Southern African Linguistics and Applied Language Studies 2007, 25(1): 91–105
Printed in South Africa — All rights reserved

SOUTHERN AFRICAN LINGUISTICS AND APPLIED LANGUAGE STUDIES
ISSN 1607–3614

The perception and identification of accent in spoken Black South African English

Philippa Louw* and Febe de Wet
Centre for Language and Speech Technology, University of Stellenbosch,
Private Bag X1, Matieland, 7200, South Africa
** Corresponding author, e-mail: phlouw@sun.ac.za*

Abstract: Can mother tongue speakers of the Nguni and Sotho languages determine each other's first language (L1) based on their English accent? Contrasting claims have been made in this regard. While Black South African English (BSAE) can be distinguished clearly from Standard South African English (SSAE) on the levels of both production and perception, insufficient evidence exists that such a distinction can be made between Nguni-English and Sotho-English. This study investigates the question of perceivable differences in BSAE accents by means of two perceptual experiments. The first aim of the experiments is to ascertain whether participants from either the Nguni or Sotho language group can determine whether a particular speaker has an SSAE or a BSAE accent. The second aim is to determine whether L1 Nguni and Sotho listeners can identify a speaker's L1 group by listening to English words and sentences pronounced by Nguni and Sotho L1 speakers. Lastly, we investigate whether there is any correlation between listeners' judgement of speakers' accent and their ability to determine a speaker's L1. The results of both perceptual experiments contradict the notion that different mother tongues influence BSAE to such a degree that the speaker's L1 is easily perceived. However, some correlation was found between perception of accentedness and the correct identification of L1.

Introduction

In the interest of developing speech technology in a multilingual country like South Africa, pronunciation variation and non-native pronunciation in particular, is a critical factor to consider. The success of applications such as voice operated systems, automatic dialogue systems and computer-assisted language learning, where large vocabulary speaker-independent Automatic Speech Recognition (ASR) is the underlying technology, depends largely on the ability of such applications to deal with interspeaker pronunciation variation (Mayfield Tomokiyo, 2001). Strik and Cucchiarini (1999: 226) consider interspeaker variation (as opposed to intraspeaker variation) to be one of the two major sources of pronunciation variation. Speakers of the same language have different dialects and accents depending on their region of origin, socioeconomic background, level of education, gender, age and — the issue under discussion in this article — mother tongue (Strik & Cucchiarini, 1999: 226).

South Africa has eleven official languages, of which English serves as the *lingua franca*. The accents of South African English vary mainly according to mother tongue (alternatively referred to as first language (L1)) groups (for example Afrikaans, English and the African Languages). This study focuses on one of these varieties, namely Black South African English[1] (BSAE), and its relation to Standard South African English[1] (SSAE) as well as its coherence as a variety of English. In order to do proper pronunciation modelling for South African English ASR systems, we need to know if BSAE can be modelled as a single accent or whether multiple models should be trained on BSAE data that is split up according to mother tongue groups.

Research has shown that a clear phonetic distinction can be made between SSAE and BSAE (Gough, 1996; Van Rooy, 2000a, 2004; Van Rooy & Van Huyssteen, 2000; Wissing, 2002). The

vowel phonology of BSAE is influenced by the native five-vowel system (as in the Nguni languages) or seven-vowel system (as in the Sotho languages).[2] In BSAE the contrast between tense and lax vowels is neutralised and there is a tendency to avoid central vowels such as the schwa. The contrast between long and short vowels may also be lost. Some diphthongs may be realised as monophthongs while others are extended over two syllables. Consonants are realised similar to the native varieties with the exception of some cluster simplification. The voiced and voiceless dental fricatives are sometimes realised as stops and the liquid /r/ sound is trilled.

Within the context of vowel production and perception at segmental level, Van Rooy and Van Huyssteen (2000: 30) maintain that, because of the similar diphthong/tense vowel-structure of Sotho-English and Nguni-English, these two varieties of English are fairly coherent and further that there is little variation within these varieties that can be ascribed to the differences between the mother tongues of the speakers. Wissing (2002: 21) goes further to say that 'perceivable differences, if any, between the English of speakers of different Bantu languages will most probably be on the suprasegmental, i.e. prosodic level'.

To date, no study has explicitly tested whether first language (L1) speakers of the Bantu languages can identify a BSAE speaker's mother tongue, or at least perceive differences in BSAE accents. Contrasting observations have been made in this regard, but mostly as passing remarks in studies on related issues. For example, in a study on the comprehensibility of varieties of South African English, Van der Walt (2000: 143) reported that black language teachers claimed to be able to distinguish between Xhosa-English, Zulu-English and so on. In contrast, a study investigating the types of labels given to BSAE speech reported that, when listeners tried to pinpoint a person's mother tongue based on the speaker's English accent, they were mostly unable to do so (Coetzee-Van Rooy & Van Rooy, 2005: 12).

In this study, the validity of these claims was investigated by conducting two perceptual experiments. The first aim of the perceptual experiments was to determine if participants from either the Nguni or Sotho L1 groups can determine whether a particular speaker has a SSAE or a BSAE accent. Secondly, we wanted to determine whether L1 Nguni and Sotho listeners can identify a speaker's L1 language group by listening to English words and sentences pronounced by Nguni and Sotho L1 speakers. One of the experiments was further designed to establish whether there is any correlation between listeners' judgements of speakers' accent and their ability to determine a speaker's mother tongue. The outcome of this study can provide valuable guidelines for the implementation of non-native pronunciation variation in ASR systems.

Experiment 1: Perceptual experiment using telephone data

Speakers

Telephone speech data collected in the African Speech Technology (AST) project were sourced as stimuli for the first perceptual experiment (Roux *et al.*, 2004). During the AST project, eleven speech databases consisting of five different languages were collected for the purpose of developing ASR systems. Five English databases were collected, including mother tongue English (SSAE), as well as Black South African English (BSAE), Indian-English, Coloured-English and Afrikaans-English. The words and sentences for this experiment were selected from the BSAE and SSAE databases.

