Natural Language Processing

Speech Recognition Technique

Learning Objectives

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

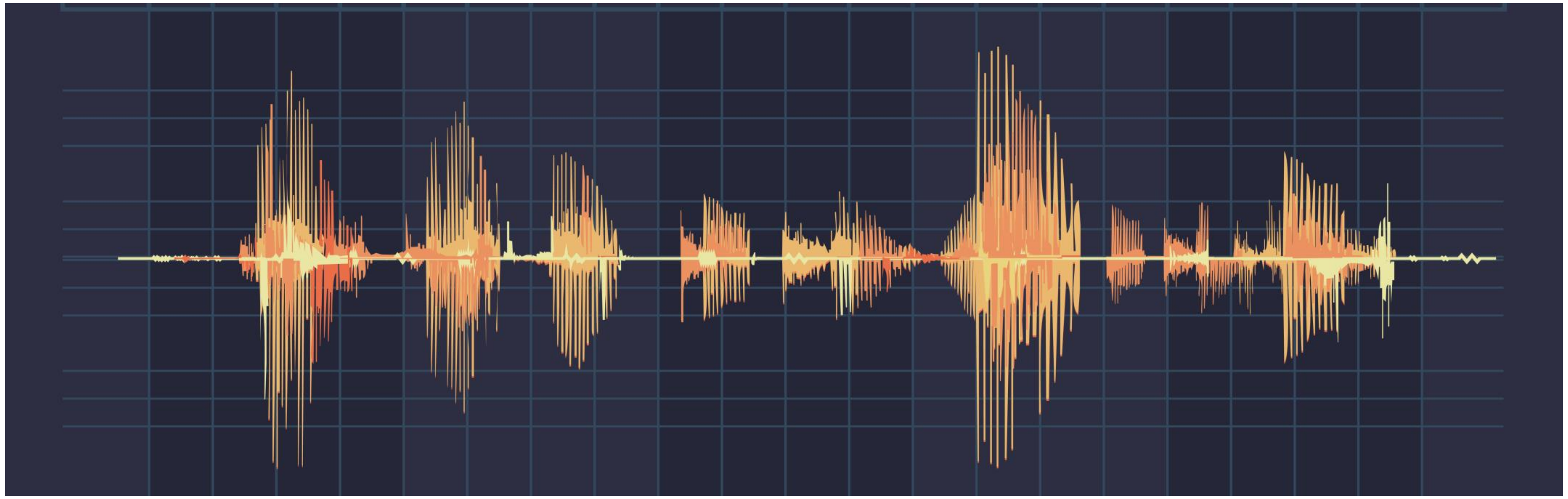
- Explain the basic concepts of speech
- Describe how to read, load, and process the data
- Explain how to create speech models
- Identify the types of speech libraries
- Demonstrate the conversion of text to speech for a paragraph

Basic Concepts of Speech

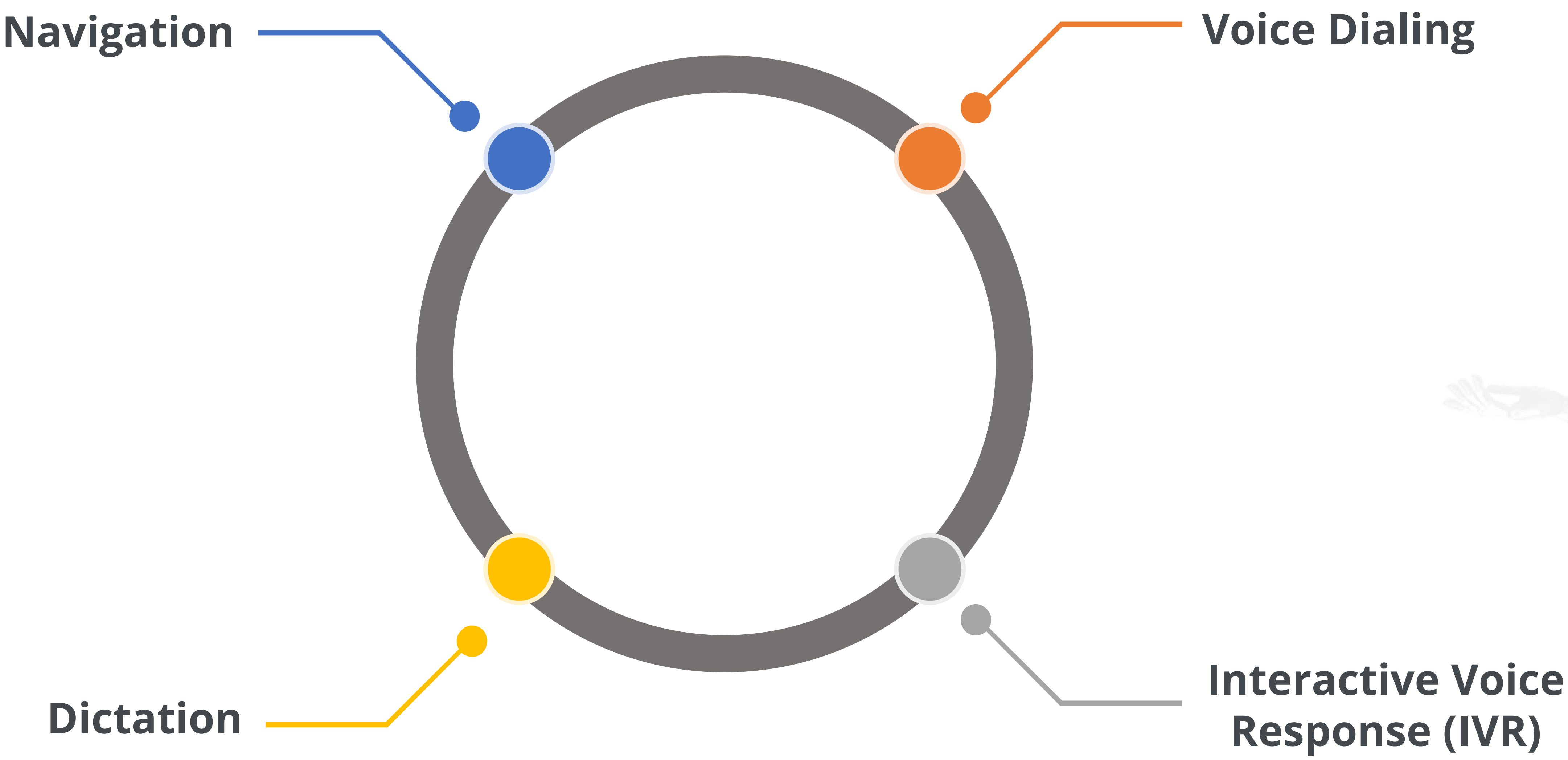
What is Speech Recognition?

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.

