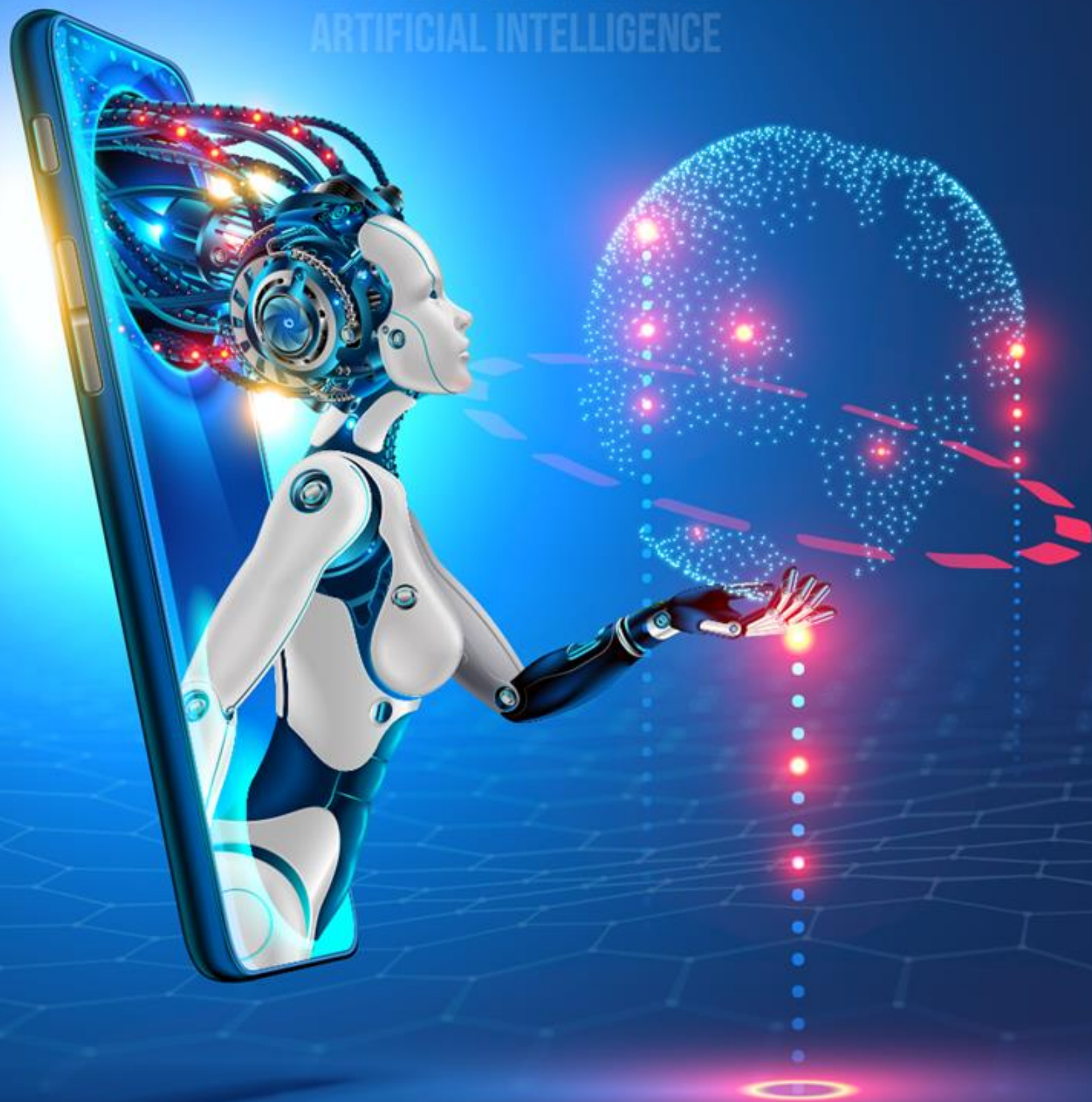


DATA AND
ARTIFICIAL INTELLIGENCE



simplilearn

PURDUE
UNIVERSITY

Natural Language Processing and Speech Recognition



Introduction to Speech Recognition

Learning Objectives

By the end of this lesson, you will be able to:

- Describe speech and its components
- Interpret speech signal processing
- List speech digitization mechanisms



Introduction to Speech

What Is Speech?

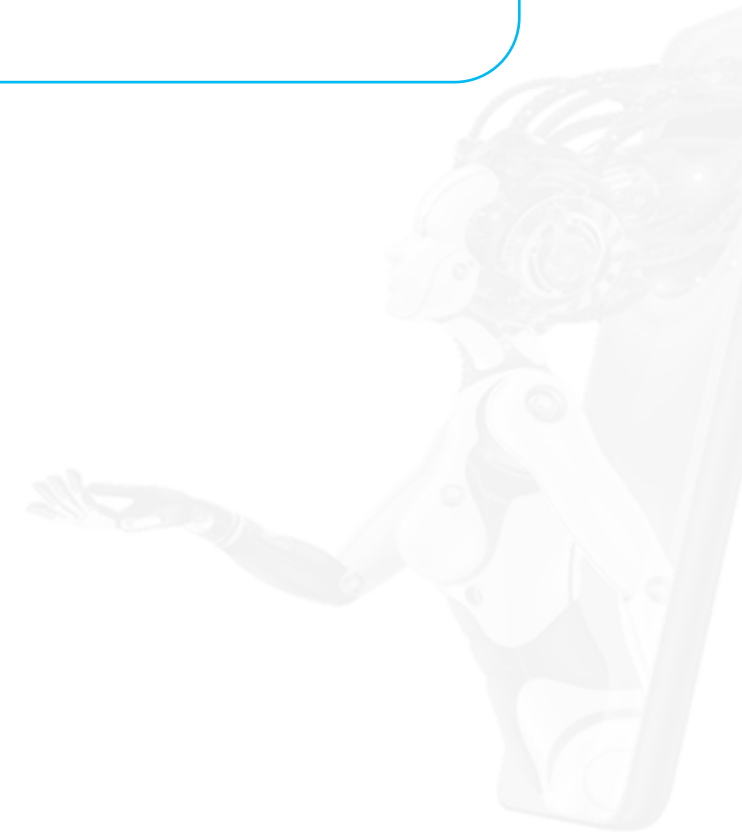
- Expression of thought by production of articulate sound, bearing a definite meaning
- Verbally produced sound



The Speech Database

- A speech database or a speech corpus is a collection of speech audio files and text transcriptions of these files.
- A speech corpus can be used to create acoustic models.

