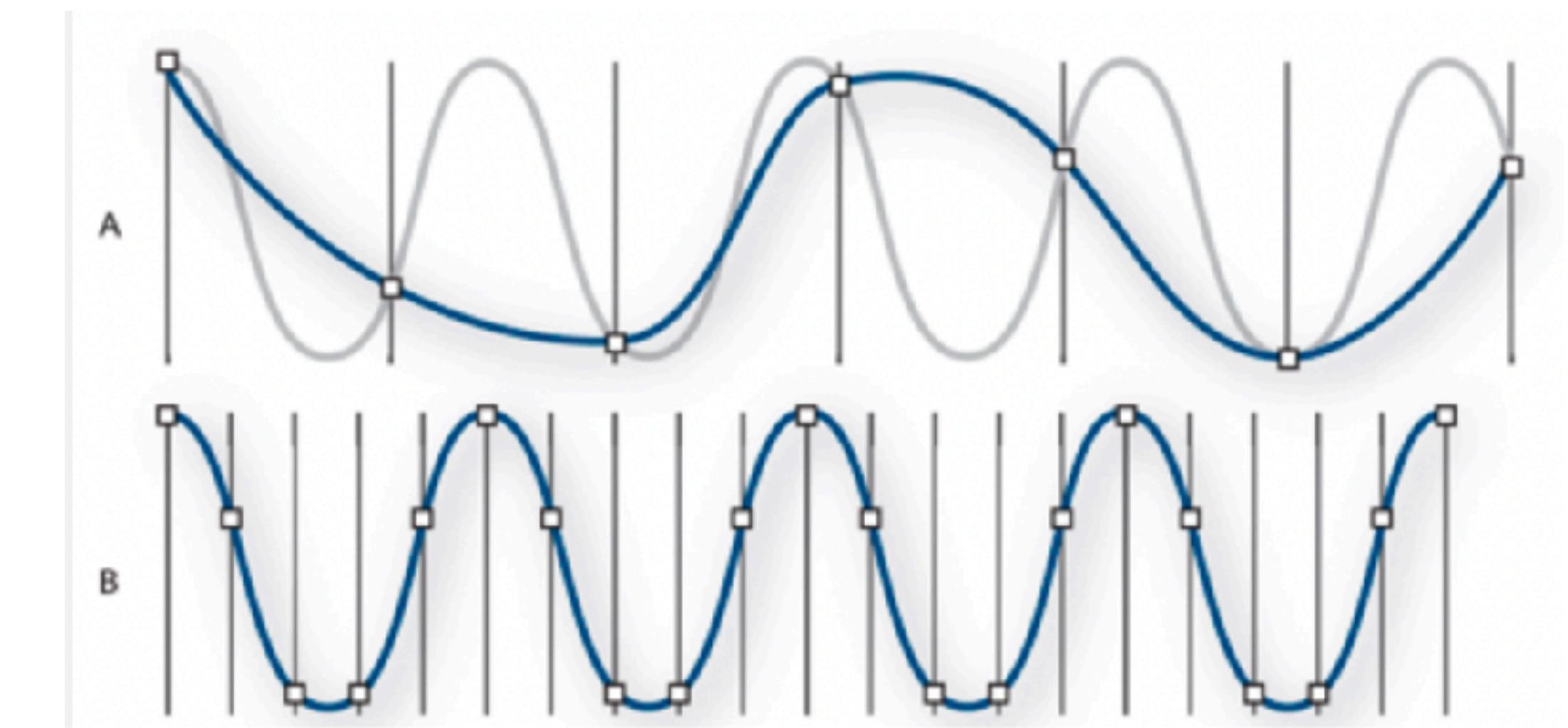
Speech sciences

Acoustic phonetics

Physical entity of sounds



Digitize speech signal



5. SEOUL Corpus: The Korean Corpus of Spontaneous Speech

- 한국어 자연발화 음성 코퍼스 (Yun, et al. 2015)

