

# Collecting and evaluating speech recognition corpora for nine Southern Bantu languages

Jaco Badenhorst, Charl van Heerden, Marelle Davel and  
Etienne Barnard

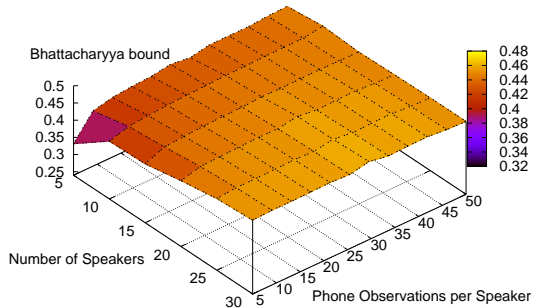
January 27, 2010

# Introduction

- Lwazi ASR corpus:
  - Approximately 200 speakers per language
  - About 350 speech minutes per language
- Goal:
  - Understand data requirements to develop a minimal system that is practically usable
- Investigated:
  - Stability of acoustic models
  - Potential for cross-language sharing
  - Phone recognition results

# Analysis of phone variability

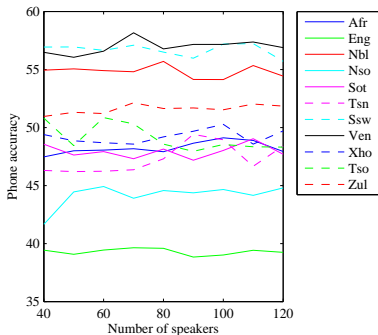
*Speaker-and-utterance three-dimensional plot for the siSwati nasal /m/*



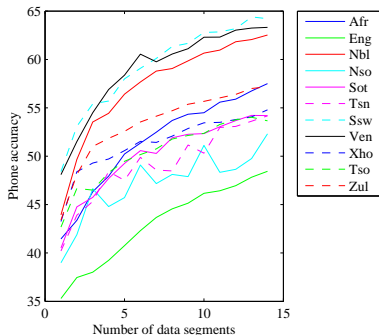
- We report only on single-mixture context-independent models (similar trends observed for more complex models)

# Phone recognition accuracy

## Number of speakers

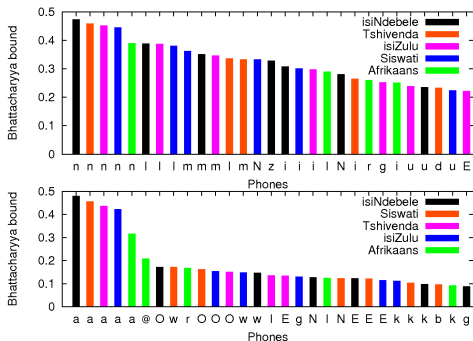


## Amount of data



# Distances between phones

*Effective distances for isiNdebele phone /a/ and /n/ and their closest matches*



- Based upon proven stability of our phone models:
  - Phone similarity between phones across languages

# Conclusion

- New method to determine data sufficiency
- Confirmed that different phone classes have different data requirements
- Our results suggest that similar phone accuracies may be achievable by using more speech from fewer speakers
- Based upon proven model stability we performed successful measurements of distances between phones of different languages
- Publications available:  
[http://www.meraka.org.za/hlt\\_publications.htm](http://www.meraka.org.za/hlt_publications.htm)