Only the speech of mesolect speakers was used to compile the BSAE data set, i.e. the minimum level of education per speaker was Grade 12 (Matric). Some of the speakers had a tertiary qualification, but they were not considered to have reached the acrolectal level as described by Wissing (2002: 24). The speech of 37 females and 35 males was used. Table 1 presents the demographic details of the speakers that were involved.

Stimuli

The set of stimuli consisted of 30 single words and 30 sentences pronounced by L1 speakers of each of the three language groups Sotho, Nguni and English. Thus, a total of 180 stimuli consisting of native South African English (SSAE), Sotho-English (SE) and Nguni-English (NE) were used.

The AST databases were designed such that each speaker read a unique set of items. It was therefore not possible to select all 30 words or sentences from the same speaker's phonecall. In some cases only one item per speaker was used and the availability of the items themselves dictated the variety of speakers represented in the final set of stimuli.

Stimuli were of two types. Firstly, single words were used so that the listeners could focus on a limited number of sounds in a limited context. Secondly, sentences provided the listeners with a variety of sounds as well as prosodic cues which could influence their judgement of accent. We would not be able to determine which specific sounds influenced the listeners' judgement if only sentences were used as stimuli. We used the typical phonetic/phonological properties of BSAE as described in Gough (1996), Van Rooy (2000a, 2004), Van Rooy and Van Huyssteen (2000) and Wissing (2002) as guidelines to select stimuli. We attempted to represent in our word list as many of the relevant phenomena as possible. However, it was not possible to use the exact example words given by these authors, since these words did not occur in the AST databases.

Table 2 gives an overview of the contexts and realisations of the BSAE sounds that were used in the first experiment. Examples of the type of sentence (or, in some cases, phrase) in which these words occurred are given in Table A in the Appendix. In most cases, the single words were selected from utterances where they occurred in sentences. Thus, an isolated word as well as the same word as it occurred in a sentence was presented as stimuli. As far as possible, the words and

Table 1: Speaker demographics for experiment 1

L1	# Speakers	Educational Level					Gender	
		Matric	Under-Graduate	Honours	Masters	PhD	Male (35)	Female (37)
S.Sotho	13	10	1	1	1		9	4
Tswana	24	16	6	1	1		9	15
Xhosa	16	7	2	6	1		8	8
Zulu	18	5	7	2	3	1	8	10
Ndebele	1		1				1	

Table 2: Realisation of sounds in BSAE and example words used in experiment 1

Realisation of sounds in BSAE	Words	
Vowels		
Neutralisation of tense/lax vowels or long/short vowels	list task half	information into mini
Avoidance of central vowels	sentence	matric
[æ] replaced by [ɛ]	help	
Schwa in open syllables replaced with [a]	operator	
Diphthongs		
Narrower diphthongs realised as single monophthongs	telephone make, day	change only, over
Consonants		
Affricate [tʃ] becomes fricative [ʃ]	opportunities	
Trilled [r] as allophone for liquid	directory	
Velar plosive devoicing	goodbye	
General	cancel check continue development	eleven goodbye perhaps repeat

sentences were selected from the SSAE and BSAE databases in such a way that the contents were the same for each language group.

Listeners

A total of 36 participants (none of whom participated in the AST project) were recruited on the Stellenbosch University campus. Table 3 gives an overview of the distribution of L1, level of education and gender amongst the listeners.

Test administration

A perception test was set up using the Praat software package (www.praat.org). The 180 stimuli were played in random sequence, but the same sequence was used for all participants. The question 'Can you identify to which language group this speaker belongs?' was displayed on the computer screen together with four clickable buttons. The four buttons represented the four options available, i.e. 'Sotho', 'Nguni', 'English' and 'I don't know'. Each stimulus was played only once and as soon as the participant had made his/her choice by clicking on one of the four buttons, s/he would hear the next stimulus.

Instructions were given verbally to the participants before commencing the experiment. The participants listened to the stimuli using earphones. A short pre-test, presenting three test utterances, served to demonstrate the procedure as well as to ensure that the participants could hear the stimuli clearly. When the participants indicated that they understood the procedure and what was expected of them, they were presented with the test stimuli. The stimuli were presented in three sets of 60 and short breaks were allowed between sets. The participants took an average of 20 minutes to complete the perception test.[3]

Results[4]

The responses to the two types of stimuli, sentences and words, were first analysed as a single data set and then split into two data sets so that the responses to the isolated words could be compared to those of the sentences. Furthermore, the responses to all stimuli (SSAE, NE and SE) were first analysed as a whole and then the responses to SSAE and BSAE stimuli were analysed separately. A total of 10% of all responses in this perception test were 'I don't know'. Most of these responses (77%) were to the single word stimuli. The 'I don't know' responses were removed from the data set before the percentage of correct results was calculated.

The results for all stimuli combined in one data set are illustrated in Figure 1. The figure shows that only 57% of the responses were correct, i.e. listeners were not able to distinguish clearly between the three different accents.

The percentage of correct responses to stimuli produced by SSAE speakers was significantly higher (71%) than the overall score, which indicates that the listeners could distinguish between the BSAE accent and the SSAE accent.

Figure 2 shows the results for only the BSAE stimuli (with responses to the SSAE stimuli

Table 3: Listener demographics for experiment 1

L1	# Listeners	Educational Level				Gender	
		Undergraduate	Honours	Masters	PhD	Male (16)	Female (20)
N.Sotho	7	7				3	4
S.Sotho	7	6		1		4	3
Tswana	4	4				2	2
Xhosa	9	5		3	1	3	6
Zulu	9	6	2	1		4	5

removed from the data set). According to the data in Figure 2, the percentage of correct responses for all types of stimuli is 49%. Listeners performed slightly better on the sentence stimuli than on the word stimuli, but this difference was not significant. This result seems to indicate that the listeners could not determine whether a speaker's mother tongue was from the Nguni or Sotho language group, irrespective of whether supra-segmental information was present or not.