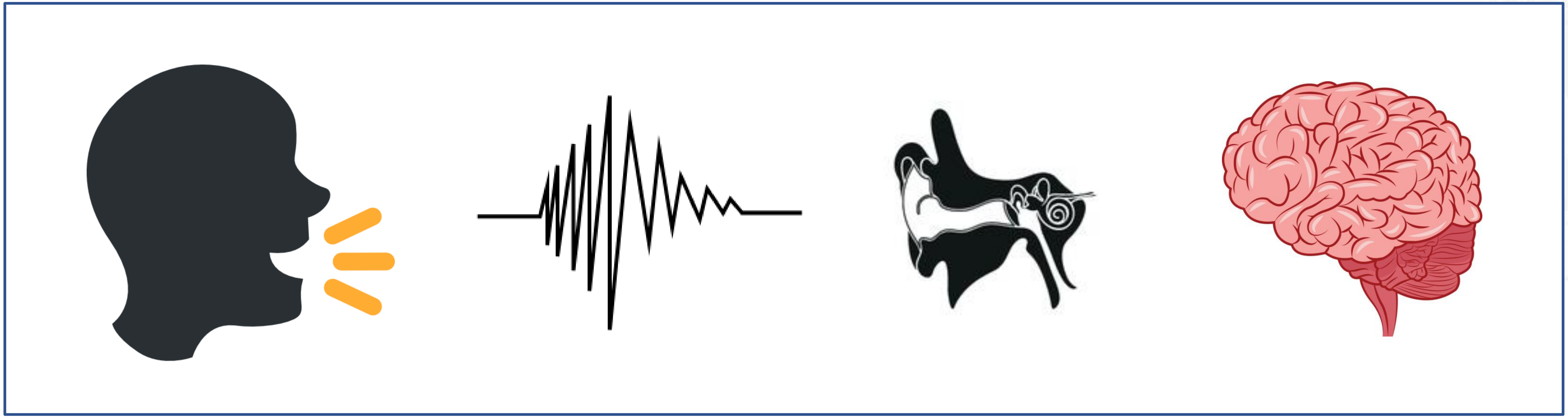


Speech Recognition: Uses

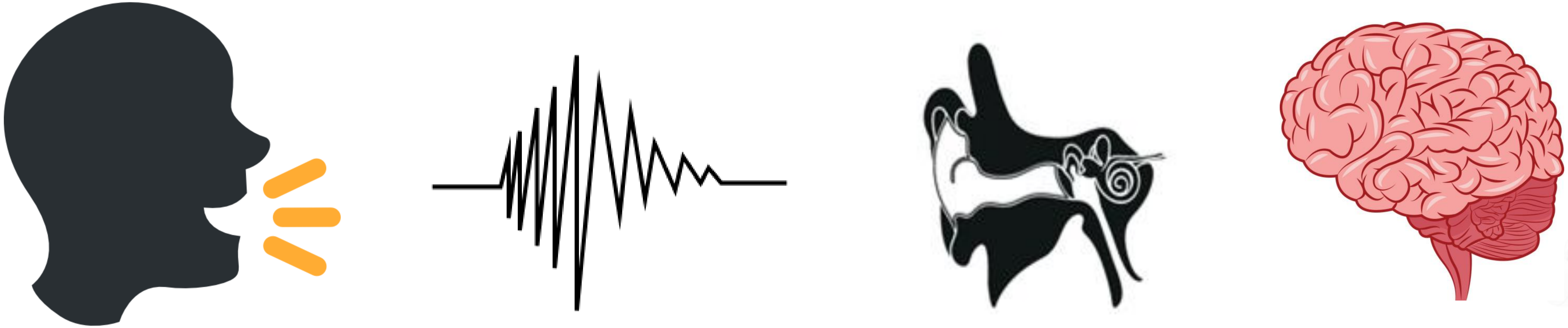


Speech Recognition by Humans

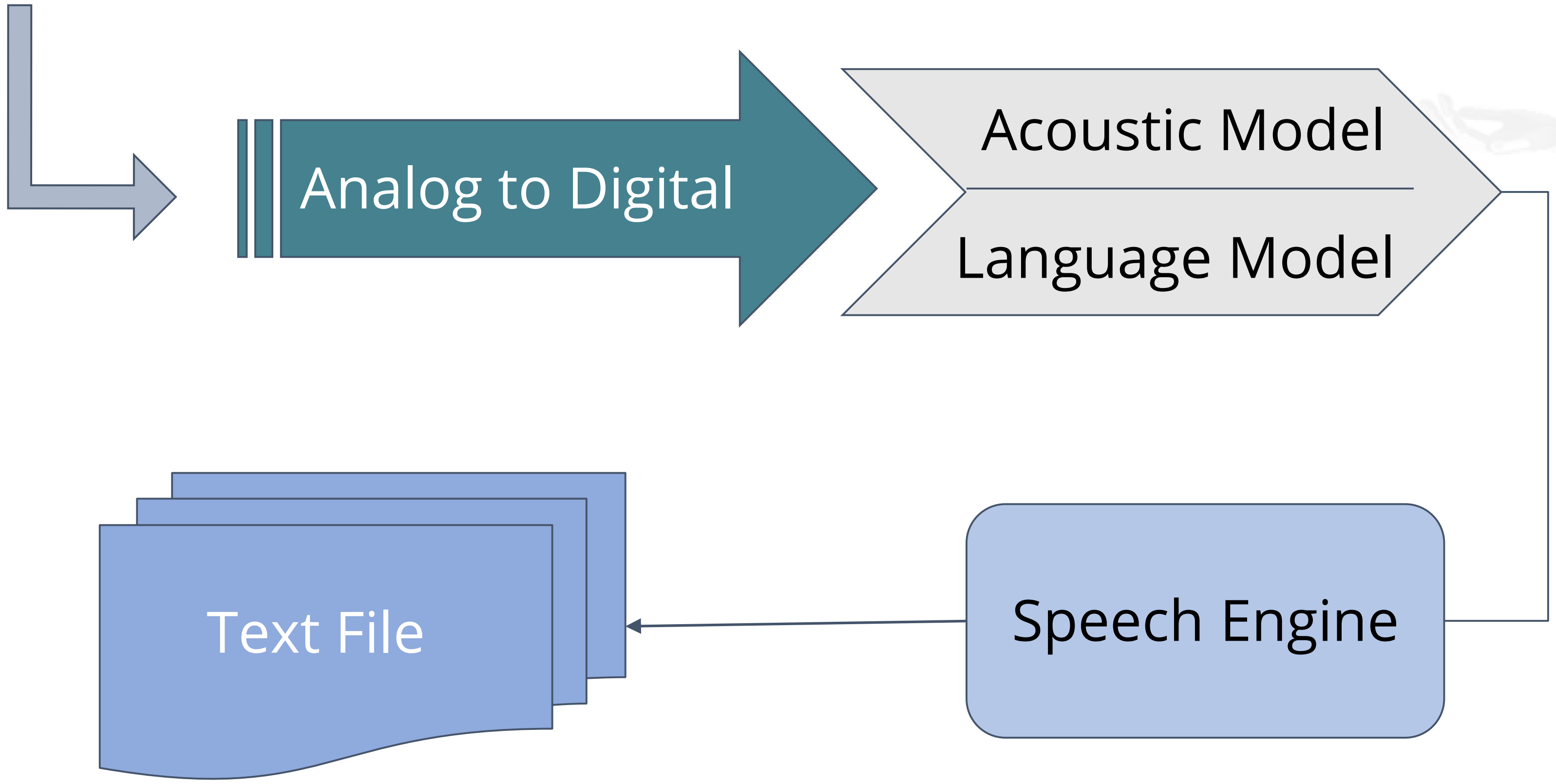
Articulation produces sound waves and the ear transmits it to the brain for processing.



Speech Recognition by Computers



Sound waves are fed into conversion system and then for processing, after applying the acoustic model or language model.



Acoustic and Language Models

Acoustic Model

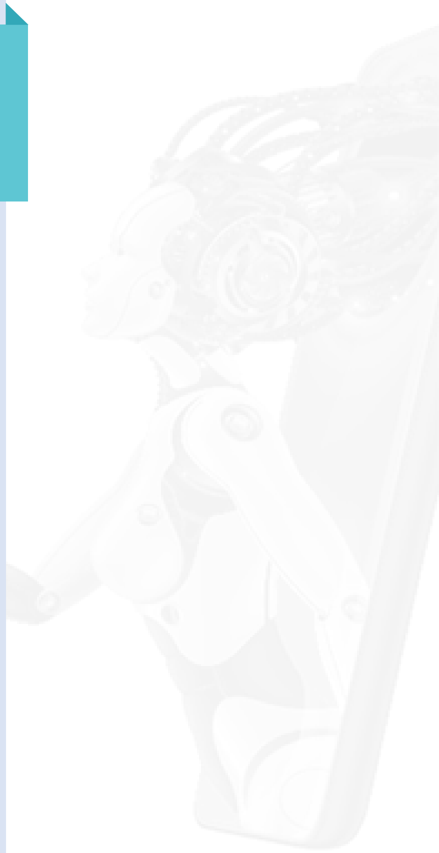
This kind of model is created by using transcripts of speech and processing it with the software for creating statistical representation of the sounds.

It is mainly used in speech recognition engine to recognize speech.

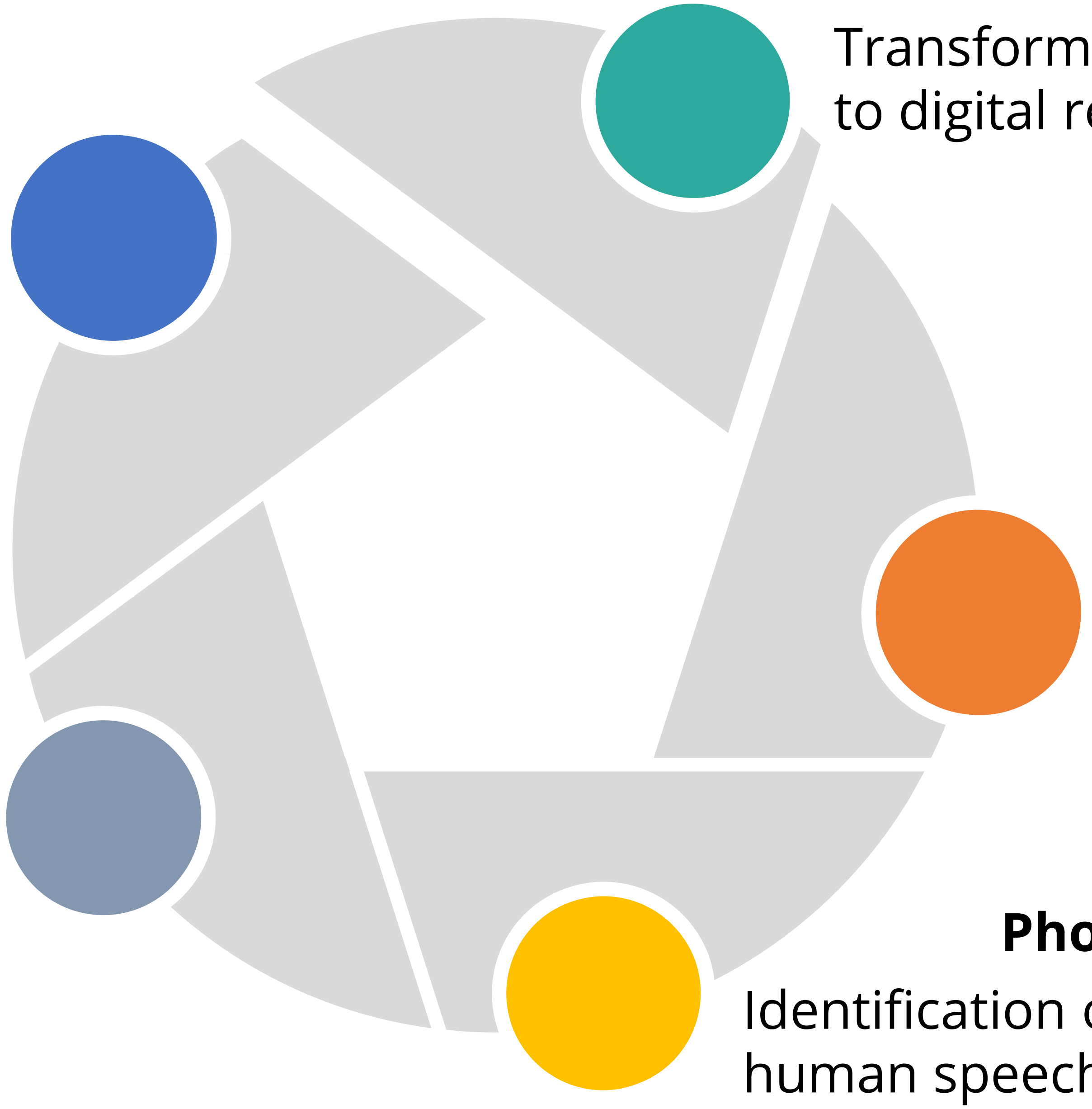
Language Model

It is used in Natural Language Processing applications to capture the properties of language and processing.

Example: To predict the next word in a sequence



Processes Involved in Speech Recognition



Digitization

Transformation of analog signal to digital representation

Signal Processing

Noise reduction from speech

Phonetics

Identification of variations in human speech

Lexicology and Syntax

Understanding word, their nature, and meaning

Phonology

Recognition of each constituent sound distinctions

Sequence of States

- Speech is a continuous audio stream.

