Multi-Accent Speech Recognition of Afrikaans, Black and White Varieties of South African English

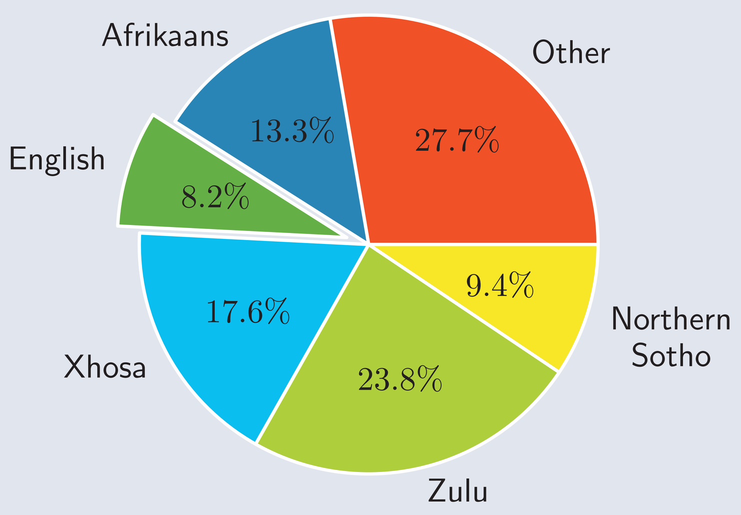
Herman Kamper and Thomas Niesler
Department of Electrical and Electronic Engineering, Stellenbosch University, South Africa

Interspeech 2011, Florence, Italy
12th Annual Conference of the International Speech Communication Association

Introduction

In South Africa, English is the lingua franca as well as the language of government, commerce and science.

Only 8.2% of the population use English as first language. Accented speech is therefore highly prevalent.



Speech recognition systems must be robust to multiple accents of South African English (SAE), an under-resourced language.

Aims of Research:

1. The development of a system able to simultaneously recognise multiple accents of SAE.
2. Determine the degree of performance degradation caused by accent misclassifications.

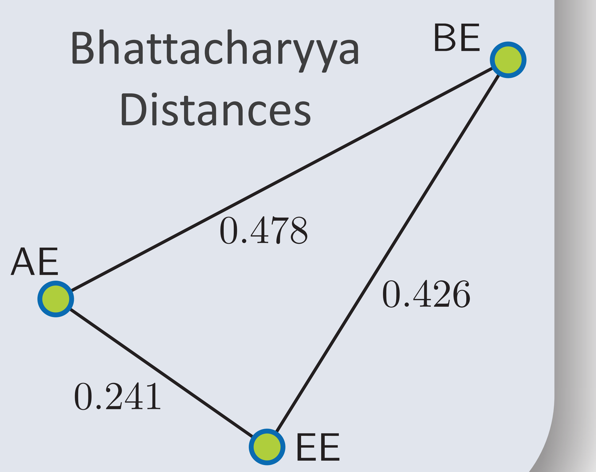
South African English Accents

Accents of South African English Considered:

- Afrikaans English (AE) is the second language used by White first language Afrikaans speakers (5.7%).
- Black South African English (BE) is non-mother-tongue English used by Black South Africans; prominent in government, commerce and media (77.8%).
- White South African English (EE) is the first language used by White South African speakers, chiefly of British descent (3.8%).

Considered in Two Pairs:

