Institute of Phonetic Sciences,
University of Amsterdam,
Proceedings 20 (1996), 65 - 80.

TOWARDS MODELLING ACOUSTIC DIFFERENCES BETWEEN L1 AND L2 SPEECH: THE SHORT VOWELS OF AFRIKAANS AND SOUTH-AFRICAN ENGLISH

*Liesbeth Botha**

Abstract

The acoustic differences between Afrikaans and South-African English, spoken as L1 and L2, are measured for nine short vowels. The spoken language data base of 22 male speakers, collected for comparative studies, is described. The features used in an initial comparison of the isolated vowels and vowels in CVC words are the first three formant values and ratios. Significant differences are found between the production of /ɛ/ and /y/ by Afrikaans and English mother-tongue speakers, and to a lesser extent between /i/, /ə/ and /u/. Several interesting trends that seem to contradict popular beliefs concerning South African accents are observed. Directions for future research and the application of the envisioned L1-L2 model in speech technology are given.

1. Introduction

Second-language (L2) speech is interesting from a number of linguistic as well as technical points of view. Linguistic issues for instance involve L2 perception and production in acquisition, while technical aspects are mainly concerned with speech technology, i.e. how to make artificial speech "sound like" an L2 (in speech generation systems), or how to improve automatic speech recognition (ASR) system performance on L2 speech.

Humans can recognise and identify a foreign accent (i.e. a pronunciation of the same words that is different from their own pronunciation, and which is often characteristic of the mother tongue of the speaker) quite easily in informal circumstances, and some studies have quantized this (Flege, 1984; Flege et al., 1995a; Flege et al., 1995b; Flege, 1987). Although humans probably use many cues (for instance linguistic contexts and knowledge of the world) to identify an accent, most of the scientific studies are performed purely on acoustic prompts. It can probably be safely assumed that most of the information regarding the foreign accentedness is embedded in the

* Visiting research scientist from: Department of Electrical and Electronic Engineering, University of Pretoria, Pretoria 0002, South Africa.
liesbeth.botha@ee.up.ac.za

acoustic signal. In addition, state-of-the-art speech recognition systems (Bourlard et al., 1996; Barnard et al., 1995; Young, 1996) use only an acoustic signal as input.

In this study we aim to model the acoustic differences between L1 and L2 speech signals. The long-term objectives are to model these acoustic differences and to apply this model to speech recognition systems, so that they can deal with foreign accents in a structured way. Currently ASR systems have to be trained specifically on data of all the accents (variations) with which they are expected to deal.

Our short-term goal is to quantify the acoustic differences between two of the many languages spoken in South-Africa, namely Afrikaans and English, spoken as L1 and L2. Although we have no empirical data to support this statement, it is generally possible to distinguish between an Afrikaans and English (mother-tongue) accent in South Africa, i.e. it is possible to tell whether the language (Afrikaans or English) is spoken as an L1 or an L2. It is this distinction that we now want to quantify by studying the production of the nine short vowels /a/, /æ/, /ɛ/, /i/, /ə/, /œ/, /u/, /y/ and /ɔ/, when spoken as L1 and as L2. The /y/ does not occur in English, and can thus only be studied as an L2 sound when pronounced by English mother-tongue speakers and as an L1 sound when produced by Afrikaans mother-tongue speakers. The four conditions, namely Afrikaans and English texts spoken as L1 and L2, respectively, are visualised as follows:

	L1	L2
Afrikaans text	Cond-Afr-L1	Cond-Afr-L2
English text	Cond-Eng-L1	Cond-Eng-L2

The paper is organised as follows: in Section 2 we discuss some of the relevant literature and give some background information about the languages of South Africa. Sections 3 and 4 contain a description of the database and of the methods used, respectively. In Section 5 we discuss our preliminary results and we conclude with suggestions for projects as a continuation of this work in Section 6.

2. Background

2.1. Second language studies

From a linguistic point of view, second languages have been studied on many levels, and still it seems that the mechanisms are not well-understood. An excellent overview of the issues involved in second language speech can be found in Leather and James (1996). In the empirical research of psycho-acoustics, Flege has by far been the most prolific. The literature abounds with perceptual studies (Fox et al., 1995; Flege et al., 1995b; Schouten, 1975) as well as studies on the production of L2 sounds (Flege et al., 1995c; Flege, 1987), often in a comparative framework (Flege and Eefting, 1987). However, a systematic comparison between all the sounds of two languages involved (L1 and L2) can generally not be found. This is probably due to the magnitude of such a task, if the detailed methods of the phonetic sciences (Shearer, 1982) are used.

The field of automatic speech recognition has matured to the point where useful commercial applications are showing up. These successful systems are built using statistical methods, both for acoustics and linguistics. The fact that phonetic and linguistic knowledge is not explicitly used in these systems, and the reasons and remedies for this situation are topics of lively discussion in the speech community, see for instance Bourlard et al. (1996). The issue in this paper is not to debate the un/desirability of this situation, but to speculate whether a statistical comparison of an L1 and L2 would lead to a useful model - one that can be used to improve the performance of an automatic speech recognition system on an accent not represented in its set of training data. It is our aim to investigate this, and the current project is the first small step in that direction. The formant values used in our comparison are not used explicitly in ASR systems either, but they form a useful tool for our investigation. The long-term goal would be to find a transformation (based on the L1-L2 model) for the parameters of an ASR system trained on L1 data, so that the ASR system can deal with L2 speech. This goal has been formulated before with reference to the robustness of ASR systems, but has been elusive.

2.2. Languages used in this study

The following paragraphs are somewhat speculative, but a lack of empirical data dictates this style. White urban South Africans are mostly bilingual in Afrikaans and South-African English. The latter language will sometimes be referred to merely as English in the scope of this paper. This bilingualism is not symmetric, in that Afrikaans mother-tongue speakers are better in their L2 (English) than the SA English mother-tongue speakers are in their L2 (Afrikaans). The hypotheses for this situation are beyond the scope of this paper. Both languages are taught for all twelve years of primary and secondary school.