Figure 3 shows the percentage of correct responses per listener for the BSAE stimuli. Only 17 out of 36 listeners achieved a score higher than 50%, the highest of which was 67%.

It was also observed that the listeners' responses showed a bias towards their own mother tongue: Nguni L1 listeners classified the majority of the stimuli as Nguni, irrespective of the language group to which the speakers actually belonged. A similar trend was observed in the Sotho L1 listeners' responses, although the tendency to choose their own mother tongue was not as strong as the bias observed for the Nguni L1 listeners.

Since almost every stimulus was produced by a different speaker, we did not attempt to determine whether there were any speaker-specific attributes that may have influenced the listeners' judgements.

To summarise, the results obtained in this experiment do not support the claim made by some L1 speakers of South African Bantu languages that they can derive a speaker's mother tongue from his/her English accent. However, this result cannot be regarded as conclusive, because the stimuli that were used in the experiment suffer from a number of shortcomings. Firstly, the stimuli

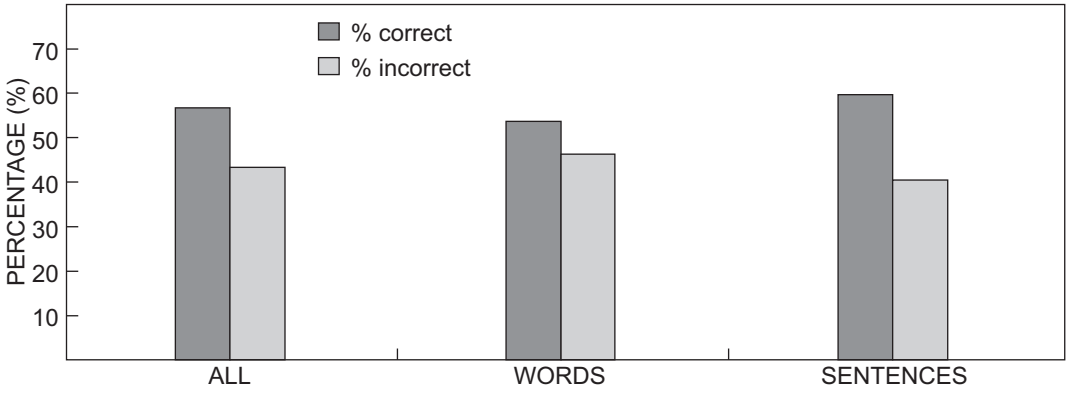

Figure 1: Percentage of correct and incorrect responses for all stimuli and all accents in experiment 1

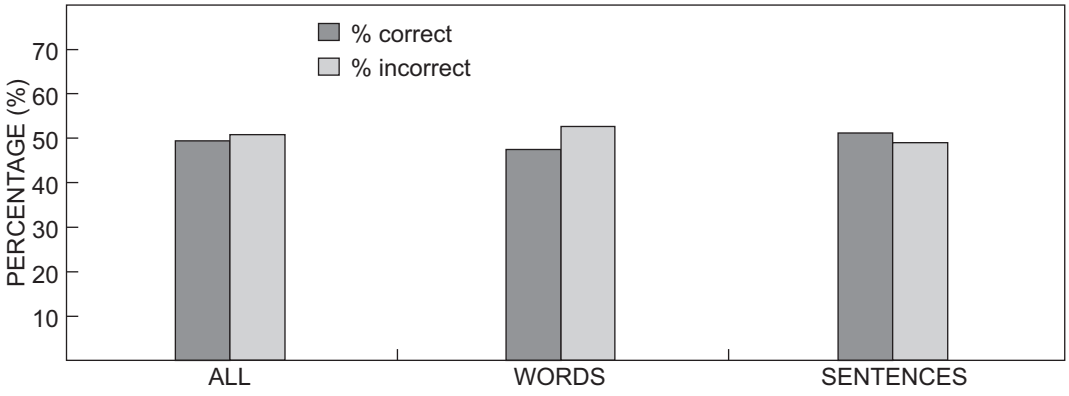

Figure 2: Percentage of correct and incorrect responses for all BSAE stimuli in experiment 1

were collected to develop automatic speech recognition technology and not to be used in perceptual experiments. The utterances were recorded over telephone lines as read prompts and it may not be possible to perceive the distinctive suprasegmental features of the different BSAE varieties in this kind of data. Secondly, the data contain very few examples of individual words. Most of the isolated word stimuli that were used in this experiment were cut out from sentences. As a consequence, many of the single word stimuli were very short. Finally, although the experimental design did take level of education into account, there was still enormous variation in the level of English proficiency among the BSAE speakers and different degrees of accent may have influenced the participants' opinions.

Experiment 2: Perceptual experiment using studio data

Given the shortcomings of the data used in the first experiment, it was decided to do a follow-up experiment in which the stimuli could be designed and recorded specifically for the purpose of conducting a perceptual experiment. The elicited utterances therefore included examples of both read and spontaneous speech and the data was recorded in a studio instead of over the telephone. The data set also included more examples of words pronounced in isolation. These words were selected such that they contained phones that could provide cues on a speaker's L1 background. The speakers as well as the listeners also provided more detail about the languages they speak and the language environments they encounter in their everyday lives. During the experiment, the recorded speech was scored for accentedness by the participants in order to determine whether perceived accent influences second or third language (L2 or L3) listeners' ability to determine what the L1 background of an L2/L3 speaker of English is.

Speakers

Five male and six female speakers were recruited on the Stellenbosch University campus. Representatives of the Nguni and Sotho mother tongue groups could only be recruited on the basis of availability. The distribution of mother tongue speakers in each group, as well as other demographic information, is shown in Table 4. The speakers were between 18 and 30 years of age and most were registered for a Bachelor's degree at the university. All participants could be described as mesolectal speakers. They speak between one and six languages other than their L1. On average they speak three of these languages on a weekly basis. About 66% of the speakers indicated that they are able to determine a person's mother tongue (Nguni or Sotho), based on his/her English accent. One male and one female native English speaker were also recruited. Each speaker participated in a 30-minute recording session.