1. AE+EE (similar)
2. BE+EE (different)



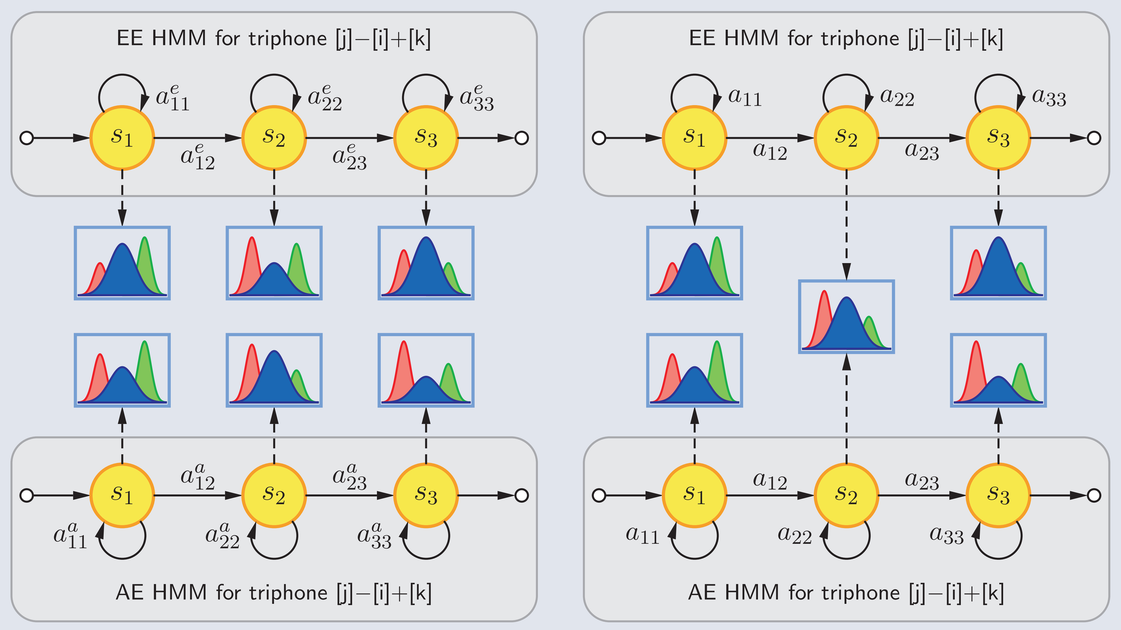
Speech Databases

Accents	Training Set: Speech (h)	Training Set: Number of Speakers	Evaluation Set: Speech (min)	Evaluation Set: Number of Speakers
AE	7.02	276	24.16	21
BE	5.45	193	25.77	20
EE	5.95	245	23.96	18

Language Model (LM) Perplexities

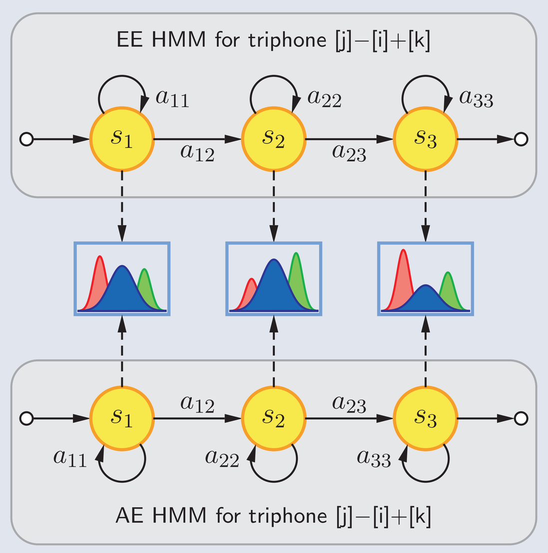
Accents	Matched LM	AE+EE LM	BE+EE LM
AE	25.81	25.46	-
BE	30.30	-	29.63
EE	28.97	27.16	26.67

Acoustic Modelling



Considered Three Approaches:

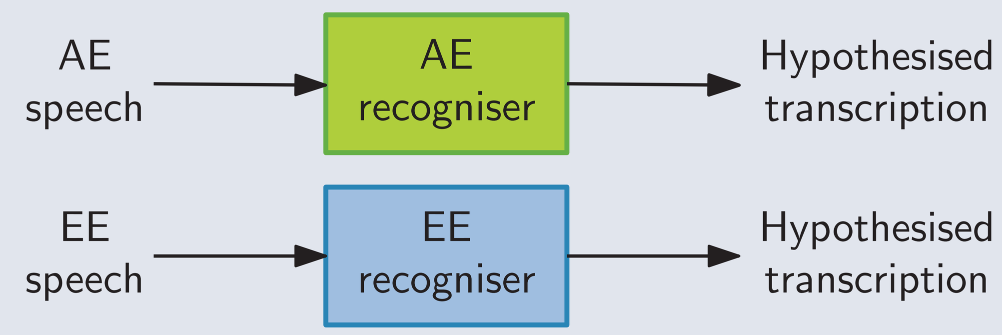
1. Accent-specific modelling: Separate modelling of each accent (top left).
2. Multi-accent modelling: Selective sharing across accents when models are similar (top right).
3. Accent-independent modelling: Data is pooled across accents yielding a single model set (bottom right).