- Interview speech recordings of 40 Seoul dialect speakersK (20F + 20M; 10s, 20s, 30s, 40s)
- Total 240 sound files + 240 label files: e.g., 1 hr per speaker (6 speech files)

The screenshot shows the OpenSLR website interface. At the top, there's a navigation bar with 'OpenSLR' and 'Open Speech and Language Resources'. Below that, a sub-navigation bar has 'Home' and 'Resources' links. The main content area is titled 'SEOUL CORPUS' and includes the URL <http://www.openslr.org/113/>. It provides detailed metadata: Identifier: SLR113, Summary: The Korean Corpus of Spontaneous Speech (aka, Seoul Corpus), created from the NRF(Korea)-funded project, Category: Speech, License: Attribution-NonCommercial 2.0 Generic (CC BY-NC 2.0). It also lists download links for 'readme.tgz' (27M), 'label.tgz' (57M), and 'sound.tgz' (2.5G), along with mirrors for US, EU, and CN. A section titled 'About this resource' contains a BibTeX entry for the corpus: @article{yunetal2015, title={The Korean Corpus of Spontaneous Speech}, author={Yun, W., Yeon, K., Park, S., Lee, J., Cho, S., Kang, D., Byun, K., Hahn, H., Kim, J.I.}, journal={Journal of the Korean Society of Speech Sciences}, volume={7(2)}, pages={103–109}, year={2015}, publisher={} }

- Recording period: 2012~2013
- 230,000 word tokens (51,000 types): 1.1M phoneme tokens
- Labeling (7 tiers): including phoneme, pronunciation form, syllables, words, transcription, etc.
- Used Forced alignment process
- Free distribution for research

Part 2

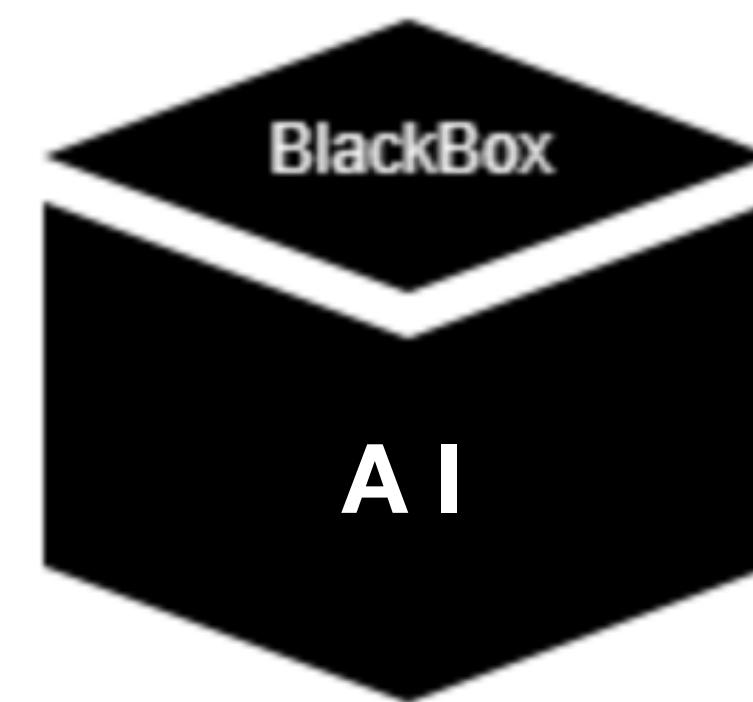
6. Information transfer in the digital age