Afrikaans originates from 17th century Dutch. The language was introduced when Jan van Riebeeck arrived at the Cape in 1652. During the past 350 years many other languages such as English, French, German and Malaysian have influenced Afrikaans, which was officially recognised as a separate language (apart from Dutch) in 1925 in South Africa. Although there are nine other official languages in South Africa today (from the Nguni, Sotho and Venda language groups), and although these were also spoken in SA throughout the past 350 years, they have had a small influence on Afrikaans.

There are approximately six million mother-tongue speakers of Afrikaans and three million mother-tongue speakers of English presently living in SA. With such a large number of speakers to draw from, a comprehensive comparison of these accents seems plausible.

The literature on acoustic phonetics of both Afrikaans and SA English is sparse, but several introductory texts in Afrikaans phonetics exist, e.g. Coetzee (1981) and De Villiers and Ponelis (1987). Values for the formants of the Afrikaans vowels were published in Taylor and Uys (1988) and Van der Merwe et al. (1993). In De Villiers and Ponelis the authors remark that at that time (1987) almost no acoustic phonetic research (instrumental, or computer-based) had been performed in South Africa. This situation was also mentioned in Van der Merwe et al. (1993), and to the best of my knowledge, still exists today.

Very little has also been done in the field of Afrikaans and SA English speech technology. In the 1980's a text-to-speech system for Afrikaans (De Stadler, 1988) was constructed and some research on automatic syllabification and labelling (Prinsloo and Coetzer, 1990) was conducted. The 1990's in SA has so far been marked by very little research activity and few publications in this area. Some papers can be found in the proceedings of the yearly IEEE-supported South African Symposium on Communications and Signal Processing (COMSIG).

The present study is, inter alia, also part of an effort to revive speech technology research in SA, as commercial products are not and will not be fully language independent for a long time, and will have to be adapted for the SA languages and markets.

3. Data

3.1. Comprehensive data set

Due to the logistics of this project, conducted as part of a sabbatical at the Institute of Phonetic Sciences at the University of Amsterdam, all data had to be captured in South Africa before the actual research started in The Netherlands. A large amount of data was assembled, in order to provide for different contexts and in an attempt to capture all the phones used in standard Afrikaans and South-African English.

3.1.1. Subjects

Male speech is generally accepted to be easier to analyse than female voices, due to its lower pitch and smaller pitch variation. Male voices thus seemed to be a logical starting point. Twenty-two male speakers, of which eleven have Afrikaans and eleven English as native-language, all between the ages of 22 and 46, were invited to participate.

3.1.2. Texts

The data consist of the set of read texts as shown in Table 1. The type of text is shown (prose, sentence word list or sound list), a description of the text is given and the language of the text (Afrikaans or English) is indicated. Each text is pronounced and analysed as L1 when the language of the text corresponds to the mother-tongue of the speaker, and as L2 when the mother-tongue of the speaker differs from the language of the text. In this way, all four conditions (Cond-Afr-L1, Cond-Afr-L2, Cond-Eng-L1 and Cond-Eng-L2) as mentioned in Section 1 could be studied.

Table 1. Texts used in comprehensive database.

	Type	Description	Language
1.	Prose	Noordewind en die son	Afrikaans
2.	Prose	North wind and the sun	English
3.	Sentences	6 semantically nonsense	Afrikaans
4.	Sentences	10 TIMIT (ARPA, 1990): 2sa+3si+5sx	English
5.	Word list	39 words	Afrikaans
6.	Word list	52 words	English
7.	Word list	100 phonetically balanced	Afrikaans
8.	Word list	100 phonetically balanced	English
9.	Sound list	all vowels in context of /r,s,l,t,n/	Afrikaans
10.	Sound list	all consonants in context of /u,i,a/	Afrikaans
11.	Sound list	all vowels in /hVt/ context	Afrikaans
12.	Sound list	all vowels isolated and pronounced short	Afrikaans
13.	Sound list	all vowels isolated and pronounced long	Afrikaans

The *North wind and the sun* passages (IPA, 1966) in both English and Afrikaans were chosen for its popularity and widespread use in phonetic research. The 6 Afrikaans nonsense sentences were made up by the author to contain all the phonemes found in an introductory text on Afrikaans phonetics (Coetzee, 1981). They are syntactically correct, but semantically meaningless. The English TIMIT (ARPA, 1990) sentences are those spoken by subject mrds0 from dialect region 3 (a random choice). They contain two of the so-called "dialect" (sa) sentences, three phonetically diverse (si) as well as five phonetically compact (sx) sentences. The English and Afrikaans short word lists were taken from class notes of an introductory course in phonetics, used to illustrate the sound differences between English and Afrikaans. The lists of 100 Afrikaans and 100 English words were taken from material used in clinical speech pathology at the University of Pretoria. These words are considered "phonetically balanced" by the clinicians. The sound lists were made up by the author in an attempt to cover all individual sounds in Afrikaans, and with some regard to context and coarticulation effects. Clearly, only a small set of diphones and triphones are covered. The speakers were instructed to pronounce the "words" formed by the /r,s,l,t,n/, /u,i,a/ and /hVt/ contexts (lists 9, 10 and 11) as if they were Afrikaans words.

3.1.3. Recording

The data were recorded in the anechoic room of the Language Laboratory at the University of Pretoria, using a Beyer Dynamics M201N(C) microphone. The data were directly digitised at 11.125 kHz (16 bits) by a SoundBlaster card in an IBM compatible PC, using Microsoft Windows' Wave Studio software. The total database consists of about 350 MByte of data.

The data of each subject was captured in a single, one hour, recording session. The subjects received verbal instructions from the author regarding posture, voice loudness, paper noise, nature of the texts and goals of the research. They were allowed time to scan the texts before recording. A limited amount of stuttering followed by self-correction of the subjects was allowed, otherwise the individual text was rerecorded.

The prose, sentences and word lists texts were typed in normal orthographic symbols. For the sound lists, orthographic symbols of Afrikaans were mostly used, with a few phonetic symbols (e.g. ə) to resolve ambiguities. These were explained to the subjects.

3.2. Data used in experiments

In the three experiments conducted and subsequently described, only a small subset of the comprehensive database was used. Experiments planned with the rest of the data will be described in Section 6. In the selections described below, data of all 22 speakers were used.

