Accent identification in the presence of code mixing

T.R. Niesler[1] F. de Wet[2]

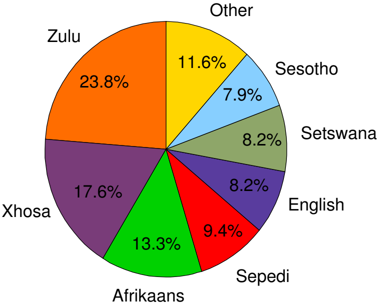
[1] Department of Electrical and Electronic Engineering
University of Stellenbosch

[2] Centre for Language and Speech Technology
University of Stellenbosch

Odyssey 2008 Speaker and Language Recognition Workshop
24 January 2008

INTRODUCTION

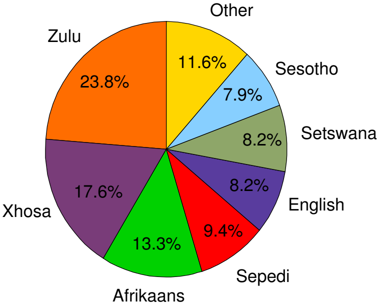
- English is the 5th most common mother tongue among 11 official languages



- However English is used as the *lingua franca*
- Hence non mother tongue English is extremely common

INTRODUCTION

- English is the 5th most common mother tongue among 11 official languages



- However English is used as the *lingua franca*
- Hence non mother tongue English is extremely common

CODE MIXING AND SWITCHING

- Including English words or phrases as part of an utterance is accepted practice in several African languages
- Especially numbers, dates and money amounts
- English alternative is often shorter
    - The number “2353”
      (*Two thousand three hundred and fifty three*)
    - In Xhosa this is:
      Amawaku amabini namakhulu amathathu namashumi amahlanu nantathu
    - Which means literally:
      Thousands that-are-two and hundreds-that-are-three and tens-that-are-five and three

CODE MIXING AND SWITCHING

- Including English words or phrases as part of an utterance is accepted practice in several African languages
- Especially numbers, dates and money amounts
- English alternative is often shorter
  - The number “2353”
    (*Two thousand three hundred and fifty three*)
  - In Xhosa this is:
    Amawaku amabini namakhulu amathathu namashumi amahlanu nantathu
  - Which means literally:
    Thousands that-are-two and hundreds-that-are-three and tens-that-are-five and three

CODE MIXING AND SWITCHING

- Including English words or phrases as part of an utterance is accepted practice in several African languages
- Especially numbers, dates and money amounts
- English alternative is often shorter
    - The number “2353”
      (*Two thousand three hundred and fifty three*)
    - In Xhosa this is:
      Amawaku amabini namakhulu amathathu namashumi amahlanu nantathu
    - Which means literally:
      Thousands that-are-two and hundreds-that-are-three and tens-that-are-five and three

CODE MIXING AND SWITCHING

- Including English words or phrases as part of an utterance is accepted practice in several African languages
- Especially numbers, dates and money amounts
- English alternative is often shorter
  - The number “2353”
    (*Two thousand three hundred and fifty three*)
  - In Xhosa this is:
    Amawaku amabini namakhulu amathathu namashumi amahlanu nantathu
  - Which means literally:
    Thousands that-are-two and hundreds-that-are-three and tens-that-are-five and three

CODE MIXING AND SWITCHING

- Including English words or phrases as part of an utterance is accepted practice in several African languages
- Especially numbers, dates and money amounts
- English alternative is often shorter
  - The number “2353”
    (*Two thousand three hundred and fifty three*)
  - In Xhosa this is:
    Amawaku amabini namakhulu amathathu namashumi amahlanu nantathu
  - Which means literally:
    Thousands that-are-two and hundreds-that-are-three and tens-that-are-five and three

