Limited-vocabulary “small" language automatic speech recognition using Machine Learning

George Vlad Stan
Master Project Presentation
Academic support

Supervisor: Anna Sampaio Bon
2nd supervisor: Hans Akkermans
Technical expertise: André Baart
Technical expertise: Francis Dittoh
Example use cases for Automatic Speech Recognition (ASR)

- Collecting information and providing knowledge to farmers in Sub-Saharan Africa
- Philippines tuberculosis patients health information collection during the SARS-CoV-2 pandemic
Context and technical research

• Low-resource environment requires software with a small hardware footprint
  • Limited bandwidth internet connection
  • Devices with low processing power and storage

• Cultural differences
  • Oral communication preferred over written communication

• Choice of data types and machine learning model
  • Mel spectrograms to represent audio data
  • Convolutional neural networks: the state-of-the-art ML model for speech recognition
Data collection

- Vocesrares.nl
- Built an audio data collection web application using minimum overhead and code (410 kB)
- Written using barebone Javascript, HTML and CSS
- Currently able to collect the words for “yes” and “no” in 16 different languages
- Can be easily expanded to other words in order to increase vocabulary (numbers 0 to 9)
- More languages can be easily added, as needed
- Interface changes with language selection
- Audio files have an average size of 14 kB and can be uploaded fast on a 2G connection
Data collection

Home page with the Twi language selected
Data Processing

- Audio files verified one by one to meet quality standards
- Files are normalized and their format is standardised for use in the machine learning model
- Audio signal is converted to a Mel spectrogram
Data Processing

- Visualisation of mel spectrograms of random files in our dataset
Machine learning model

- Convolutional Neural Network using Mel Spectrograms as input data
- Training-Validation-Test 70:20:10 ratio

<table>
<thead>
<tr>
<th>Samples</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>97% - 98% accuracy</td>
</tr>
<tr>
<td>400</td>
<td>90% accuracy</td>
</tr>
<tr>
<td>200</td>
<td>84% accuracy</td>
</tr>
</tbody>
</table>
Future work

• Expand the application to collect and train on more words in order to increase the vocabulary
• Explore the performance of the vast number of machine learning models available
• Further parameter tuning experimentation can be done
• More languages can be added from different communities around the world
• Software can be made open source so that anyone who wants to build a not-for-profit Automatic Speech Recognition system is able to
Thank you!