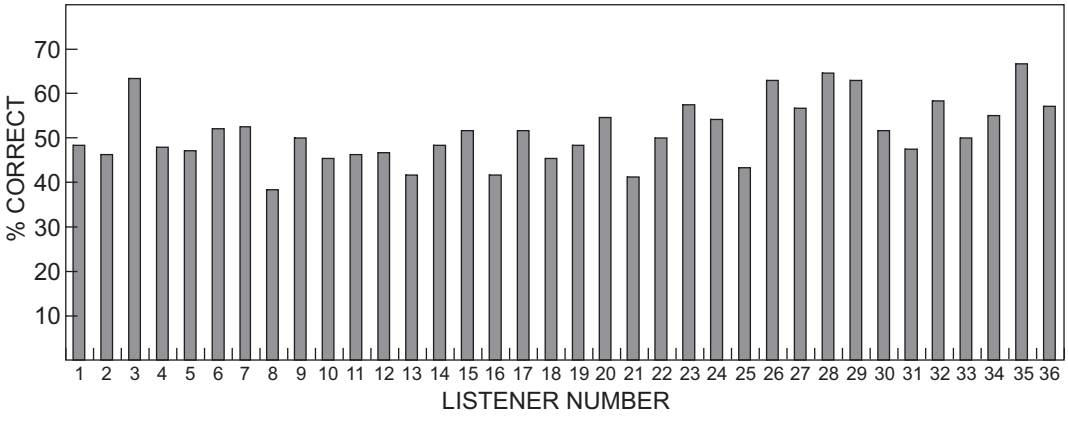

Figure 3: Percentage of correct responses per listener (BSAE sentences only) in experiment 1

Stimuli

The stimuli for this experiment were recorded in a soundproof room on a computer using a Creative sound card (Sound Blaster, Audigy2 ZS). Monophonic speech was captured at a sampling rate of 16kHz. The recordings included read speech as well as spontaneous speech. Each speaker read 36 words and 18 sentences that were constructed in order to contain vowels, diphthongs and consonants in contexts that were considered to be significant with regards to the BSAE accent. The same references that were consulted in the first experiment were consulted again in the process of constructing the word list and sentences. However, for this experiment we were able to use a greater number of the example words and contexts presented by these authors. Table 5 summarises the contexts and realisations of sounds in BSAE that were focused on and presents the words chosen for experiment 2.

Presenting carefully constructed read sentences (Table B, Appendix) in the experiment had the advantage of limiting grammatical errors. In this way we tried to prevent grammatical errors from interfering with the listeners' evaluation of a speaker's accent. On the other hand, reading sentences could limit the spontaneity of the speech. To compensate for the lack of spontaneity in the read speech, we recorded each speaker's reply when asked what s/he would do if they won the national lottery.

The words and sentences were first presented to each speaker on paper so that they could familiarise themselves with the contents of what they were about to read. During the recording itself, the items were presented in the form of a slide show on a computer screen. This made it easier for the speakers to maintain their posture while reading into the microphone. The question about the lottery was presented as the final item in the slide show.

The perceptual experiment consisted of two parts. In part A, words and sentences were presented, while part B consisted of sentences and spontaneous speech.

In part A, the 36 words were divided into two groups of 18. The 18 sentences were divided into two groups of nine. The words and sentences were subsequently organised into four data sets as shown in Table 6.

The stimuli were selected such that each data set contained one sentence and at least one word per speaker. Each set also included one word and one sentence produced by the two native speakers of English, one male and one female. The total number of stimuli in each data set was 31: 18 BSAE words, nine BSAE sentences, two SSAE words and two SSAE sentences. The amount of data per mother tongue and language group is presented in Table 7.

Part B included all the spontaneous lottery sentences as well as some of the sentences (one per speaker) used in part A. In addition to the data from the 11 BSAE speakers, Part B also included speech produced by one native English speaker (female).

All stimuli were screened before the perceptual experiment and only stimuli that were regarded as clearly audible were included in the final version of the experiment. During the experiment, the words and sentences in each set were presented in random order. The same random order was used for each listener.

Table 4: Speaker demographics for experiment 2

L1	# Listeners	Educational Level			Gender	
		Under-graduate	Honours	Masters	Male (16)	Female (20)
S.Sotho	1			1		1
N.Sotho	3	3			2	1
Tswana	2	2			1	1
Xhosa	3	2		1	1	2
Zulu	2	1	1		1	1

Listeners

The listeners were recruited to take part in experiment 2 in the same manner as for experiment 1, but students who had participated in the recording sessions or the previous experiment were excluded. A total of 20 participants between the ages of 18 and 40 years were recruited. The participants were registered for a Bachelors, Honours or Masters degree. Table 8 shows details of the listener demographics.

After completing the perception test, the participants were given a questionnaire to complete. Here they indicated that they spoke between two and nine languages other than their L1. On average, they speak three of these languages on a weekly basis. Figure 4 shows their responses when asked if they believed that they were able to identify a person's L1 based on his/her English accent when listening to him/her on the telephone, the radio or the television. Between 65 and 70% of the listeners indicated that they were able to do this and only 10% said that they could not. When asked which aspects of BSAE speech are most indicative of a speaker's L1, they gave varying responses such as 'character and facial features', 'gestures while speaking', 'stress patterns on

