What's going on with voice recognition?

S. Arburola

Abstract

Voice Recognition is so unknown from how it works to how it will be working in future. A lot of technologies are using it nowadays voice recognition algorithms for facility on interactions and all user have just the see of the iceberg. On this paper is write a view of what is internet informs to users about voice recognition.

Categories and Subject Descriptors (according to ACM CCS):   [Artificial intelligence]: Natural language processing—Speech recognition

1. Introduction

Voice recognition is completely invisible to the developer as to the user, it is one of the forms of communication with machines that is positioning itself to other more traditional forms of interaction such as digital buttons that have already displaced physical buttons, even the voice is displacing tactile functionality in some cases.
Voice technologies have recently been implemented from mobile devices to vehicles, incorporating voice command systems, so that it is not necessary to digitally dial someone's phone number and even look for it in the contacts application or in case of cars take off the hands of the steering wheel to answer calls and have a route by means of GPS.

2. Where speech recognition is focused?

Recognition of voice has a great challenge, not only is responding a command, in this already has enough experience in the industry of technology where all so algorithms that respond to previously studied and categorized environments, this is where the participation of artificial intelligence it is noted, since to respond optimally to a voice command one must know the context and even the slang and jargon of the user, in addition to noticing the linguistic variations referring to geography and social status.
Understanding natural language, that is the center of all voice recognition, since it is not only recognizing faces, objects and settings but understanding the context, and emphasizing comprehension and not only knowledge of the user's environment derives a series of computational effort similar to the natural when two unknown people interact.

In this, Amazon has been based on its voice recognition projects for household task facilitators, such as the Eco device that uses the Alexa software, which has positioned itself in a great way and has maintained the satisfaction in the clients of Amazon.

3. The challenges

When asking a phone number we can say: what is your phone? Can I have your phone number? or can you pass me your phone number? in English and Spanish it is extended to ¿cuál es tu móvil? ¿me das tu número? o ¿ me pasas tu número de teléfono? , among many other formulas. The goal of LEX; The recognition software used for Alexa by Amazon is that by classifying keywords, Alexa can understand what the user wants to say in whatever form.

A solution developed for this is the one presented by AWS, which is Amazon Polly, a new technology to understand the written text in a natural way when analyzing the context. For example, if you write, when do the Athletics play? Polly understands that it is a baseball team and offers the results by voice. It provides access to a suite of resources to quickly and easily build Alexa-enabled products, including APIs, hardware and software development tools, and documentation.
Alexa is always getting smarter with new capabilities and services through machine learning, regular API updates, feature launches, and custom skills. The AVS API is a programming language agnostic service that makes it easy to integrate Alexa into your devices, services, and applications.

3.1. Vocal Footprint

Several companies and universities have developed studies on the unique identification of people through a technology called Vocal Footprint.

Its field of development has been concentrated in a dynamic capture while the person speaks, although a phrase can also be established as a password and these precise words are those that are compared for authentication.

Several factors affect it which give it rough indexes of difficulty; we can take into account the timbre, the sharpness, the age, whether it is masculine or feminine, the channel used to transmit it, distortion of the environment and even if it is a recording or not.

3.2. Classification of sounds

The basic units of sound are called phonemes and there are two main types: vowels and consonants. The first are produced when the vocal tract is excited by pulses of air caused by the vocal cords. The shape of the vocal tract determines the resonant frequencies of the tract, called formants. The vowels have three formants between frequencies from 200 Hz to 3 kHz, which vary from person to person. In the production of consonants, the vocal cords are relaxed, but there are exceptions.

Digital filtering can improve audio signals because it separates the signal frequencies from the noise frequencies, right here is where the neural networks intervene as processing systems, hardware and software, since being able to learn from the experience from signals coming from abroad, within a framework of parallel and distributed computing they provide the tools so that voice recognition becomes possible.

The digital filtering, the processing provided by the neural networks, the applied artificial intelligence and the storage of the variety of individuals for study makes possible the collection, signal processing, pattern recognition and start-up of the algorithm, where they are detected from commands of direction until the pause of a final point. With such affinity to make an ideal assistant or interactive system that provides the comity of a human interlocutor with the knowledge that lies on the network, with calculation capacity and information model to facilitate, prevent and provide not only for comfort but also for improve the daily activities of sectors so much personal, social but even scientific.