CODE MIXING AND SWITCHING

- Including English words or phrases as part of an utterance is accepted practice in several African languages
- Especially numbers, dates and money amounts
- English alternative is often shorter
    - The number “2353”
      (*Two thousand three hundred and fifty three*)
    - In Xhosa this is:
      Amawaku amabini namakhulu amathathu namashumi amahlanu nantathu
    - Which means literally:
      Thousands that-are-two and hundreds-that-are-three and tens-that-are-five and three

- This study considers the effect of code-mixing on the accuracy of automatic accent-identification systems
- How accurately can the mother tongue of Xhosa and Zulu speakers be determined:
  1. When the English is part of a mixed code.
  2. When the English is part of a monolingual dialogue

BACKGROUND

This work was motivated by two previous studies:

1. Language identification for Xhosa and Zulu
    - Even in mixed-code utterances with mostly English words language could be identified with 70% accuracy
2. Accent identification for Nguni and Sotho MT speakers
    - Two largest language families
    - Share similar vowel systems
    - Automatic and perceptual tests showed that neither humans nor machines were able to classify accents reliably

Contradiction?

- Nguni and Sotho accents could not be accurately classified
- Xhosa and Zulu accents (both Nguni) could be distinguished

In the second case the English was embedded in a mixed code, in the first it was not.

Contradiction?

- Nguni and Sotho accents could not be accurately classified
- Xhosa and Zulu accents (both Nguni) could be distinguished

In the second case the English was embedded in a mixed code, in the first it was not.

DATABASES

- AST corpora: recorded and annotated telephone speech
- Phonetically-diverse mix of read and spontaneous speech
- Speakers from targeted language & accent groups
- Five languages: Afrikaans, English, Sesotho, Xhosa, Zulu
- Separate English corpora for five accent groups: Afrikaans, English, Coloured, Indian and Black speakers
- We have used the Xhosa and the Zulu corpora, as well as the English by Black speakers
- Code-mixing and switching is very common in the Xhosa and Zulu corpora

BLACK ENGLISH CORPUS

- Mother tongues in Black English (BE) corpus are known:

Mother tongue	% of speakers
Xhosa	23
Zulu	18
Sesotho	23
Tswana	32
Other	4

- Extract subsets due to Xhosa and Zulu speakers: XBE & ZBE

Name	Mins.	Utts.	Spkrs.	Phones
XBE	23.8	614	17	9 112
ZBE	25.8	643	16	9 841

- This testing material is free of code switching and mixing

BLACK ENGLISH CORPUS

- Mother tongues in Black English (BE) corpus are known:

Mother tongue	% of speakers
Xhosa	23
Zulu	18
Sesotho	23
Tswana	32
Other	4

- Extract subsets due to Xhosa and Zulu speakers: XBE & ZBE

Name	Mins.	Utts.	Spkrs.	Phones
XBE	23.8	614	17	9 112
ZBE	25.8	643	16	9 841

- This testing material is free of code switching and mixing

XHOSA AND ZULU CORPORA

- Divide Xhosa (XX) and Zulu (ZZ) corpora into:
- Training sets ...

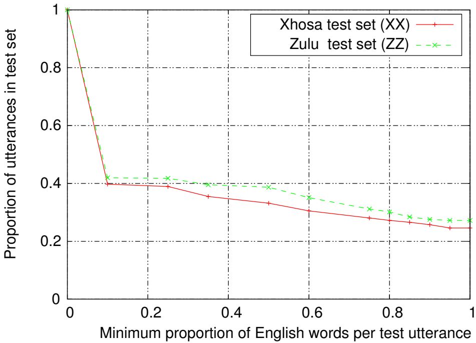
Corpus name	Speech (h)	No. of utts.	No. of spkrs.	Phone tokens
XX	6.98	8 538	219	177 843
ZZ	7.03	8 295	203	187 249

- Test sets ...