Table 5: Realisation of sounds in BSAE and example words used in experiment 2

Realisation of sounds in BSAE	Words		
Vowels			
Neutralisation of tense/lax vowels or long/short vowels	his	loose	passed
	sit	lot	start
	foot	force	teams
Avoidance of central vowels	nursing	village	workers
	birthday	telephone	grievances
[æ] replaced by [ɛ]	manager		
Schwa in open syllables replaced with [a]	weather		
No vowel reduction in unstressed syllables	seventy		
	diploma		
Diphthongs			
Raising diphthongs: tensing of second element	voice		
	twice		
Centering diphthongs: second element replaced	poor		
by lower vowel	sure		
	year		
Narrower diphthongs realised as single monophthongs	boat		
	great, birthday		
Consonants			
Devoicing of final plosives	bad		
Velar plosive devoicing	game		
Replacing of dental fricatives with dental plosives	brother, their		
	south		
Voiceless glottal fricative becomes voiced	hear		
Affricate [tʃ] becomes fricative [ʃ]	watch		
	opportunities		
Trilled [r] as allophone for liquid	really		
Deletion of plosives in consonant clusters	consonants		

Table 6: The four data sets used in part A of experiment 2

Set 1	Set 2	Set 3	Set 4
words I	words II	words II	words I
sentences I	sentences II	sentences I	sentences II

words', 'the way they pronounce their r's' and 'tempo and pitch'. Finally the listeners were asked to rate their own English accent on a scale of 1 to 5 (1 meaning 'strong accent' and 5 meaning 'English L1'), to which the average response was 3.4.

Test administration

Experiment 2 was executed in the same way as experiment 1 with regard to the instructions given to the participants, the computers and earphones used and the presentation in Praat.

In Part A of experiment 2, participants were asked to identify the mother tongue group of the speaker. This part of the experiment was presented in the same way as in the previous perception test.

In Part B, participants were asked to identify the mother tongue group of the speaker and to rate his/her English accent. This part included a second row of buttons on which the participant could click. These included buttons numbered from 1 to 5 — 1 meaning 'strong accent' and 5 meaning 'English L1'.

Table 7: Data per L1 for part A of experiment 2

L1	Words	Total words	Sentences	Total sentences
Northern Sotho	20		15	
Southern Sotho	3	Sotho	5	Sotho
Tswana	13	36	10	30
Xhosa	21	Nguni	17	Nguni
Zulu	15	36	11	28

Table 8: Listener demographics for experiment 2

L1	# Listeners	Educational Level			Gender	
		Under-graduate	Honours	Masters	Male (10)	Female (10)
S.Sotho	5	5			4	1
Tswana	4	4				4
Xhosa	8	3	3	2	5	3
Zulu	2	2			1	1
Ndebele	1	1				1

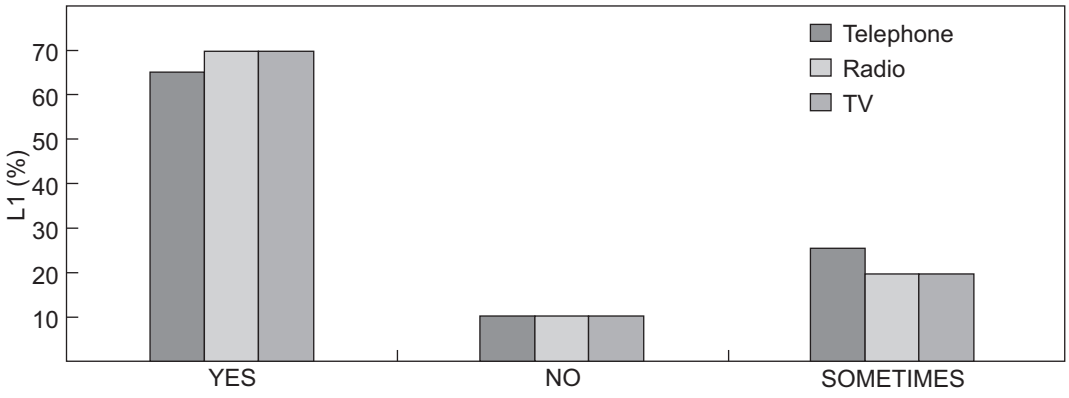

Figure 4: Determining a speaker's L1 via various media

Participants took short breaks after each set of 31 stimuli and they took an average of 20 minutes to complete the perception test.

Results

The results for experiment 2 were calculated and analysed in the same way as for experiment 1. However, the relevant graphs show a fourth percentage correct/incorrect bar representing the results for the 'lottery' stimuli. Responses to the repeated sentences that occurred in Part B are not displayed here, but will be briefly discussed.

When combining all stimuli in one data set, it was found that 59% of the responses were correct. Seven per cent of the listeners' responses were 'I don't know', and these were mostly (80%) responses to the single word stimuli. As Figure 5 shows, listeners achieved a significantly higher score for the sentences and lottery items, but they did not seem to be able to determine the accents for the word stimuli.

As was the case in experiment 1, the percentage of correct responses to stimuli produced by SSAE speakers was significantly higher (91%) than the corresponding value for the combined data set, which indicates that the listeners could clearly distinguish between the BSAE and SSAE accents.

Figure 6 shows that, when the responses to the SSAE stimuli are removed from the data set, the percentage of correct responses for all stimuli decreases to 54%. Listeners did, however, perform significantly better when judged only on their responses to sentences and lottery items. Here their responses show some correlation to that shown in Figure 4, where just over 60% of the listeners claimed that they could distinguish between different L1 accents over the telephone. We expected listeners to perform better when judging the lottery items compared to the sentences, since the spontaneous lottery items must have contained more prosodic cues than the sentences (cf. Wissing's (2002: 21) statement that perceived differences between the English of speakers of different Bantu languages would most probably be on the suprasegmental level). However, according to the results in Figure 6, there is almost no difference between the listeners' responses to these two types of stimuli. From these results it was inferred that, on average, the listeners could not confidently determine whether a speaker's mother tongue was from the Nguni or Sotho language group.

Figure 7 shows the percentage of correct responses per listener. For this analysis, only the responses to BSAE sentences were calculated. The 'best' listener achieved 81% correct, while the lowest scoring listener classified 42% of the stimuli correctly.

