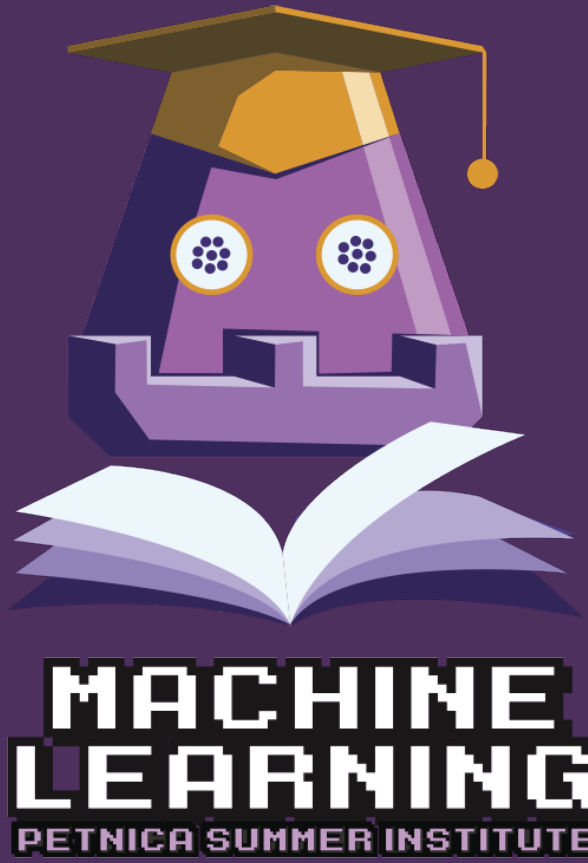
Vowel Classification using Machine learning

Ivan
Stošić

29/07/2017



Microsoft | Development Center Serbia

PETNICA Science Center

Problem overview

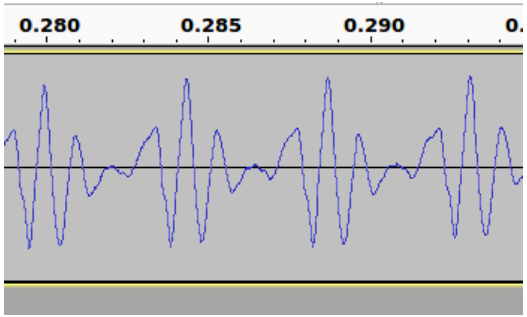
- Classify vowels of some language based on sound
- Input consists of an audio stream or a file
- Output should consist of a sequence of vowels detected

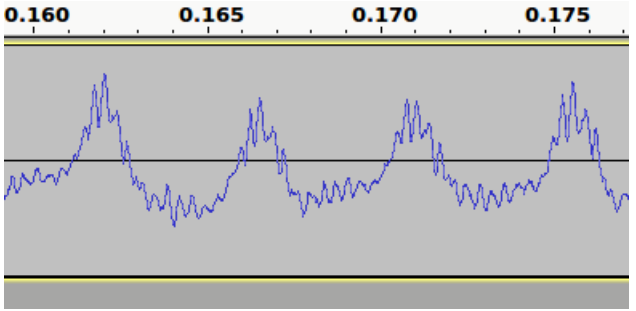
Approaches

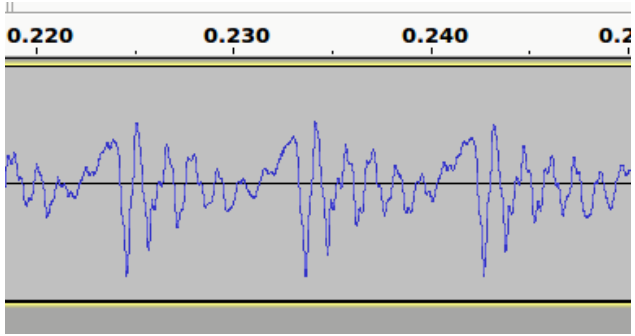
- Use a neural network as a classifier
- What should the features be?