The scope of this study was restricted to the nine short vowels of Afrikaans /a, i, ə, œ, u, ɛ, æ, y, ɔ/, of which all but /y/ also occur in English. The pronunciation of these vowels as L1 and L2 sounds were studied. This choice of short vowels in itself seems to be debatable. Taylor and Uys (1988) do not include /æ/ in their list of "steady-state" Afrikaans vowels, while Van der Merwe et al. (1993) do not consider /y/. In texts on Afrikaans phonetics, /e/, /ø/ and /o/ are also classified as monophthongs (Coetzee, 1981; De Villiers and Ponelis, 1987), but the acoustics of these sounds clearly show them as diphthongized, so they were not included for the moment.

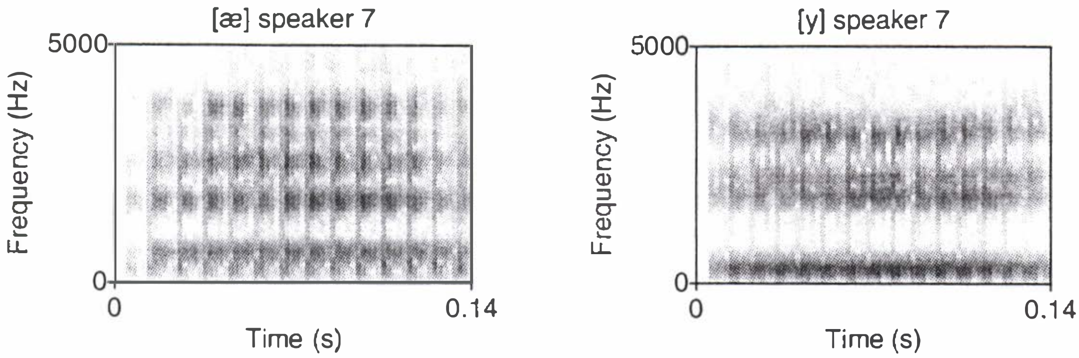

Figure 1. Spectrograms of /æ/ and /y/, showing them as steady-state vowels in Afrikaans.

Our inclusion of /æ/ and /y/ as steady-state vowels, as well as our omission of /e/, /ø/ and /o/ is motivated by the graphical display of their spectrograms, as shown in Figures 1 and 2.

In the first experiment, the vowels spoken short and in isolation were used. This is sound list 12 in Table 1. (Most of the vowels in Afrikaans occur in both long and short forms, and were recorded as such in sound lists 12 and 13). Since the speakers were instructed that these are Afrikaans sounds, only the two conditions Cond-Afr-L1 (the Afrikaans mother-tongue speakers) and Cond-Afr-L2 (the English mother-tongue speakers) were tested. The sounds were manually segmented from the background noise, by inspection of the waveform/oscillogram only. Since the sounds were spoken with pauses in between, there were clear energy changes on which to segment.

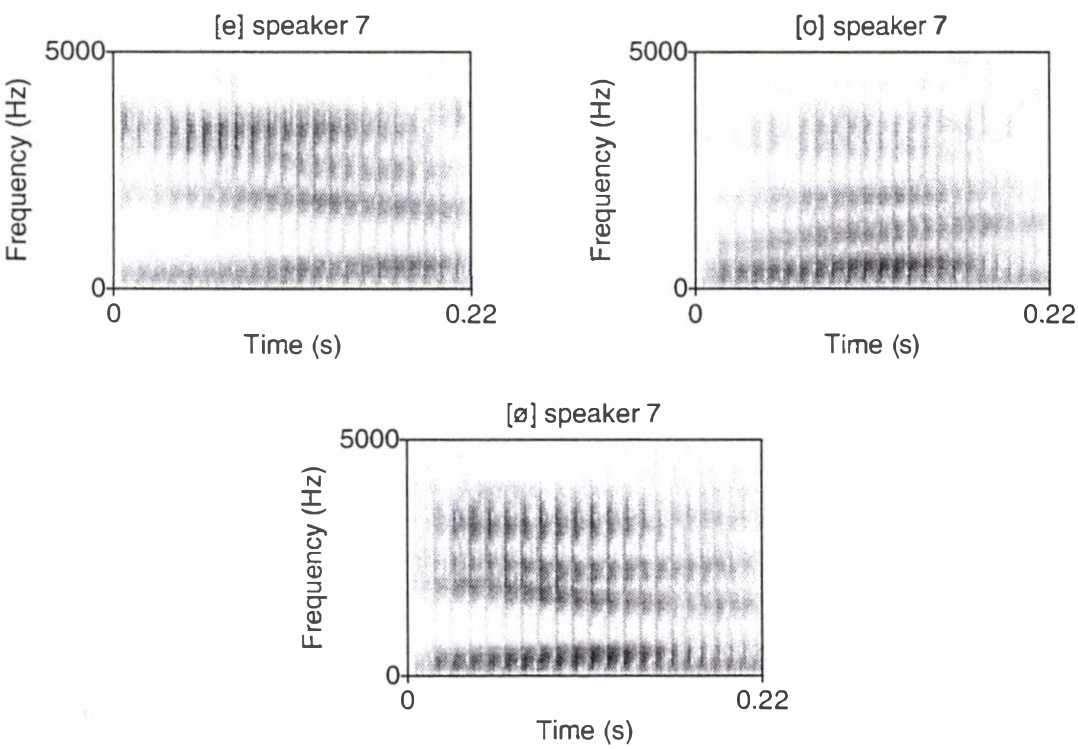

Figure 2. Spectrograms of /e/, /ø/ and /o/, showing their diphthongized nature in Afrikaans.

In the second experiment, Afrikaans words (from word list 5, Table 1), containing these nine vowels were chosen. These are: *was* [vas], *kiem* [kim], *is* [əs], *guns* [xœns], *roep* [rup], *ses* [sɛs], *vel* [fæl], *kos* [kɔs] and *debuut* [dəbyt]. Again, since these are Afrikaans words, Cond-Afr-L1 and Cond-Afr-L2 were investigated. The vowels were again manually segmented by inspecting both the waveform/oscillogram and spectrogram (as shown by the "show sonogram" of the Praat (Boersma, 1996) speech analysis software), to select the steady-state section of the vowel. In the case of coarticulation effects, as for instance in the word [fæl], only the initial steady-state of the vowel was retained. The segmentation was checked by displaying the formants (see Section 4) of each individual segmentation. Hardly any complications occurred.