4. Relevant current applications

One of the most used applications for voice commands is the attendees on mobile devices, in the last period Alexa, Google Home, Ok Google, Siri and Cortana has been in everyone's lips. In addition to these technologies, others with the service of dictation have taken positioning as they are Windows Recognition, Dragon Naturally Speaking, Braina and VoxCommando.

Basically it has tried to minimize the writing action, making faster the production of editorial texts and publications thanks to the speed of writing with which is given to dictate a software that is previously configured to recognize a voice.

Medicine has also played a leading role in including voice commands in assertive technologies for the disabled and limited mobility patients during their recovery, thanks to the support of non-linear techniques.

Rosetta Stone proxies a frequency of 100 times per second of the database engine for comparison with millions of examples of voices of native speakers, with the aim of accurately assessing pronunciation, providing an authentic service to learn a language.This expands the database with each dialect and way of speaking of each user and at the same time collaborates with the globalization of services for interpersonal relations in the different areas.

5. Conclusions

Artificial intelligence is one of the great interpreters in today's technology, much of what is currently being developed requires the processing of BigData and who is more efficient to do it well than an assistance system capable of understanding what the user needs. and not exactly what you request.

Many applications have tried to answer the need for a voice insulator, and there are few that successfully satisfy users. An important factor is the security of information that moves when the user is interacting with a voice assistant, but the most alarming is the misinformation towards the user, since those within the scientific and development community prevent these cases. How much the user knows how his assistant recognizes the voice to unlock an electronic device and why only with the voice of the owner serves and not with the voice of a relative who may have a voice footprint is quite limited.

Even for developers it is a new world, and the gap to it has been reduced but this needs more effort in truthful research.

In the end, what it sounds is that an algorithm learns how it speaks, how it says each word, how it is expressed and determines a context as accurate as possible according to their customs, for a developer what happens is very different since it must implement all this in a world where voice recognition has been used for years but commercial selfishness reduces the acquisition of previous discoveries and where the latest in voice recognition is focused on intelligent assistance to users of which very few they

will develop technological solutions to optimally use the technology with which they have acquired knowledge.

6. References

(2018). Support.microsoft.com. Retrieved 24 February 2018, from https://support.microsoft.com/es-cr/help/12427/windows-speech-recognition-commands

A.Llorca, A. (2018). Cuatro herramientas de reconocimiento de voz que permiten prescindir del teclado (entre otros) y agilizar las tareas. Genbeta.com. Retrieved 22 February 2018, from https://www.genbeta.com/herramientas/cuatro-herramientas-de-reconocimiento-de-voz-que-permiten-prescindir-del-teclado-entre-otros-y-agilizar-las-tareas

Alexa Voice Service. (2018). Developer.amazon.com. Retrieved 24 February 2018, from https://developer.amazon.com/alexa-voice-service

Alexa Voice Service Overview (v20160207) | Alexa Voice Service. (2018). Developer.amazon.com. Retrieved 25 February 2018, from https://developer.amazon.com/docs/alexa-voice-service/api-overview.html

Como funciona un sistema de reconocimiento de voz. (2018). eldiario.es. Retrieved 25 February 2018, from http://www.eldiario.es/turing/reconocimiento-voz-biometria_0_201230680.html

La tecnologia de reconocimiento de voz - Philips. (2018). Dictation.philips.com. Retrieved 23 February 2018, from https://www.dictation.philips.com/co/acerca/notas-de-prensa/press/la-tecnologia-de-reconocimiento-de-voz/

Las 5 Mejores aplicaciones de reconocimiento de voz. (2018). Profesional Review. Retrieved 24 February 2018, from https://www.profesionalreview.com/2016/08/17/5-mejores-aplicaciones-reconocimiento-voz/

Reconocimiento de Voz. (2018). Rosettastone.es. Retrieved 24 February 2018, from http://www.rosettastone.es/reconocimiento-de-voz

Reconocimiento de voz, texto e imagenes, lo nuevo de Amazon. (2018). Cromo. Retrieved 24 February 2018, from https://www.cromo.com.uy/reconocimiento-voz-texto-e-imagenes-lo-nuevo-amazon-n1006115

Reconocimiento de voz, texto e imagenes, lo nuevo de Amazon. (2018). Cromo. Retrieved 25 February 2018, from https://www.cromo.com.uy/reconocimiento-voz-texto-e-imagenes-lo-nuevo-amazon-n1006115