- It is a sequence of states:
  - diphones
  - triphones
  - quinphones

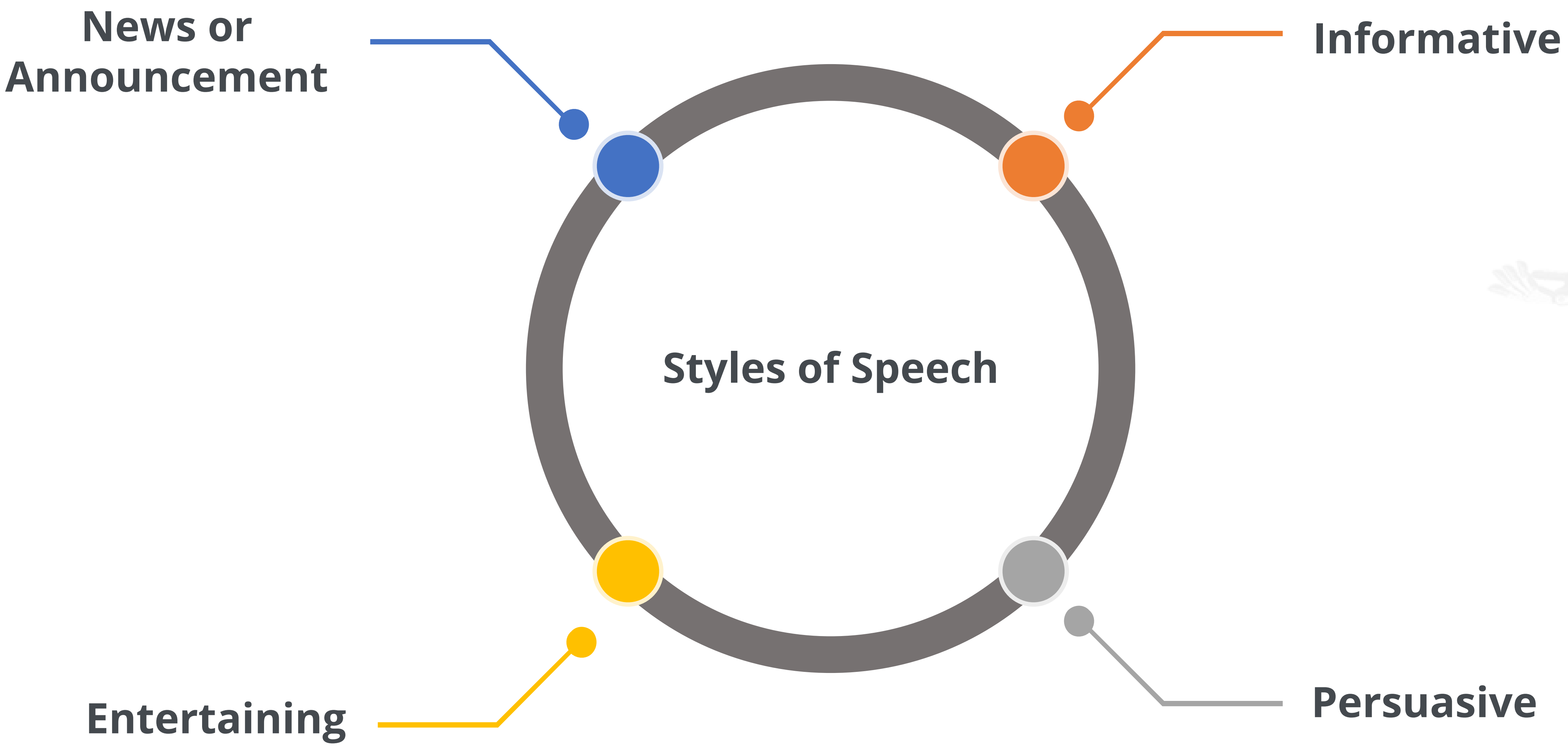
  Example: TRIM
  TRIM has 4 phones: [T], [R], [I], and [M]

- Phones are used to drive the words.

Styles of Speech

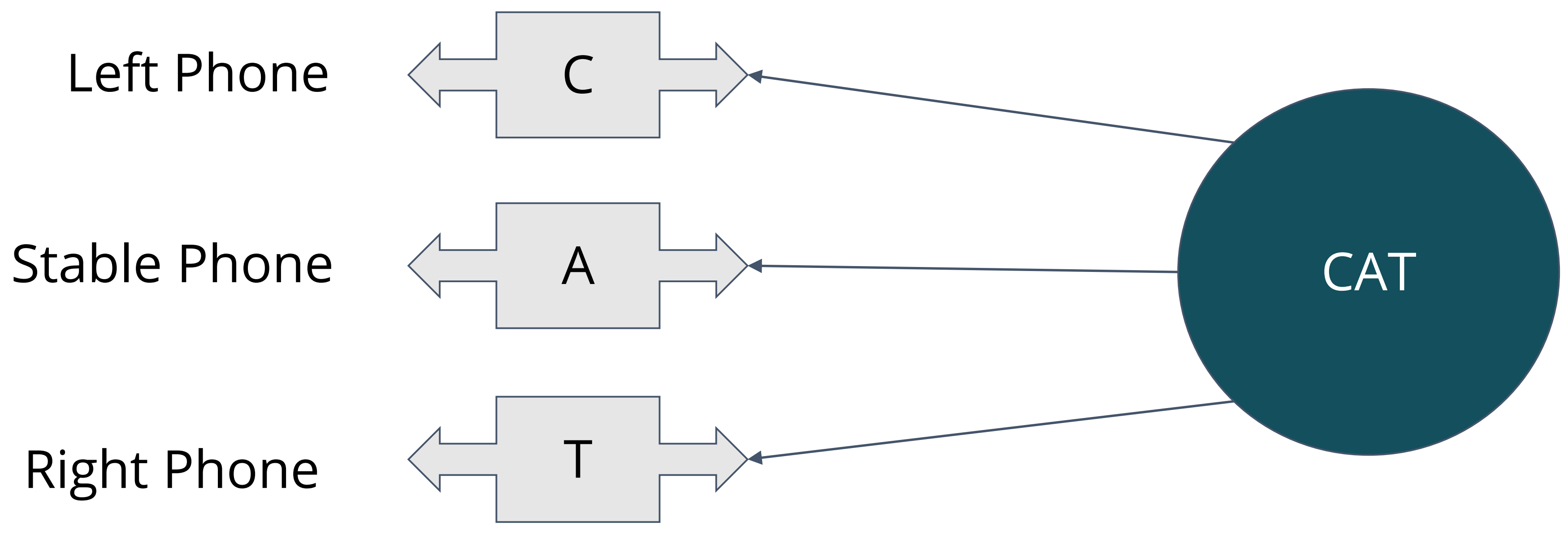
Acoustic property of waveform depends on the phone-context (in which context it is spoken). Example:

- **Tell** me a story
- This is a fairy **tail**

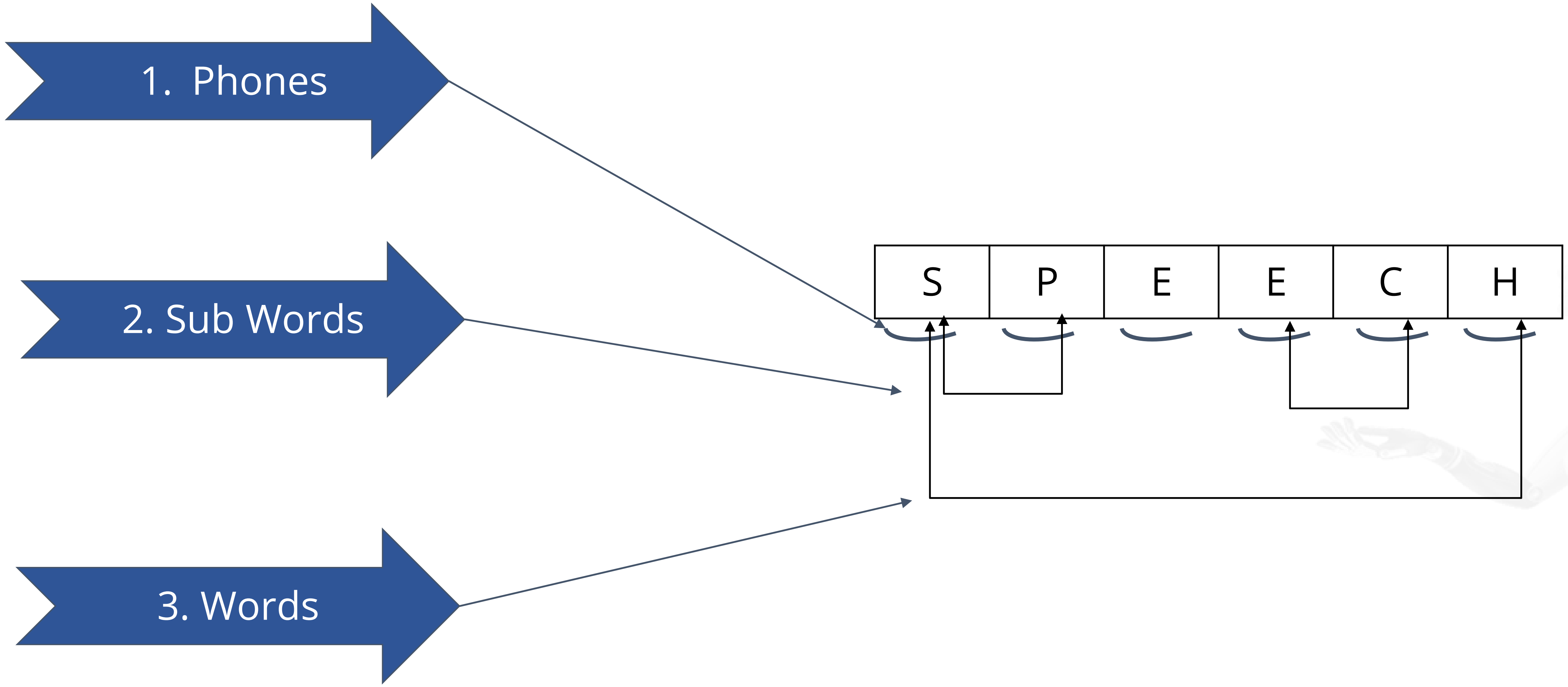


Detectors

- Detectors are used in speech processing for identifying the presence or absence of human speech.
- In computation, parts of phones must be processed rather than the whole.
- Example: Triphones (left, stable, and right phone) for CAT



