Speech recognition

From Wikipedia, the free encyclopedia

In computer science and electrical engineering, speech recognition (SR) is the translation of spoken words into text. It is also known as "automatic speech recognition" (ASR), "computer speech recognition", or just "speech to text" (STT).

Some SR systems use "training" (also called "enrolment") where an individual speaker reads text or isolated vocabulary into the system. The system analyzes the person's specific voice and uses it to fine-tune the recognition of that person's speech, resulting in increased accuracy. Systems that do not use training are called "speaker independent"[1] systems. Systems that use training are called "speaker dependent".

Speech recognition applications include voice user interfaces such as voice dialling (e.g. "Call home"), call routing (e.g. "I would like to make a collect call"), domotic appliance control, search (e.g. find a podcast where particular words were spoken), simple data entry (e.g., entering a credit card number), preparation of structured documents (e.g. a radiology report), speech-to-text processing (e.g., word processors or emails), and aircraft (usually termed Direct Voice Input).

The term voice recognition[2][3][4] or speaker identification[5][6] refers to identifying the speaker, rather than what they are saying. Recognizing the speaker can simplify the task of translating speech in systems that have been trained on a specific person's voice or it can be used to authenticate or verify the identity of a speaker as part of a security process.

From the technology perspective, speech recognition has a long history with several waves of major innovations. Most recently, the field has benefited from advances in deep learning and big data. The advances are evidenced not only by the surge of academic papers published in the field, but more importantly by the world-wide industry adoption of a variety of deep learning methods in designing and deploying speech recognition systems. These speech industry players include Microsoft, Google, IBM, Baidu (China), Apple, Amazon, Nuance, IflyTek (China), many of which have publicized the core technology in their speech recognition systems being based on deep learning.

Applications

In-car systems

Typically a manual control input, for example by means of a finger control on the steering-wheel, enables the speech recognition system and this is signalled to the driver by an audio prompt. Following the audio prompt, the system has a "listening window" during which it may accept a speech input for recognition.

Simple voice commands may be used to initiate phone calls, select radio stations or play music from a compatible smartphone, MP3 player or music-loaded flash drive. Voice recognition capabilities vary between car make and model. Some of the most recent car models offer natural-language speech recognition in place of a fixed set of commands. allowing the driver to use full sentences and common phrases. With such systems there is, therefore, no need for the user to memorize a set of fixed command words.

Usage in education and daily life

For language learning, speech recognition can be useful for learning a second language. It can teach proper pronunciation, in addition to helping a person develop fluency with their speaking skills.[61]

Students who are blind (see Blindness and education) or have very low vision can benefit from using the technology to convey words and then hear the computer recite them, as well as use a computer by commanding with their voice, instead of having to look at the screen and keyboard.[62]

Students who are physically disabled or suffer from Repetitive strain injury/other injuries to the upper extremities can be relieved from having to worry about handwriting, typing, or working with scribe on school assignments by using speech-to-text programs. They can also utilize speech recognition technology to freely enjoy searching the Internet or using a computer at home without having to physically operate a mouse and keyboard.[62]

Speech recognition can allow students with learning disabilities to become better writers. By saying the words aloud, they can increase the fluidity of their writing, and be alleviated of concerns regarding spelling, punctuation, and other mechanics of writing.[63] Also, see Learning disability.

Voice Recognition Software's use, in conjunction with a digital audio recorder, a personal computer and Microsoft Word has proven to be positive for restoring damaged short-term-memory capacity, in stroke and craniotomy individuals.

People with disabilities

People with disabilities can benefit from speech recognition programs. For individuals that are Deaf or Hard of Hearing, speech recognition software is used to automatically generate a closed-captioning of conversations such as discussions in conference rooms, classroom lectures, and/or religious services. [64]

Speech recognition is also very useful for people who have difficulty using their hands, ranging from mild repetitive stress injuries to involved disabilities that preclude using conventional computer input devices. In fact, people who used the keyboard a lot and developed RSI became an urgent early market for speech recognition.[65][66] Speech recognition is used in deaf telephony, such as voicemail to text, relay services, and captioned telephone.