Bootstrapping Spoken Information Retrieval for Unwritten Languages

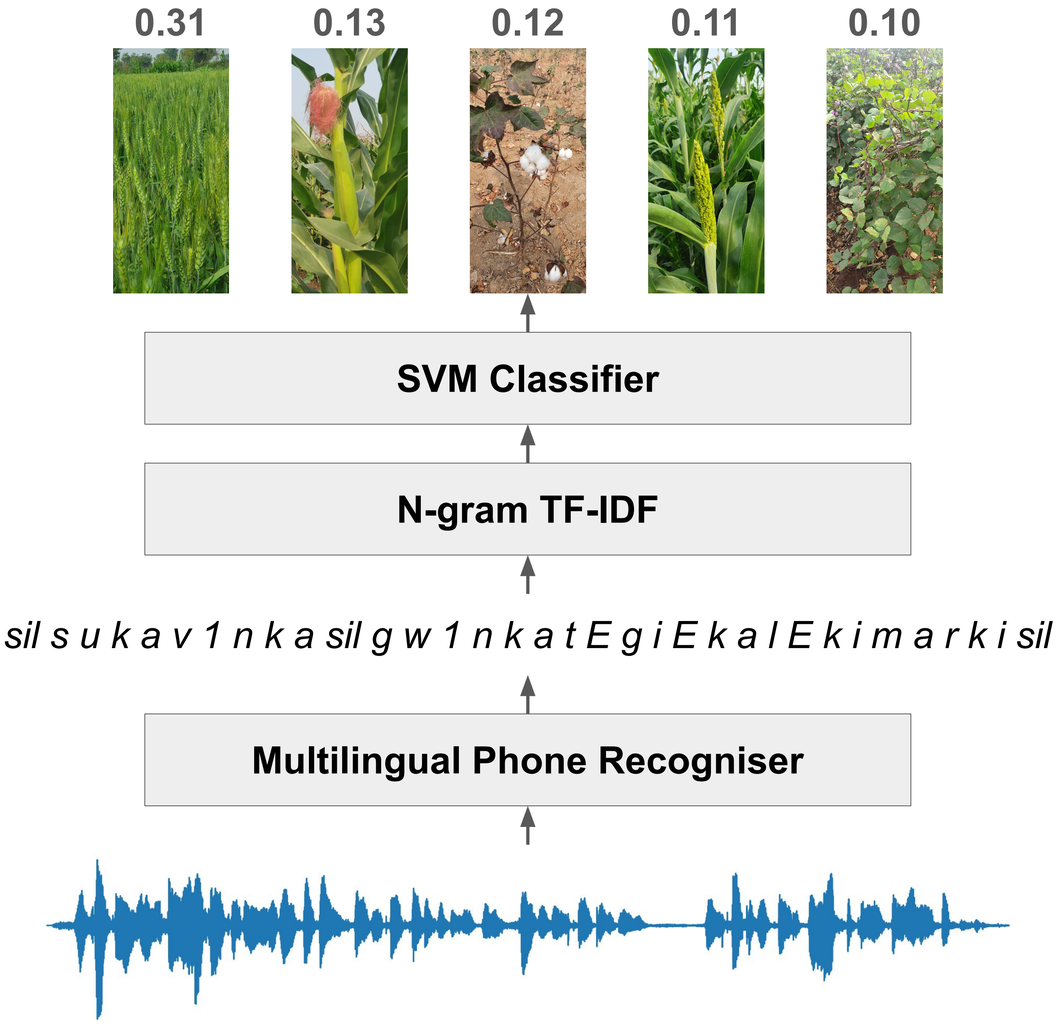
Ondrej Klejch, Electra Wallington, Thomas Reitmaier, Emily Nielsen, Dani Kalarikalayil Raju, Nina Markl, Gavin Bailey, Jennifer Pearson, Matt Jones, Simon Robinson, Peter Bell

How can we bootstrap SIR for an unseen unwritten language?
- We do not know how people that never used Google search for things.
- We need to collect data, but we want to provide value from day zero.
- We need to design a task that is:
  - **Relevant** for local people
  - **Interesting** to collect data to train more advanced methods
  - **Simple** to show early wins

Spoken Information Retrieval Architecture
- Multilingual phone recogniser
- Phone n-gram TF-IDF feature extractor
- SVM classifier
- https://github.com/unmute-tech/voice-search-server

Benefits of our approach:
- It can be deployed from day zero to collect more data.
- The SVM classifier can be easily updated with more data.
- The whole system can run locally on a Raspberry PI.

