



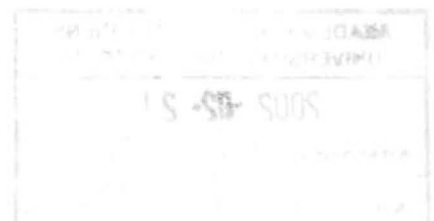
A comparative acoustic analysis of the long vowels and diphthongs of Afrikaans and South African English

by

Claude Pierre Prinsloo

Submitted in partial fulfilment of the requirements for the degree
Master of Engineering (Computer Engineering)
in the Faculty of Engineering
UNIVERSITY OF PRETORIA

August 2000



Abstract

Keywords: Automatic speech recognition, formant extraction, statistical formant analysis, acoustic phonetics, Afrikaans and South African English vowels and diphthongs, pronunciation dictionaries.

In this study, the long vowels and diphthongs of Afrikaans and South African English are acoustically compared.

The results of this study are important to linguists in the understanding, teaching and correction of language and the development of pronunciation dictionaries. This is particularly important in the South African context where many speakers use more than one of the eleven official languages on a regular basis¹. Further importance lies in the use of these acoustical models as a means of improving automatic speech recognition, realistic computer-speech generation and automatic accent recognition. The study will also aid in the study of acoustic phonetics.

The experimental work was performed on a database that was collected of the long vowels and diphthongs of the two languages as spoken by mother-tongue (L1) and second language (L2) speakers for four cases as shown in the table.

The vowels and diphthongs were recorded in single utterances of normal words, in isolated form and in a “pseudo-word” (h-vowel-t or h-diphthong-t) form. The data was

¹According to the 1996 census: 23% of South Africans speak isiZulu, 18% isiXhosa, 14% Afrikaans and 9% English as their first home language.



Afrikaans spoken by Afrikaans speakers	L1 Afrikaans
Afrikaans spoken by SA English speakers	L2 Afrikaans
SA English spoken by SA English speakers	L1 SA English
SA English spoken by Afrikaans speakers	L2 SA English

then verified, segmented and labelled. The relevant vowels and diphthongs were then extracted and compared.

The comparison between L1 and L2 speech is based on the formant locations and the formant tracks. This involved the calculation of the resonance peak tracks (formants) of the voiced speech, visual verification of the formant tracks and then producing a graphical representation of the locations/trajectories of the vowels/diphthongs. The significance of any difference in mean location/trajectory was tested using the analysis of variance (ANOVA) statistical test.

Comparative experiments based on the pitch trajectories of the vowels and diphthongs were also performed. Finally, the level of diphthongization of the vowels and diphthongs was analysed.

The comparative experiments mostly confirm common hypotheses on the equivalence and difference between the two South African accent/language groups, but some of the findings challenge traditional views. One of these challenges arise over the classification of <o:> and <e:> as diphthongs and not as vowels as is commonly done.

Uittreksel

Sleutelwoorde: Outomatiese spraakherkenning, formant ontrekking, statistiese formant analise, akoestiese fonetiek, Afrikaanse en Suid-Afrikaanse Engelse vokale en diftonge, uitspraak woordeboeke.

In hierdie studie word die lang vokale en diftonge van Afrikaans en Suid-Afrikaanse Engels akoesties vergelyk.

Die resultate van die studie is belangrik vir taalkundiges in die verstaan, onderrig en korreksie van taal en die ontwikkeling van uitspraak-woordeboeke. Dit is van besondere belang in die Suid-Afrikaanse konteks waar baie sprekers meer as een van die elf amptelike tale op 'n gereelde basis gebruik². Verdere belang lê in die gebruik van die akoestiese modelle as 'n manier van verbetering van outomatiese spraakherkenning, realistiese rekenaar-spraak generasie en outomatiese aksentherkenning. Die studie sal ook bydra tot die bestudering van akoestiese fonetiek.

Die eksperimentele werk is uitgevoer op 'n versamelde databasis van die lang vokale en diftonge van die twee tale soos gepraat deur moedertaal (L1) en tweedetaal (L2) sprekers in vier gevalle soos in die tabel aangedui.