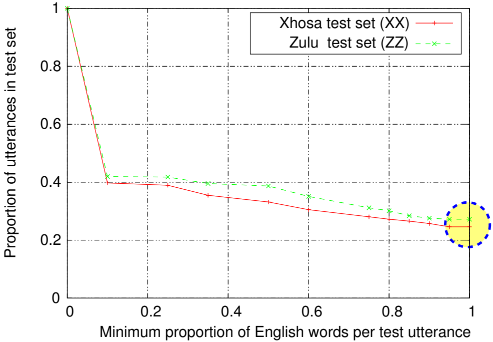
Corpus name	Speech (min)	No. of utts.	No. of spkrs.	Phone tokens
XX	26.8	609	17	10 925
ZZ	27.1	583	16	11 008

- Code mixing and switching is frequent in these test sets.

CODE MIXING IN XHOSA AND ZULU CORPORA



CODE MIXING IN XHOSA AND ZULU CORPORA



- Extract subsets containing only English words: XXE and ZZE

BE

English utterances spoken by Black mother–tongue speakers

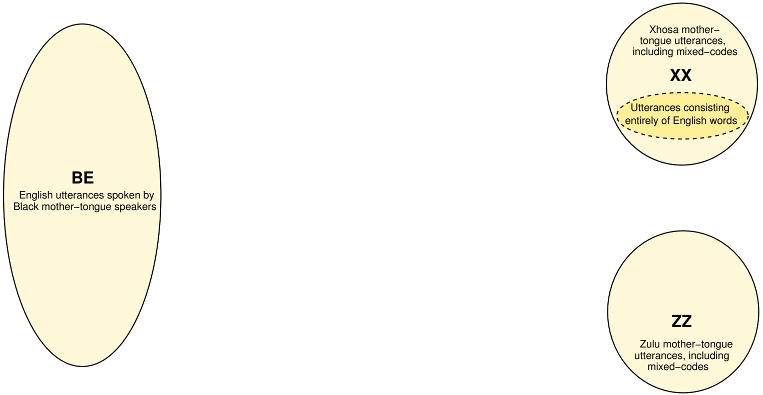
Xhosa mother–tongue utterances, including mixed–codes

XX

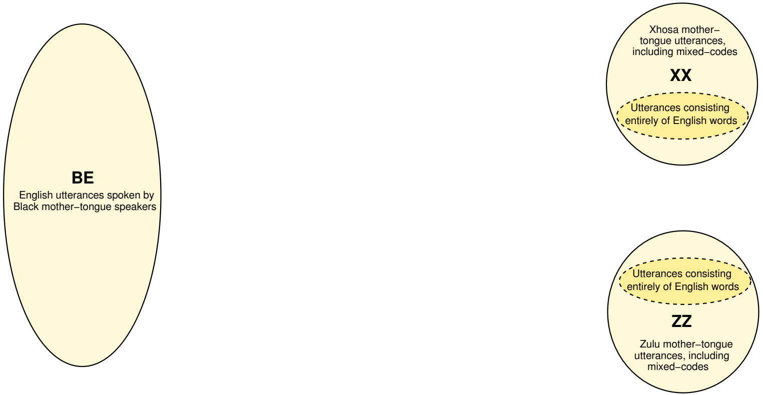
ZZ

Zulu mother–tongue utterances, including mixed–codes

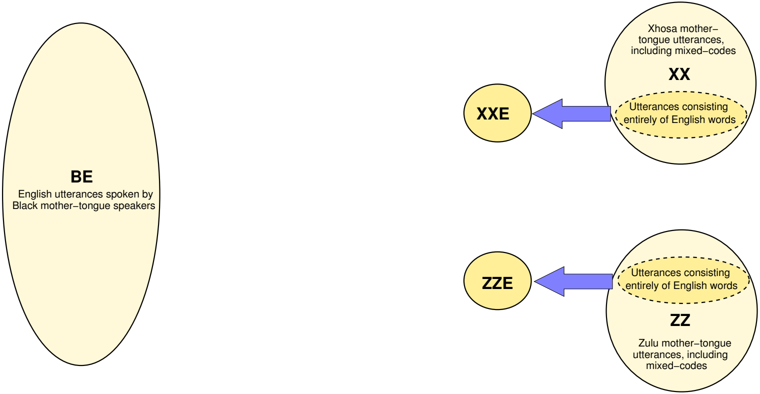
Start off with three test sets ...