How would you bootstrap SIR for an unseen language?



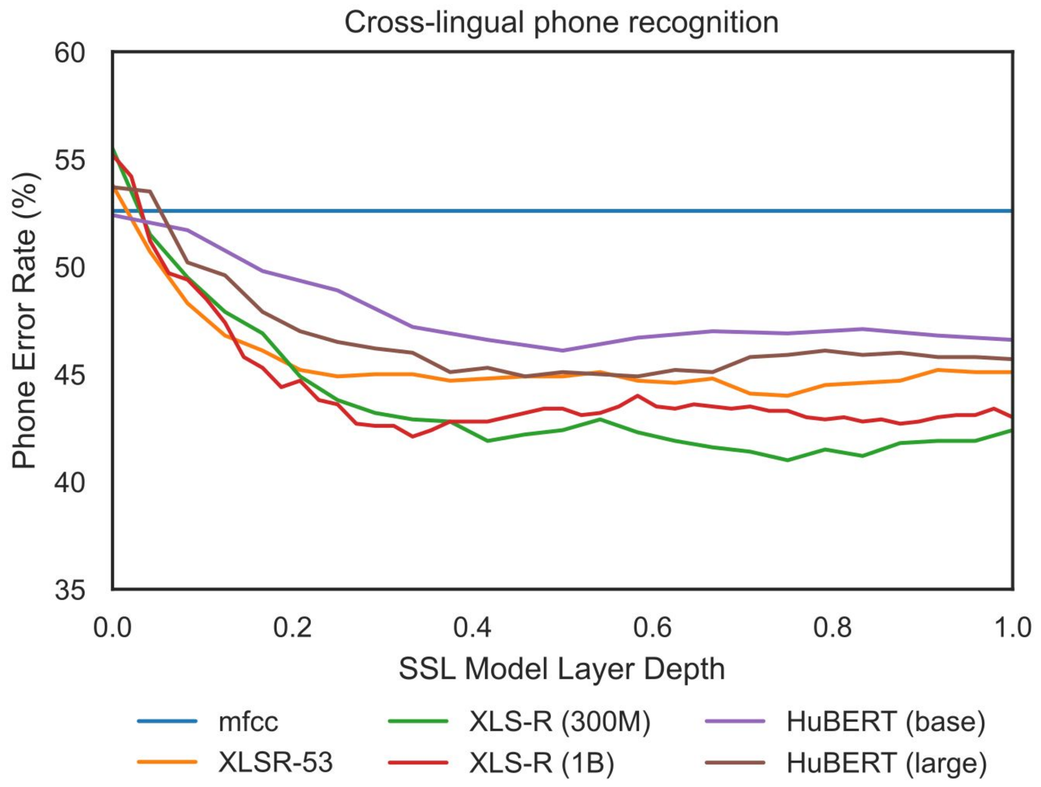
Multilingual Phone Recognition

How can we do phone recognition without any training data?
We train a multilingual phone recogniser for well-resourced languages and we use it to transcribe speech from the unseen language.

Multilingual phone recogniser
- Small Kaldi TDNN-F acoustic model
- Phone bi-gram trained on training transcripts used for decoding
- XLS-R 300M activations from the 18th layer instead of MFCC features

Training data
- 20 hours of training data per language
- English, Spanish, German, French, Polish, Russian

Test languages
- Bulgarian, Czech, Hausa, Portuguese, Swahili, Swedish, Ukrainian



Banjara Results