Input



Input



Output

Output

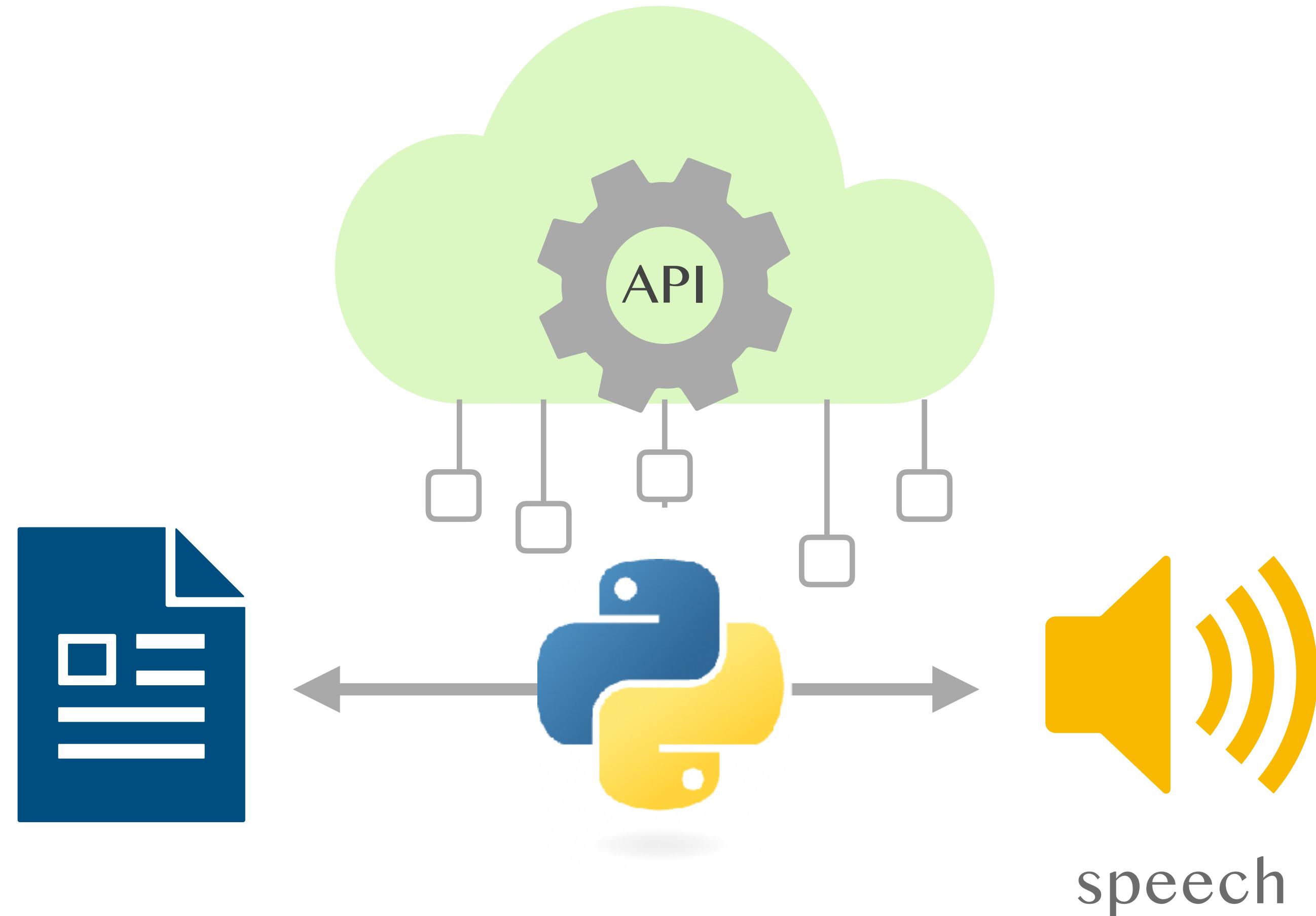
Output



