AUTOMATIC LEARNING OF A PHONOLOGICAL SYSTEM: A CASE STUDY ON MBOSHI LANGUAGE

LUCAS ONDEL AND LUKÁŠ BURGET
BRNO UNIVERSITY OF TECHNOLOGY

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

ABSTRACT

Over the last decade, a lot of research focused on automatic learning of basic acoustic units, a.k.a "pseudo-phones", for low-resource languages. In this work, we investigate the potential and the limits of this research on a real case scenario for documenting a low-resource language. We performed our experiments on Mboshi, an African language from the Bantu family. Results show that despite some progress, automatic learning from under-resourced languages remains a very challenging task and requires further research.

ST4ALL

ST4ALL, "Speech Technologies for All":

- speech-to-text, speech synthesis, data mining, ...
- requires a lot of human-annotated data !!
- necessary amount of data can be reduced by injecting "knowledge": phonetics, dictionary, grammar, ...

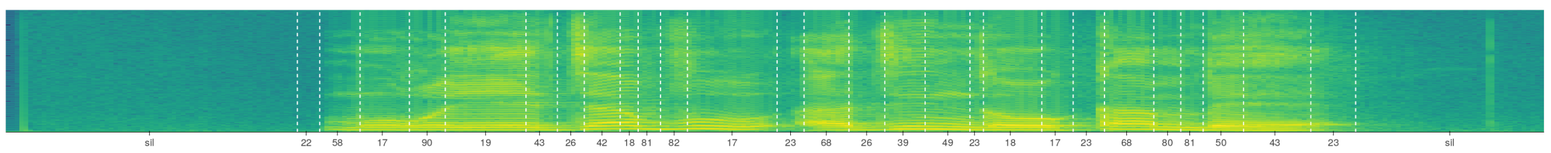
Can we learn this knowledge automatically ?

CORPUS

MBOSHI corpus:

- 5131 recordings from the Mboshi language (4.5 hours)
- 4 speakers
- clean speech
- close to a "real case" scenario of documenting a language
- word/phonetic transcriptions

TASK



Discovering a set of "pseudo-phones"/acoustic units from unlabeled recordings:

- segmentation: when a unit starts and ends
- clustering : which segments are similar ?

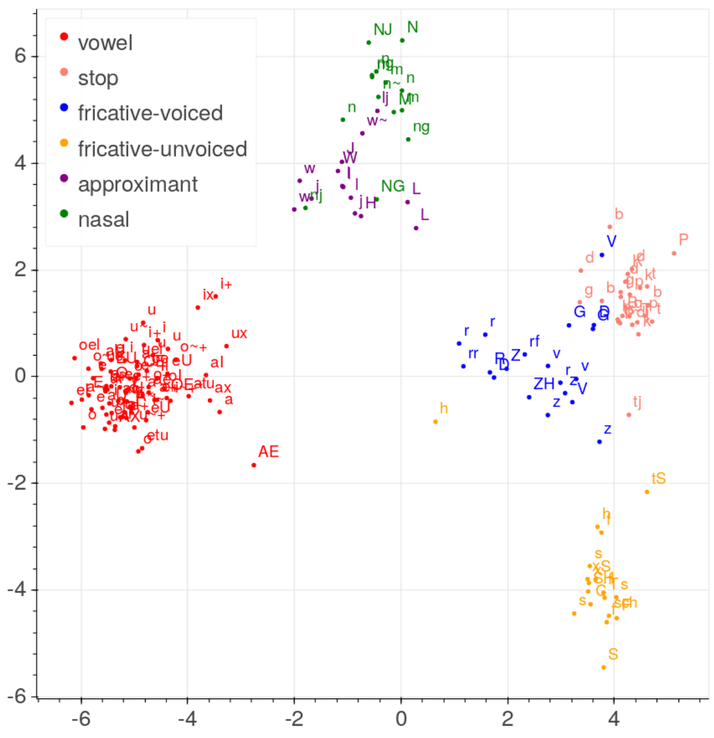
METRICS

Two metrics:

- Normalized Mutual Information: measures the clustering quality
- F-Score: measures the segmentation quality.

