

Éric Le Ferrand^{1,2}, Steven Bird¹ and Laurent Besacier²

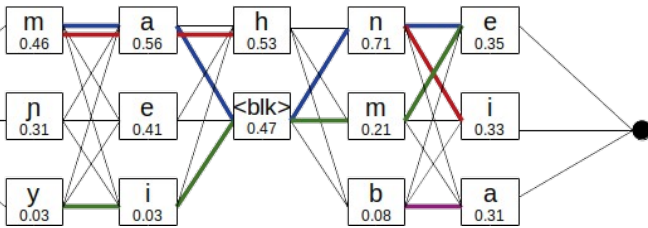
¹Charles Darwin University, ²Université Grenoble Alpes

Background

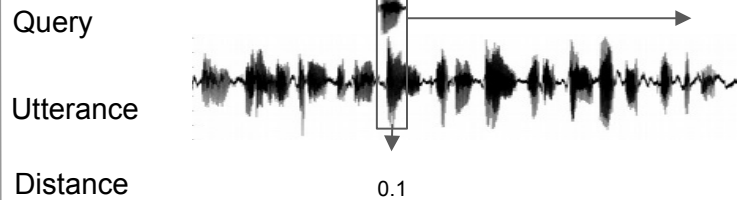
- ➔ Traditional spoken terms detection methods are data greedy and not adapted for Indigenous languages[1]
- ➔ DTW, traditionally used for very low-resource languages, lacks of precision and rely on the quality of the spoken queries[2]
- ➔ Recent research made phone recognition approaches accessible to very low-resource languages[3]

Spoken Term Detection for very low-resource languages

Phones exploration in confusion network (P2W)



Dynamic Time Warping with multilingual Bottleneck features (DTW)



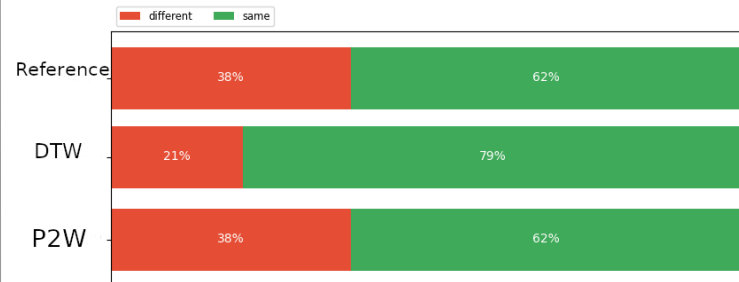
Experiments on two very low-resource languages:

Mboshi (mb): 21min of training data
Kunwinjku (kun): 35min of training data

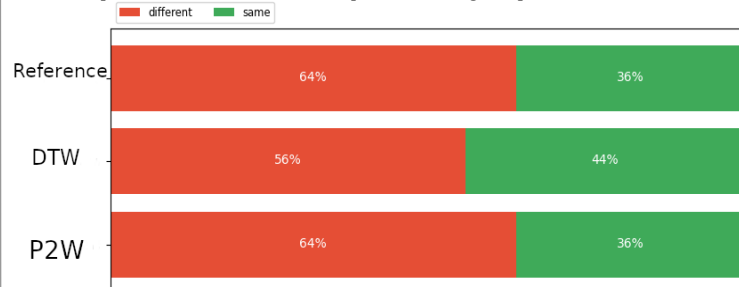
Results based on optimized threshold

	recall	precision	F-score
dtw_mb	14.55%	20.46%	17.01%
p2w_mb	22.61%	45.97%	30.31%
dtw_kun	42.09%	22.81%	29.59%
p2w_kun	17.41%	62.50%	27.23%

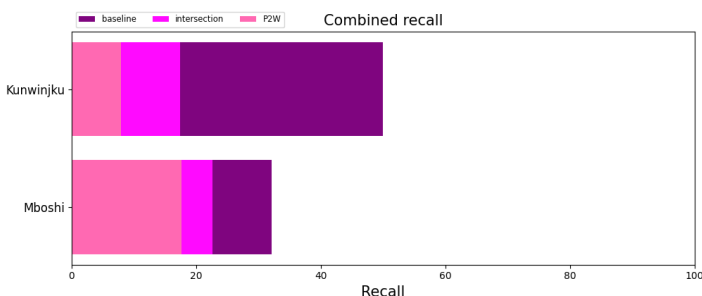
Interspeaker retrieval (Mboshi)



Interspeaker retrieval (Kunwinjku)



Intersection between the methods' results



References

- [1] Murat Saraclar and Richard Sproat. 2004. Lattice-based search for spoken utterance retrieval. In Proceedings HLT-NAACL. pages 129–136.
- [2] Eric Le Ferrand, Steven Bird, and Laurent Besacier. 2020. Enabling interactive transcription in an indigenous community. In COLING 2020.
- [3] Xinjian Li, Siddharth Dalmia, Juncheng Li, Matthew Lee, Patrick Littell, Jiali Yao, Antonios Anastasopoulos, David R Mortensen, Graham Neubig, Alan W Black, et al. 2020. Universal phone recognition with a multilingual allophone system. In ICASSP 2020