Speech Keyword Spotting

Implementation using Edge Impulse

José Bagur
jabagur@uvg.edu.gt

Agenda

1. What is speech recognition?
2. Data collection insights
3. Hands on exercise

Sobre mí

Profesor e investigador en la Universidad del Valle de Guatemala (UVG). Creador de contenido en Arduino®.

Encargado del Laboratorio de Aeroespacial de la UVG.

Me gusta caminar, leer, coleccionar discos de vinilo, ver películas y tomar café.

What is speech recognition?

Speech recognition is a powerful tool that allows humans to **interact** with electronic devices using human voice.

For example, this tool is implemented in smart devices like Amazon's Alexa smart speaker.

What is speech recognition? (2)

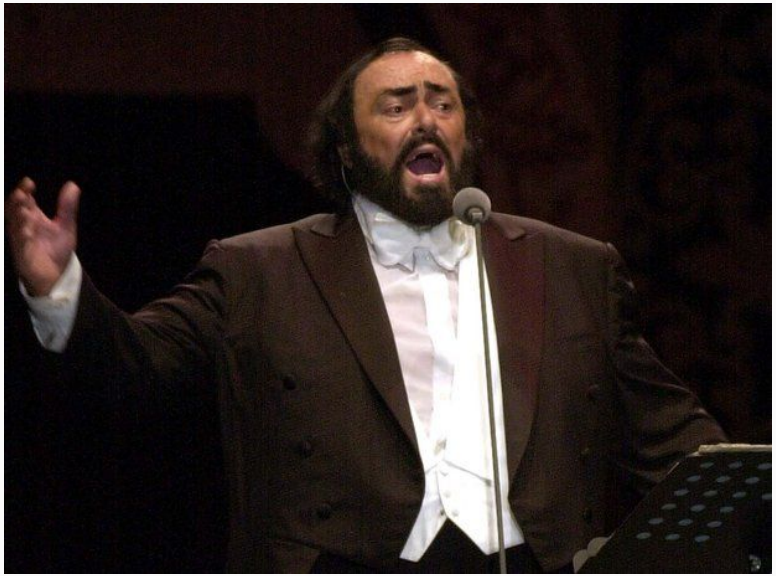
Machine learning is used for speech recognition in embedded devices. In the Amazon's Alexa example, there are **two forms** of machine learning going on:

1. **Inference** (performed locally)
2. NLP (using a remote server)

Data collection

There are two ways we can use to collect data for creating a keyword spotting system:

1. Use existing datasets
2. Create our own data set



Data collection (2)

There are several speech recognition datasets.

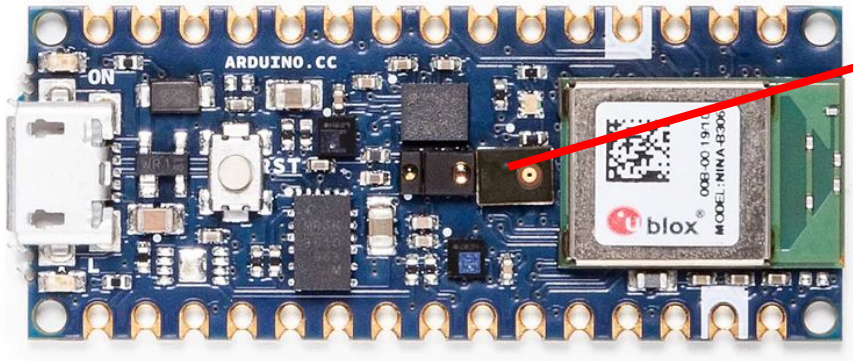
A good starting point is the Google Speech Commands Dataset. This is a great open source dataset that has 65,000 one-second long utterances of 30 short words, by thousands of different people.

Data collection (3)

We can use any audio recording device to record our own custom keyword dataset. But there are some important considerations:

- Aim for at least 50 samples (ideally you would want to get thousands of samples from different people with different voices, genders, accents, etc. to create a more robust model)
- Edit captured audio to match the sample rate of the target device
- Record data as a 32-bit floating point WAV file

Data collection (4)



MP34DT05 MEMS microphone

Sample rate: **16kHz** in the Arduino Nano 33 BLE Sense board

Data collection (5)

Audacity is a free, open source, cross-platform audio software that can be used to edit captured audio:

- Resample and change the sampling rate
- Define a time-defined snippet of audio around the utterance of the keyword