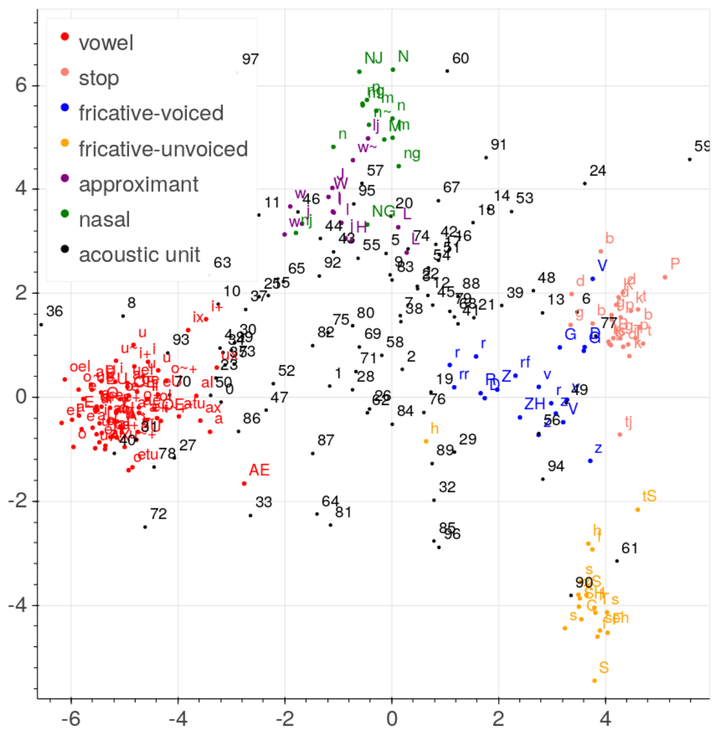
METHOD (STEP I)

Learn a "phonetic space" from well documented languages: French, German, Polish and Spanish



METHOD (STEP II)

Discovering the units is equivalent to "projecting" the Mboshi data into the phonetic space.



SOFTWARE

All results are reproducible with the BEER software:

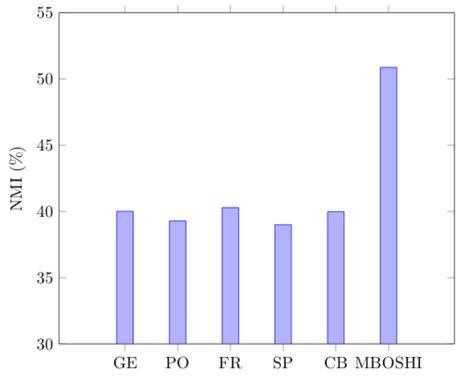
- https://github.com/BUTSpeechFIT/beer
- python implementation using pytorch
- BSD-2 license

CONCLUSION

Despite progress over the past few years, learning the phonetics of language in a data-driven fashion remains a very challenging task. As explored in this work, knowledge transfer across languages is a promising direction as it allows to re-use the outcome of several decades of speech research to our problem. Yet, this technology still requires further research as we have no ideal way yet to extract the phonetic information of a language. Current research focuses on deriving better phonetic space as well as more "robust" projection method.

RESULTS

Model	F-score	NMI (%)
DP-HMM [1]	49.20	34.41
DP-HMM [2]	46.89	35.98
DP-SHMM [3]	**57.65**	**39.98**



REFERENCES

[1] Chia-ying Lee and James Glass. A nonparametric bayesian approach to acoustic model discovery. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 40–49. Association for Computational Linguistics, 2012.

[2] Lucas Ondel, Lukáš Burget, and Jan Černockỳ. Variational inference for acoustic unit discovery. *Procedia Computer Science*, 81:80–86, 2016.

[3] Lucas Ondel, Hari Krishna Vydana, Lukáš Burget, and Jan Černockỳ. Bayesian subspace hidden markov model for acoustic unit discovery. *arXiv preprint arXiv:1904.03876*, 2019.

[4] Pierre Godard, Gilles Adda, Martine Adda-Decker, Juan Benjumea, Laurent Besacier, Jamison Cooper-Leavitt, Guy-Noël Kouarata, Lori Lamel, Hélene Maynard, Markus Müller, et al. A very low resource language speech corpus for computational language documentation experiments. *arXiv preprint arXiv:1710.03501*, 2017.