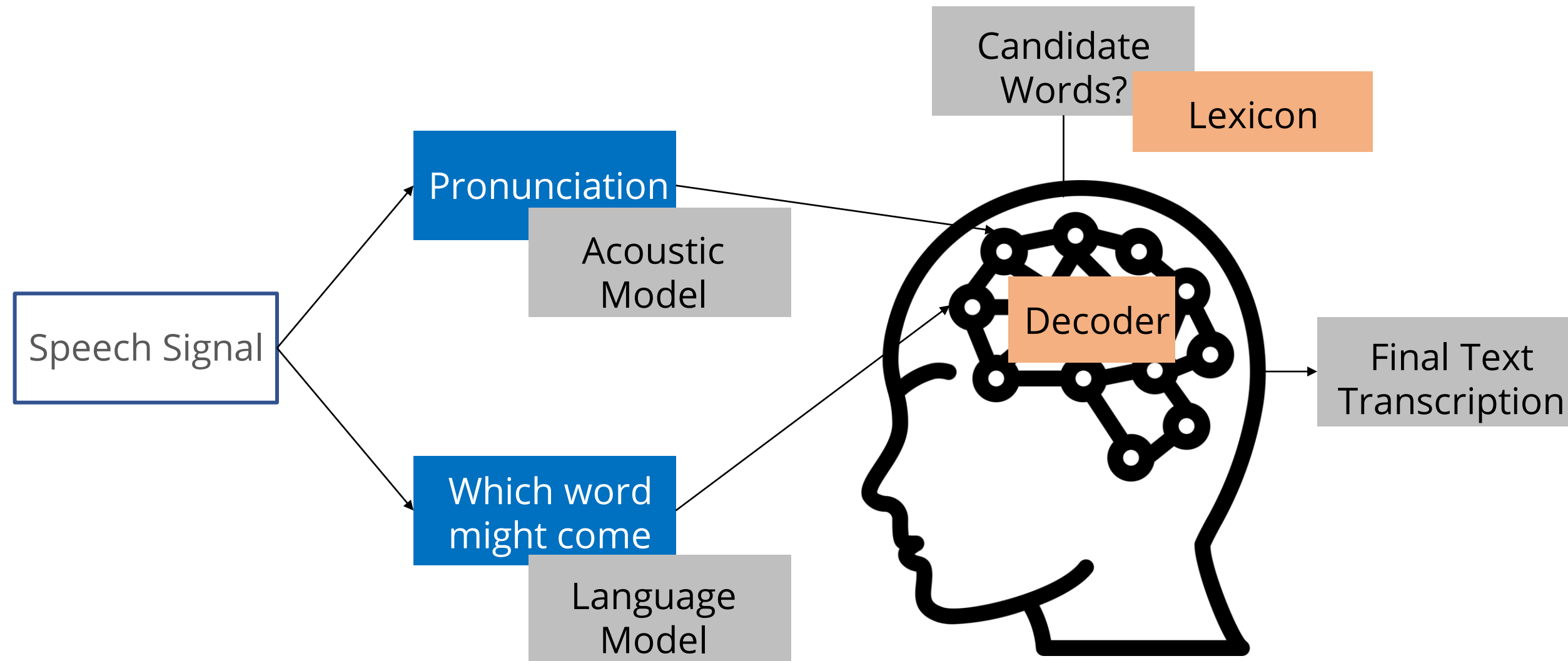
Genre	Size in hours	Number of sentences	Number of phonemes
ARCTIC	0.868	1,132	35,970
BTEC	7.122	7,633	267,007
NEWS	8.157	4,904	302,325
Miscellaneous	0.563	652	26,836
Total	16.71	14,321	632,165



Human Speech and Phonetics

How Humans Do It?

Articulation produces sound waves which the ear conveys to the brain for processing.



How Computers Do It?

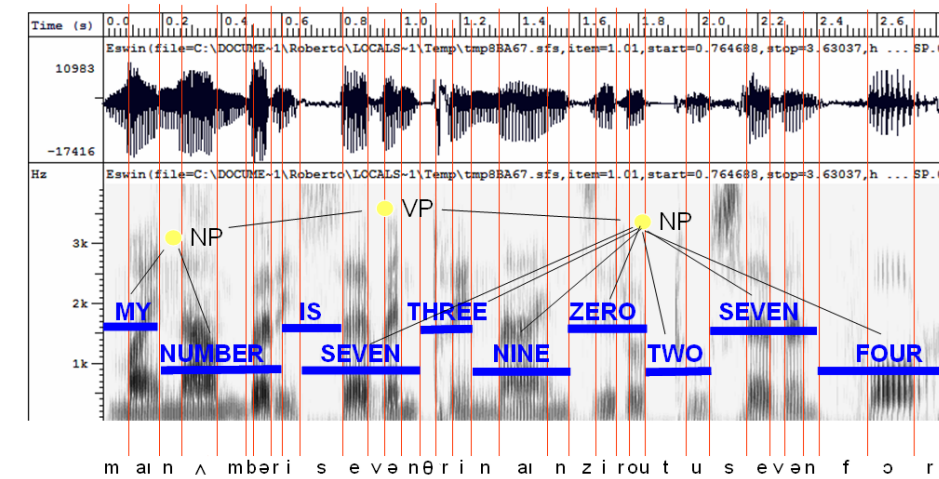
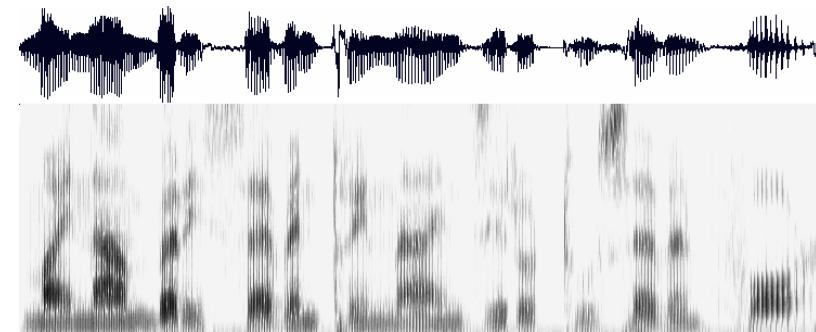
Digitization acoustic analysis of the speech signal leads to linguistic interpretation.