Word Formation



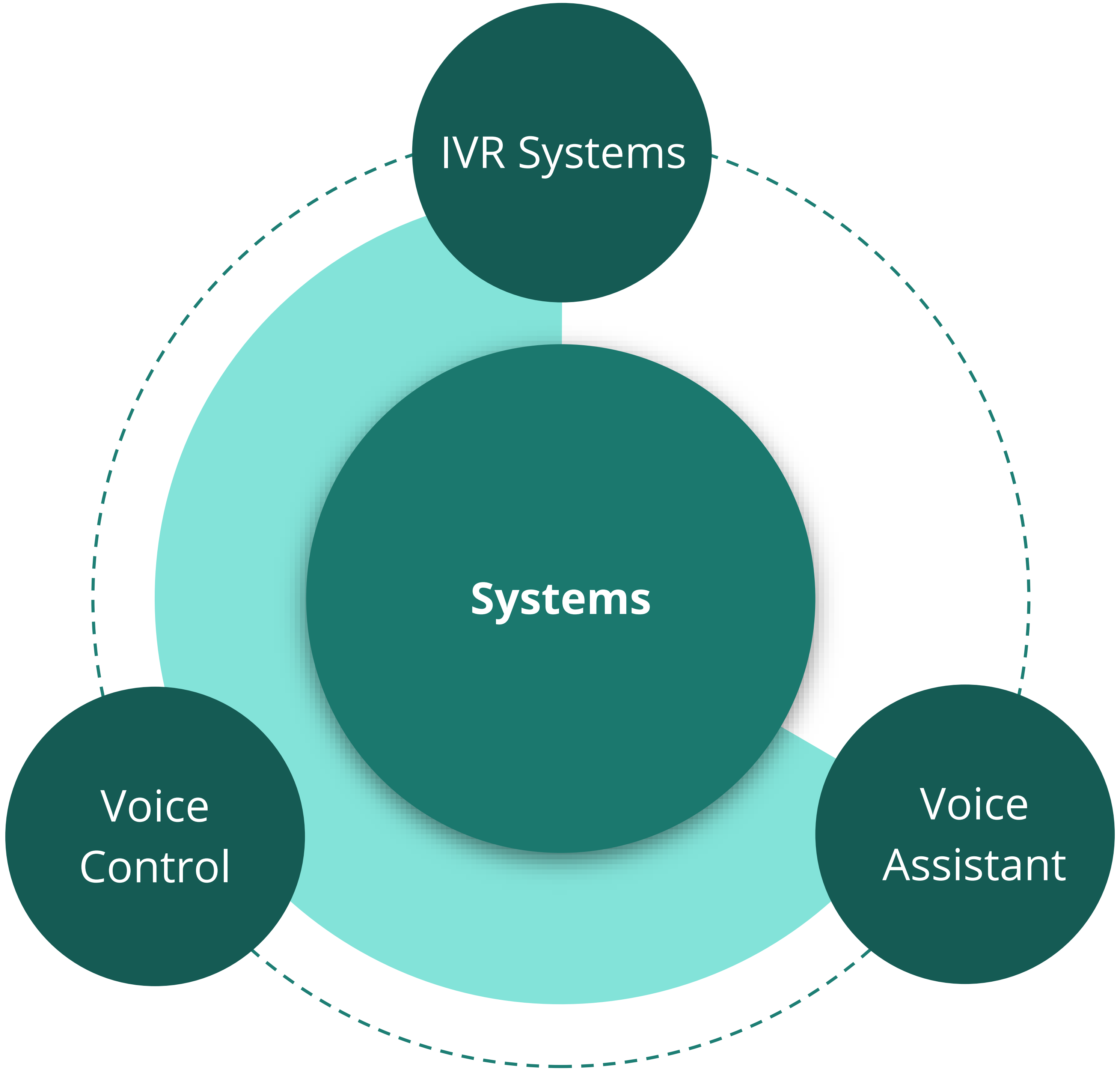
1. Phones

2. Sub Words

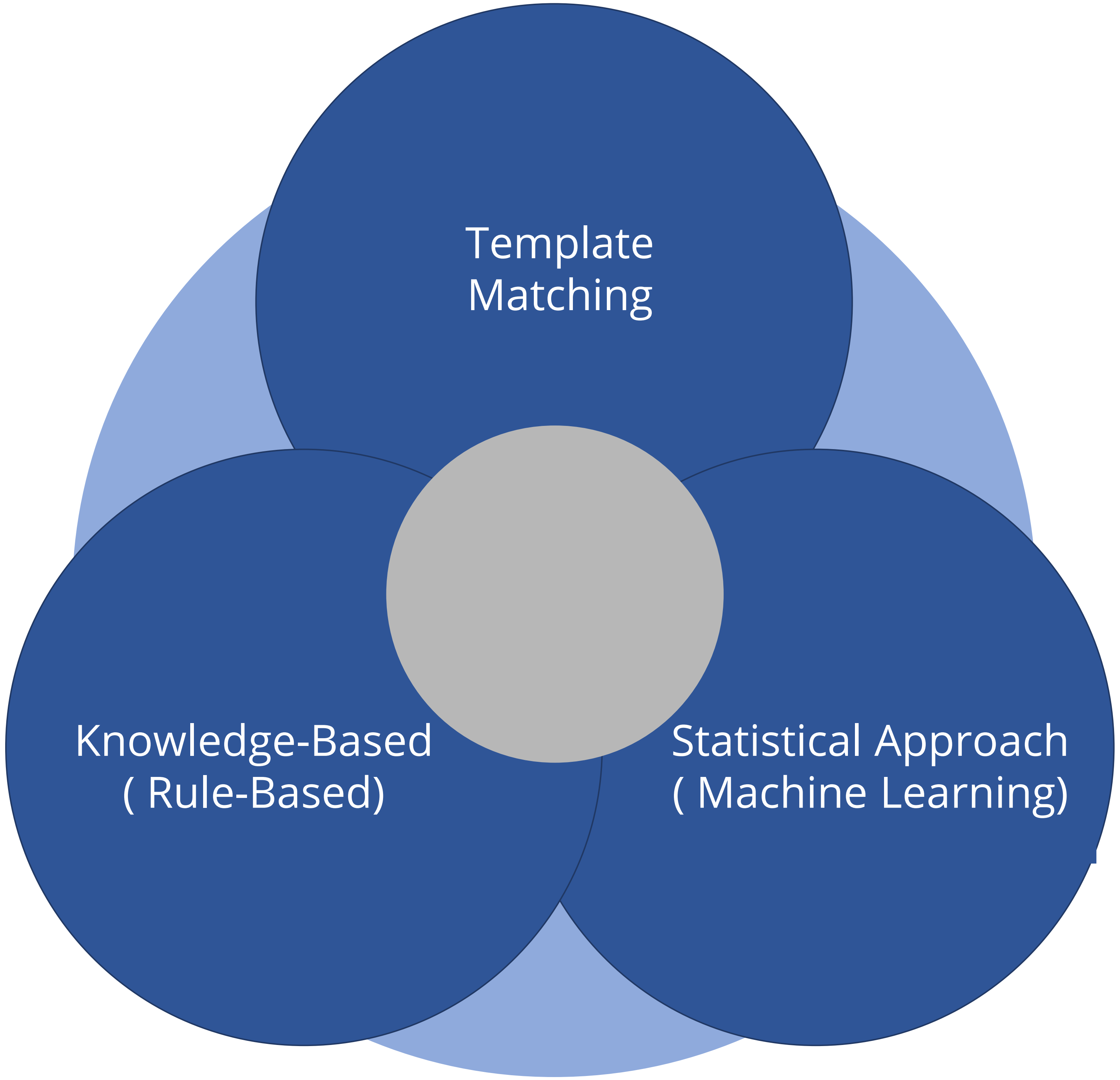
3. Words

S	P	E	E	C	H

Speech Recognition System

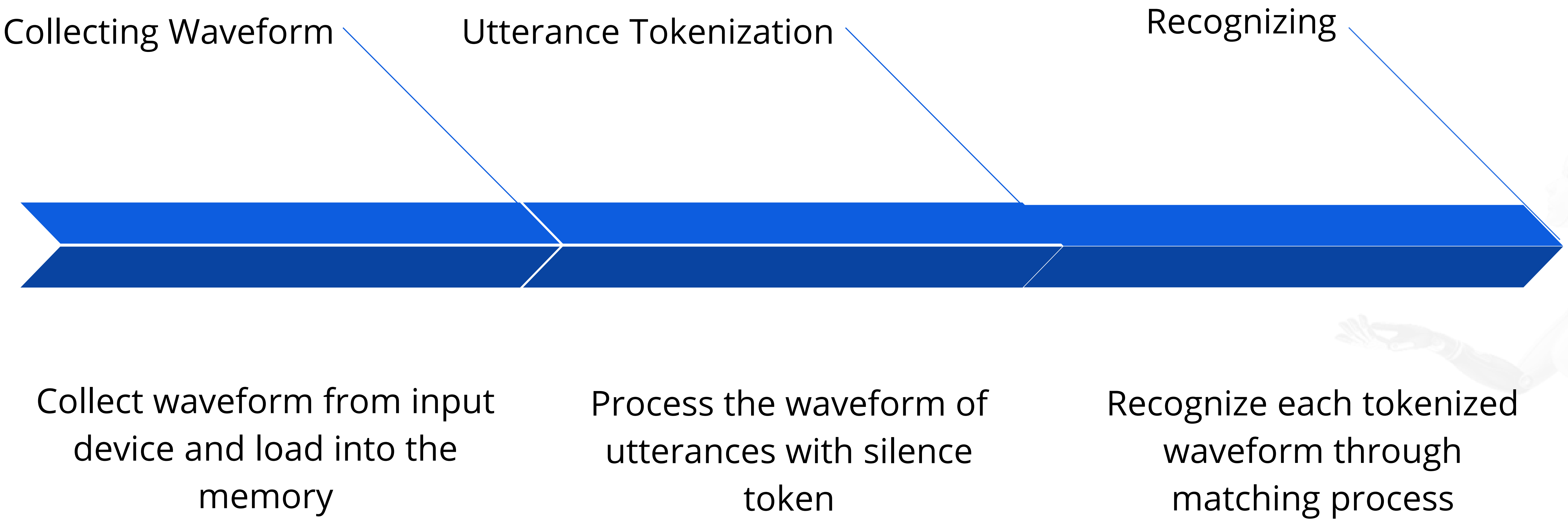


Speech Recognition: Approaches


Template Matching
Knowledge-Based (Rule-Based)
Statistical Approach (Machine Learning)