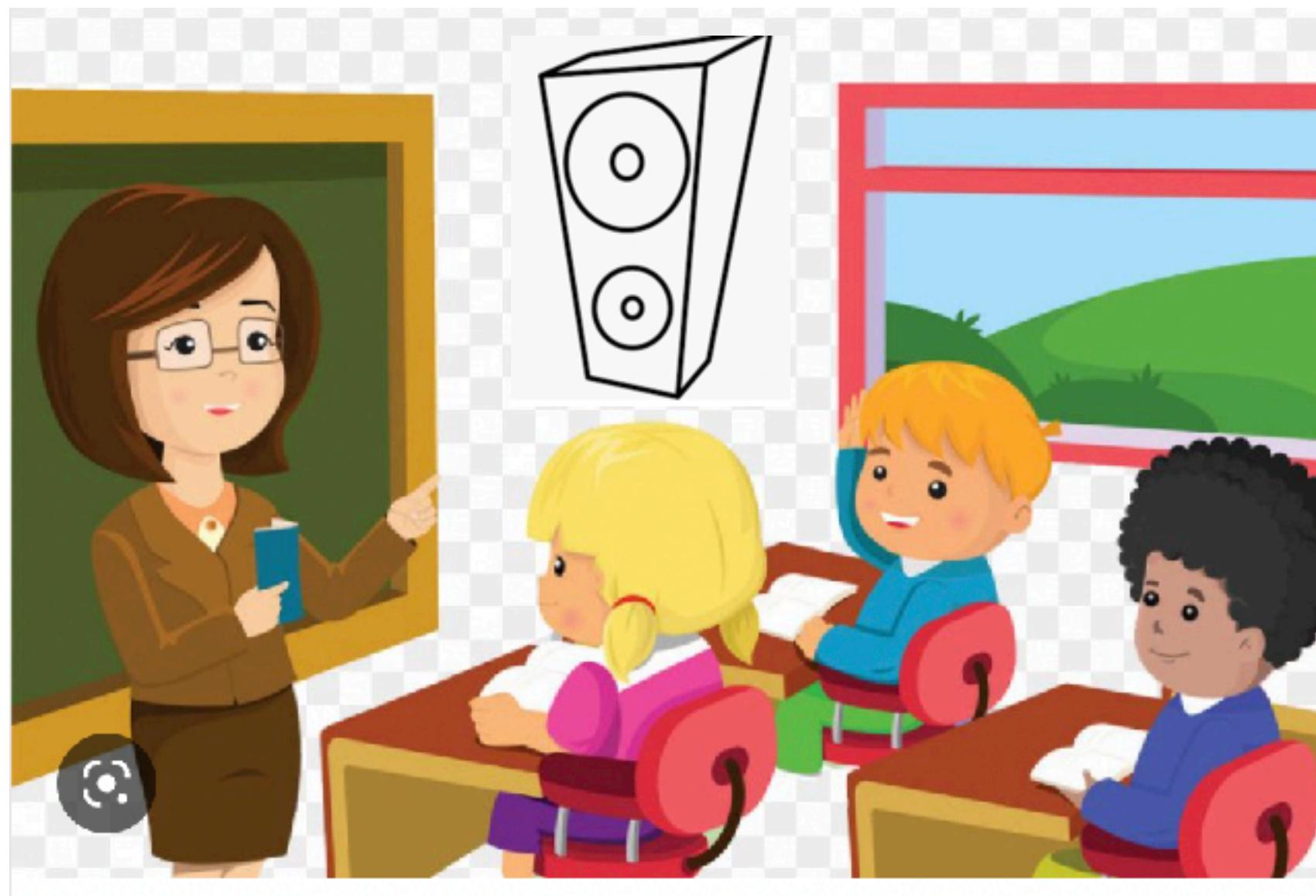
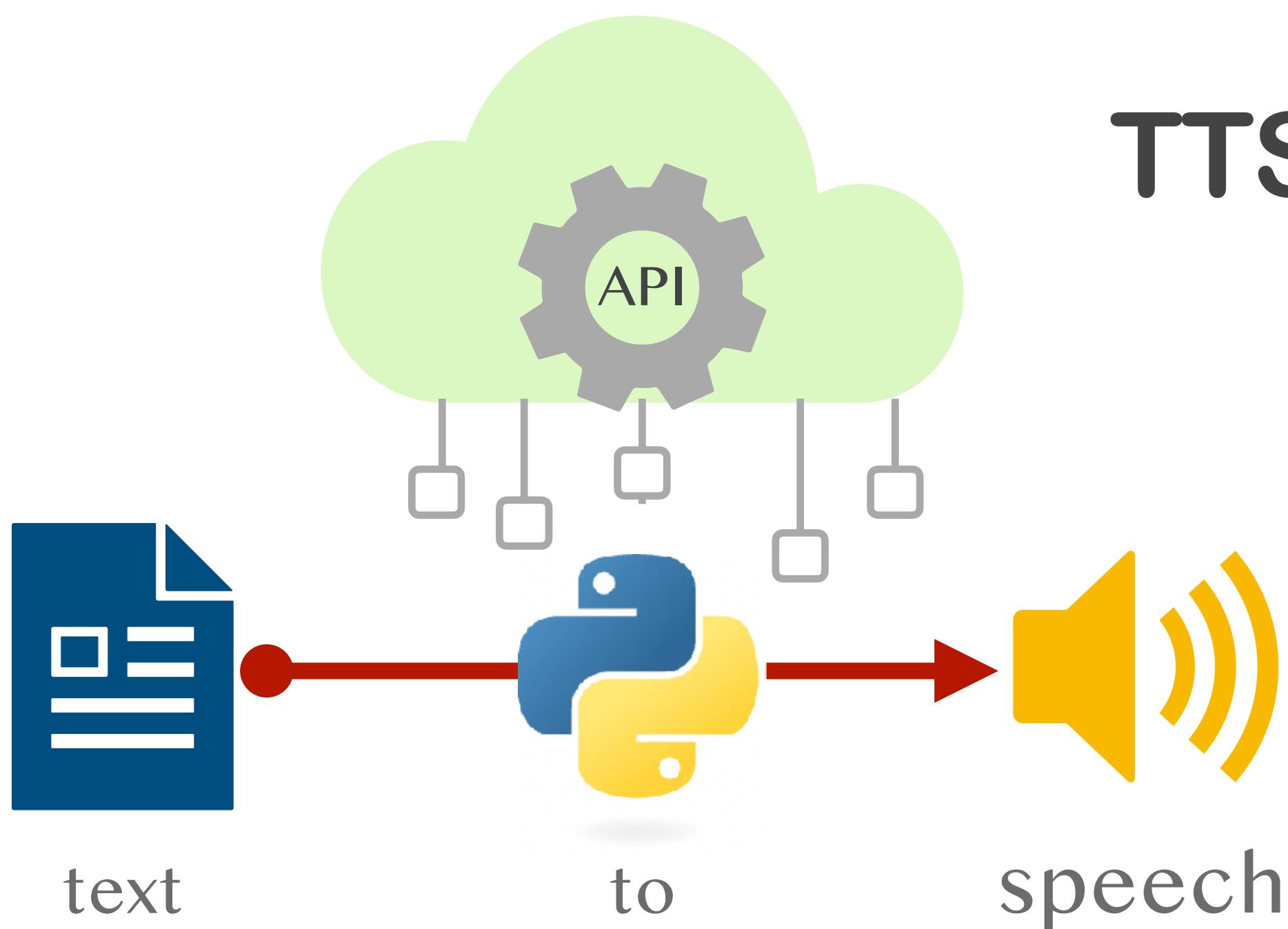
AI-Generated Art Won a Prize. Artists Aren't Happy. - The New York Times