Acoustic
Waveform



Acoustic
Signal



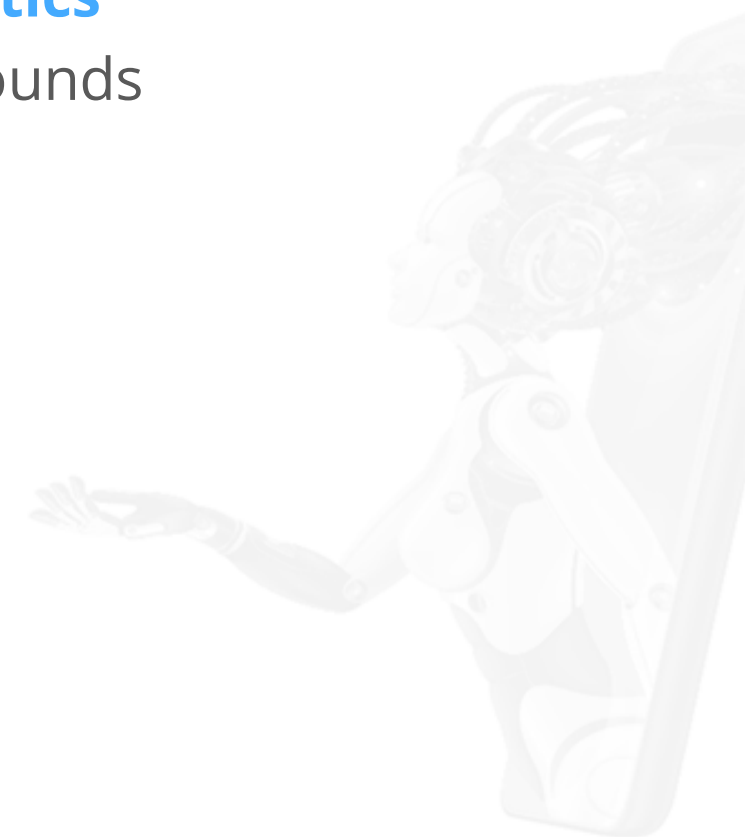
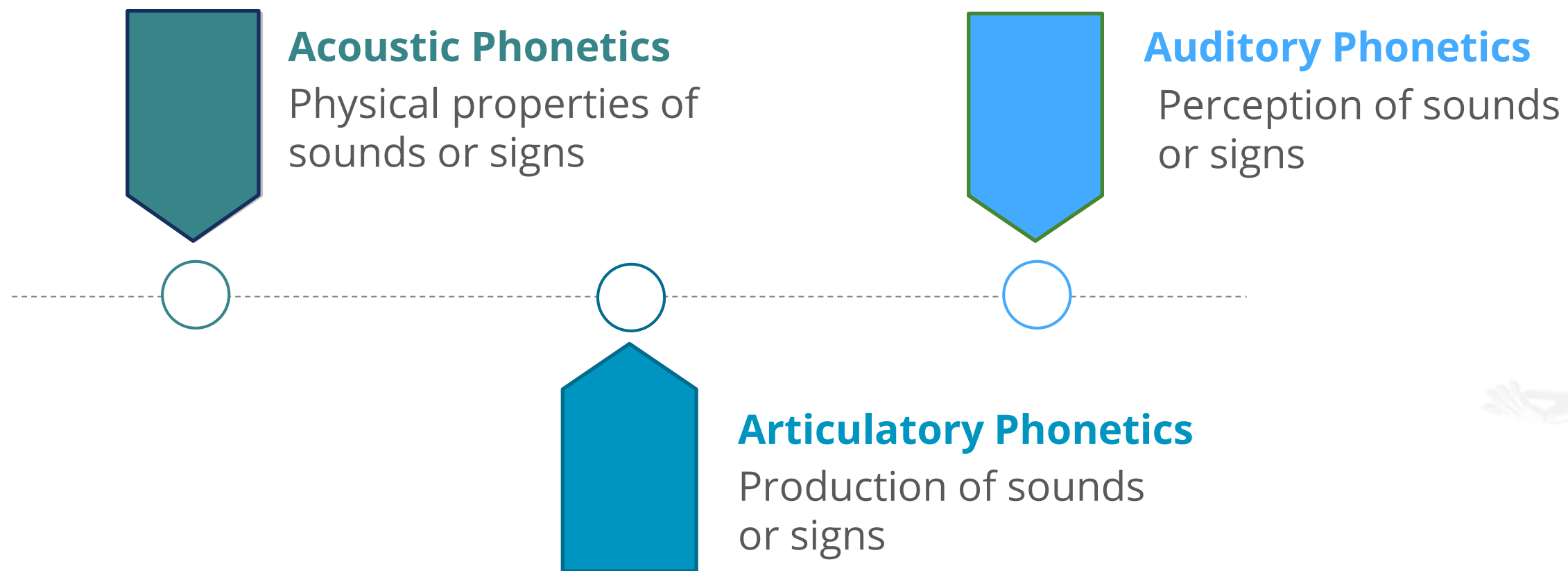
Speech
Recognition

What Is Phonetics?

