ALGOSUP 2022 Project 5 B - Language detector

Functional specifications by Léo CHARTIER

Project requirements

The goal of this project is to create a device that would listen to conversations and detect whether the language is French or English.
It will make use of Artificial Intelligence (Deep Learning) to detect the language.

Requirements specifications

The device will additionally be equipped with a light for user feedback:
- If English is detected, the light color and/or brightness will change depending on the correctness of the English level.
- If another language is detected, the light will color red and/or pulse.

Main issues

➢ Privacy and security

The device should process the data in a way that is not used for anything else than detecting the language. This means that the data should not be stored, nor be used for anything else than the language detection.
Moreover, there should be no way for the data to leak to the outside world. This means that the device should run autonomously and have no means of connection other than its interface. As such the device should be made by our own means and not just run on a computer or a device with an internet or bluetooth connection.
Furthermore, for a high level of security, sensors may be added to the device that, when moved, opened, or when an unknown device tries to plug in, would erase all the data stored on the device. The rehabilitation of the device would require an operator (developer or another capable hand) to reupload the program to the device using an authorized computer.

➢ Training data collection

We first of all need to collect data to train the model.
There are two ways to do this, each with its advantages and disadvantages:

- Collecting data from the students.
  This would be a more or less long and tedious process depending on the environment of recording and the desired degree of accuracy.
  It would be very specific to the school and would not work outside of it but would produce more accurate results. This means that the model should be retrained often as new students arrive.
- Finding data on the internet.
  This method is more general and much more scalable data and usage-wise. It would indeed consider every type of accent and would not be limited to the school and its students.
  This would also mean we could directly start the training of the model as soon as it is ready at the cost of some accuracy.

Risks

As we will be working with artificial intelligence, the training of the model will take a long time. This might affect the schedule of the project and needs to be properly handled and planned early.
The training data is not provided and might have to be collected (more on this below). As such, the acquisition of the data would also delay the project and we need to be certain this process will not impact the rest of the project.
Finally, the device should be bought and the box created, being the only cost of the project, adding yet again to the schedule.