Visit

API (Application Programming Interface)



TTS (Text-to-Speech)

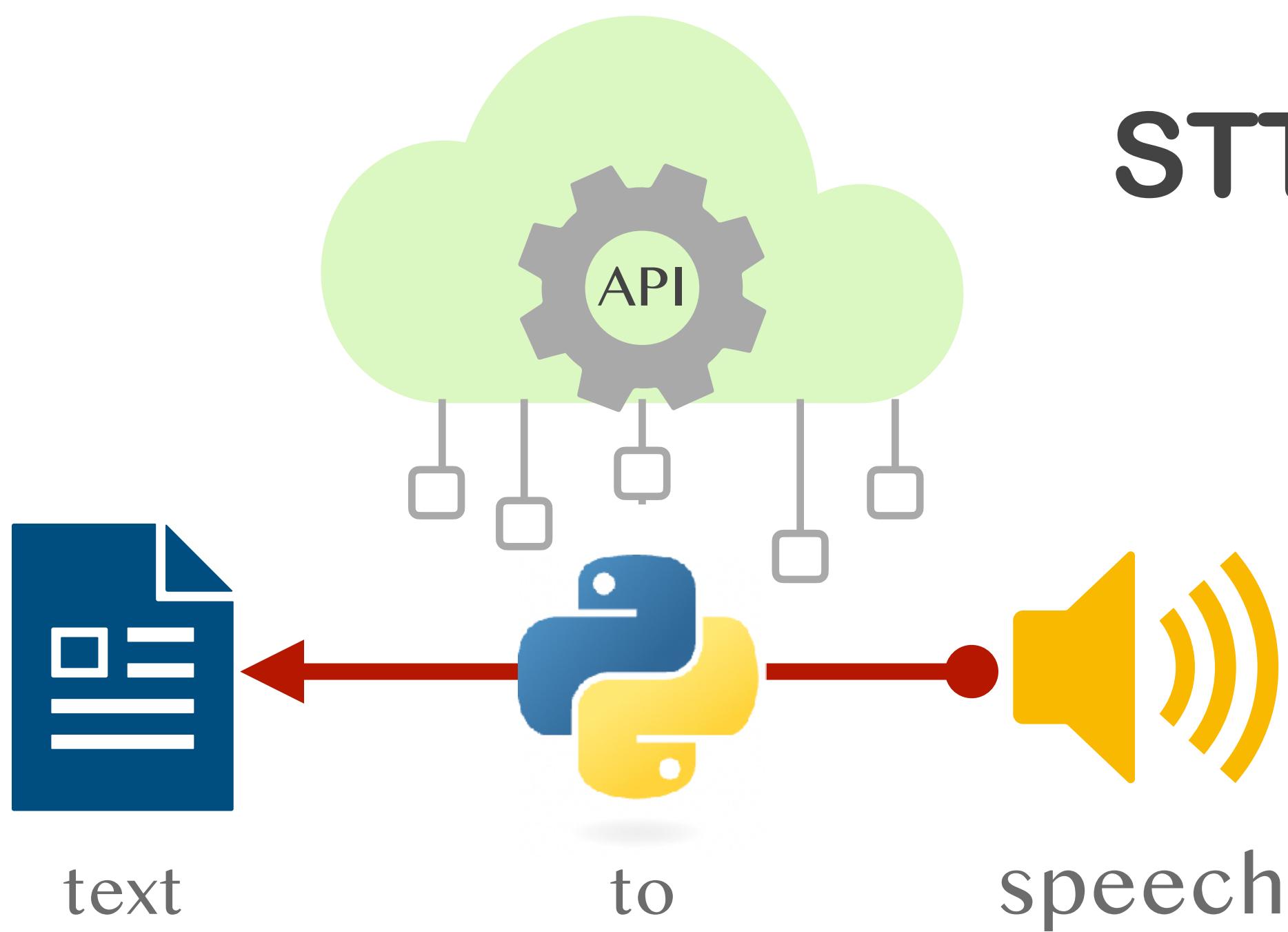


OffiDocs