BE
English utterances spoken by Black mother–tongue speakers
Xhosa mother–tongue utterances, including mixed–codes
XX
Utterances consisting entirely of English words
ZZ
Zulu mother–tongue utterances, including mixed–codes

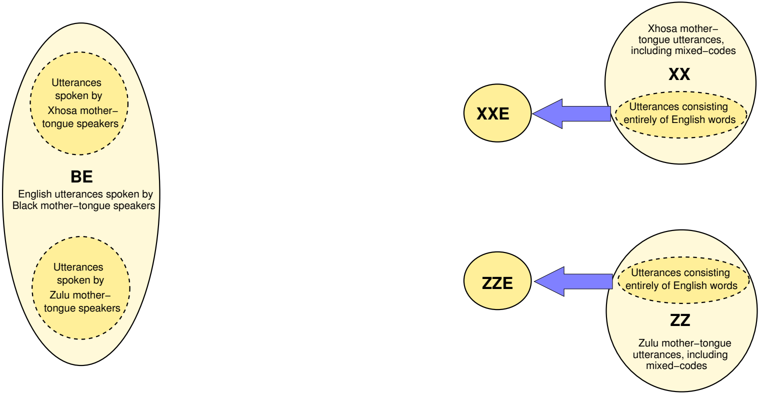
Identify English within Xhosa ...


BE
English utterances spoken by Black mother–tongue speakers
Xhosa mother–tongue utterances, including mixed–codes
XX
Utterances consisting entirely of English words
Utterances consisting entirely of English words
ZZ
Zulu mother–tongue utterances, including mixed–codes

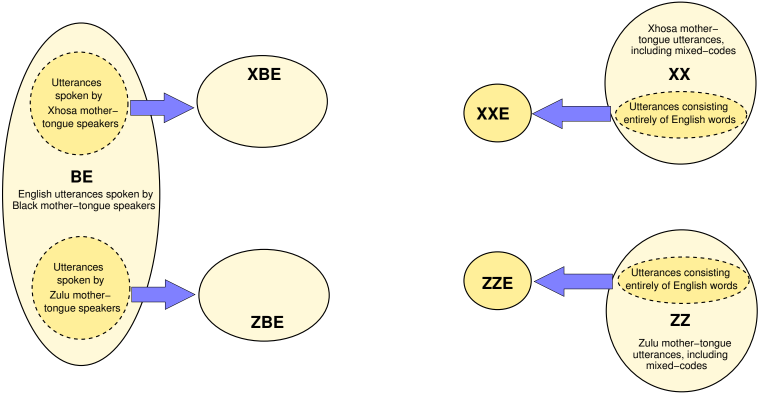
Identify English within Zulu ...


BE
English utterances spoken by Black mother–tongue speakers
Xhosa mother–tongue utterances, including mixed–codes
XX
Utterances consisting entirely of English words
XXE
Utterances consisting entirely of English words
ZZ
Zulu mother–tongue utterances, including mixed–codes
ZZE

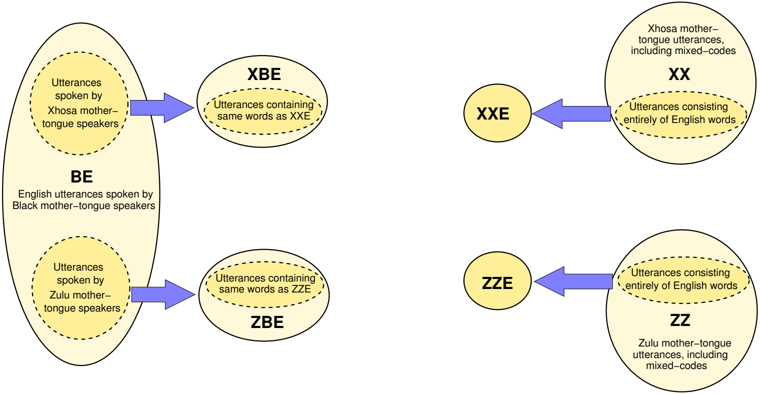
Name these subsets XXE and ZZE.


