

Radio-browsing for Developmental Monitoring in Uganda

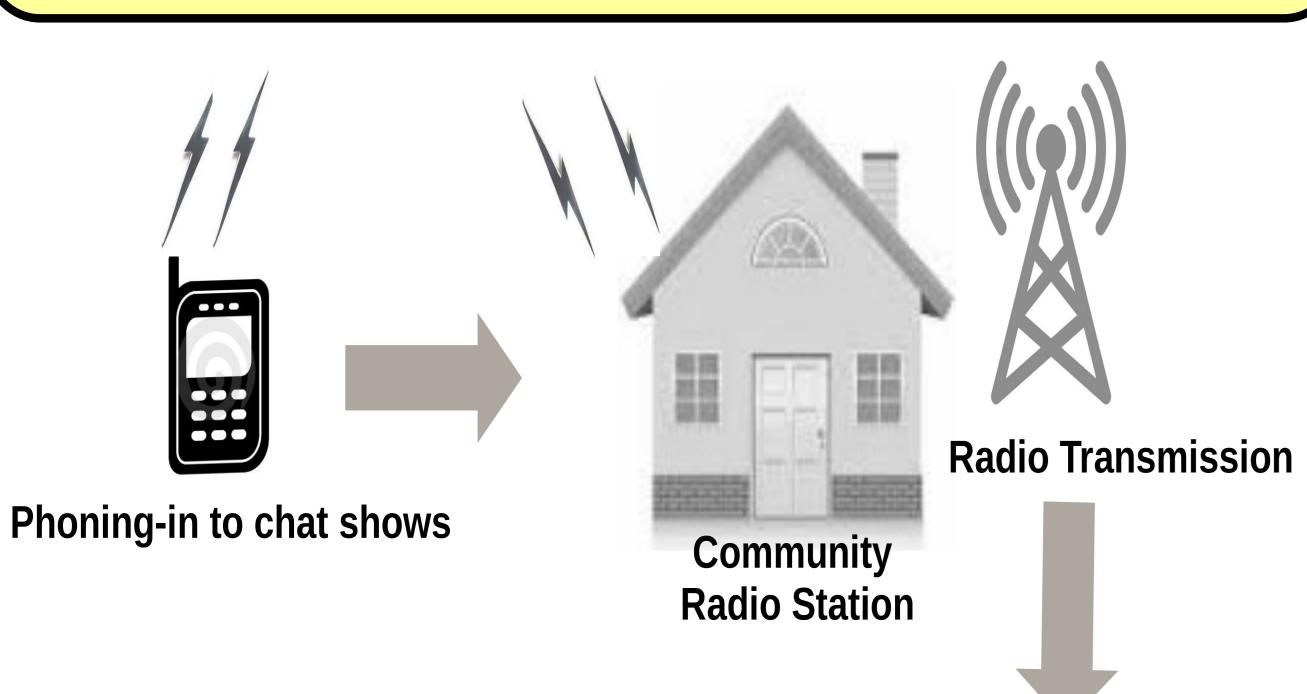
Raghav Menon¹ Armin Saeb¹ Hugh Cameron² William Kibira² John Quinn² Thomas Niesler¹

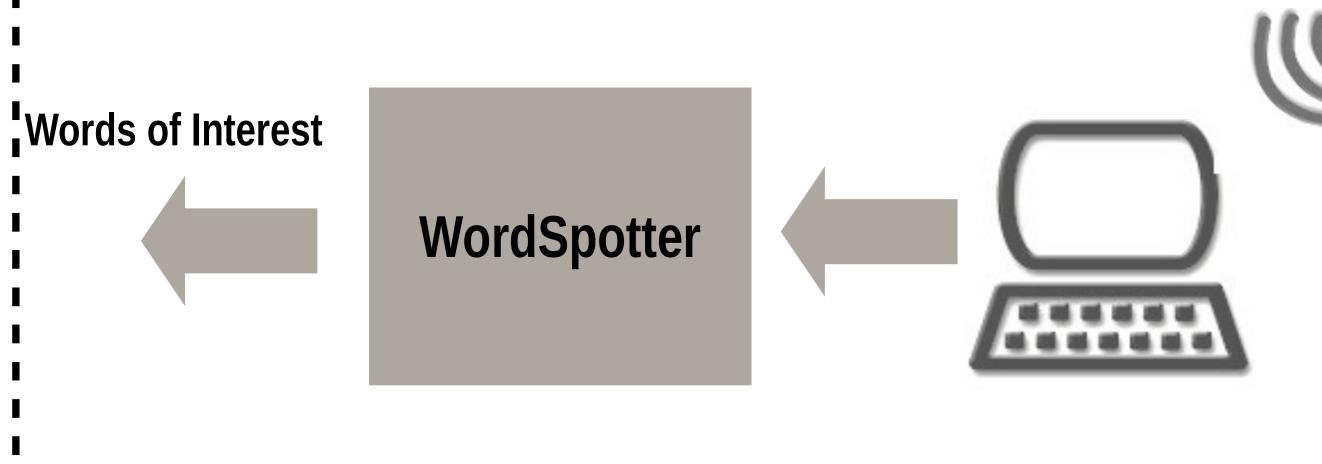
¹Stellenbosch University, South Africa ²Pulse Labs Kampala, Uganda