The scientific study of speech sound



Types of Phonetics



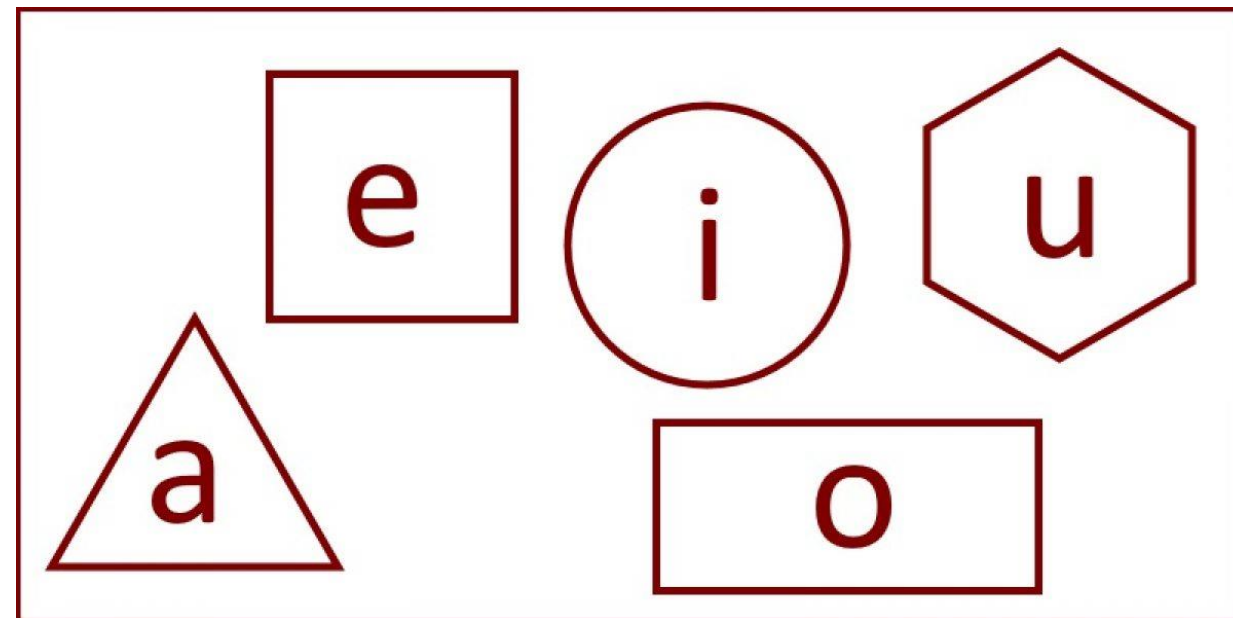
Types of Speech Sounds

Vowels are the class of sound which creates the least obstruction to the flow of air when it passes from the larynx to the lips.

Vowels

Consonants

Cardinal Vowels



Types of Speech Sounds

Consonants are sounds that obstruct the flow of air through the vocal tract.

Vowels

Consonants

Cardinal Vowels

B, C, D, F,
G, H, J, K,
L, M, N, P,
Q, R, S, T,
V, W, X,
Y, Z



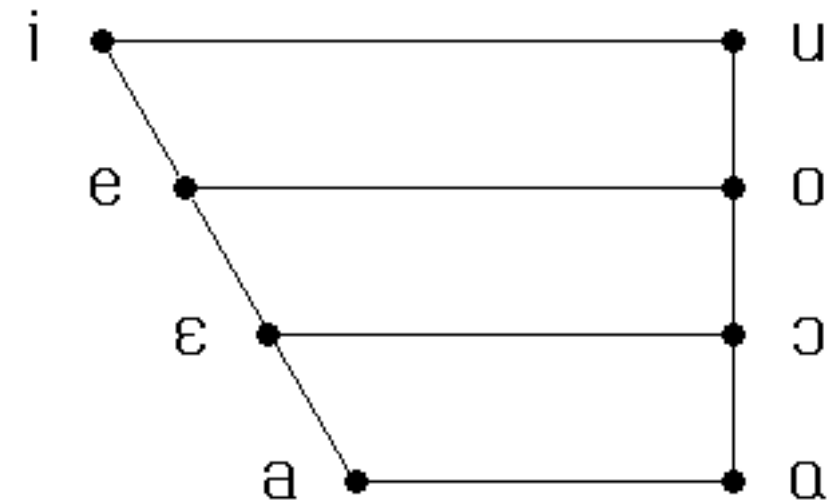
Types of Speech Sounds

Cardinal vowels refer to the way of classifying vowels independently from the vowel system of any language. They serve as reference points that other vowels can relate to.

Vowels

Consonants

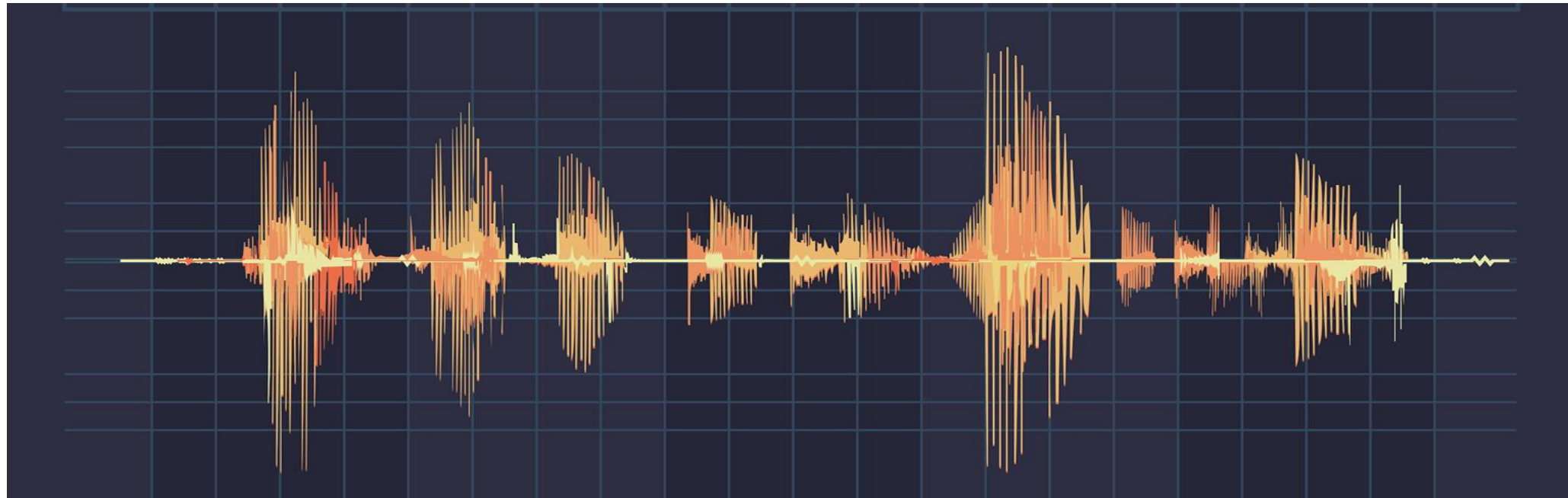
Cardinal Vowels



What Is Speech Recognition?

Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format.

It is also known as automatic speech recognition or computer speech recognition.



Speech Recognition Techniques: Evolution

The Audrey System



John R. Pierce

- The first serious speech recognizer (AUDREY) was developed in 1952 by John, Biddulph, and Balashek of Bell Labs.
- Using a simple frequency splitter, it generated plots of the first two formants, which it identified by matching them against prestored patterns in an analog memory.
- With training, it was reported, the machine achieved 97 percent accuracy on the spoken forms of 10 digits.

Speech Processing and Digital Computers



James Flanagan joined Bell Labs in 1957 and introduced AD/DA and started pioneering new research in speech processing using computers.

