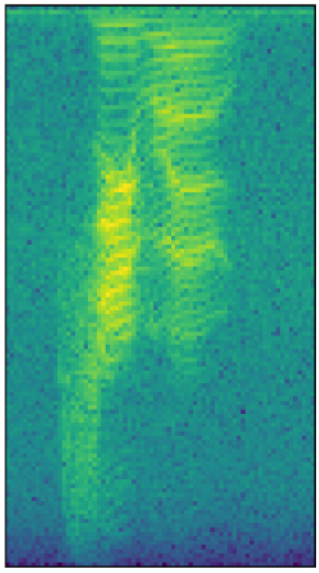
Recognizing speech

Portfolio assignment 1

MAL2 Autumn 2024



In this assignment, you will implement and compare two neural networks to classify spoken digits from "zero" to "nine" using the Google Speech Commands Dataset. You will build both a **fully connected neural network** and a **convolutional neural network** and tune hyperparameters to maximize accuracy and minimize overfitting.

The input data is in the form of log-spectrograms (2D plots of the logarithm of the audio intensity as a function of time and sound frequency). The log-spectrogram to the right is an instance of the spoken digit “zero”.

The dataset – `x_digits.npy` and `y_digits.npy` – is preprocessed for you and can be loaded using

```
X = np.load('x_digits.npy')
```

and similar for y.[1]

You are to hand in a notebook with

- output (all cells must be run)
- relevant comments describing your approach, experiments, and findings
- both neural networks
- the accuracy of train, validation and test sets for both neural networks
- at least two interesting figures or animations

[1] To access the files from Colab, the easiest way is uploading it to Google Drive and then using

```
from google.colab import drive
drive.mount('/content/drive')
```