Reading, Loading, and Processing the Voice Data

Reading, Loading, and Processing the Voice Data



Collecting Waveform

Collect waveform from input device and load into the memory

Utterance Tokenization

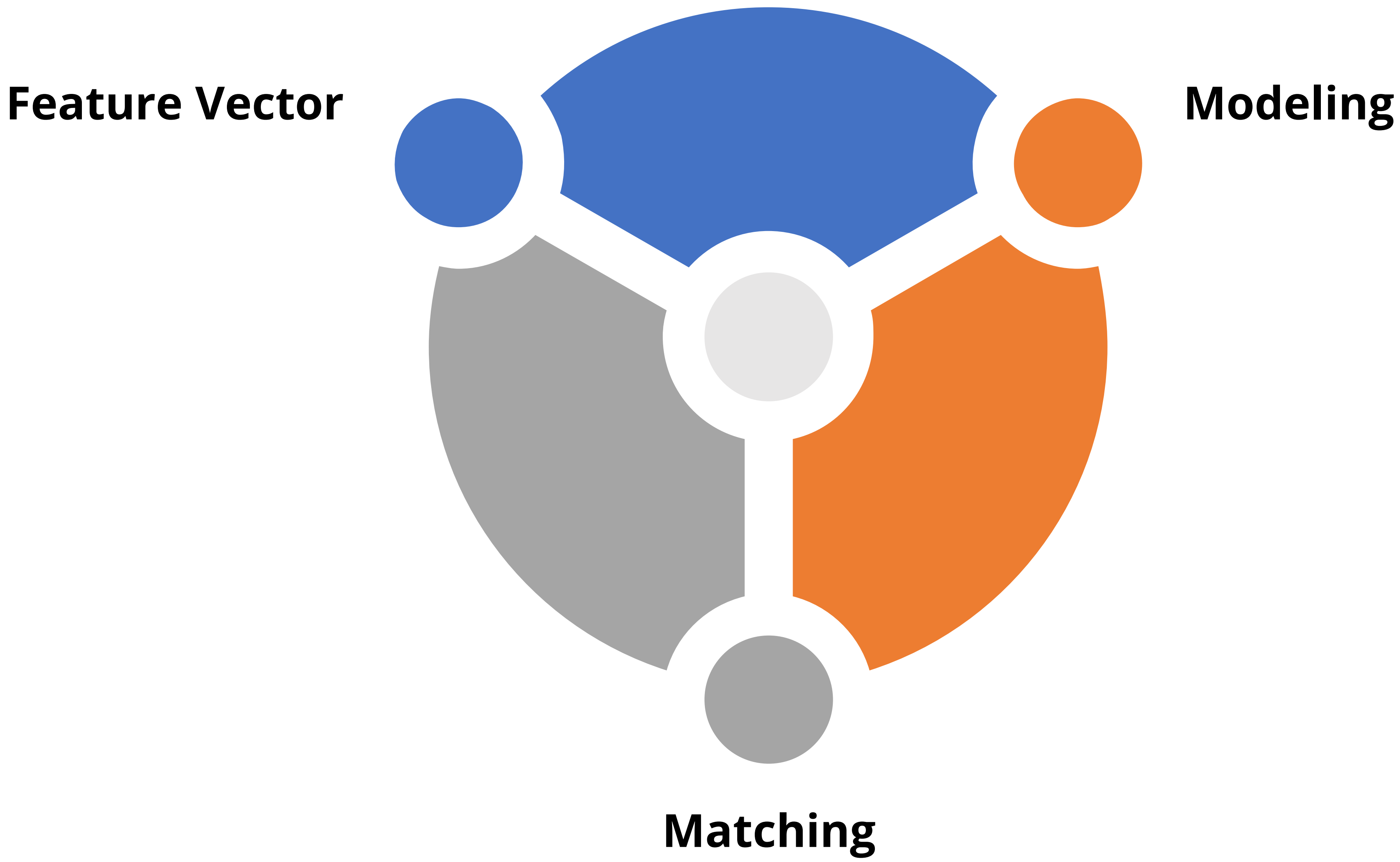
Process the waveform of utterances with silence token

Recognizing

Recognize each tokenized waveform through matching process

Reading, Loading, and Processing the Voice Data

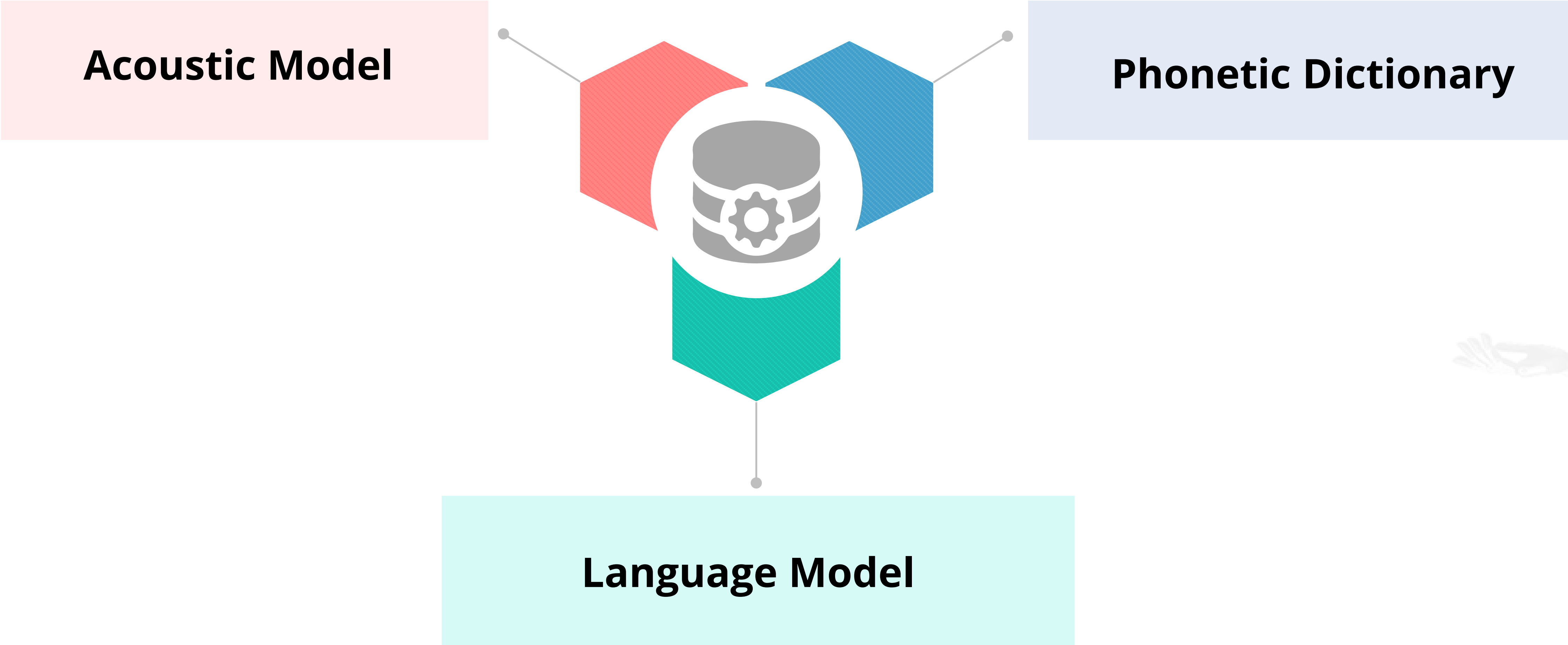
The following are the matching processes:



Creating Speech Model

Creating Speech Model

Based on the speech structure, the following three models are used in speech recognition:



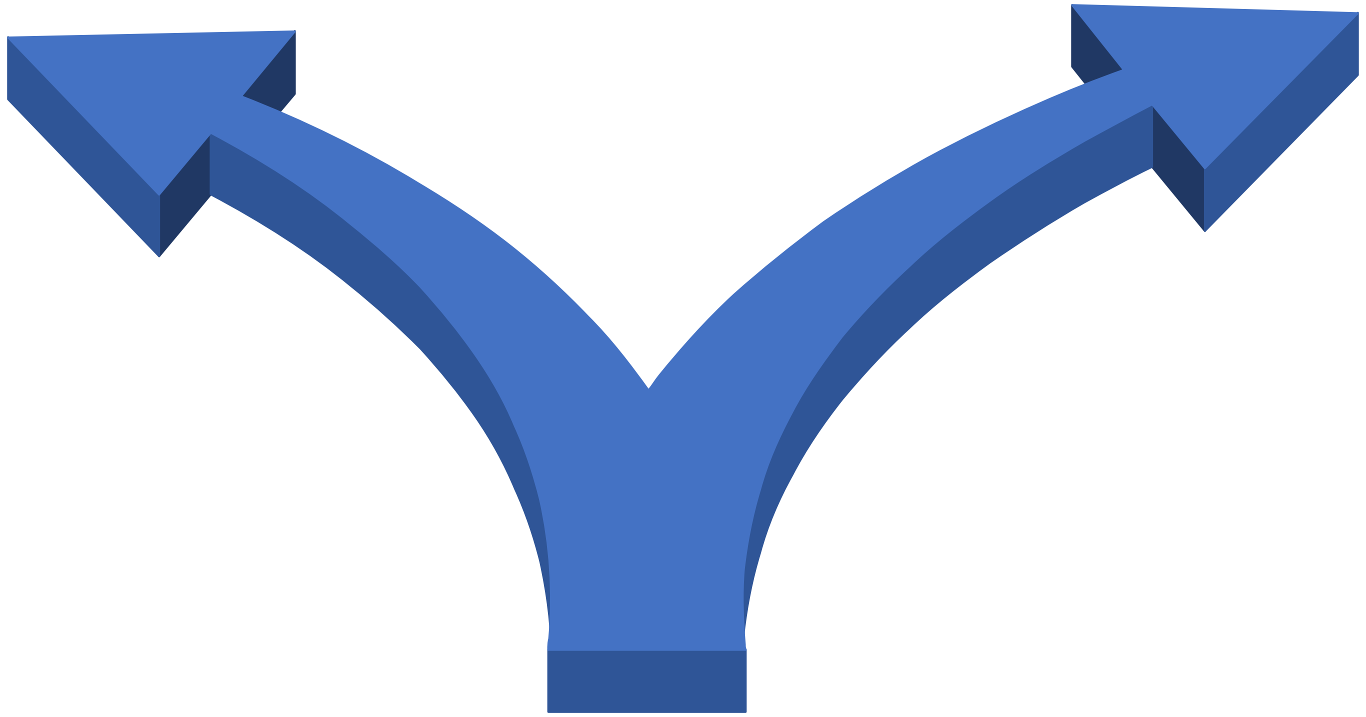
Acoustic Model

Phonetic Dictionary

Language Model

Creating Speech Model: Acoustic Model

Speech model contains acoustic properties for each senone. Hidden Markov Model (HMM) is one of the most common types of acoustic model.