The ARPA SUR Project



Raj Reddy - CMU

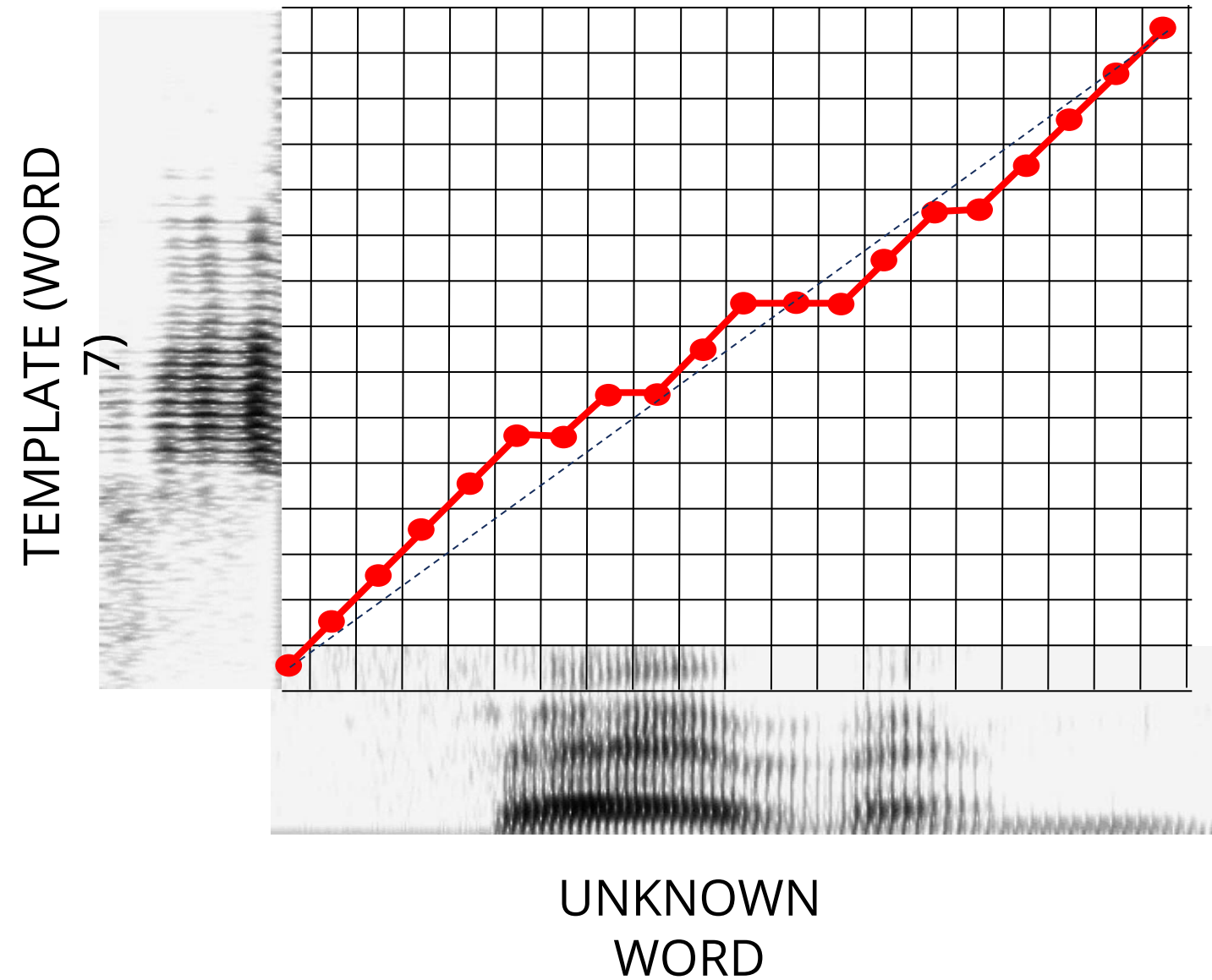
- During 1971 to 1976 ARPA launched 5 year Spoken Understanding Research program.
- The goal of ARPA SUR was to build a recognition system with 90 percent sentence accuracy for continuous speech sentences, using thousand-word vocabularies, not in real time.
- One of the ways the CMU team achieved the goal was by limiting spoken words to certain sequences in the sentence.

Limitations of the ARPA SUR Project

- Lack of clear evaluation criteria
- Speech understanding was way ahead of its time
- Project was not extended



The Engineering Approach



- In 1970's, concepts from dynamic programming were used to calculate the least cost path of best matches of feature vectors between unknown word and templates.
- Recognizing one word could take seconds to hours, depending on vocabulary size and word length.