The results of experiment 1 suggested that listeners had a bias towards their own mother tongue. This trend was not observed in the current experiment. However, the listeners were not

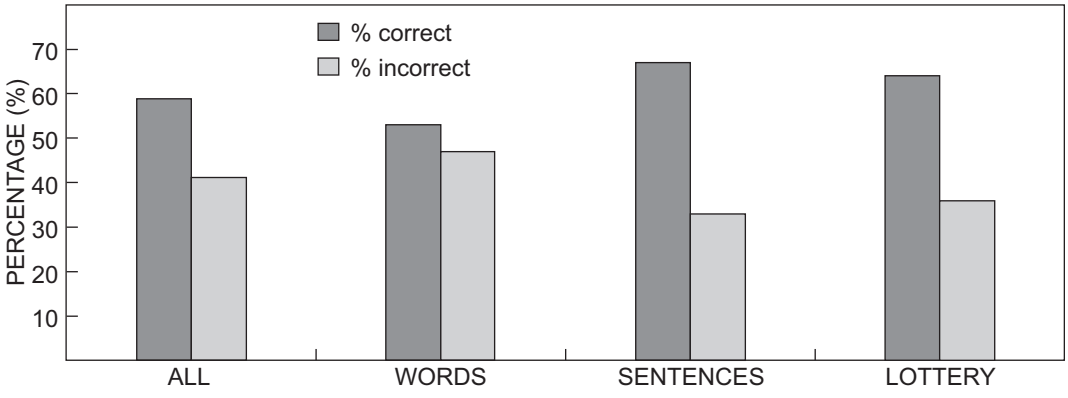

Figure 5: Percentage of correct and incorrect responses for all stimuli and all accents in experiment 2

very consistent in judging the accents: for the eleven sentences that were repeated in part B of the test, listeners chose the same accent they had chosen in part A only seven times (on average). Not one listener chose the same accent twice for all eleven repeated stimuli.

The accuracy with which the mother tongue group of each speaker could be determined, is illustrated in Figure 8. The graph shows that Xhosa speaker X02M's English accent was influenced by his mother tongue to such a degree that the listeners could easily determine his accent. On the other hand, listeners had great difficulty in correctly identifying the accent of Northern Sotho speaker NS02F.

Figure 9 shows the accent ratings for each speaker. For some of the speakers there seems to be a clear correspondence between their accent rating and the accuracy with which their accent could be determined. For example, speaker X02M whose L1 was mostly determined correctly also has the lowest accent rating, while the speaker with the lowest percentage of correct responses (NS02F) also has the highest accent rating. However, for the other speakers the correlation between their accent rating and the accuracy with which their accent could be determined, is weak (Pearson's r = –0.5).

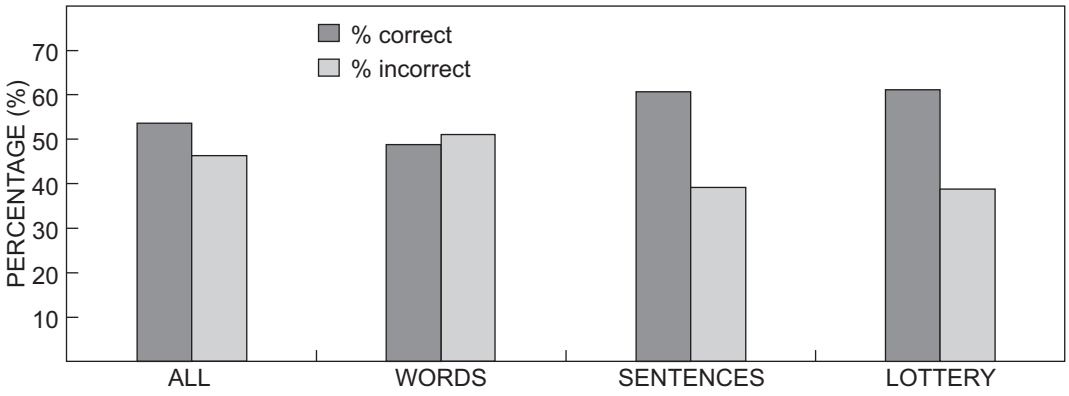

Figure 6: Percentage of correct and incorrect responses for all BSAE stimuli in experiment 2

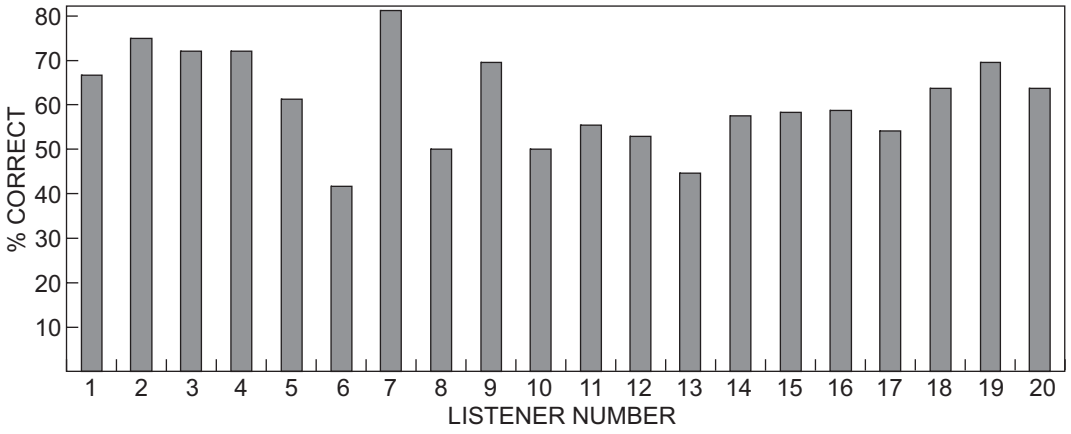

Figure 7: Percentage of correct responses *per listener* (BSAE sentences only) in experiment 2

Discussion

Native speakers of South Africa's Bantu languages are often optimistic about their ability to determine a person's mother tongue background based on his/her English accent. Although many admit that physical features may assist them in determining a person's L1, most are convinced that they can also identify accents over the telephone or on the radio. However, the results of this study indicate that, while the majority of the participants have no difficulty distinguishing between SSAE and BSAE accents, no unconditional statements can be made about their ability to correctly identify a person's mother tongue background based on his/her English accent.