Context-Independent
(Most probable feature vector for each phone)

Context-Dependent
(Made from senones with context)

Creating Speech Model: Acoustic Model

Hidden Markov Model

1 The Hidden Markov Model (HMM) is a statistical Markov model.

2 It allow us to predict a sequence of unknown variables from a set of observed variables. Example: Predicting the marks obtained based on the subject chosen

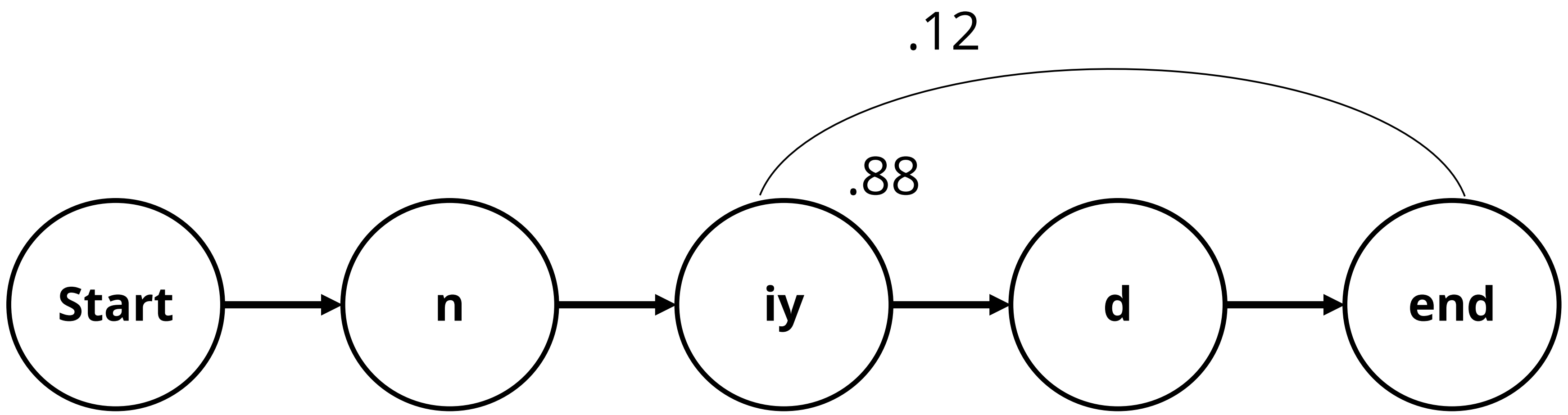
3 The systems are modeled with Markov process in hidden states.

4 The Markov process assumption is that the future is independent of the past, given the present.

Creating Speech Model: Acoustic Model

Hidden Markov Model

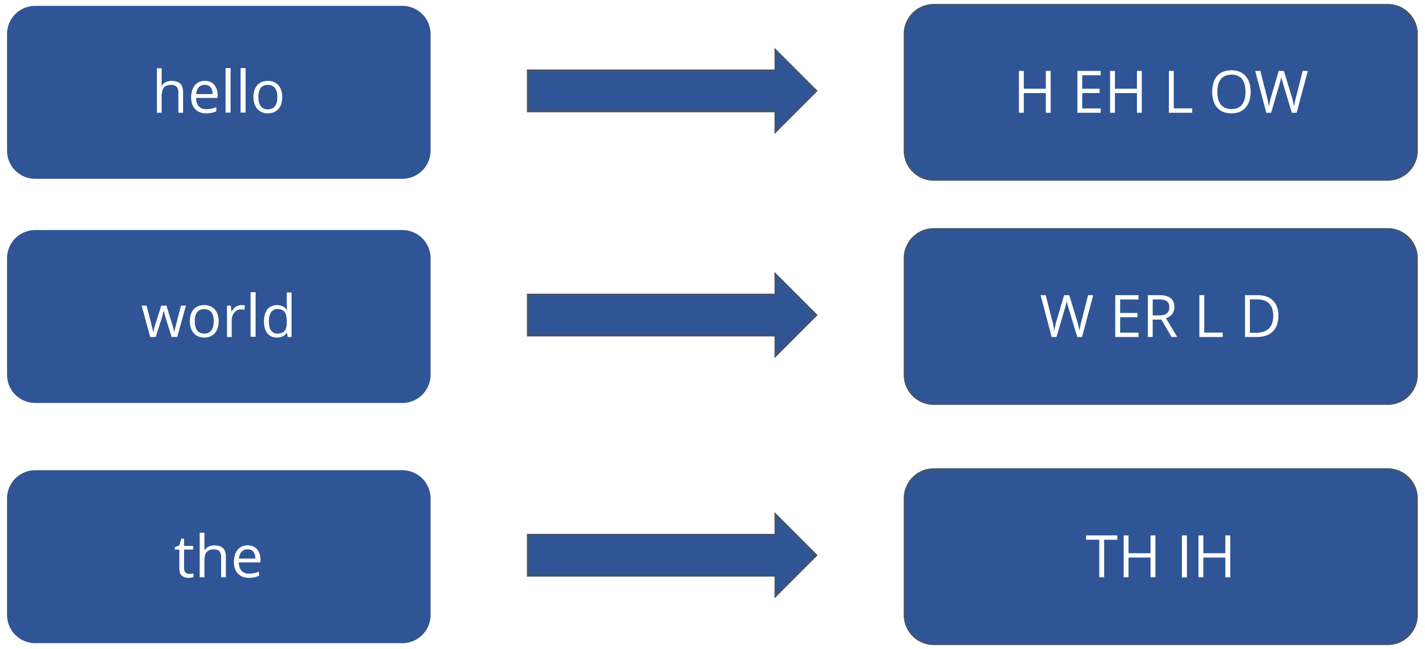
- Markov chains are modeling sequences with discrete states.
- If given a sequence, HMM is used to identify the most likely character to come next or the probability of a given sequence.
- Discrete states are required to work upon Markov chains.



Word model for "need"

Creating Speech Model: Phonetic Dictionary

- It provides the system with mapping of vocabulary words to sequences of phonemes.
- Phonetic Dictionary contains mapping of words to phones.
- In this model, many words and phones are created which are used for matching process and to act like a model.



hello → H EH L OW

world → W ER L D

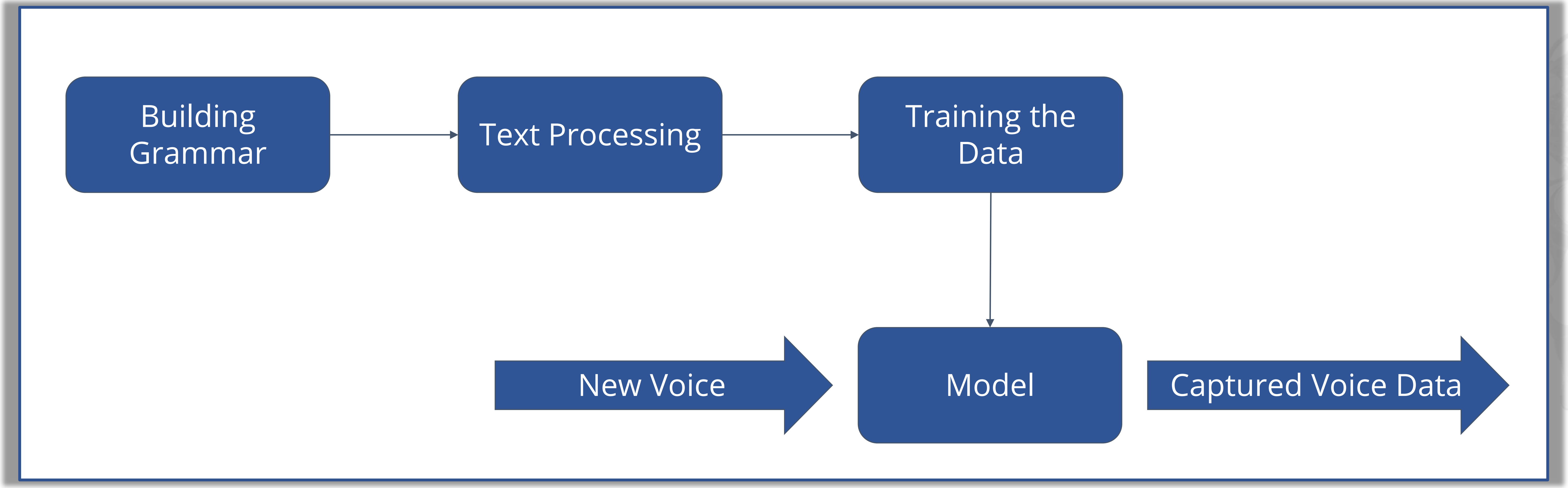
the → TH IH

Creating Speech Model: Language Model

- It defines which word can follow the previously recognized word.
- It restricts the matching process for non-probable words.
- The most common language model is the n-gram language model.