The Statistical Approach



Fred Jelinek



Jim
Baker

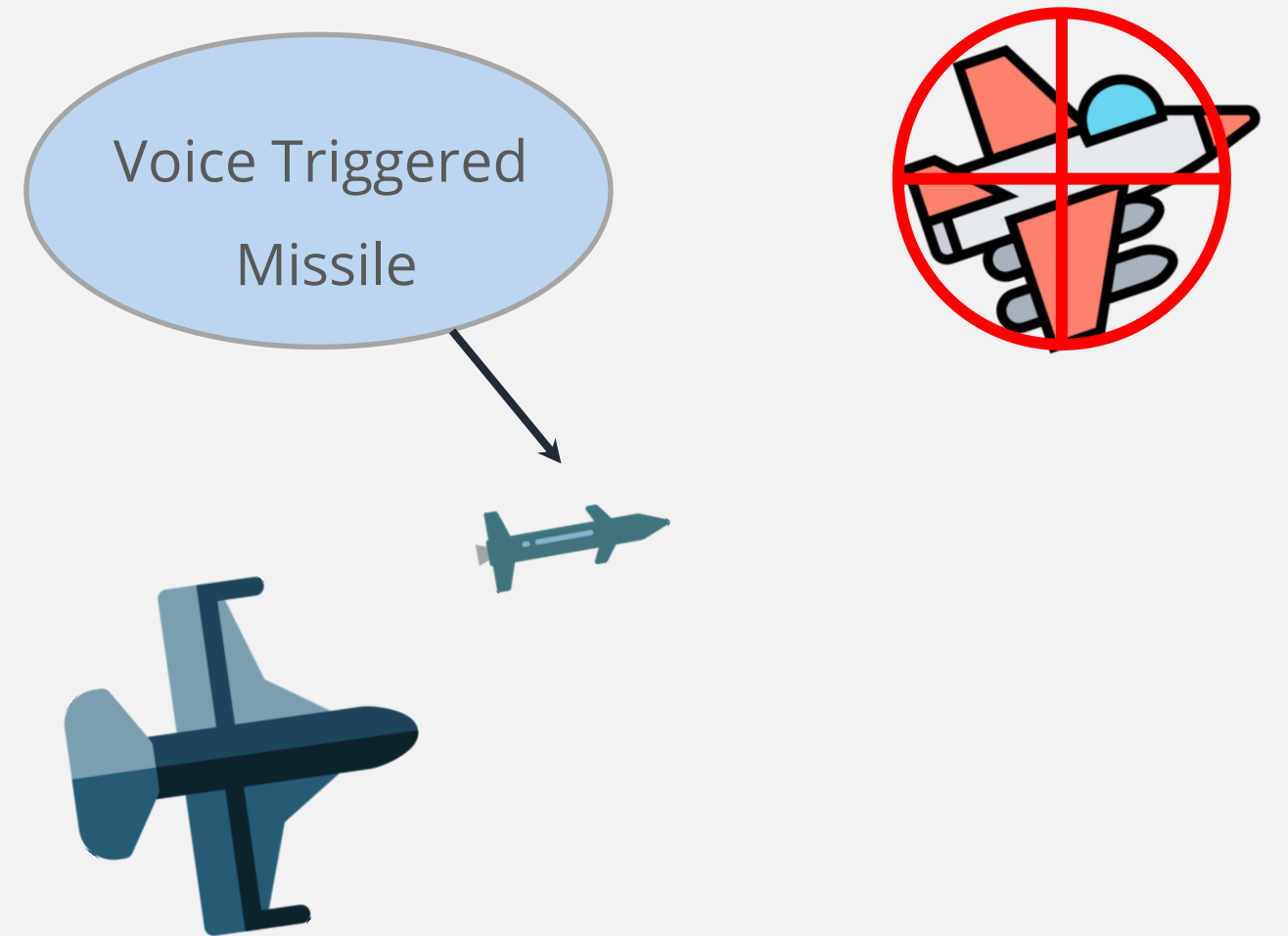
- In 1980s, purely statistical approach was pursued by Fred Jelinek and Jim Baker for IBM's T. J. Watson.
- This approach was based on the work of **Hidden Markov Models** by Leonard Baum at IDA, Princeton in the late 1960s.
- This led to the foundation of modern speech recognition engines.

Applications of Automated Speech Recognition (ASR)