In the third experiment, eight English words containing similar vowels (only /y/ does not occur in English) were chosen from word list 6 (Table 1) to test conditions Cond-Eng-L1 and Cond-Eng-L2, and similarly segmented. These words are: *dart* [dart], *feel* [fi:l], *about* [əbaut], *purr* [pœr], *fool* [fu:l], *get* [gɛt], *cat* [kæt], and *box* [bɔks].

4. Method and Results

The objective of the study is to model the acoustic differences between the production of nine short vowels as L1 and L2 by 11 Afrikaans (group 1) and 11 English (group 2) mother-tongue speakers. In order to do this, the first five formant frequencies of each utterance were calculated in consecutive time frames, using the Split-Levinson algorithm (Willems, 1986). Then the difference between F1 of the group 1 utterances and group 2 utterances was calculated. Subsequently the difference calculation was repeated for two other formant frequencies (F2 and F3), as well as for the formant ratios (F1/F2, F2/F3 and F1/F3). A statistical t-test was performed (Shearer, 1982) to determine the significance of the differences. The concept of representing vowels by the formant ratios (for speaker normalisation purposes) was taken from Van der Merwe et al. (1993) and Minifi et al. (1973).

Manipulation of the speech data was done using Praat, a speech analysis software package (Boersma, 1996) and statistical analysis was performed using Systat (Wilkinson et al. 1992).

The window and time step sizes of the formant frequency calculations were 25ms and 10ms respectively for the isolated vowels and 12.5ms and 5ms for the vowels in context (within the English and Afrikaans words). The smaller values for the in-context vowels were necessary due to their shorter duration. Pre-emphasis was employed from 50 Hz in order to enhance high frequencies and to improve the formant analysis. A Hamming window was used to minimise the errors at time-frame boundaries.

Since only steady-state vowels were analysed, a true dynamic analysis of the formants was not required. A single value to represent the formant was sought. Several options were considered about what part of the utterance to use for formant calculation. Choosing for average formants over the whole vowel segment was rejected due to the variation towards the end. The formant values at maximum intensity of the sound also proved to be impractical, due to the variation of the location of the maximum intensity of the isolated vowels.

A measurement of the maximum, minimum and average duration of the vowels were made for isolated, Afrikaans and English word vowels, in order to choose usable regions for formant calculations. From the duration values and from visual inspection of the formants, three regions were chosen for the isolated vowels (Experiment 1): (1) 0 to 120 ms, (2) 60 to 120 ms and (3) 80 to 160 ms. This roughly constitutes the first half, second quarter and second third, respectively, of an utterance of average duration. In subsequent experiments, such a detailed analysis was not deemed necessary, and for the Afrikaans word vowels (Experiment 2), formants were calculated in the (1) centre half and (2) first 75% of the vowel. The formant values of the English word vowels (Experiment 3) were only determined in the first 75% of the vowel. In these regions, the average, median and standard deviation of each formant frequency were determined.

The standard deviation (σ) of the formants were not used as features to determine differences, but rather to check whether the formant trajectories could indeed be considered steady-state. The σ's showed that the values do not, on average, deviate more than 10% in the selected region.

In order to check for F0-Fi interaction (or to prevent it, if necessary), the pitch of the different speakers was also calculated using the autocorrelation method (Boersma, 1996), a 10ms step size, and a 25ms Hamming window. It was found that the average pitch measured in the Afrikaans word list was 134 Hz and 137 Hz for groups 1 (Afrikaans mother-tongue speakers) and 2 (English mother-tongue speakers), respectively, and 130 Hz and 128 Hz when measured in the English word list. This indicates that voice pitch does not bias our results in any significant way.

The distribution of the first two formants (F1 and F2) of experiment 1 (isolated Afrikaans vowels) are graphically displayed for Afrikaans and English mother-tongue speakers in Figures 3 and 4, respectively. The values shown are the median values of the measurements in the centre 33% of the utterances. The values in Figure 3 constitute Cond-Afr-L1 and those in Figure 4 constitute Cond-Afr-L2.

The *average* formant values of the isolated Afrikaans vowels are shown separately for Afrikaans and English mother-tongue speakers on an F1-F2 plane in Figure 5. The differences between L1 (Afrikaans mother-tongue speakers) and L2 (English mother-tongue speakers) values are clearly visible. The L1 values correspond quite accurately with the formant values published in Van der Merwe et al. (1993), as well as those in Taylor and Uys (1988).

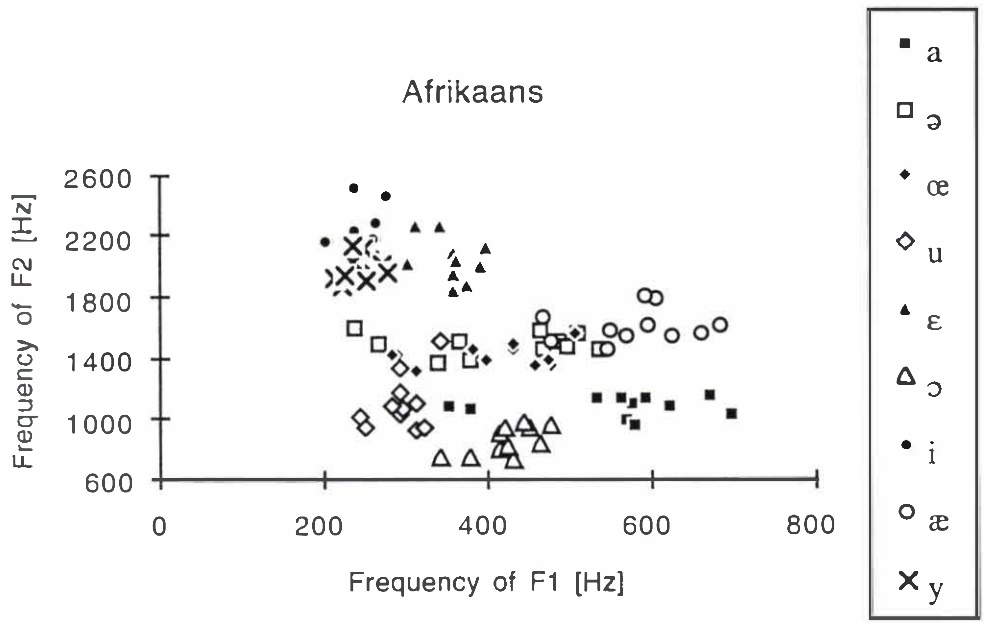

Figure 3. Distribution of the median F1 and F2 values in the centre 33% of the isolated vowel utterance of Afrikaans mother-tongue speakers.