Introduction

- Local African radio broadcasts often contain actionable information for development and relief programmes.
- Radio-browsing system developed for Ugandan English and two indigenous languages, Luganda and
- Automatic speech recognition includes combination of HMM/GMM, SGMM and DNN/HMM architectures.
- The system already provides practically useful results for relief efforts.

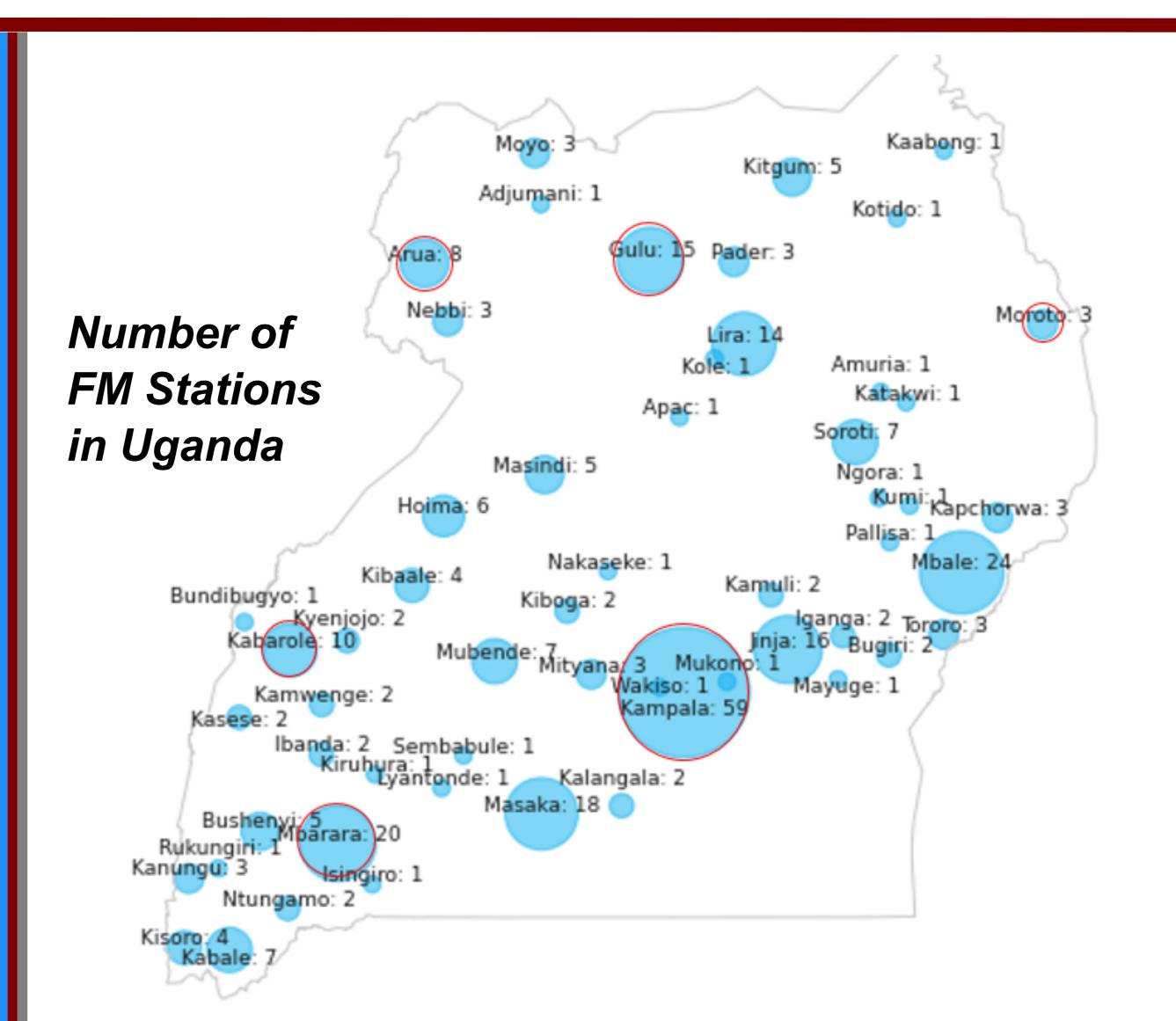




Radio Reception

System Configuration

- Radio stations host phone-in shows.
- Listeners call in, discuss issues of importance.