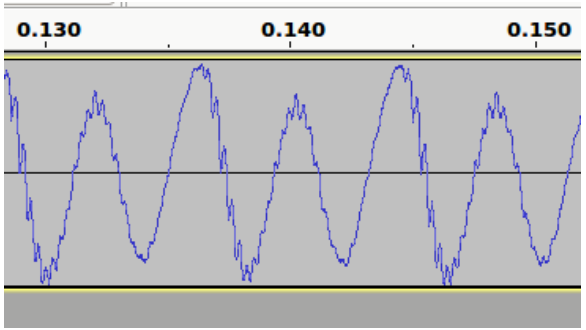
Approaches

- Raw signal



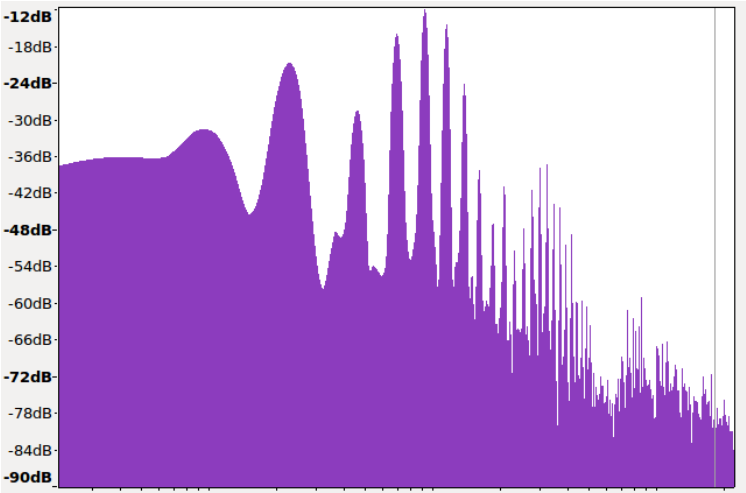


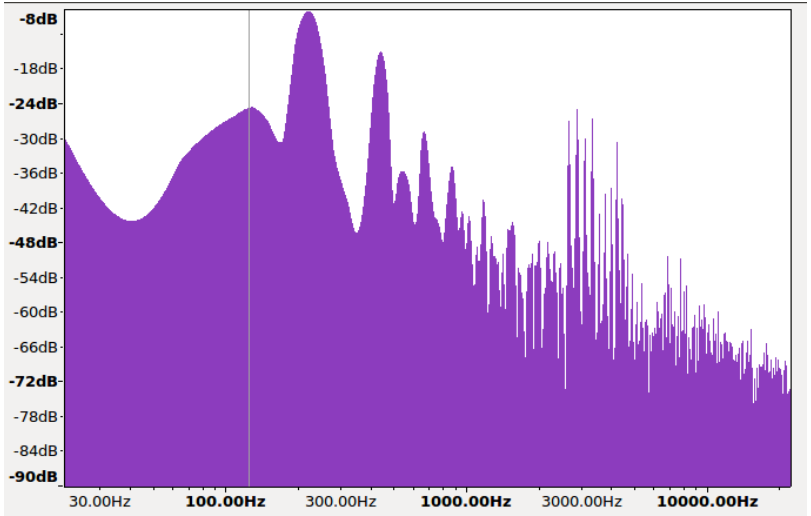




Approaches

- Discrete Fourier Transform, computed using FFT
- Autocorrelation





Approaches

- FFT was used in the final design
- Fully connected neural network with two hidden layers
- [1024, 128, 32, 5]
- Tensorflow, Adam Optimizer, Tanh activation, Softmax

Training

- Small dataset
- Each input file is split into windows of 1024 samples each, which may overlap
- Augment the training set by pitch-shifting and adding noise
- Around 62000 such windows were in the final training set

Results

- >97% accuracy on the validation test
- Confusion matrix:

t\g	a	e	i	o	u
a	3560	3	0	45	8
e	0	4137	2	4	1
i	0	48	3977	0	7
o	15	36	0	3808	13
u	0	10	336	68	2964

Problems

- The dataset is very small
- Low quality audio equipment
- In speech, vowels are often reduced

Conclusion

- Simple neural networks can easily perform the task of classifying vowels with decent accuracy, even when trained on low-quality data
- If you don’t have a lot of good data, just throwing it into the network will not produce good results

MACHINE LEARNING

PETNICA SUMMER INSTITUTE

Thank you!