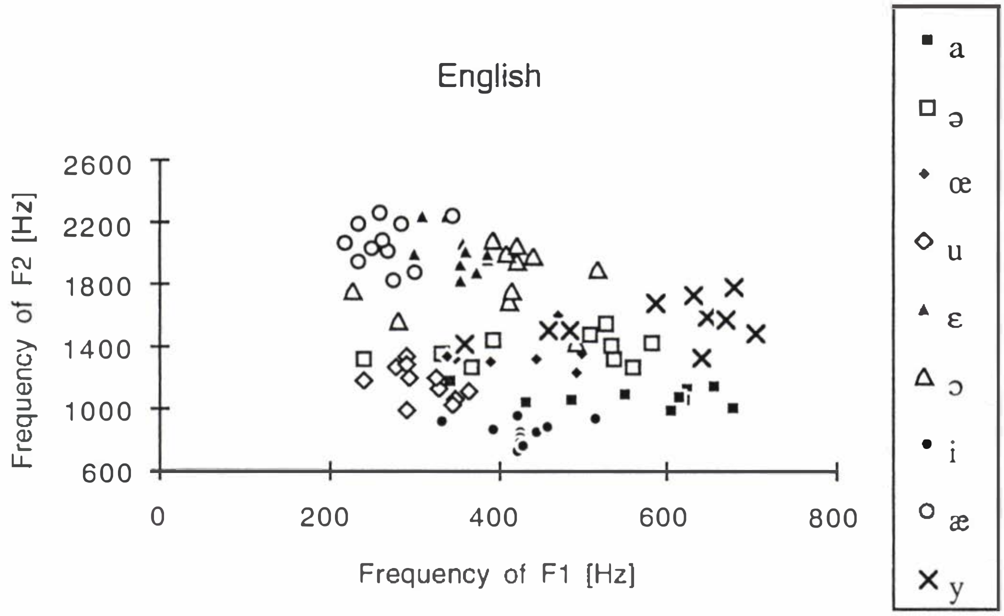

Figure 4. Distribution of the median F1 and F2 values in the centre 33% of the isolated vowel utterance of English mother-tongue speakers.

The average formant values of the isolated Afrikaans vowels are shown separately for Afrikaans and English mother-tongue speakers on an F1-F2 plane in Figure 5. The differences between L1 (Afrikaans mother-tongue speakers) and L2 (English mother-tongue speakers) values are clearly visible. The L1 values correspond approximately with the formant values published in Van der Merwe et al. (1993), as well as those in Taylor and Uys (1988).

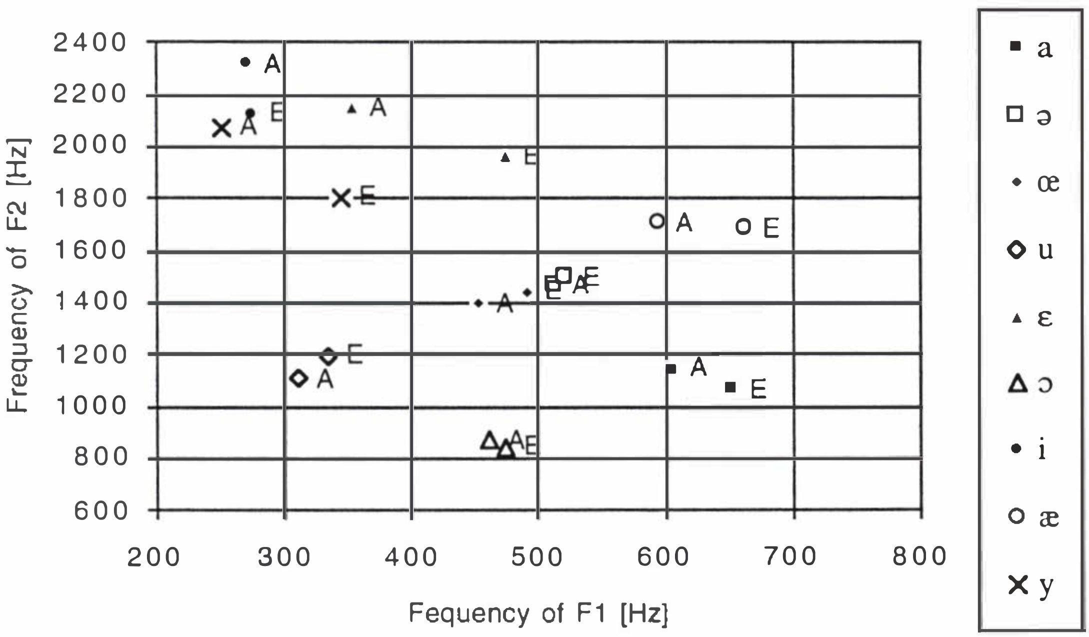

Figure 5. Average F1 and F2 formant values of the isolated Afrikaans vowels. English mother-tongue speakers indicated with 'E' and Afrikaans mother-tongue speakers indicated with 'A'.

In Tables 2 (isolated Afrikaans vowels, i.e. Cond-Afr-L1 and Cond-Afr-L2), 3 (Afrikaans words, i.e. Cond-Afr-L1 and Cond-Afr-L2) and 4 (English words, i.e. Cond-Eng-L1 and Cond-Eng-L2), the parameters which display significant differences (probability less than 10%) are indicated.

Table 2. Significant differences between production of isolated vowels by Afrikaans and English mother-tongue speakers. Formant ratios are indicated by F1:2 (F1/F2), etc.