- ◆Broadcasts recorded using RTL2832 based DVB-T receiver, Raspberry-pi.
- Keyword spotter used for searching words of interest.
- Errorful system output filtered and selected positives passed to human screeners.



Data Compilation

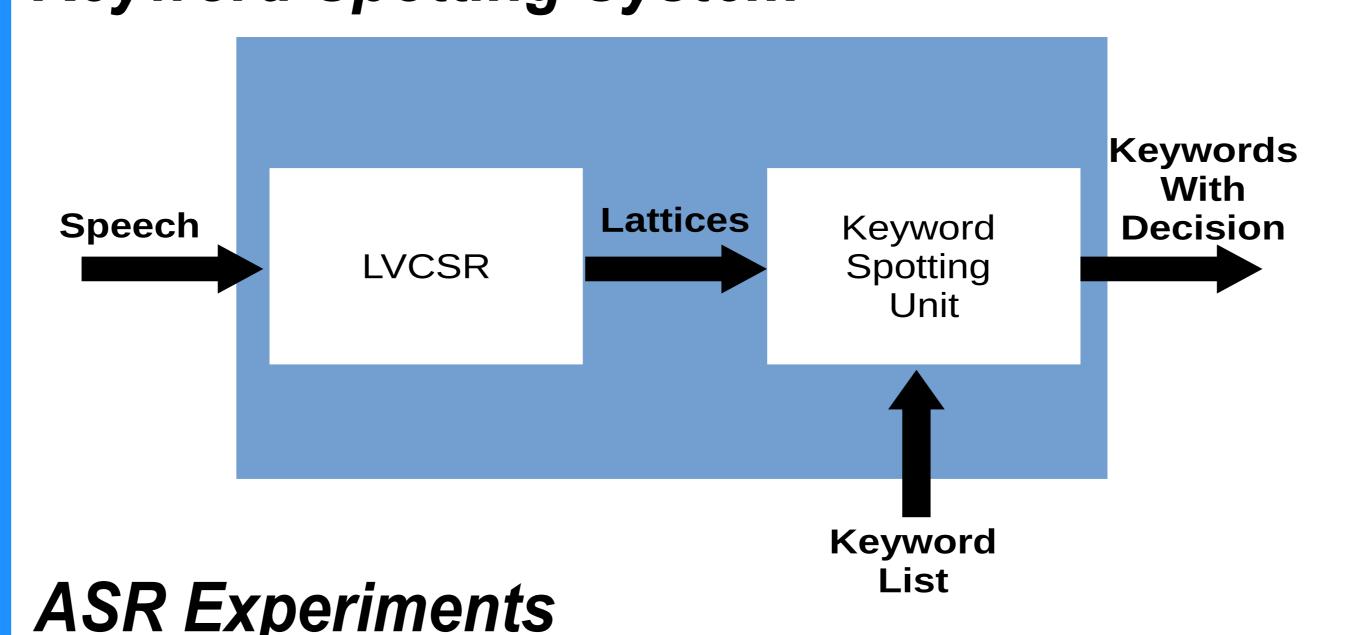
- In Uganda, 216 FM transmitters broadcasting on 299 frequencies (2014).
- ◆First recorded broadcasts in Kampala area.
- conflict, poverty and instability are recurrent (2016). oracle measures, ATWV and MTWV.
- ◆Further posts to be added where relief and development programs are concentrated.