Applications of ASR

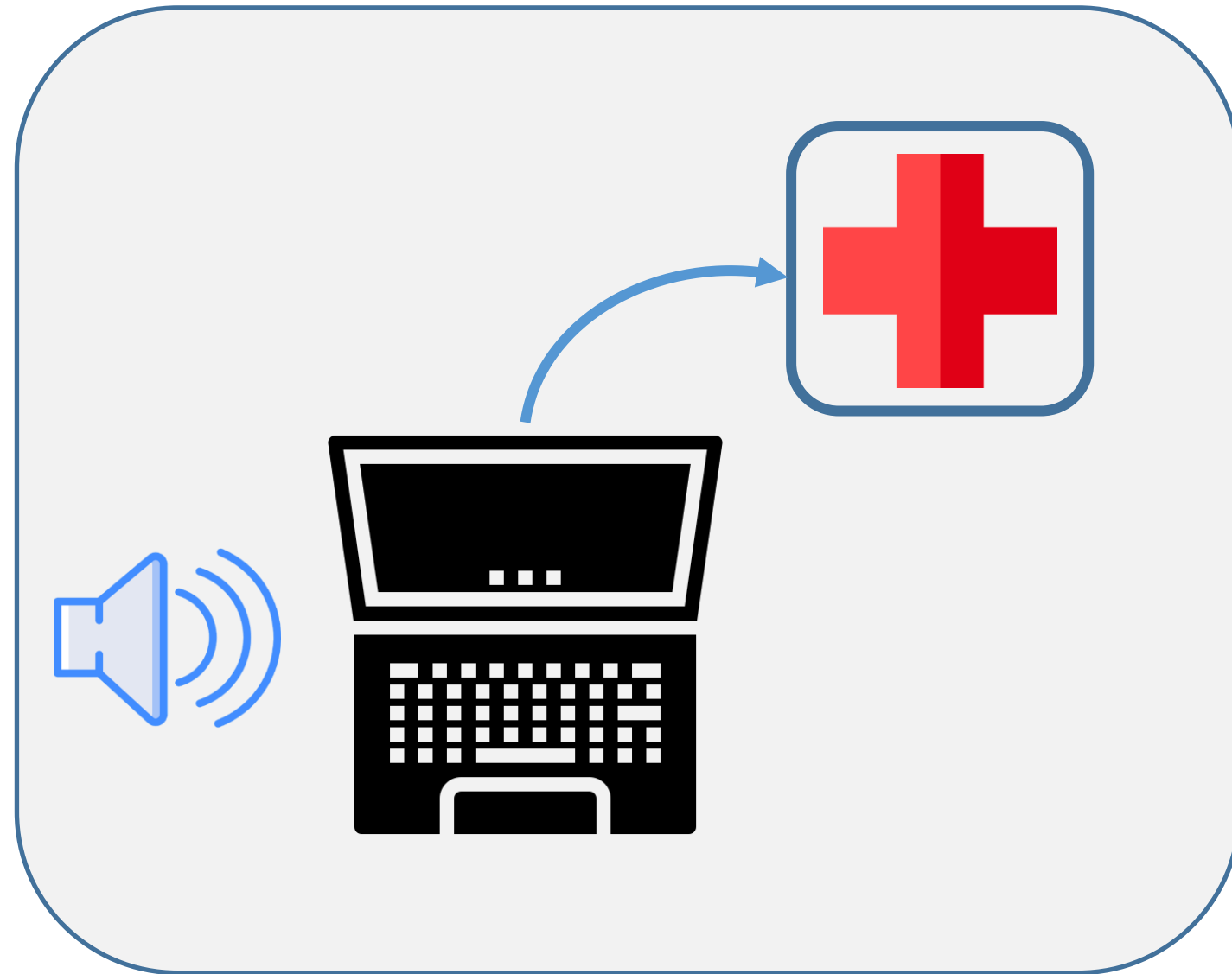


Forensic and Law Enforcement

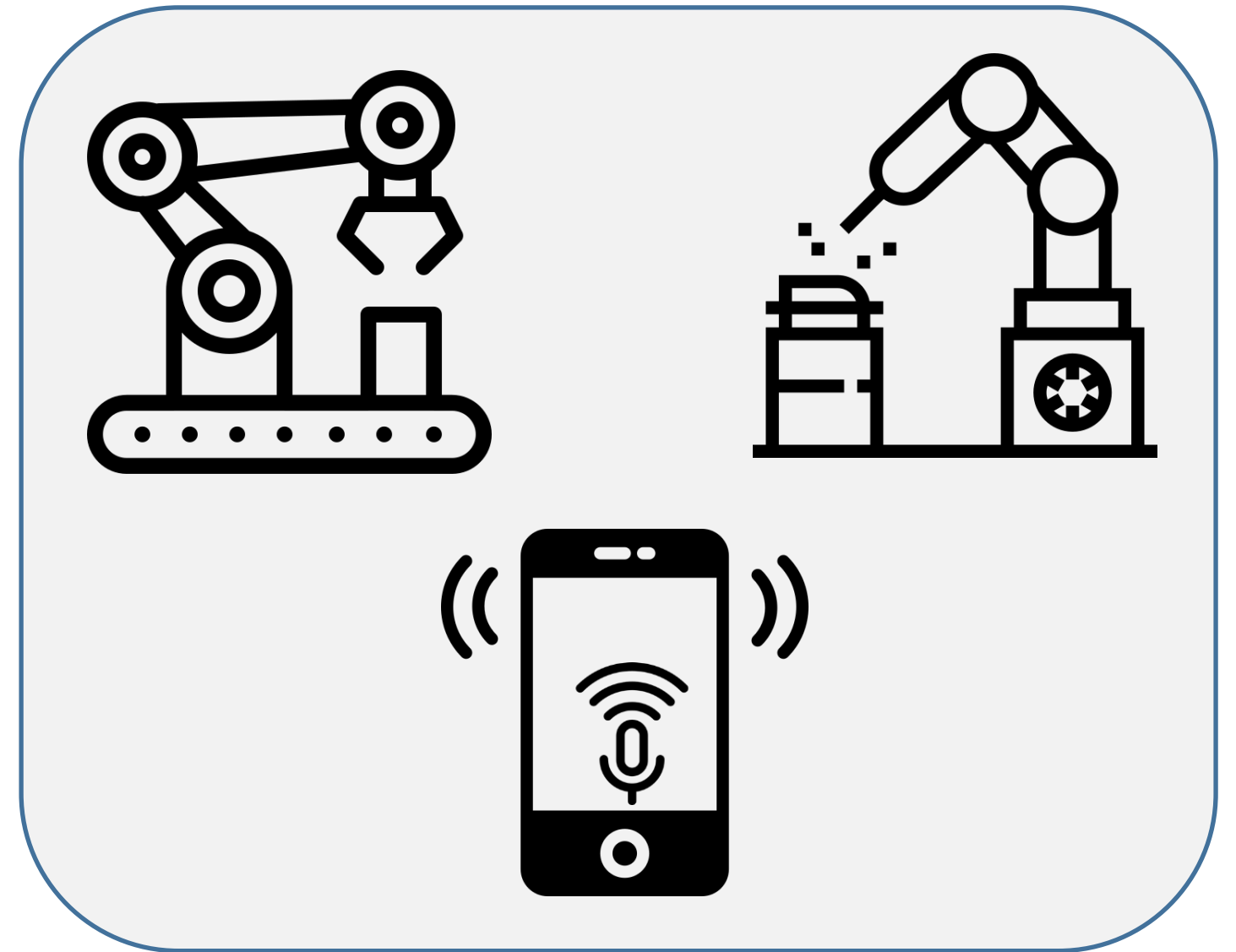


Defense and Aviation

Applications of ASR

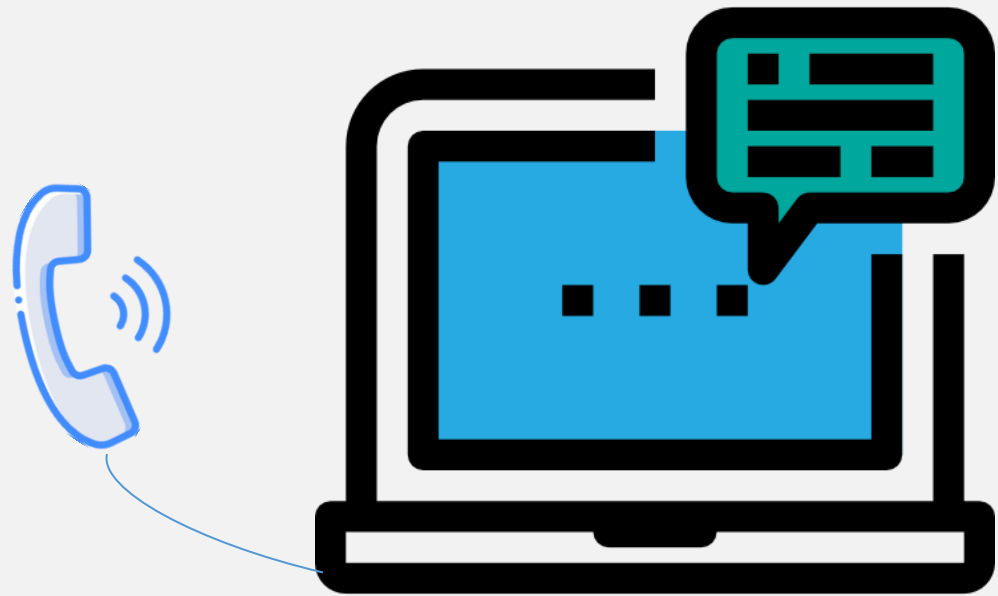


Medical Assistance

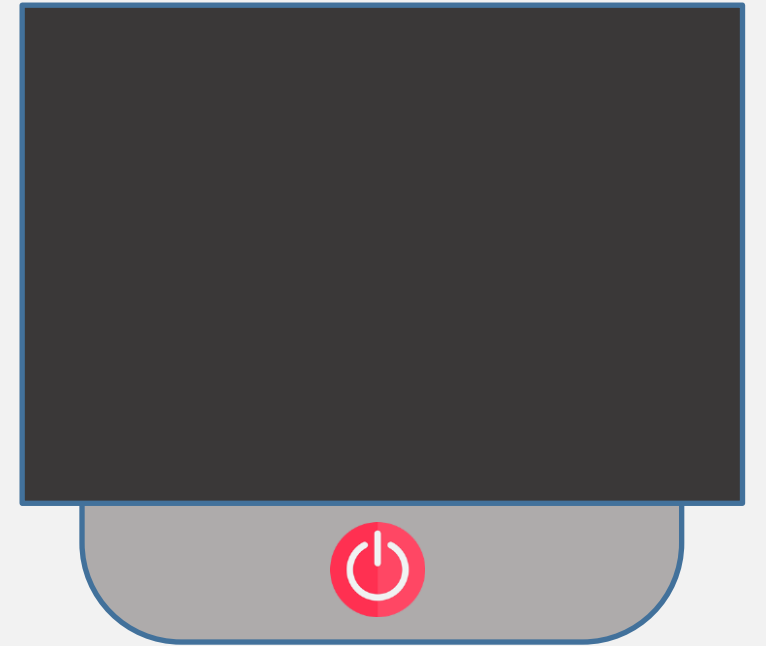


Industrial Robotics

Applications of ASR



Telecommunications Industry



I.T. and Consumer Electronics

Applications of ASR



Home Automation and Security Access Control

Use Cases of ASR

Use Cases of ASR: Finance

- Royal Bank of Canada (RBC) has the facility to pay bills using Apple's Siri.
- Members of United Services Automobile Association (USAA) accesses information about account balances, transactions, and spending patterns using Amazon Alexa.
- Customers of U.S. Bank perform banking tasks using Amazon Alexa.

Use Cases of ASR: HR and Marketing

- HR and marketing department handles lot of work which are repetitive. Therefore, instead of using conventional HR tools, they can opt for ASR.
- For example, they can scour employee details for a certain parameter from the database using ASR.

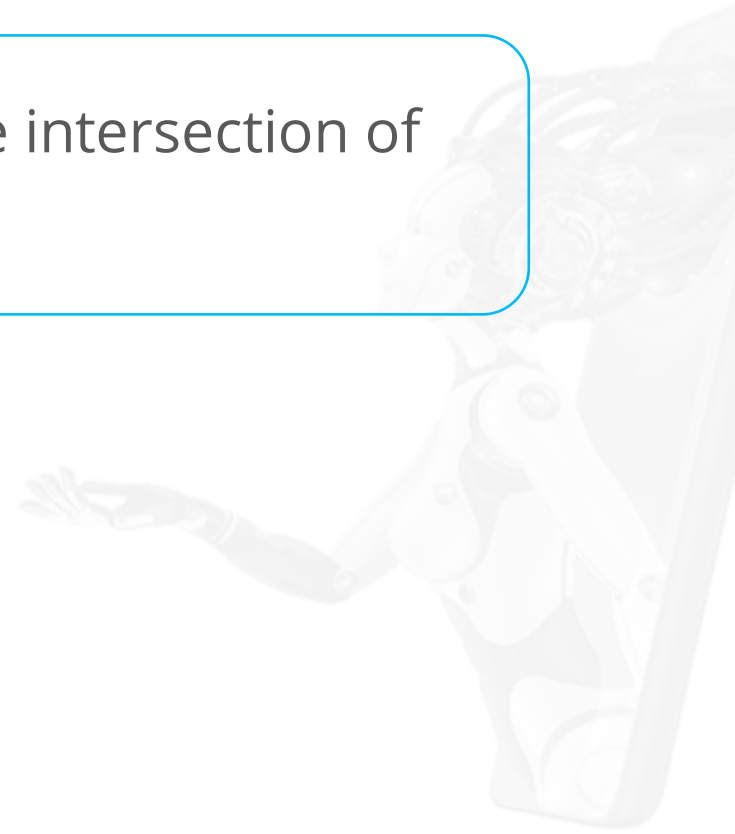
Use Cases of ASR: Retail

- 👁 Customers can sort items in Amazon.com using Alexa.
- 👁 Amazon is investing billions of dollars in shops, where customers can purchase items with ASR technology.

Speech Processing

What Is Speech Processing?

It is the study of speech signals, the processing methods of these signals, and also the intersection of digital signal processing and natural language processing.



Components of Speech Processing



Speech Coding

- Compresses a speech file
- Uses standard compression techniques
- For example: In MP3 format, you use perceptual coding exploiting sensory organ biases



Speech Synthesis

- Constructs speech waveform from words
- Depends on speaker quality and accent
- Faces challenge with prosodies



Speech Recognition

- Converts a sound waveform to words
- Is the most relevant task in the industry
- For example: Sphinx by CMU, ViaVoice by IBM, and SDK by Microsoft

Speech Verification

- Concerned with biometrics
- Accepted as a verification technique