		First 50% of utterance				
Vowel		ɛ	i	ə	u	y
Values	median	F1 F2		F2	F3	F2
	mean	F1 F2		F2	F3	F2
Ratio's	median	F1:2 F2:3 F1:3	F1:2	F1:2 F2:3		F1:2 F2:3
	mean	F1:2 F2:3 F1:3	F1:2 F1:3	F1:2 F2:3		F1:2 F2:3
		Second 25% of utterance				
Vowel		ɛ	i	ə	u	y
Values	median	F1 F2 F3		F2		F2
	mean	F1 F2 F3		F2	F3	F2
Ratio's	median	F1:2 F2:3 F1:3	F1:2 F1:3	F1:2 F2:3		F1:2 F2:3
	mean	F1:2 F2:3 F1:3	F1:2	F1:2 F2:3		F1:2 F2:3
		Second 33% of utterance				
Vowel		ɛ	i	ə	u	y
Values	median	F1 F2 F3		F2	F3	F1 F2
	mean	F1 F2 F3		F2	F3	F2
Ratio's	median	F1:2 F2:3 F1:3	F1:2	F1:2 F2:3	F1:3 F2:3	F1:2 F2:3 F1:3
	mean	F1:2 F2:3 F1:3	F1:2	F1:2 F2:3	F1:3	F1:2 F2:3

Table 3. Significant differences between production of vowels in Afrikaans words by Afrikaans and English mother-tongue speakers.

		Centre 50% of utterance							
Vowel		a	æ	ɛ	i	ə	œ	u	y
Values	median	F1	F2	F1 F2			F2		
	mean		F2	F2			F2		
Ratio's	median		F2:3	F1:2 F2:3 F1:3			F2:3		F2:3
	mean			F1:2 F2:3 F1:3			F2:3		F2:3
		First 75% of utterance							
Vowel		a	æ	ɛ	i	ə	œ	u	y
Values	median			F1					
	mean			F1					
Ratio's	median			F1:2 F1:3	F1:2 F1:3	F2:3		F1:3	F2:3
	mean			F1:2 F1:3	F1:2 F1:3			F1:3	F2:3

Table 4. Significant differences between production of vowels in English words by Afrikaans and English mother-tongue speakers.

		First 75% of utterance		
Vowel		ɛ	u	ɔ
Values	median	F1 F2	F1	F2
	mean	F1 F2		F2
Ratio's	median	F1:2 F2:3 F1:3	F1:2	
	mean	F1:2 F2:3 F1:3	F1:2	

In Table 5, the shading of the rectangles is a subjective judgement, indicating how much the vowels of the L1 and L2 pronunciations differ in each of the experiments (isolated Afrikaans vowels, Afrikaans words and English words). Since the formant ratios of /ɛ/, /y/ and /i/ seem to differ consistently, at least in the isolated vowels and Afrikaans words, these values are worth closer inspection. The average values of the F1/F2, F2/F3 and F1/F3 ratios of the isolated vowels (centre 33% of the vowel, median features) are shown graphically in Figures 6 and 7. The ratios of the vowels in the Afrikaans words are given in Table 6. Also shown for comparison in the table are the formant ratios published by Van der Merwe et al. (1993), as well as the values of the neutral vowel /ə/ and the vowel /æ/.

Table 5. Subjective degree of difference between production of vowels in isolation, Afrikaans and English words. The dark rectangle ■ indicates a significant difference in a number of parameters and the clear rectangle ❐ indicates a difference in only one or two parameters.

Formant values

Vowel	a	æ	ɛ	i	ə	œ	u	y	ɔ
Isolated			■		■		■	■	
Afrikaans word	❐	❐	■			❐			
English word			■				❐	NA	❐

Formant ratios

Vowel	a	æ	ɛ	i	ə	œ	u	y	ɔ
Isolated			■	■	■		❐	■	
Afrikaans word		❐	■	■	❐	❐	❐	■	
English word			■				❐	NA	

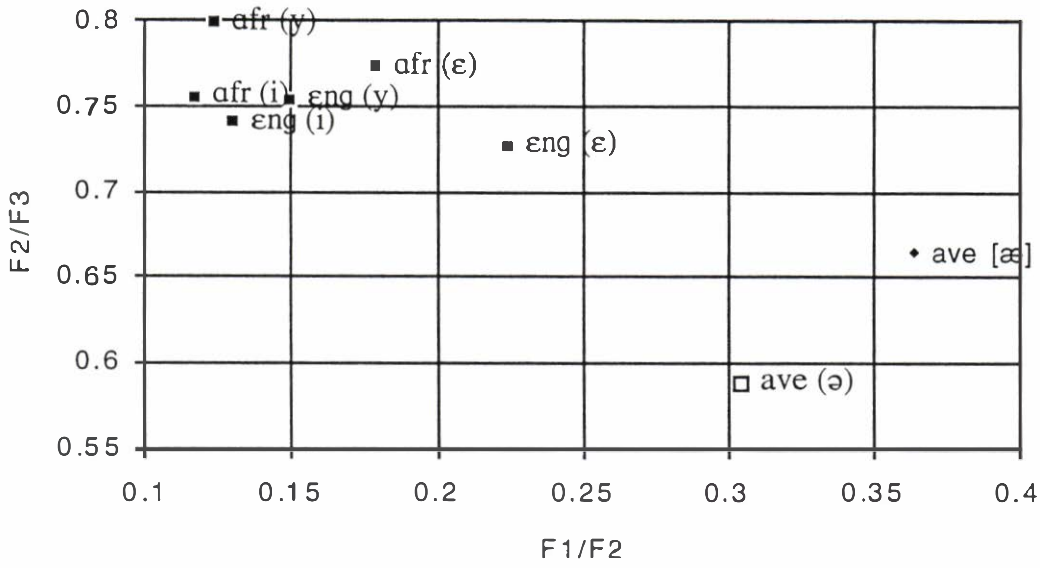

Figure 6. Average formant ratios F2/F3 vs. F1/F2 of the isolated /ε/, /i/, /y/, /ə/ and /æ/ vowels.