Datasets

- Datasets compiled in Ugandan English, Luganda and
- Ugandan English augmented with SA English which was found to be closer than US English.
- ◆Trigram language model using SRILM.
- ◆Pronounciation dictionaries developed by phoenetic experts but limited for Luganda and Acholi.

	English		Luga	ında	Acholi		
	Train	Test	Train	Test	Train	Test	
Utterances	12966	490	8172	794	4863	184	
Speech	23h	37m	9h	62m	9h	18m	
Speakers	758	48	380	75	203	NA	
LM Words	109 million		1 mil	llion	1 million		

Keyword Spotting System



- Lattices generated using various acoustic models.
- ◆Best results are exhibited by DNN/HMM and SGMM-BMMI models.

Cristom	%WER	%WER	%WER	
System	(English)	(Luganda)	(Acholi)	
HMM/GMM	46.71	57.15	59.04	
SGMM	43.32	54.36	58.89	
SGMM-BMMI	40.39	52.47	57.50	
DNN/HMM	38.88	53.54	57.75	

Keyword Spotting Experiments

◆Additional posts added in Gulu and Moroto where Performance evaluated using DET curves and NIST

	English		Luganda		Acholi	
	ATWV	MTWV	ATWV	MTWV	ATWV	MTWV
HMM/GMM	0.5169	0.5278	0.2199	0.3095	0.2852	0.4146
SGMM	0.5860	0.6188	0.2584	0.3523	0.2845	0.4455
SGMM-BMMI	0.6018	0.6188	0.2357	0.3300	0.3554	0.4690
DNN/HMM	0.6115	0.6225	0.2807	0.3617	0.5210	0.5773
Combination (Avg)	0.5958	0.6500	0.3195	0.3660	0.4771	0.4828
Combination (Geo Avg)	0.6018	0.6134	0.3743	0.3743	0.4604	0.5157
Combination (Har Avg)	0.5946	0.6134	0.3717	0.3723	0.4788	0.5157
Combination (Condorcet)	0.6151	0.6565	0.2607	0.3587	0.3180	0.5034

