



# Very Low Resource Radio Browsing for Agile Developmental and Humanitarian Monitoring

Armin Saeb<sup>1</sup> Raghav Menon<sup>1</sup> Hugh Cameron<sup>2</sup>  
William Kibira<sup>2</sup> John Quinn<sup>2</sup> Thomas Niesler<sup>1</sup>

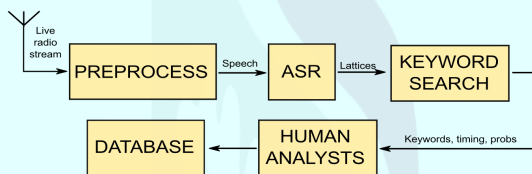
<sup>1</sup>Stellenbosch University, South Africa <sup>2</sup>UN Global Pulse, Kampala



## Introduction

- ♦Rural Uganda has poor internet connectivity so people communicate their problems through phone in talk shows hosted by small community radio stations.
- ♦The UN has piloted a radio browsing system which monitors such radio discussions to support their relief and development programs.
- ♦This work considers what can best be done using resources assembled in a very short time.
- ♦We assume availability of 12 minutes of annotated data in the target language.

## Radio Browsing System



- ♦Live audio is preprocessed to detect acceptable speech.
- ♦Lattices searched for keywords of interest.
- ♦Results passed to human operators to obtain a structured, categorised and searchable information.

Topic	Analyst translation
natural-disaster, food-security	"Elephants that are suspected to have come from South Sudan went and attacked Abalo Kodi village and destroyed food [crops] about 20 acres."
refugees.camps	"I stand with my two legs and say that staying in the camps is very very good [...] those days when people were not in the camps they used to keep money in anthills and under the beds, but after coming out of the camps they have knowledge about banking."
health.service-delivery	"The road here is so bad that the ambulance got stuck in a ditch and could not reach the hospital. People came and had to collect the medicine and carry it on foot to the hospital."
health.malaria-prevention	"People are using mosquito nets in the wrong way, for example scrubbing their bodies, washing dishes, making fences around chicken houses, some even turkey houses or pigsties."

Examples of relevant discussions discovered by the radio browsing system

## Data

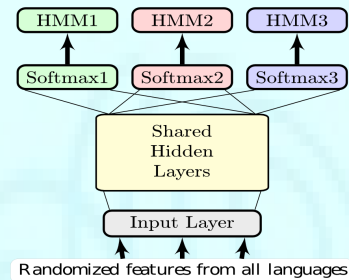
- ♦Target language is Acholi, a severely under-resourced Ugandan language.
- ♦Luganda and English serve as the source languages.

	Acholi	Luganda	English
Transcribed train sentences	80	8773	14898
Transcribed train speakers	4	380	889
Transcribed train speech	12m	9.6h	26h
Untranscribed train sentences	4782	—	—
Untranscribed train speakers	199	—	—
Untranscribed train speech	9h	—	—
Test sentences	184	—	—
Test speech	18m	—	—
OOV rate	2.6%	—	—
Vocabulary words	15750	35098	77140
LM sentences	83831	—	—
LM words	1.3M	—	—

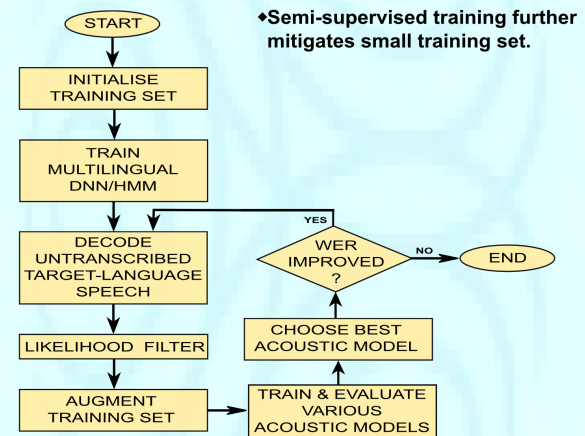
Dataset used for experimentation

## Multilingual DNN/HMM Models

- ♦Multilingual DNN/HMM AM mitigates the small target dataset.



## Experimental Setup



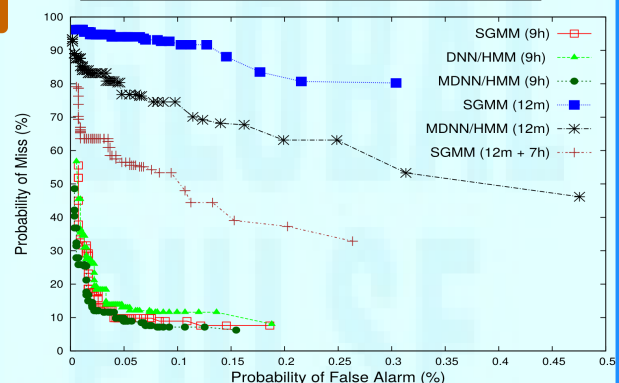
- ♦Semi-supervised training further mitigates small training set.

## ASR Results

Acoustic model	% WER (12m)	% WER (12m+7h)	% WER (9h)
HMM/GMM	96.06	64.32	48.63
SGMM	92.07	62.60	47.09
DNN/HMM	99.30	65.07	46.99
MDNN/HMM	77.92	64.91	43.56

## Keyword Spotting Results

Acoustic model	ATWV	MTWV
SGMM (9h)	0.5844	0.6484
DNN/HMM (9h)	0.5221	0.6857
MDNN/HMM (9h)	0.6400	0.6843
SGMM (12m)	-0.1872	-0.0356
MDNN/HMM (12m)	0.0415	0.0726
SGMM (12m+7h)	0.1430	0.2541



## Conclusion

- ♦Semi-supervised learning combined with multilingual acoustic modelling achieves substantial improvements in a low resource scenario.
- ♦The best system exhibited a WER 62.6% using just 12 minutes of transcribed speech in the target language.