Utterances spoken by Xhosa mother–tongue speakers
BE
English utterances spoken by Black mother–tongue speakers
Utterances spoken by Zulu mother–tongue speakers
Xhosa mother–tongue utterances, including mixed–codes
XX
Utterances consisting entirely of English words
XXE
ZZE
Utterances consisting entirely of English words
ZZ
Zulu mother–tongue utterances, including mixed–codes

Identify Xhosa and Zulu speakers in BE ...


Utterances spoken by Xhosa mother–tongue speakers
XBE
BE
English utterances spoken by Black mother–tongue speakers
Utterances spoken by Zulu mother–tongue speakers
ZBE
Xhosa mother–tongue utterances, including mixed–codes
XX
XXE
Utterances consisting entirely of English words
ZZE
Utterances consisting entirely of English words
ZZ
Zulu mother–tongue utterances, including mixed–codes

Name these subsets XBE and ZBE.



Utterances spoken by Xhosa mother–tongue speakers

BE

English utterances spoken by Black mother–tongue speakers

Utterances spoken by Zulu mother–tongue speakers

XBE

Utterances containing same words as XXE

Utterances containing same words as ZZE

ZBE

Xhosa mother–tongue utterances, including mixed–codes

XX

Utterances consisting entirely of English words

XXE

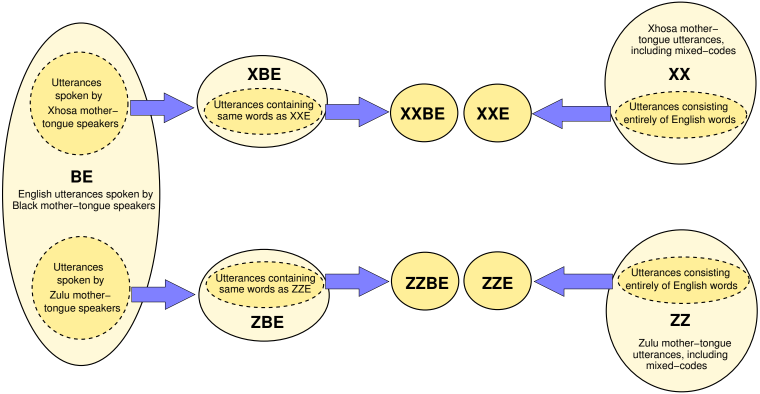
ZZE

Utterances consisting entirely of English words

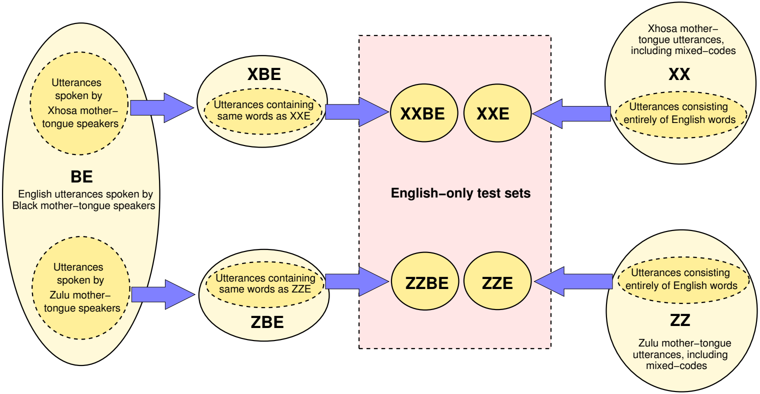
ZZ

Zulu mother–tongue utterances, including mixed–codes

Identify subsets containing same words as XXE and ZZE ...