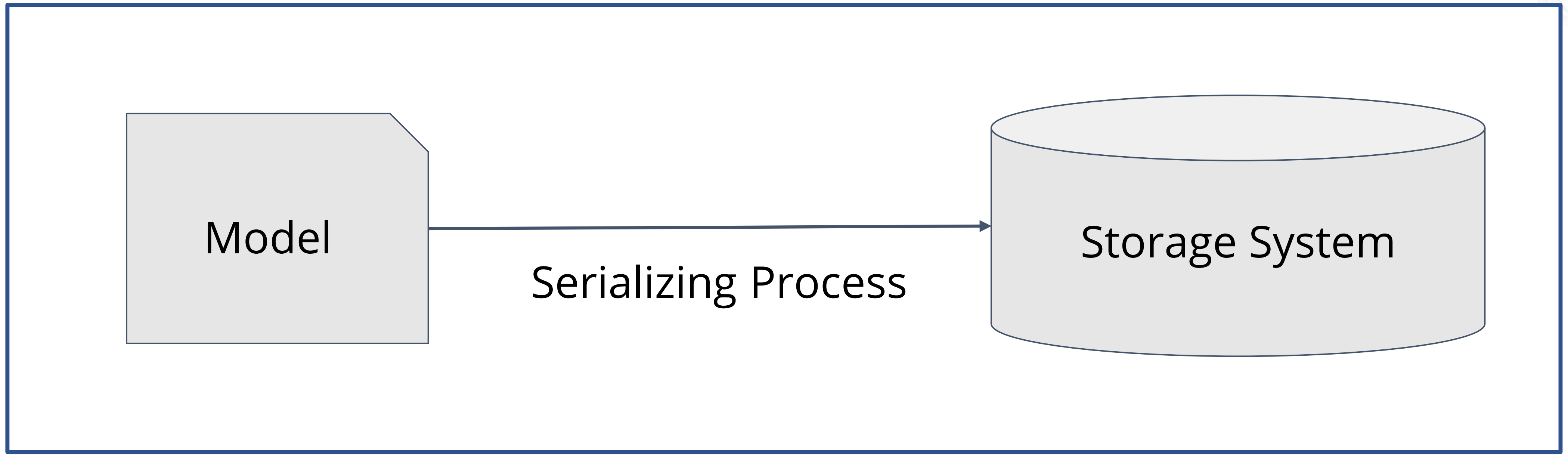
Creating Speech Model: Language Model

Language Model Building Process:



Saving a Model

- Once a model is created, it must be saved in the file system for reutilization.
- The models are saved based on the library used to create them.
- It uses the serialization technique.
- The final weights are saved when used by the neural network.



Use Cases

Use Cases

Speech recognition is implemented in many real-life scenarios.

Siri

Google Assistant

amazon alexa

Use Cases

Developers can use APIs provided by vendors to implement in the software or devices.

CMU Sphinx

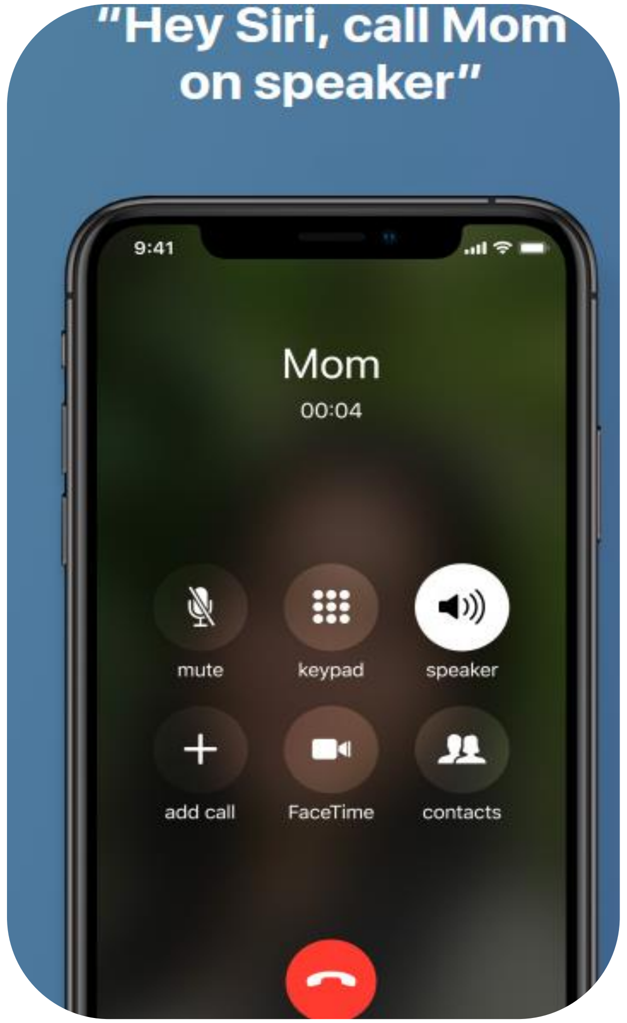
Microsoft Web API

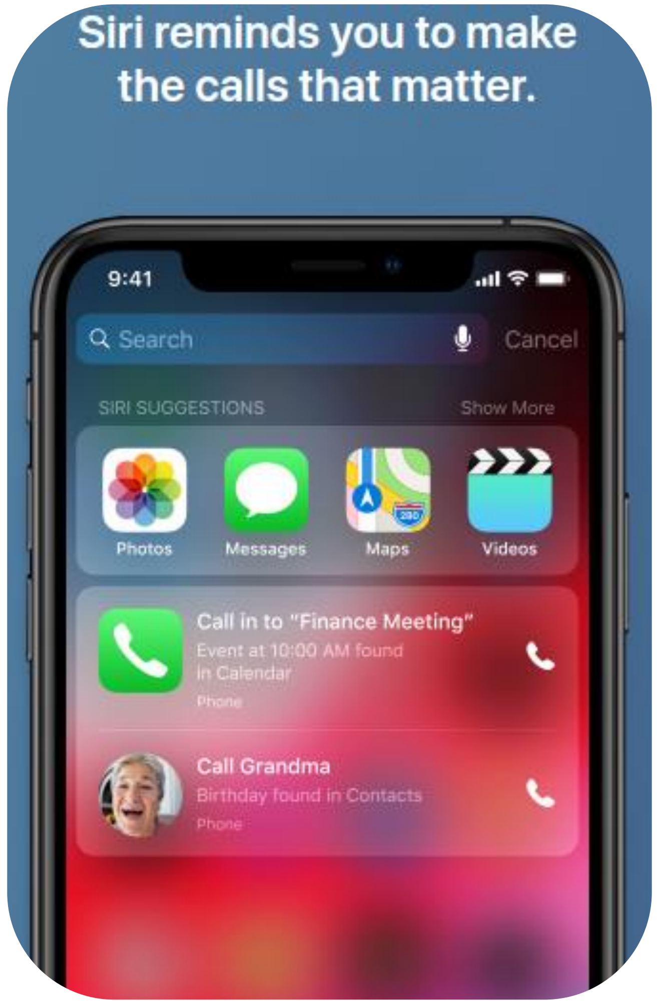
Baidu

Google Web Speech

Use Cases: Apple Siri

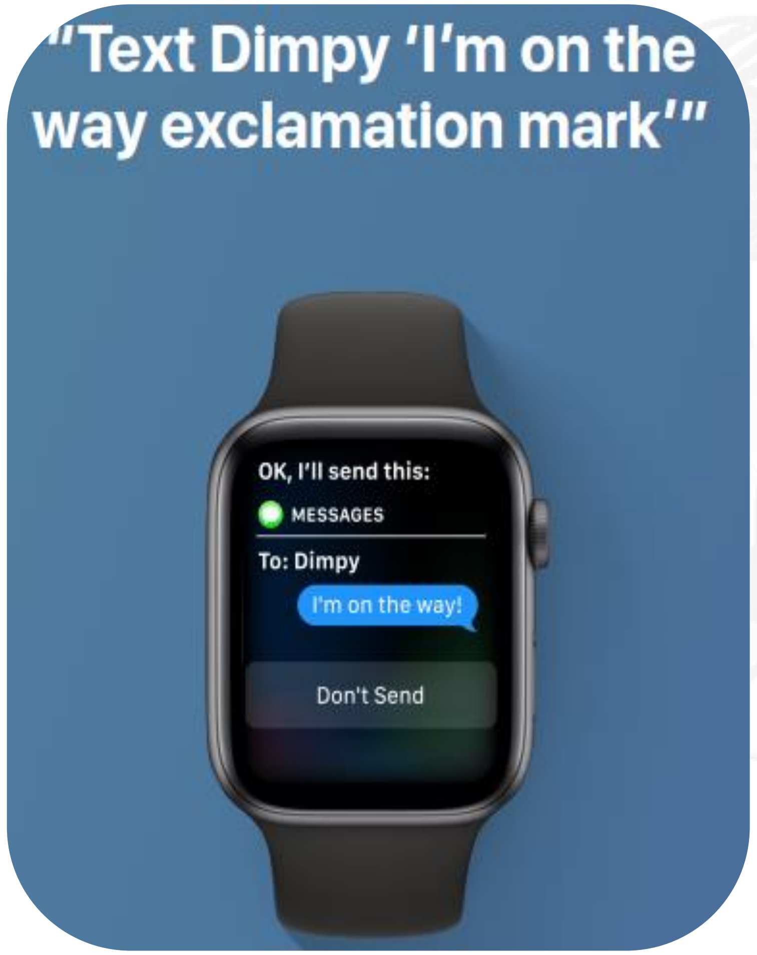
iPhone users can experience Siri, the voice assistant by Apple. It helps you simplify navigating through your iPhone by merely being attentive to your voice and performing the task you would like it to do.





Call in to "Finance Meeting"
Event at 10:00 AM found in Calendar
Phone

Call Grandma
Birthday found in Contacts
Phone

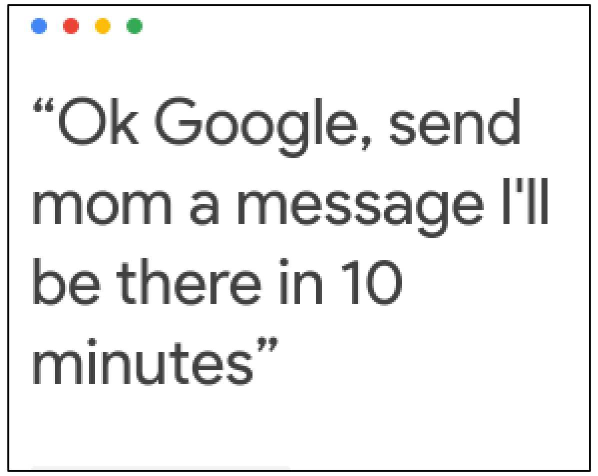


OK, I'll send this:
MESSAGES
To: Dimpy
I'm on the way!

Source

Use Cases: Google Assistant

Like Siri, Google Assistant will act with your phone to do a range of tasks like setting alarms or playing music.