- Modeled differently from speech recognition in terms of speaker quality, prosody, pitch, and accent



Dyslexia and Auditory Problems

- Deal with voice and ear defects
- Detect and correct speech disfluencies
- Lead to develop better ear substitutes such as Cochlear Implants



Audio Engineering

- Adds effects to sound
- Adds clarity to reproduction
- Used in big industries like Dolby, Bose, and Philips
- Includes voice morphing





Knowledge Check

Knowledge Check

1

Which of the following answers correctly defines speech recognition?

- a. A process where a human speaks to a computer and the computer recognizes what is being said
- b. A process where a computer generates human speech via a machine or robot
- c. A process where a speech pathologist recognizes a child's speech despite difficulty in verbalization
- d. A process where a person of a different country recognizes the speech of a different language



Knowledge Check

1

Which of the following answers correctly defines speech recognition?

- a. A process where a human speaks to a computer and the computer recognizes what is being said
- b. A process where a computer generates human speech via a machine or robot
- c. A process where a speech pathologist recognizes a child's speech despite difficulty in verbalization
- d. A process where a person of a different country recognizes the speech of a different language



The correct answer is **b**

Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format.

Knowledge Check

2

Which of the following is not an example of a program that uses speech recognition technology?

- a. Apple's Siri on the iPhone
- b. Microsoft Word
- c. Dragon Dictation
- d. Google Chrome



**Knowledge
Check**

2

Which of the following is not an example of a program that uses speech recognition technology?

- a. Apple's Siri on the iPhone
- b. Microsoft Word
- c. Dragon Dictation
- d. Google Chrome



The correct answer is **b**

Microsoft Word does not use the speech recognition technology.

Key Takeaways

- 👁 Speech processing is the study of speech signals and their processing methods.
- 👁 Digitization acoustic analysis of the speech signal leads to linguistic interpretation.