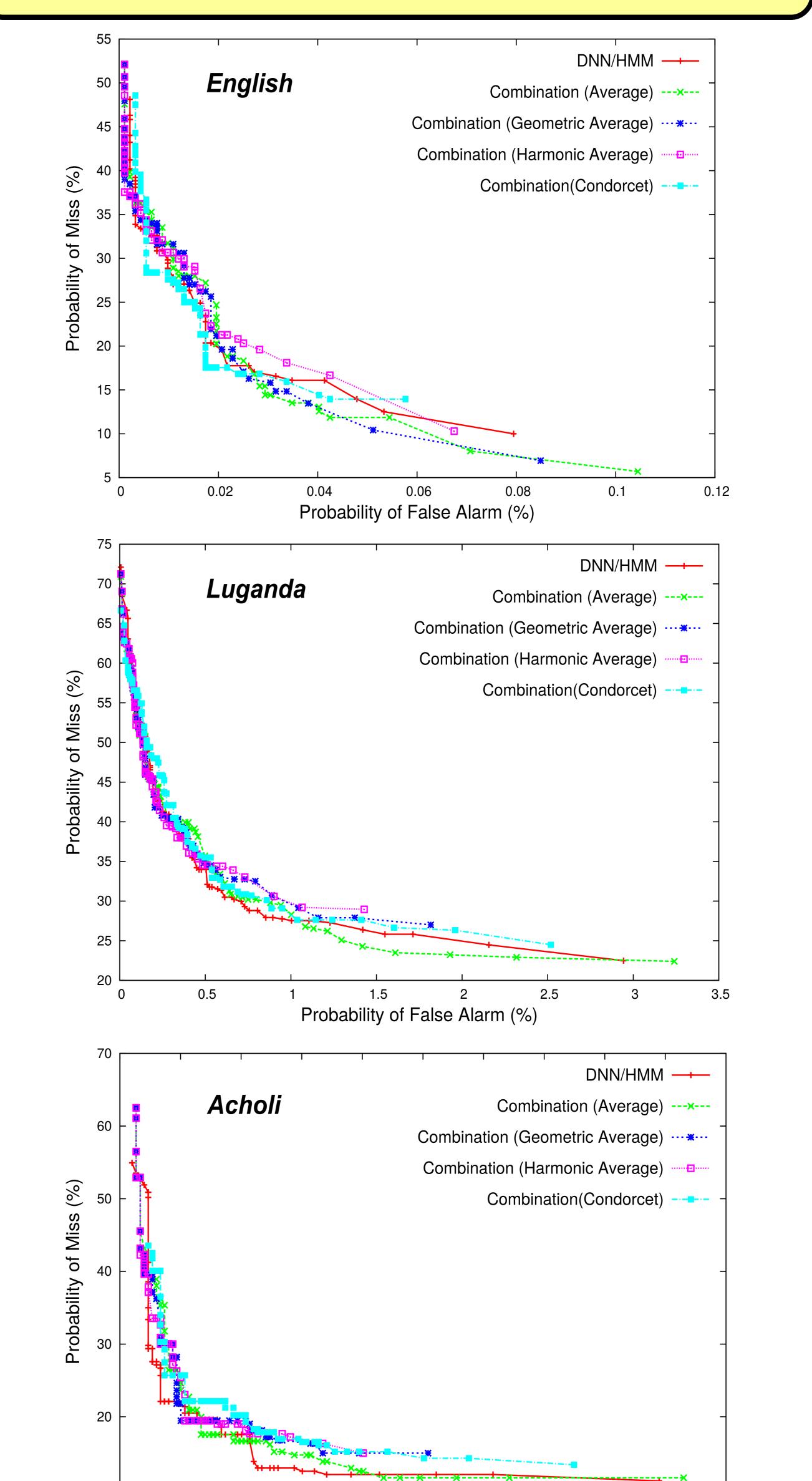
System Combination

◆The four systems were first aligned and combined using

Arithmetic Average
$$scoreA_f = \frac{1}{N} \sum_{n=1}^{N} score_n$$

Geometric Average $scoreG_f = \exp\left\{\frac{1}{N} \sum_{n=1}^{N} ln(score_n)\right\}$
Harmonic Average $scoreH_f = \frac{N}{\sum_{n=1}^{N} \frac{1}{score_n}}$

- System combination improves performance in certain Radio browsing system has been applied to over areas of DET curve.
- ◆ Condorcet voting principle shows better result for English Luganda and 450 hours of Acholi. similar performances in other cases.



Probability of False Alarm (%)

5500 hours of audio including 650 hours of

Keyword	Context of detection
English	
Flood	the residents were warned to vacate the areas that constantly f ood when it rains but they say
Malaria	if they shift their sugar cane will be stolen most health centers in Agago district are now without anti malaria drugs
Health	Awac health center three in Gulu district has been operating without toilets
Luganda (Context translated to English)	
Akawuka (means insect but common usage means AIDS)	Me if I have a house maid I make sure I check her HIV status, if she is positive I chase her and get another one because I cannot risk leaving my baby with her.
Omukyala (Wife or woman)	Police in Buloopa in Kamuli District has arrested a wife (woman). It is said that she stabbed her husband in the neck and the main reason is yet to been known.
Ababunda (Refugees)	There has been a Cholera outbreak in the camp for refugees known as Eregu Bidi-Bidi Camp. This is because of the level of sanitation which is very poor.
Acholi (Context translated to English)	is very poor.
Olong kot [Lightning]	Lightning has struck 20 cows to death in Guda Palwo village in Madopei sub county in Lamwo district.
Yat [Medicine]	We have a problem in Ongango sub county, you find that if they have brought medicine today, but tomorrow you go to the hospital you will hear that ((NGs)) has not signed.
Kwan [Education]	People are objecting the Universal University education since the government has already failed to run both the Universal Primary education and Universal Secondary Education projects.

◆The probability of error was observed 50-60% for Luganda, 40-50% for Acholi and 30-40% for \ Ugandan English.

Conclusion

- First results for a radio-browsing system applied to Ugandan English, Luganda and Acholi.
- Provides managers of relief and development programmes at the UN with current information.
- ◆The radio-browsing system is actively and successfully being deployed in Uganda.