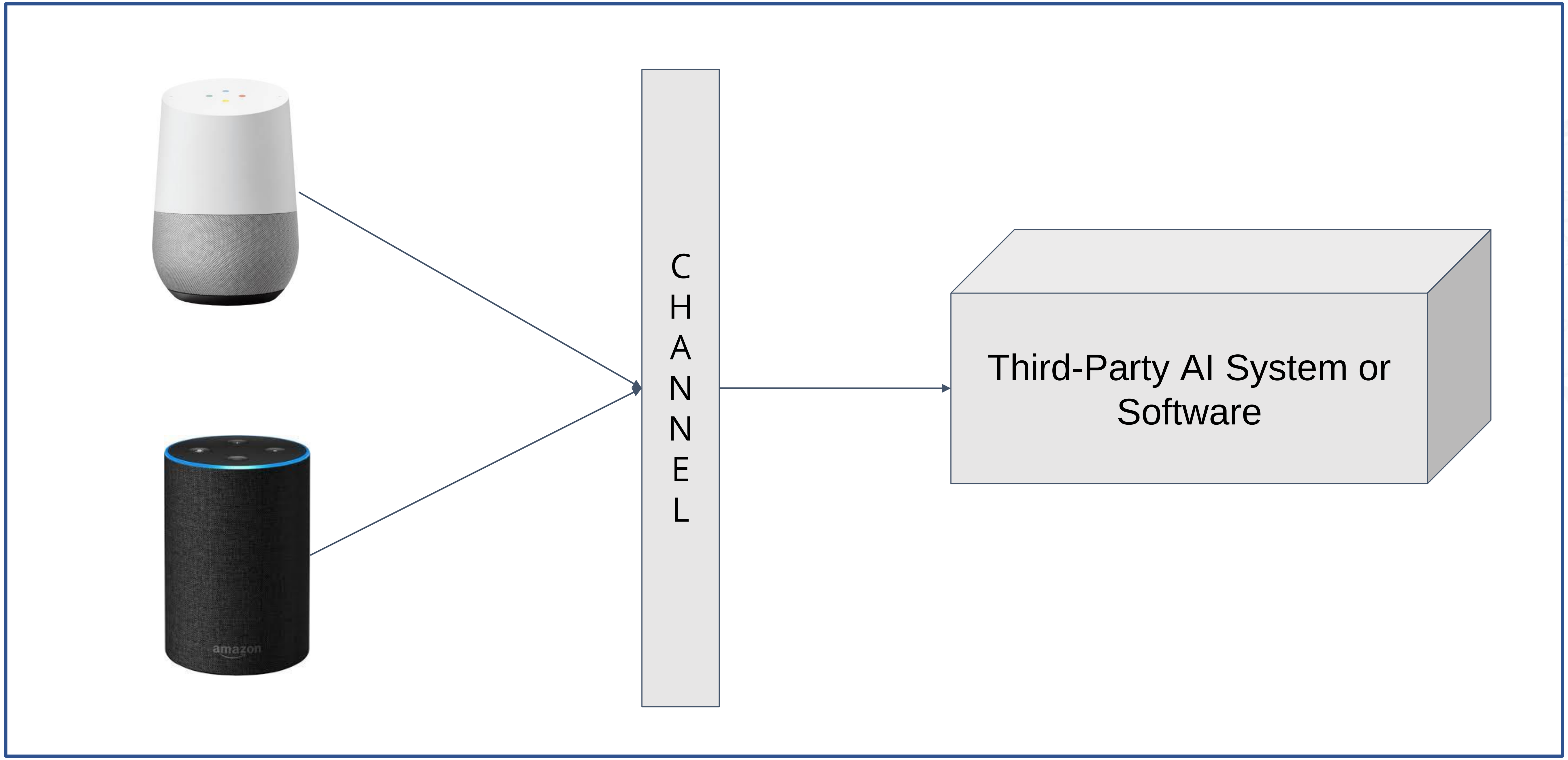
"Ok Google, send mom a message I'll be there in 10 minutes"

Use Cases: Amazon Echo

Amazon Echo is a home control chatbot device that responds to humans, according to what they say. It responds by playing music, movies, and more.

Use Cases: Third-Party Application

Using voice channels for interacting with third-party system



Speech Libraries

Types of Speech Libraries

Pyaudio

pip install Pyaudio

Speech Recognition

pip install SpeechRecognition

Google-Speech API

pip install google-api-python-client

Speech Libraries: Pyaudio

```
""PyAudio Example: Play a wave file."""


import pyaudio
import wave
import sys

CHUNK = 1024
if len(sys.argv) < 2:
    print("Plays a wave file.\n\nUsage: %s filename.wav" % sys.argv[0])
    sys.exit(-1)
wf = wave.open(sys.argv[1], 'rb')
# instantiate PyAudio (1)
p = pyaudio.PyAudio()
# open stream (2)
stream = p.open(format=p.get_format_from_width(wf.getsampwidth()),
channels=wf.getnchannels(),
                rate=wf.getframerate(), output=True)
```

Speech Libraries: Pyaudio

```
# read data
data = wf.readframes(CHUNK)
# play stream (3)
while len(data) > 0:
    stream.write(data)
    data = wf.readframes(CHUNK)
# stop stream (4)
stream.stop_stream()
stream.close()
# close PyAudio (5)
p.terminate()
```

Speech Libraries: Speech Recognition

```
import speech_recognition as sr

# get audio from the microphone

r = sr.Recognizer()

with sr.Microphone() as source:
    print("Speak:")
    audio = r.listen(source)

try:

    print("You said " + r.recognize_google(audio))

except sr.UnknownValueError:

    print("Could not understand audio")

except sr.RequestError as e:

    print("Could not request results; {0}".format(e))
```

Speech Libraries: Google-Speech API

credentials.json

Save the file credentials.json to your working directory

```
from __future__ import print_function
import pickle
import os.path
from googleapiclient.discovery import build
from google_auth_oauthlib.flow import
InstalledAppFlow
from google.auth.transport.requests import
Request
# Delete the file token.pickle. if you need to
modify these files
SCOPES =
['https://www.googleapis.com/auth/documents.re
adonly’]
# Required document id of sample document
DOCUMENT_ID =
'195j9eDD3ccgjQRttHhJPymLJUCOUjs-jmwTrekvdjFE’
```

Speech Libraries: Google-Speech API

```
def main():
creds = None
    if os.path.exists('token.pickle'):
        with open('token.pickle', 'rb')
as token:
            creds = pickle.load(token)
    # let the user log in, when there
aren't any (valid) credentials
available.
    if not creds or not creds.valid:
        if creds and creds.expired and
creds.refresh_token:
            creds.refresh(Request())
```

```
else:
flow =
InstalledAppFlow.from_client_secrets_file
(‘credentials.json', SCOPES)
creds = flow.run_local_server(port=0)
with open('token.pickle', 'wb') as token:
pickle.dump(creds, token)
service = build('docs', 'v1',
credentials=creds)
# Extract the documents contents from the
Docs service.
document =
service.documents().get(documentId=DOCUME
NT_ID).execute()
    print('The title of the document is:
{}'.format(document.get('title’)))
if __name__ == '__main__':
main()
```

Speech to Text

Problem Statement: Speech recognition is an important feature in several applications such as home automation and artificial intelligence. We have an audio file; your task is to translate the audio to text.

Access: Click on the **Practice Labs** tab on the left side panel of the LMS. Copy or note the username and password that is generated. Click on the **Launch Lab** button. On the page that appears, enter the username and password in the respective fields, and click **Login**.

ASSISTED PRACTICE

Speech to Text: Extract Keywords from Audio Reviews

Objective: Convert audio reviews of a product to text and identify which features of the product are being discussed.

Problem Statement:

After the success of Tap Portable Bluetooth Speaker, Amazon is about to launch a new version in the market. To understand which of the features customers liked, Amazon did a focus group discussion with some selected customers. The audio in the discussions has been recorded. The reviews that the panelists gave to Tap is what interests Amazon. As the data scientist, your task is to convert the given audio file to text, assess which features of the Bluetooth speaker are being talked about in the audio reviews. In module 3, we extracted the top 15 features from the reviews. We can use this as our feature list, assess which of these are present in the audio reviews. Also, for future utility and for immediate analysis, you need to make a process or function that captures audio from the microphone, converts to text, analyses, and returns the features discussed in the audio.

LESSON-END PROJECT

Key Takeaways

You are now able to:

- Explain the basic concepts of speech
- Describe how to read, load, and process the data
- Explain how to create speech models
- Identify the types of speech libraries
- Demonstrate the conversion of text to speech for a paragraph

Knowledge Check

Knowledge Check

1

What is speech recognition?

a. Process of understanding the words that are spoken by human beings

b. Understanding the wave

c. Both a and b

d. None of the above

Knowledge Check

1

What is speech recognition?

a. Process of understanding the words that are spoken by human beings

b. Understanding the wave

c. Both a and b

d. None of the above

The correct answer is **c.**

The basic process of speech recognition is understanding of waveform of the words spoken by human beings.

Knowledge Check

2

What kind of signal is used in speech recognition?

a. Electromagnetic signal

b. Electric signal

c. Acoustic signal

d. Radar

Knowledge Check

2

What kind of signal is used in speech recognition?

a. Electromagnetic signal

b. Electric signal

c. Acoustic signal

d. Radar

The correct answer is **c.**

Acoustic signal is used to identify a sequence of words uttered by a speaker.

Knowledge Check

3

Which of these is not an audio file type?

a. WAV

b. MP3

c. OGG

d. MP4

Knowledge Check

3

Which of these is not an audio file type?

a. WAV

b. MP3

c. OGG

d. MP4

The correct answer is **d.**

MP4 is a video file type.

Knowledge Check

4

Computer microphone converts audio signals into _______.

a. Electrical waves

b. Electromagnetic waves

c. Digital signals

d. Analog signals

Knowledge Check

4

Computer microphone converts audio signals into _______.

a. Electrical waves

b. Electromagnetic waves

c. Digital signals

d. Analog signals

The correct answer is **a.**

Computer microphone converts audio signals into electrical waves.

Knowledge Check

5

Which of the following models uses probabilistic approach?

a. Acoustic Model

b. Phonetic Dictionary

c. Language Model

d. Matching Process

Knowledge Check

5

Which of the following models uses probabilistic approach?

a. Acoustic Model

b. Phonetic Dictionary

c. Language Model

d. Matching Process

The correct answer is **a.**

Acoustic model uses probabilistic approach.