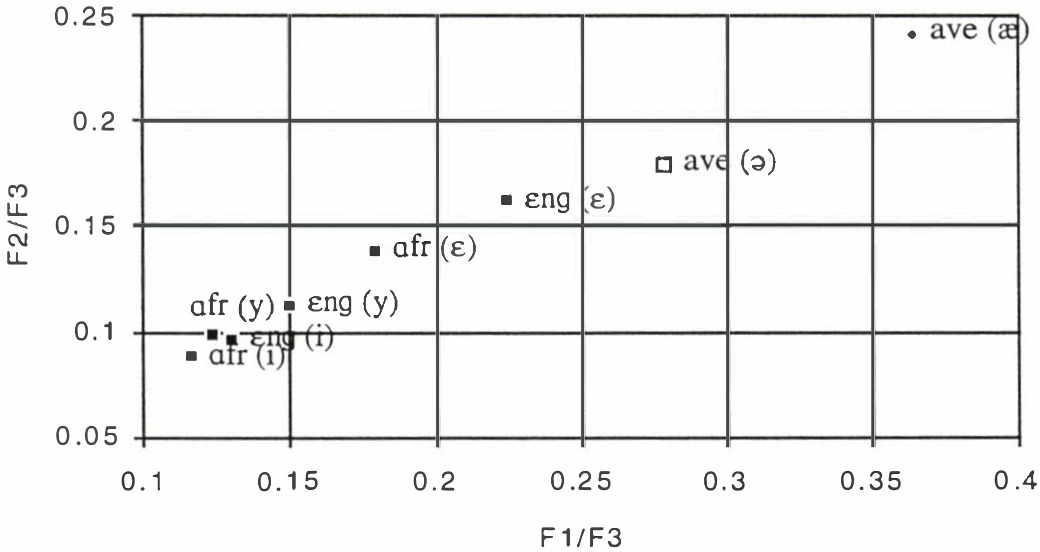

Figure 7. Average formant ratios F2/F3 vs. F1/F3 of the isolated /ε/, /i/, /y/, /ə/ and /æ/ vowels.

Table 6. Formant ratio values of the /ɛ/, /i/, /y/, /ə/ and /æ/ vowels. Significant differences are indicated by the * and the average of the Afrikaans and English word values are also given.

Vowel	Source	Mother tongue	F1/F2	F2/F3	F1/F3
/ɛ/	Vd Merwe et al., 1993	Afrikaans	0.17	0.76	0.13
	/sɛs/	Afrikaans	0.21*	0.75	0.16*
		English	0.23*	0.73	0.17*
	average		0.22	0.74	0.16
/i/	Vd Merwe et al., 1993	Afrikaans	0.11	0.75	0.08
	/kim/	Afrikaans	0.14*	0.75	0.10*
		English	0.15*	0.75	0.11*
	average		0.14	0.75	0.11
/y/	Vd Merwe et al., 1993	Afrikaans	NA	NA	NA
	/dəbyt/	Afrikaans	0.14	0.80*	0.11
		English	0.15	0.76*	0.11
	average		0.14	0.78	0.11
/ə/	Vd Merwe et al., 1993	Afrikaans	0.35	0.62	0.21
	/əs/	Afrikaans	0.29	0.64	0.19
		English	0.32	0.61	0.19
	average		0.31	0.62	0.19
/æ/	Vd Merwe et al., 1993	Afrikaans	0.35	0.69	0.23
	/fæl/	Afrikaans	0.35	0.66	0.23
		English	0.38	0.63	0.24
	average		0.37	0.64	0.23

5. Discussion

In terms of modelling the acoustic differences between Afrikaans and English, produced as L1 and L2, the steady-state vowels show few differences. The most prominent differences are in the production of /ɛ/, which were observed in all the tests, namely in isolated vowels, English and Afrikaans words. More differences were observed in the production of the isolated vowels (where a number of significant differences between /ɛ/, /ə/, /u/ and /y/ occurred) than in the production of the same vowels in the context of Afrikaans and English words. See Table 5 for a summary and Tables 2, 3 and 4 for the details of these differences.

The graphical representation (Figures 6 and 7) and tabulation (Table 6) of the formant ratios of the most different vowels (/ɛ/, /i/ and /y/), show some interesting trends. The ratios of these vowels are displayed together with the ratios of the /æ/ and /ə/. The /æ/ can be viewed as a "stronger" relative of /ɛ/, because the tongue is lower and tenser and the mouth is open wider. The neutral vowel /ə/ is viewed as a "weaker" relative of all the other vowels. Against this background, the trends for the English mother tongue speakers seem to be that the production of the /ɛ/ is "stronger" than their Afrikaans counterparts, since the values are closer to those of the /æ/. In contrast, the /i/ is pronounced "weaker" by the English speakers, as their /i/-values tend towards the /ə/-values. Inspecting the values of Table 6, it can be seen that the /y/ is pronounced more like /i/, its unrounded counterpart.

The trend concerning /i/ and /ə/ is contrary to the popular notion that Afrikaans speakers tend to neutralise the /i/, relative to their English neighbours. This seems to be worth an in depth separate study.

The former trend concerning the /æ/ and /ɛ/ is interesting in relation to a general belief about Afrikaans speakers in the northern part of South Africa, formerly *Transvaal*. It is typical of Transvaal speakers to pronounce the word *ek* as [æk], while Southern Cape speakers would pronounce it as [ɛk]. The interesting part is that the Transvaal English speakers of this study produce an even stronger /æ/ than the Afrikaans speakers in the same region.

The third interesting observation relates to the theory of *equivalence classification* (Flege, 1987), which states that L2 speakers learn to produce unknown sounds similar to the L1 speakers of the same sounds, because these sounds form a new category for L2-speakers. The results of this study contradict this theory, as the /y/ which is unknown in English, shows a number of significant differences in both the isolated vowels as well as the Afrikaans word [dəbyt]. (See Tables 2, 3 and 5.)

6. Conclusions and Further Projects

It is generally believed that an Afrikaans or English accent (for the language spoken as L1 or L2) can be easily detected in the bilingual society of South Africa. The results of this study show that the steady-state vowels is not the only place to look if this accent is to be quantified. We did learn, however, that /ɛ/ and /y/ are consistently pronounced differently, and to a lesser extent also /i/, /ə/ and /u/.

Since it is common knowledge that prosody contributes largely to foreign accentedness, an obvious continuation of this study would be to investigate dynamic differences, such as found in the diphthongs and controversial/ambiguous sounds /e/, /ø/ and /o/ in Afrikaans. Appropriate dynamic features for the comparison of these sounds will have to be found.