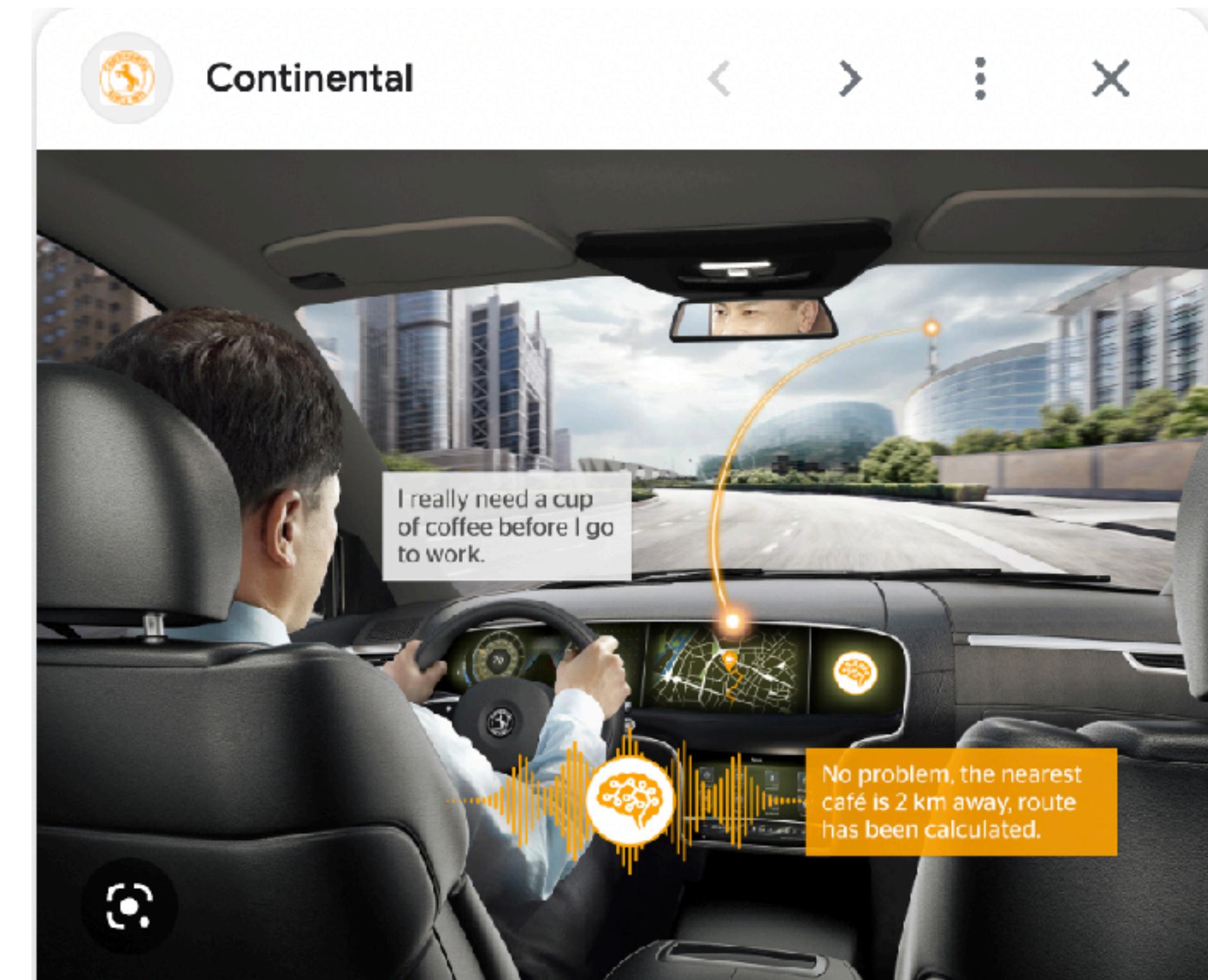
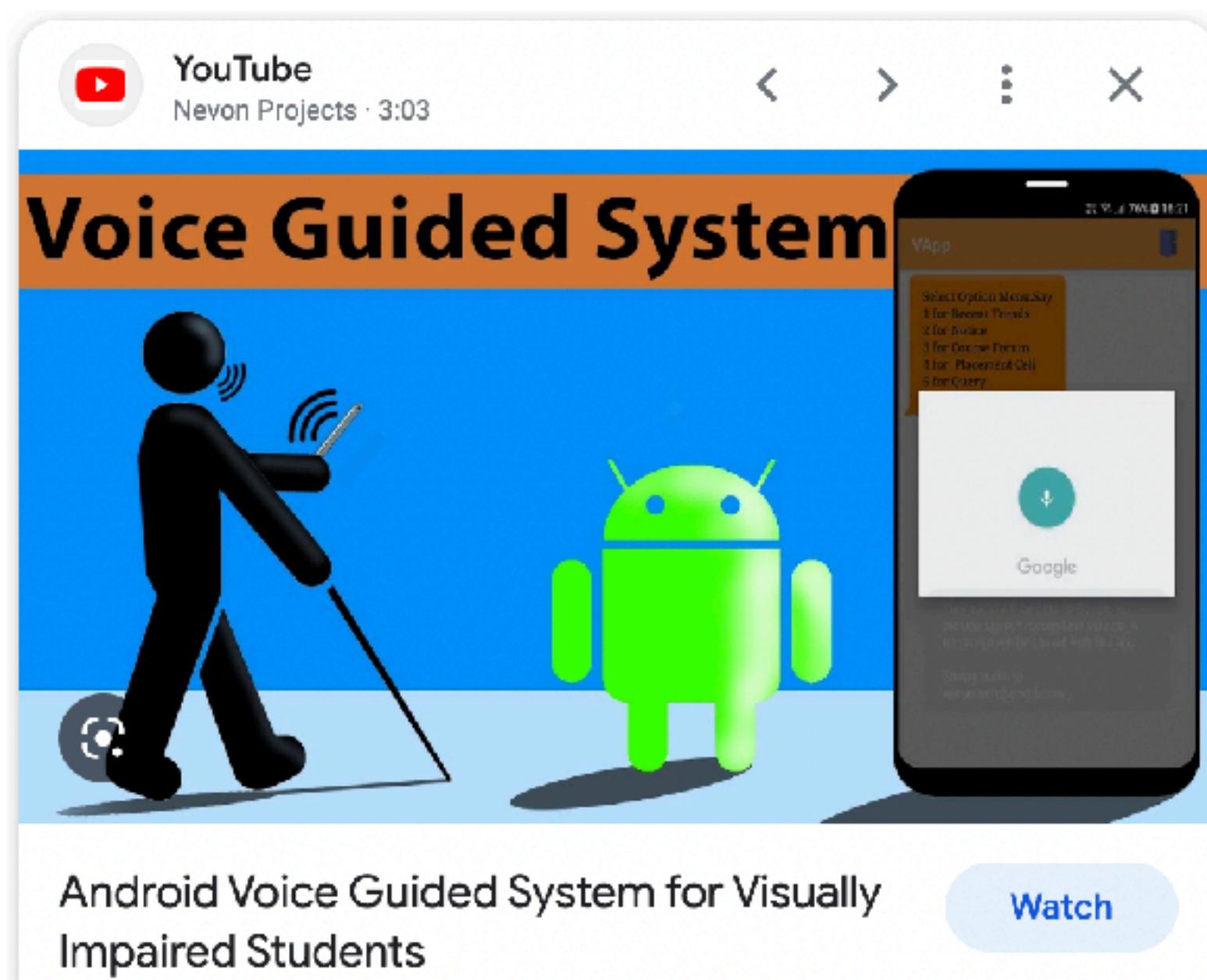
Voice Reader: Text to Speech

Text to Speech: Voice Reader TTS in Chrome with [Visit](#)

STT (Speech-to-Text)



text to speech



Smart Voice Assistant: Continental Develops Adaptive Voice-Activated...

Visit