In the first perceptual experiment, which was conducted using telephone speech, none of the participants were able to recognise the BSAE speakers' accents beyond the chance level of 50/50 Nguni/Sotho. Furthermore, we expected that sentences would be easier to classify than words, but the results showed no significant difference between the listeners' responses to word and sentence stimuli.

Because the telephone data suffered from a number of shortcomings, a second perceptual

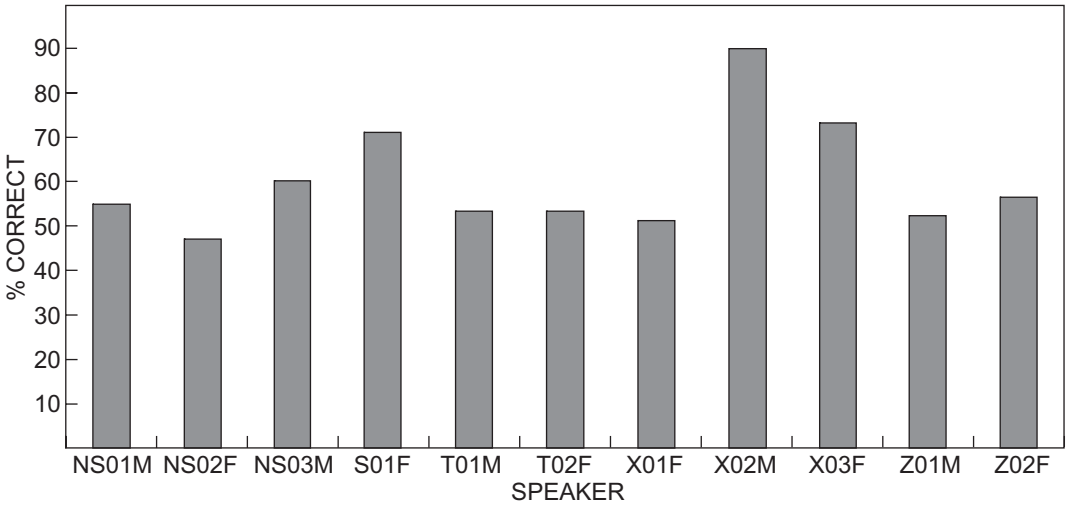

Figure 8: Percentage of correct responses *per speaker* (BSAE speakers, sentences only) in experiment 2

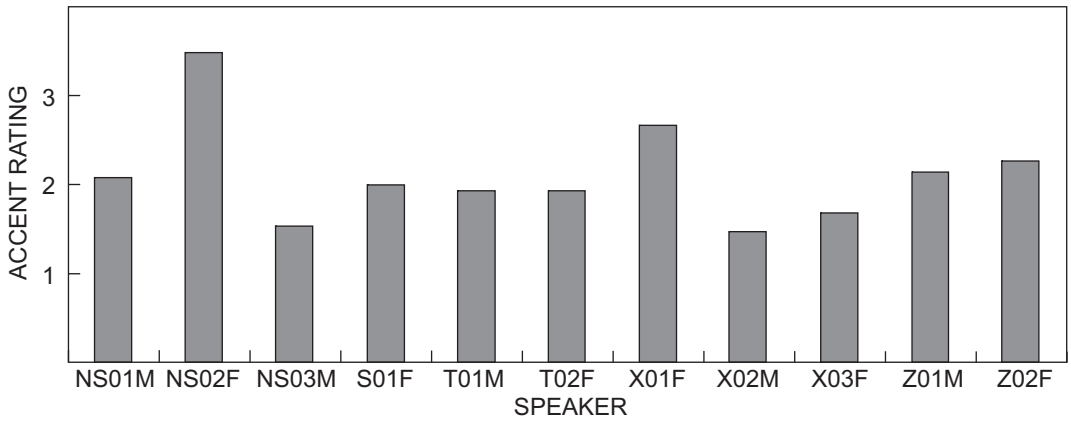

Figure 9: Accent rating per speaker on a scale of 1 to 5 for experiment 2 (BSAE speakers, sentences only)

experiment was conducted using good quality recorded speech, specifically designed to include acoustic cues relevant to different BSAE accents. In this instance the listeners performed slightly better in classifying sentence stimuli than in classifying word stimuli, but, as was the case in experiment 1, the overall results indicate that none of them were consistently able to correctly distinguish the Nguni accent from the Sotho accent. Some of the sentence stimuli comprised spontaneously produced speech, which we included in an attempt to present the listeners with examples of spontaneously produced prosodic patterns. Contrary to our expectations, there was almost no difference between the accuracy with which the listeners identified the accents of the read and spontaneous sentence stimuli. However, for a number of speakers, most listeners were able to correctly determine their mother tongue group from their English accent. For these speakers, the listeners were also in agreement that the particular speaker had a strong accent, which indicates some correlation between perception of accentedness and the correct identification of L1.

The results of both experiment 1 and 2 contradict the notion that different mother tongues influence BSAE to such a degree that the speaker's L1 is easily perceived. Within the limitations of our experimental set-up, we can therefore concur with the findings of Van Rooy & Van Huyssteen (2000) that Nguni-English and Sotho-English are not distinctly different.

Both the listeners and speakers who participated in the experiments were from the mesolectal group, mostly originating from a multilingual environment such as the university. There are many aspects that may influence a speaker's accent that were not necessarily represented in this study. A follow-up study with participants who are carefully selected to represent different educational backgrounds and different levels of proficiency and exposure to English, might be considered. However, the results of this study are quite relevant if we take into account that the demographics of the current participants are similar to those who will eventually use the speech input applications that we aim to develop.

