Speech recognition

- Speech recognition, also known as automatic speech recognition (ASR), computer speech recognition or speech-to-text, is a capability that enables a program to process human speech into a written format.

- Speech recognition, or speech-to-text, is the ability of a machine or program to identify words spoken aloud and convert them into readable text.

- Many speech recognition applications and devices are available, but the more advanced solutions use AI and machine learning. They integrate grammar, syntax, structure, and composition of audio and voice signals to understand and process human speech.

- **Speech recognition** is used to identify words in spoken language.

How does speech recognition work?

- Speech recognition systems use computer [algorithms](#) to process and interpret spoken words and convert them into text. A software program turns the sound a microphone records into written language that computers and humans can understand, following these four steps:

- analyze the audio;

- break it into parts;

- digitize it into a computer-readable format; and

- use an algorithm to match it to the most suitable text representation.

- Speech recognition software must adapt to the highly variable and context-specific nature of human speech.
- The software algorithms that process and organize audio into text are trained on different speech patterns, speaking styles, languages, dialects, accents and phrasings. The software also separates spoken audio from background noise that often accompanies the signal.
- To meet these requirements, speech recognition systems use two types of models:
- **Acoustic models.** These represent the relationship between linguistic units of speech and audio signals.
- **Language models.** Here, sounds are matched with word sequences to distinguish between words that sound similar.

Applications of speech recognition

- **Education.** Speech recognition software is used in language instruction. The software hears the user's speech and offers help with pronunciation.

- **Customer service.** Automated voice assistants listen to customer queries and provides helpful resources.

- **Healthcare applications.** Doctors can use speech recognition software to transcribe notes in real time into healthcare records.

- **Disability assistance.** Speech recognition software can translate spoken words into text using closed captions to enable a person with hearing loss to understand what others are saying. Speech recognition can also enable those with limited use of their hands to work with computers, using voice commands instead of typing.

- **Court reporting.** Software can be used to transcribe courtroom proceedings, precluding the need for human transcribers.

- **Emotion recognition.** This technology can analyze certain vocal characteristics to determine what emotion the speaker is feeling. Paired with sentiment analysis, this can reveal how someone feels about a product or service.

- **Hands-free communication.** Drivers use voice control for hands-free communication, controlling phones, radios and global positioning systems, for instance.

- **Mobile devices.** Smartphones use voice commands for call routing, speech-to-text processing, voice dialing and voice search. Users can respond to a text without looking at their devices. On Apple iPhones, speech recognition powers the keyboard and Siri, the virtual assistant. Functionality is available in secondary languages, too. Speech recognition can also be found in word processing applications like Microsoft Word, where users can dictate words to be turned into text.

The features of speech recognition systems?

- Good speech recognition programs let users customize them to their needs. The features that enable this include:

- **Language weighting.** This feature tells the algorithm to give special attention to certain words, such as those spoken frequently or that are unique to the conversation or subject. For example, the software can be trained to listen for specific product references.

- **Acoustic training.** The software tunes out ambient noise that pollutes spoken audio. Software programs with acoustic training can distinguish speaking style, pace and volume amid the din of many people speaking in an office.

- **Speaker labeling.** This capability enables a program to label individual participants and identify their specific contributions to a conversation.

- **Profanity filtering.** Here, the software filters out undesirable words and language.

The advantages of speech recognition?

- There are several advantages to using speech recognition software, including the following:
- **Machine-to-human communication.** The technology enables electronic devices to communicate with humans in natural language or conversational speech.
- **Readily accessible.** This software is frequently installed in computers and mobile devices, making it accessible.
- **Easy to use.** Well-designed software is straightforward to operate and often runs in the background.
- **Continuous, automatic improvement.** Speech recognition systems that incorporate AI become more effective and easier to use over time. As systems complete speech recognition tasks, they generate more data about human speech and get better at what they do.

The disadvantages of speech recognition?

- While convenient, speech recognition technology still has a few issues to work through. Limitations include:
- **Inconsistent performance.** The systems may be unable to capture words accurately because of variations in pronunciation, lack of support for some languages and inability to sort through background noise. Ambient noise can be especially challenging. Acoustic training can help filter it out, but these programs aren't perfect. Sometimes it's impossible to isolate the human voice.
- **Speed.** Some speech recognition programs take time to deploy and master. The speech processing may feel relatively slow.
- **Source file issues.** Speech recognition success depends on the recording equipment used, not just the software.

Speech Synthesis

- Speech synthesis, in essence, is the artificial simulation of human speech by a computer or any advanced software. It's more commonly also called [text to speech](#).

- The counterpart of the voice recognition, speech synthesis is mostly used for translating text information into audio information and in applications such as voice-enabled services and mobile applications.

- Apart from this, it is also used in assistive technology for helping vision-impaired individuals in reading text content.

- It is a three-step process that involves:

  - Contextual assimilation of the typed text

  - Mapping the text to its corresponding unit of sound

  - Generating the mapped sound in the textual sequence by using synthetic voices or recorded human voices

Text-to-Speech

- Text-to-Speech allows developers to create natural-sounding, synthetic human speech as playable audio. You can use the audio data files you create using Text-to-Speech to power your applications or augment media like videos or audio recordings (in compliance with the Google Cloud Platform Terms of Service including compliance with all applicable law).

- Convert text into natural-sounding speech using an API powered by the best of Google's AI technologies.

- Improve customer interactions with intelligent, lifelike responses

- Engage users with voice user interface in your devices and applications

- Personalize your communication based on user preference of voice and language

- Text-to-Speech is ideal for any application that plays audio of human speech to users. It allows you to convert arbitrary strings, words, and sentences into the sound of a person speaking the same things.

Refer: https://cloud.google.com/text-to-speech/docs/basics

Sentiment Analysis

- Sentiment analysis is the process of analyzing digital text to determine if the emotional tone of the message is positive, negative, or neutral.

- Today, companies have large volumes of text data like emails, customer support chat transcripts, social media comments, and reviews.

- Sentiment analysis tools can scan this text to automatically determine the author's attitude towards a topic.

- Companies use the insights from sentiment analysis to improve customer service and increase brand reputation.

- Sentiment analysis focuses on the polarity of a text (*positive, negative, neutral*) but it also goes beyond polarity to detect specific feelings and emotions (*angry, happy, sad*, etc), urgency (*urgent, not urgent*) and even intentions (*interested v. not interested*).

- If polarity precision is important to your business, you might consider expanding your polarity categories to include different levels of positive and negative:

- Very positive

- Positive

- Neutral

- Negative

- Very negative

- Refer: https://www.geeksforgeeks.org/what-is-sentiment-analysis/

Segmentation and recognition

- Segmentation is a process of dividing an object/image into number of segments based on the requirements of the application.

- Segmentation of text from image is involved in applications such as document retrieving, object identification, detection of vehicle licence plate etc.