Utterances spoken by Xhosa mother–tongue speakers
BE
English utterances spoken by Black mother–tongue speakers
Utterances spoken by Zulu mother–tongue speakers
XBE
Utterances containing same words as XXE
XXBE
XXE
Xhosa mother–tongue utterances, including mixed–codes
XX
Utterances consisting entirely of English words
Utterances containing same words as ZZE
ZBE
ZZBE
ZZE
Utterances consisting entirely of English words
ZZ
Zulu mother–tongue utterances, including mixed–codes

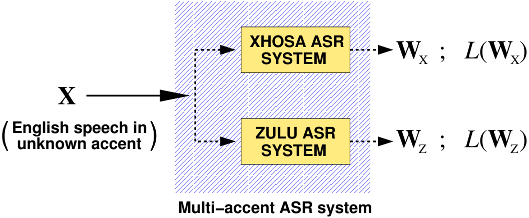
Name these XXBE and ZZBE


Utterances spoken by Xhosa mother–tongue speakers
BE
English utterances spoken by Black mother–tongue speakers
Utterances spoken by Zulu mother–tongue speakers
XBE
Utterances containing same words as XXE
Utterances containing same words as ZZE
ZBE
XXBE
XXE
English–only test sets
ZZBE
ZZE
Xhosa mother–tongue utterances, including mixed–codes
XX
Utterances consisting entirely of English words
Utterances consisting entirely of English words
ZZ
Zulu mother–tongue utterances, including mixed–codes

Comparable test sets with and without mixed codes

ACCENT IDENTIFICATION

- Use Parallel Phone Recognition followed by Language Modelling (PPRLM)


X
(English speech in unknown accent)
XHOSA ASR SYSTEM
ZULU ASR SYSTEM
$\mathbf{W}_X \; ; \; L(\mathbf{W}_X)$
$\mathbf{W}_Z \; ; \; L(\mathbf{W}_Z)$
Multi–accent ASR system

- Each recogniser has language-specific acoustic and language models

ACOUSTIC MODELS

- Use Xhosa and Zulu (XX and ZZ) training sets to obtain acoustic models with HTK
- Common set of 90 phones
- 99.5% coverage of phones in Black English (BE) corpus
- Parameterisation: 12 MFCCs and energy, with Δ and $\Delta\Delta$
- Cross-word triphones using decision-tree clustering, 8 mixtures per state and diagonal covariances
- Approximately 1250 clustered states per accent

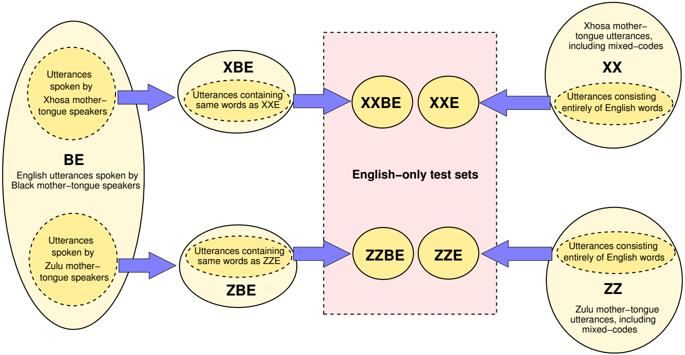
RESULTS: PHONE LOOP

Test	Classified as (%)	
corpus	Xhosa	Zulu
XXE	75.0	25.0
ZZE	29.1	70.9
Average correct	73.4%	
XXBE	48.9	51.1
ZZBE	37.4	62.6
Average correct	55.4%	

- English drawn from Xhosa and Zulu test-sets classified with 73.4% accuracy
- English drawn from Black English test data classified with 55.4% accuracy

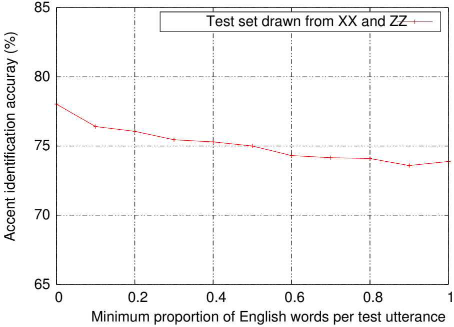
EFFECT OF % OF ENGLISH WORDS

- Proportion of English words per utterance in the test sets drawn from the XX and ZZ test-sets can be varied


