Natasha Chan

Conversing with Siri

Everyone knows who Siri is, but how exactly does it understand and get the information needed?

Conversing with Siri

Natasha Chan

How the voice of Siri is built

Gathering

Pruning

Unused Units
Used less than 10 000 times
Used more than 1 million times
Used less than 10 times

Voice recognition is something that is implemented into Apple products to provide a lifestyle that is easier for those who use their products. Apple has named their voice recognition system Siri, who is designed as a digital assistant to help the consumers life.

Siri uses two primary technologies: Speech Recognition and Natural Language Processing. This post will be focussing on Speech Recognition. Siri stands for, "Speech Interpretation and Recognition Interface" and is Apple's digital voice assistant. It can be activated with the simple phrase, "Hey Siri." Or, "Ok Siri." For the consumer, Siri is quite easy to use, but the logistics behind the makings of it are quite complex.
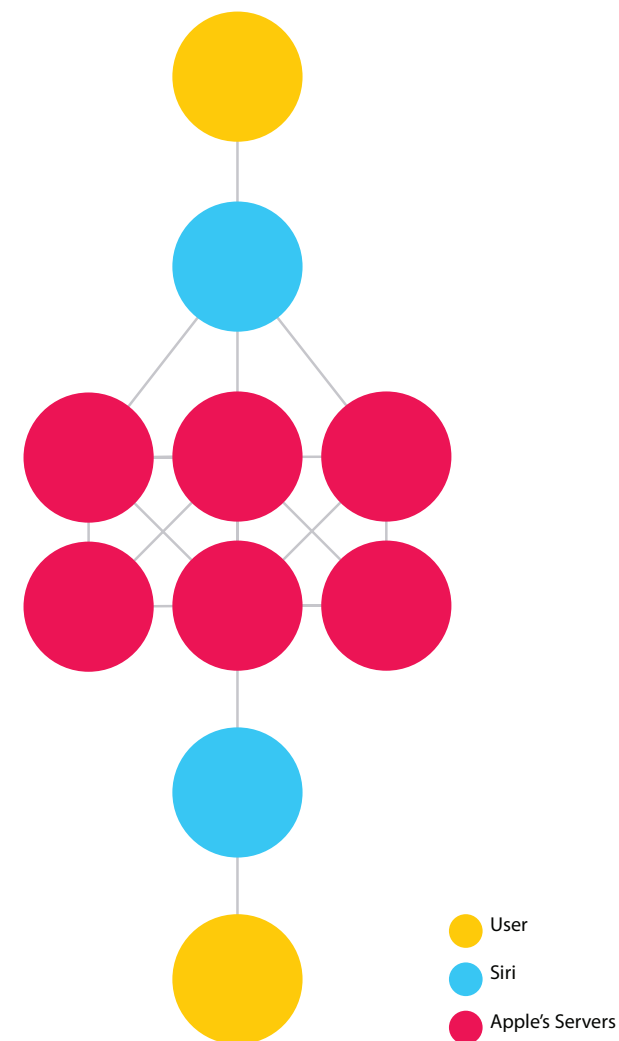
So how exactly is the voice built? The voices chosen for Siri must be able to be heard with a Siri persona. There are male and female voices along with each of the accents; the following accents are available

in English; American, Australian, British, Indian, Irish and South African. There is 10-20 hours of audio recorded at 48kHz, 16bit size. From the recordings, a language model is built up using a script.

From here on, the audio needs to be cleaned up and reduced in size. The audio is put into a pattern using a, "speaker independent deep neural network acoustic model". Each phone has features added into the database where key words are stored onto the phone which can be processed into standard codec to reduce the size of the audio files, or "units". To reduce the size of the audio data, the unused units must be removed. This process is called pruning of which there are three steps. The first step is to gather all the words of which there are over 200 million words. Statistics show that 41% of the units are never used, 8.3% of the units are used less than 10,000 times and 8.8% of the units are used more than 1 million times. All units that are used less than ten times are pruned down to under 500 bytes of units of space.

After the voice has been built, the functionality of this data needs to be used. To do this, Siri needs to be activated by the user. This can be done by holding down the "home" button on models iPhone 4s to iPhone 8, holding down the on/off button on models iPhone X and above or subsequently saying, "Hey Siri." Siri then records the user's voice and converts the sound into data files which are then sent to the apple servers. This is why an internet connection is required to use Siri. Apple's servers then send a signal back to Siri and Siri then is able to reply with direct commands for certain phrases such as, "The weather in Sydney is 26oC." If a direct command cannot be found then Siri will give the standard response, "I found this on the web." Siri picks out key words from the phrase said by the user, for example, "Hey Siri, set a reminder for 10am tomorrow to buy eggs."

Siri itself is a simple concept however the system behind it is very complex. As technology evolves, so will too voice recognition improve. The ability for Siri to understand and converse with the user is revolutionary and will be integrated more into the consumers life.

User

Siri

Apple's Servers

References:

Capes, T. Coles, P. Conkie, A. Golipour, L. Hadijtarkhani, A. Hu, Q... Zhang, H. (August 20-13, 2017). Siri On-Device Deep Learning-Guided Unit Selection Text-to-Speech System. Retrieved from: https://isca-speech.org/archive/Interspeech_2017/pdfs/1798.PDF

Tomashenko, N. (May 22, 2018). Speaker adaptation of deep neural network acoustic models using Gaussian mixture model framework in automatic speech recognition systems, Neural and Evolutionary Computing. Chapter 3. Retrieved from: https://tel.archives-ouvertes.fr/tel-01797231/document

Hoy, M. (January 12, 2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. Retrieved from https://www.tandfonline.com/doi/full/10.1080/02763869.2018.1404391

Goel, A. (February 2, 2018) How Does Siri Work? The Science Behind Siri. Retrieved from: https://magoosh.com/data-science/siri-work-science-behind-siri/

How the voice of Siri is built

1. All 200 million + units are gathered

2. 41% are never used and 8.3% are used less than 10 000 times

3. 8.8% are used more than 1 million times

4. Units used less than ten times are pruned down to under 500B of space

Unused Units

Used less than 10 000 times

Used more than 1 million times

Used less than 10 times

How Siri functions

1. User verbally says command to Siri

2. Siri collects data, processes it and sends it to Apple's servers

3. Information is processed in Apple's servers

4. Information is sent back to User's device/ Siri and is presented to User

User

Siri

Apple's Servers