

Speech controlled calculator

A project by Luka Železnik, Aditya Patel and Lars Vonk

# Introduction

In this project we have researched if it is possible to train a model to classify speech to numbers from 0 to 9 and the basic math operators which are plus, minus, over and times. We will try to train this model, create a basic application to use this model and test this for the results.

This project was created by us since we needed a challenge which would use image or speech recognition and we wanted a project which could have real world applications but also falls within our interest field. Since the three of us are all computer scientists we love math and that is why we came up with this project.

# Hypothesis

Since we want to keep our dictionary of words small with only 14 characters we think it would be very possible to create a model with a high accuracy. Collecting the data will be a challenge since with speech there are a lot of difference between male and female voice, different accents in English, background noise, etc. To classify these correctly we will need a broad set of training and test samples.

# Methodology

To complete this project we have used the programming language Python. Within this language exist many libraries that can aid with the classification of speech. Since this is also a school project which aims to improve our knowledge in how speech recognition is done we will use the libraries only where needed.

The libraries which we have used for training the model are:

* Pytorch: for creating and training the model
* Librosa: for handling audio files
* SciPy: for help with training creating the training data
* Numpy: for data handling

Some other libraries and frameworks we have used for the prototype are:

* Flask: for the webserver
* Flutter: for creating the client

Aside from these Python libraries we have used the following github repository for more training data.  
<https://github.com/ScorpionXiezi/Audio-Calculator>

Note: we have not used this repository for any inspiration to our code.

Furthermore for the training data we have created our own audio samples and asked some of our female acquitances (girlfriends) to also record some audio samples so we could have a more broad set of training and test data with multiple genders and accents. We think our group covers a wide variety of accents since Aditya Patel and his girlfriend are from India, Luka Železnik and his girlfriend are from Slovena, Lars and his sister are from the Netherlands and Lars’ girlfriend is from Poland. Thus covering three major language groups in accents (Germanic, Slavic and Indic language groups).

# Results

# Future work

Since there was a time constraint we had time to create the minimal viable product with some extras for our prototype. But there are some other nice to have additions which we did not have time for which are mentioned below:

* Create a stream of audio that is served to the backend and classified by the model so the output is shown live in the client while the user is speaking.
* Add extra operators so more advanced mathematical equations can be input
* Add a start and stop word to the dictionary so the calculator can truly be used hands-free.
* The classification quality can be increased by training the model on more samples. Also some different methods that we have not tried yet could be experimented with to increase this accuracy.

# Conclusion