Utterances spoken by Xhosa mother–tongue speakers
BE
English utterances spoken by Black mother–tongue speakers
Utterances spoken by Zulu mother–tongue speakers
XBE
Utterances containing same words as XXE
Utterances containing same words as ZZE
ZBE
XXBE
XXE
English–only test sets
ZZBE
ZZE
Xhosa mother–tongue utterances, including mixed–codes
XX
Utterances consisting entirely of English words
Utterances consisting entirely of English words
ZZ
Zulu mother–tongue utterances, including mixed–codes

EFFECT OF % OF ENGLISH WORDS

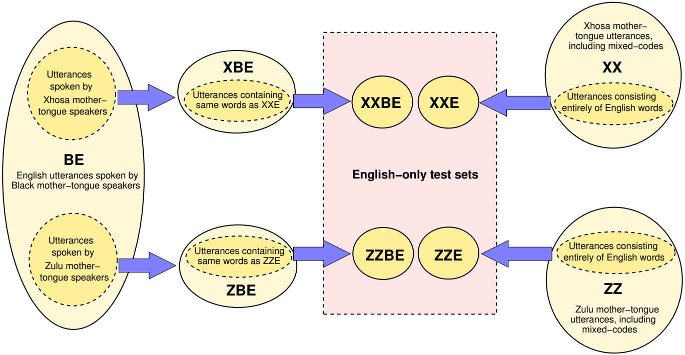
- Proportion of English words per utterance in the test sets drawn from the XX and ZZ test-sets can be varied



- Accuracy improves as proportion of Xhosa/Zulu words rises

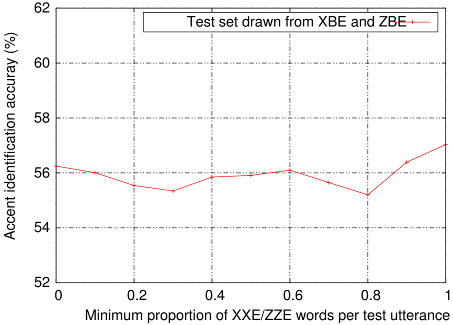
EFFECT OF WORD CHOICES

- Allow vocabularies of the mixed and unilingual test sets to differ


Utterances spoken by Xhosa mother–tongue speakers
BE
English utterances spoken by Black mother–tongue speakers
Utterances spoken by Zulu mother–tongue speakers
XBE
Utterances containing same words as XXE
Utterances containing same words as ZZE
ZBE
XXBE
XXE
English–only test sets
ZZBE
ZZE
Xhosa mother–tongue utterances, including mixed–codes
XX
Utterances consisting entirely of English words
Utterances consisting entirely of English words
ZZ
Zulu mother–tongue utterances, including mixed–codes

EFFECT OF WORD CHOICES

- Allow vocabularies of the mixed and unilingual test sets to differ



- No systematic effect on accuracy

RESULTS: BIGRAM

Test corpus	Classified as (%)	
	Xhosa	Zulu
XXE	79.7	20.3
ZZE	29.7	70.3
Average correct	74.8%	
XXBE	56.5	43.5
ZZBE	42.1	57.9
Average correct	57.2%	

- Small improvement (1-2%) in accuracy relative to phone loop
- Performance gap persists

CONCLUSIONS

- It is not possible to distinguish reliably between Xhosa and Zulu accented English when the utterances form part of a monolingual dialogue
- It is possible to distinguish between Xhosa and Zulu accented English with much better accuracy when the English is embedded in the Xhosa/Zulu as part of a mixed code
- Hence English that is part of a mixed code exhibits a much stronger accent than monolingual English produced by the same type of speaker
- For ASR this implies that:
  - Single acoustic model appropriate for monolingual English
  - Accent-specific acoustic models more appropriate for mixed codes