If the acoustic differences of Afrikaans and English are to be modelled comprehensively, a statistical approach is necessary. This will have to include all the phones, and in many contexts, such as found for American or British English in the large corpora used in ASR research. The annotation of such large corpora are gigantic tasks, which can only be attempted in South Africa, where there are limited resources, if it is done automatically. It is, however, an open question whether automatic segmentation can yield accurate enough results for this type of study.

Several studies relating to the acoustic phonetics of Afrikaans and English seem to follow logically from this work, and are possible with the comprehensive data set as described in Section 3. Some which come to mind are

- to investigate the long and frequently diphthongised vowels, to determine whether and how they differ from their short counterparts,
- to resolve the ambiguities relating to /e/, /ø/ and /o/ in Afrikaans as mentioned earlier,
- to find appropriate features on which to base a comparison of the consonants and to perform such a comparison,

- to find appropriate dynamic features for the real and semi-diphthongues and to compare them and
- to extend the comparisons to di- and triphones.

Finally it remains to be seen whether and how this model can be applied to speech technology systems. Two avenues for investigation are currently envisioned:

1. Once a rule-based speech synthesis system for English is available, the model can be applied to change the rules so that when the sound production rules are used for Afrikaans words (the new language), these are pronounced without an English accent.
2. In ASR systems that use explicit phonetic features (not the frame-based acoustic features of the state-of-the-art systems), the recognition parameters are to be adjusted to fit the new language.

Acknowledgements

The Foundation for Research Development in South Africa and the University of Pretoria made this research possible by financing my sabbatical at the Institute for Phonetic Sciences at the University of Amsterdam. Etienne Barnard inspired me to work on technical aspects of South-African speech, and helped me to plan and formulate. Louis Pols spent many hours discussing the details of this work and reading the manuscript. I thank all of them.

References

ARPA (1990): *The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus*, Training and Test Data., NIST Speech Disc CD1-1.1.

Barnard, E., Cole, R., Fanty, M. and Vermeulen, P. (1995): "Real-world speech recognition with neural networks", *SPIE Proceedings: Applications and Science of Artificial Neural Networks*, Vol. 2492, Orlando, Florida: 524-537.

Boersma, P. (1996): Praat, *A system for doing phonetics by computer*, in development, IFA, University of Amsterdam.

Bourlard, H., Hermansky, H. Morgan, N. (1996): "Towards increasing speech recognition error rates", *Speech Communication*, **18**: 205-231.

Coetzee, A.E. (1981): *Fonetiek vir eerstejaars*, Academica, Pretoria.

De Stadler, L.G. and Coetzer, M.W. (1988): "'n Afrikaanse skrif-tot-spraakomsetter", *Suid-Afrikaanse Tydskrif vir Taalkunde*, **8**: 19-35.

De Villiers, M. and Ponelis, F.A., *Afrikaanse klankleer*, Tafelberg, Kaapstad.

Flege, J.E. (1984): "Detection of Foreign Accentedness", *Proceedings of the Tenth International Congress of Phonetic Sciences.* (M.P.R. van den Broecke and A. Cohen eds.) Dordrecht, Holland: Foris Publications, 677-681.

Flege, J.E. (1987): "The production of new and similar phones in a foreign language: evidence for the effect of equivalence classification", *Journal of Phonetics*, **15**: 47-65.

Flege, J.E. and Eefting, W. (1987): "Cross-language switching in stop consonant perception and production by Dutch speakers of English", *Speech Communication*, **6**: 185-202.

Flege, J.E., Munro, M.J. and Fox, R.A. (1995b): "Auditory and categorical effects on cross-language vowel perception", *J. Acoust. Soc. Am.*, **95**(6): 3623-3641.

Flege, J.E., Munro, M.J. and MacKay, I.R.A. (1995a): "Factors affecting the strength of perceived foreign accent in a second language", *J. Acoust. Soc. Am.*, **97**(5), 3125-3134.

Flege, J.E., Munro, M.J. and MacKay, I.R.A. (1995c): "Effects of age of second-language learning on the production of English consonants", *Speech Communication*, **16**: 1-26.

Fox, R.A., Flege, J.E. and Munro, M.J. (1995): "The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis", *J. Acoust. Soc. Am.*, **97**(4): 2540-2551.

International Phonetic Association (IPA) (1966): *The Principles of the International Phonetic Association*, London.

Leather, J. and James, A. (1996): "Second language speech". In: *Handbook of Second Language Acquisition*. New York: Academic Press, 269-316

Minifie, F.D., Hixon, T.J. and Williams, F. (1973): *Normal aspects of speech, hearing and language*, Prentice Hall, New Jersey.

Prinsloo, G.J. and Coetzer, M.W. (1990): "Automatic syllabification and phoneme class labelling with a phonologically based hidden Markov model and adaptive acoustical features", *Computer Speech and Language*, **4**(3): 247-262.

Schouten, M.E.H. (1975): *Native-language interference in the perception of second-language vowels*, PhD thesis University of Utrecht, Utrecht, The Netherlands.

Shearer, W.M. (1982): *Research procedures in speech, language and hearing*, Williams & Wilkins, Baltimore/London.

Taylor, J.R. and Uys, J.Z. (1988): "Notes on the Afrikaans vowel system", *Leuvense Bijdragen*, **77**(2): 129-149.

Van der Merwe, A., Groenewald, E., van Aardt, D., Tesner, H.E.C. and Grimbeeck, R.J. (1993): "Die formantpatrone van Afrikaanse vokale soos geproduseer deur manlike sprekers", *Suid-Afrikaanse Tydskrif vir Taalkunde*, **11**(2): 71-79.

Wilkinson, L., Hill, M. and Vang, E.: *SYSTAT for the Macintosh 5.2: Statistics*, SYSTAT Inc., Evanston, Illinois.

Willems, L.F. (1986): "Robust formant analysis", *Technical Report no. 529*, Institute for Perception Research, Eindhoven, The Netherlands.

Young, S. (1996): "A review of large-vocabulary continuous-speech recognition", *IEEE Signal Processing Magazine*, September: 45-57.