System Configuration

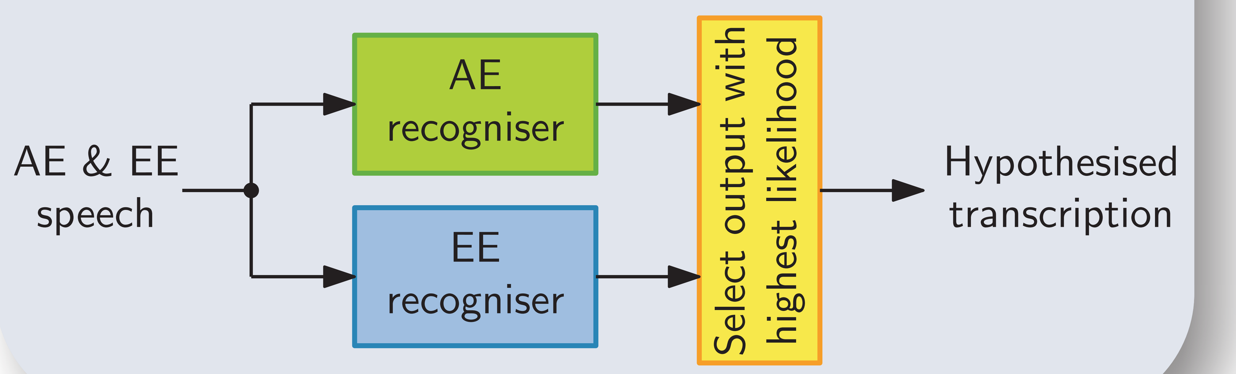
Oracle recognition:

Each test utterance is presented only to the correct accent-specific recogniser, i.e. the accent of the test data is assumed to be known.



Parallel recognition:

Each test utterance is presented to both accent-specific recognisers and the output with the highest associated likelihood is selected, i.e. test data accent is unknown. Accent identification (AID) is thus performed implicitly.



Results

AE+EE Recognition Accuracy

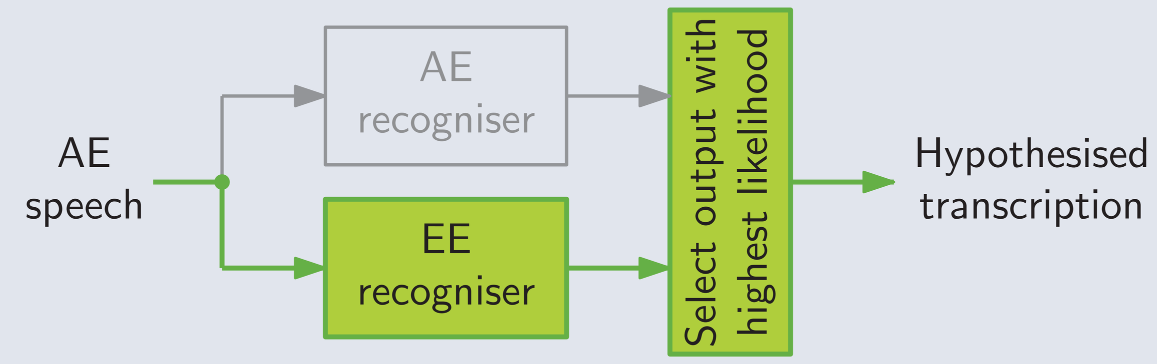
Model Set	Oracle (%)	Parallel (%)	AID (%)
Accent-specific	82.62	82.97	80.16
Accent-independent	82.97	82.97	-
Multi-accent	83.15	83.32	78.29

BE+EE Recognition Accuracy

Model Set	Oracle (%)	Parallel (%)	AID (%)
Accent-specific	74.84	74.18	93.30
Accent-independent	74.74	74.74	-
Multi-accent	75.50	74.85	93.02

Discussion

A misclassification occurs when the recogniser of the other accent is selected during recognition:



AE+EE systems: Parallel systems show small improvements over oracle systems. Accent misclassifications result in improvements and not deterioration.

BE+EE systems: Parallel systems yield deteriorated performance compared to oracle recognition systems. Misclassifications lead to deterioration.

Both accent pairs: Multi-accent acoustic models outperform accent-specific and accent-independent models.

Per-speaker processing: Parallel systems can also process speech on a per-speaker basis. This yields perfect AID for BE+EE. For AE+EE, results are poorer compared to per-utterance AID but better than oracle recognition.

Conclusions

The improvements or degradation caused by misclassifications depend on the similarity of the accents involved.

Superior identification prior to accent-specific recognition does not necessarily lead to superior performance.

It is better to employ parallel speech recognisers with multi-accent acoustic models than to pool training data.

Final Conclusions:

1. Systems using multi-accent acoustic models outperform accent-independent and accent-specific systems for the two accent pairs considered.
2. Accent misclassifications do not necessarily lead to speech recognition performance degradation.