In conjunction with this study, we conducted an ASR experiment on the AST BSAE data in order to see if recognition results improve when the BSAE data set is split into Nguni-English and Sotho-English (De Wet, Niesler & Louw, 2006). The recognition results for models trained on accent-specific data and those trained on accent-neutral data did not differ significantly. Since we found a correlation between the perception test results and the recognition results, we are confident that Nguni-English and Sotho-English need not be treated as different accents for ASR purposes.

Notes

[1] Assigning appropriate and commonly accepted labels to the different varieties of English spoken in South Africa is still a bone of contention amongst linguistic scholars (for example, Coetzee-Van Rooy & Van Rooy, 2005). Based on the arguments presented in Van Rooy (2000b) and De Klerk (2003), we have decided to use the terms 'Standard South African English' (SSAE) and 'Black South African English' (BSAE) in this paper. The term BSAE refers to the English spoken by second language speakers whose mother tongue is one of the indigenous Bantu languages of South Africa. For the purposes of this study, English as spoken in other African countries with indigenous Bantu languages, such as Zimbabwean English, is not included.

[2] The Nguni and Sotho language groups are two of the five language groups that comprise the South Eastern Bantu languages. Nguni consists of a group of sister languages that includes Xhosa, Zulu, Ndebele and Swazi. The Sotho language group includes Southern Sotho, Tswana, and Pedi (Northern Sotho) (Baily, 1995).

[3] All speakers and listeners who participated in our experiments received a monetary reward for their contribution.

[4] ANOVA tests were done on the results of both experiments. Any claims of significance were made based on the results of these tests.

Acknowledgements — This research was supported by the National Research Foundation under grant number GUN2072874.

References

Baily R. 1995. The Bantu Languages of South Africa. In: Mesthrie R (ed.), *Language and Social History*. Cape Town: David Philip Publishers, pp 19–38.

Coetzee-Van Rooy S & Van Rooy B. 2005. South African English: labels, comprehensibility and status. *World Englishes* **24**(1): 1–19.

De Klerk V. 2003. Towards a norm in South African Englishes: the case for Xhosa English. *World Englishes* **22**(4):463–481.

De Wet F, Niesler TR & Louw PH. 2006. *Nguni and Sotho varieties of South African English: Distant cousins or twins?* Presentation at the Multiling-2006 workhop. Available at: http://www.unistel.co.za/multiling2006/pdfs/dewet.pdf [accessed: 14 June 2006].

Gough D. 1996. Black English in South Africa. In: De Klerk V (ed.), *Varieties of English around the world, Focus on Africa.* Amsterdam: John Benjamins Publishing Company, pp 53–77.

Mayfield Tomokiyo L. 2001. *Recognizing Non-Native Speech: Characterizing and Adapting to Non-Native Usage in Large Vocabulary Speech Recognition.* PhD thesis. Pittsburgh: Carnegie Mellon University.

Roux JC, Louw PH & Niesler TR. 2004. The African Speech Technology Project: An Assessment. *Proceedings of the 4th International Conference on Language Resources and Evaluation* **I**:93-96. Lisbon: European Language Resources Association.

Strik H & Cucchiarini C. 1999. Modeling pronunciation variation for ASR: A survey of the literature. *Speech Communication* **29**: 225–246.

Van der Walt C. 2000. The international comprehensibility of varieties of South African English. *Word Englishes* **19**(2): 139–153.

Van Rooy B. 2000a. The consonants of BSAE: Current knowledge and future prospects. *South African Journal of Linguistics* Supplement **38**: 35–54.

Van Rooy B. 2000b. Introduction. *South African Journal of Linguistics* **38**: i–vii.

Van Rooy B. 2004. Black South African English: phonology. In: Schneider E, Burridge K, Kortmann B, Mesthrie R & Upton C (eds), *Handbook of Varieties of English 1*. Berlin: Mouton, pp 943–952.

Van Rooy B & Van Huyssteen G. 2000. The vowels of BSAE: Current knowledge and future prospects. *South African Journal of Linguistics* Supplement **38**: 15–33.

Wissing DP. 2002. Black South African English: A new English? Some observations from a phonetic viewpoint. *World Englishes* **21**(1): 129–144.

Appendix: Experimental stimuli

Table A: Sentences used in experiment 1

Can I cancel this booking?
Change the number of adults to two.
Just check to see if the dates are available.
Will you please continue?
The day before yesterday.
Scientific progress comes from the development of new techniques.
Directory enquiries please.
Eleven minutes past two.
I would like to speak English.
Half past five.
I need some help.
We just want some information.
Bring your ID with you when you book into the hotel.
Give me a list.
Make a new call.
Now I'm doing matric.
Mini-bar
Only the best players enjoy popularity.
Is there an operator I can talk to?
Helen's stint as league manager provided useful opportunities.
Be careful not to plough over the flower beds.
Perhaps I should make a call.
Can you repeat what you just said?
The elementary practical task of going to the meetings.
A telephone number.

Table B: Sentences used in experiment 2

The new manager has created many opportunities for staff development.
That girl is working very hard to pass grade 12, she wants to do a nursing diploma next year.
May I please use your telephone?
Since he had to go to town on foot, he had to make an early start.
To play the game both teams must sit on the floor.
The workers are invited to discuss their grievances with government representatives.
It was really wonderful to see my grandmother again!
Special bursaries will be available for students who want to undertake further education.
Everyone in the village was invited to the feast: rich and poor.
He will devote a great deal of interest to this project.
My brother has invited 70 people to his birthday party — that is quite a lot.
Will you be available for an appointment later today?
Bad weather is predicted for tomorrow: the wind will be gale force.
There was so much noise in the hall that it was difficult to hear his voice.
She made sure that there was no-one around before she turned the dog loose.
He must be mad to spend so much money on a watch!
This boat has already sailed to the South Pole twice.
According to the people outside the court, the judge had passed a fair sentence.