Die vokale en diftonge is in enkele uitinge van normale woorde, in geïsoleerde vorm en in “pseudo-woord” (h-vokaal-t of h-diftong-t) vorm opgeneem. Die data is dan nagegaan,

²Volgens die 1996 sensus: 23% van Suid-Afrikaners praat isiZulu, 18% isiXhosa, 14% Afrikaans en 9% Engels as hulle eerste huistaal.



Afrikaans gepraat deur Afrikaanse sprekers	L1 Afrikaans
Afrikaans gepraat deur SA Engelse sprekers	L2 Afrikaans
SA Engels gepraat deur SA Engelse sprekers	L1 SA Engels
SA Engels gepraat deur Afrikaanse sprekers	L2 SA Engels

gesegmenteer en gemerk. Die relevante vokale en diftonge is toe onttrek en vergelyk.

Die vergelyking tussen L1 en L2 spraak is gebaseer op die formantplasinge en die formanttrajekte. Dit behels die berekening van die resonansiepieke (formante) van stemhebbende spraak, visuele nagaan van die korrektheid van die formanttrajekte en dan grafiese voorstelling van die plasing/trajekte van die vokale/diftonge. Die betekenisvolheid van enige verskille in gemiddelde plasing/trajek is dan deur middel van 'n analise van variansie (ANOVA) statistiese toets, beproef.

Vergelykende eksperimente is ook op die stemtoon trajekte van die vokale en diftonge gedoen. Laastens is die mate van diftongisering van die vokale en diftonge geanaliseer.

Die vergelykende eksperimente het meerendeels die algemene hipoteses oor die ekwivalensie en verskille tussen die twee Suid-Afrikaanse aksent/taal groepe bevestig, maar sommige van die bevindings bevraagteken tradisionele standpunte. Veral die klassifikasie van <o:> en <e:> as diftonge en nie as vokale nie, word bevraagteken.



Acknowledgements

I would like to thank the following people for their help and support, without which this project would not have been possible:

- Professor Liesbeth Botha, my study leader and mentor.
- Doctor Hendrik Boshoff, for his advice and explanations on vowel space.
- My parents, for their unfailing support.
- Last, but surely not least, the Lord, who is always at my side.



Contents

1	Introduction	1
1.1	Justification	1
1.2	Background	5
1.3	Method	7
1.4	Contributions of this dissertation	8
1.5	Organisation of this dissertation	9
2	Theoretical Framework	12
2.1	Vowels	13
2.2	Diphthongs	19
2.3	Spectrograms	24
2.4	Formants	26
2.4.1	Linear prediction coefficients	31

2.5	Pitch	39
2.6	Equivalence classification	41
2.7	Cubic splines	42
2.8	Diphthongization	45
2.9	Statistics: Tests of hypotheses and significance	46
2.9.1	Analysis of variance (ANOVA) test	46
3	Experiments	49
3.1	Objectives	49
3.2	Data	51
3.2.1	Data structure	52
3.3	Method	58
3.3.1	Data recording and verification	58
3.3.2	Formant extraction	62
3.3.3	Pitch extraction	62
3.3.4	Data visualisation and comparison	64
3.4	Results	66
3.4.1	Long vowel results	66
3.4.2	Diphthong results	77

3.4.3	Diphthongization of <e:> and <o:>	86
3.4.4	Long vowel and diphthong results - ratios	87
3.4.5	Long vowel and diphthong diphthongization results	94
3.4.6	Pitch results	97
4	Summary and conclusion	105
4.1	Summary of results	106
4.2	Shortcomings and future work	110
A	Appendix	111
A.1	Formant extraction using LPC - Matlab	111
A.1.1	Autocorrelation	111
A.1.2	Durbin recursion	112
A.1.3	Formant extraction	113
A.2	Pitch extraction using autocorrelation	114
A.3	Pitch trajectories	117
A.3.1	Vowel pitch trajectories	117
A.3.2	Diphthong pitch trajectories	117
A.4	Expanded formant plots	117



A.4.1	Expanded vowel formant plots	117
A.5	Compact Disk Contents	122
	Bibliography	125