Data-driven Phonetic Comparison between South African, British and American English Pronunciations

Linsen Loots and Thomas Niesler
Department of Electronic Engineering, Stellenbosch University, South Africa

1. Introduction

Pronunciation dictionaries:

- Accent-specific: improve ASR accuracy
- Prohibitively **expensive** for under-resourced accents (like South African English)

G2P conversion supplements dictionaries, but

- has limited accuracy
- needs a large dictionary for training

Phonetic transcriptions compared in:

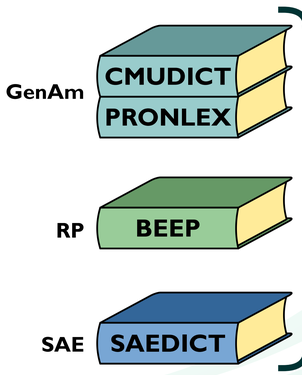
- **American** English (GenAm)
- **British** English (RP)
- **South African** English (SAE)

Decision trees are used for G2P, to analyse individual accents and to convert between them

AIM Analyse different accents and determine how best to derive pronunciations in a new accent

2. Dictionaries

Four dictionaries used for the three accents:



23,034 words in common

ARPABET

UK spelling

SAEDICT, under development at Stellenbosch University, has 36,956 entries. The others have between 90,000 and 250,000

3. G2P Conversion

Decision trees require **one-to-one alignment** between graphemes and phonemes:

Graphemes:	e	x	t	r	e	m	e
Phonemes:	eh	k+s	t	r	iy	m	-

Decision trees are **grown** recursively with node questions chosen to maximise information gain

Information entropy:

$$H(X) = -\sum p_i \log(p_i)$$

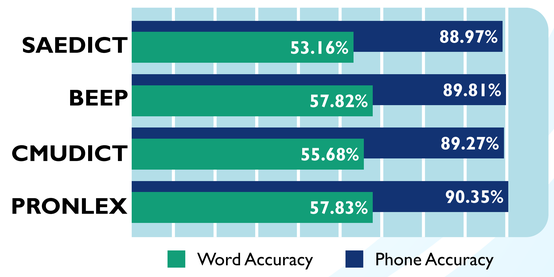
Information gain:

$$\Delta i = i(t) - p_L i(t_L) - p_R i(t_R)$$

This is equivalent to maximising:

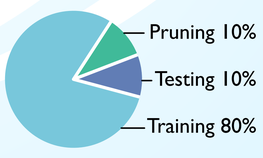
$$\Delta i = \sum_{k}^{children} \sum_{p}^{phonemes} N_{t_k,p} \log\left(\frac{N_{t_k,p}}{N_{t_k}}\right)$$

4. G2P Results



10-Fold cross-validation;

Accuracy: $\frac{N_c - N_i}{N_t}$

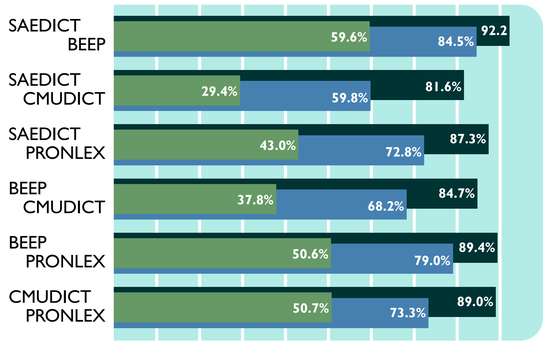


5. Phonetic Comparison

Phonetic transcriptions were aligned to compare pronunciations directly:

SAEDICT:	r	ih	ae	k	sh	ax	n	s
BEEP:	r	ih	ae	k	sh	-	n	z

Phonetic Correspondence:



Results give the average of using each of the two dictionaries as reference. Consonants are not shown: at least 93.7% match for all accent pairs.

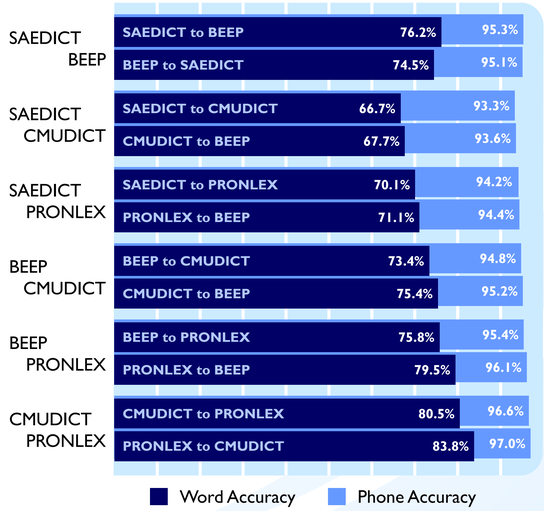
Phoneme Shifts

Phoneme confusions support linguistic theory:

- GenAm is **rhotic** - /r/'s are pronounced that are silent elsewhere: *farm, where*
- **Later Yod Dropping** in GenAm - /y/'s are dropped: *tune, duke, new*
- **Syllabic consonants** in RP, not a schwa and consonant: *bubble, sudden*
- **THOUGHT/LOT** merge in GenAm: same pronunciation for *cot, caught*
- SAE has /ih/ where RP and GenAm use /iy/: *happy, barrier*
- **Schwas** match poorly, due to stress shifts

6. Accent Conversion

G2P techniques were applied to convert directly between two accents' pronunciations:



Confusions show the same patterns as phonetic comparison. The GenAm dictionaries convert well, as do SAEDICT and BEEP.

7. Conclusions

It is better to derive pronunciations from a different accent than by G2P within an accent

For **SAE** it is clear that RP pronunciations are most similar and would provide a good source

Using this approach, almost **80%** of words are correctly predicted; for G2P this is below **60%**

Further research includes listening tests and ASR, to test the perceptual and acoustic impact

RESULT Converting from one accent to another is substantially more accurate than G2P

Presented at: Interspeech 2009, 6 - 10 September 2009, Brighton, UK