Southern African Linguistics and Applied Language Studies 2008, 26(2): 255–265
Printed in South Africa — All rights reserved

SOUTHERN AFRICAN LINGUISTICS
AND APPLIED LANGUAGE STUDIES
ISSN 1607–3614 EISSN 1727–9461
DOI: 10.2989/SALALS.2008.26.2.6.570

Vowel variation in Southern Sotho: an acoustic investigation

Etienne Barnard[1]* and Daan Wissing[2]

[1]*Meraka Institute, Human Language Technologies Research Group, CSIR, PO Box 395, Pretoria, 0001, South Africa*

[2]*Centre for Text Technology (CTexT), Private Bag X6000, North-West University, Potchefstroom, South Africa*

** Corresponding author, e-mail: ebarnard@csir.co.za*

Abstract: For the development of Human Language Technologies such as automatic speech recognition systems or text-to-speech systems, exact acoustic and phonological information of the language in question is essential. In the case of Southern Sotho, the current study is a first step in the direction of providing such information. Measurements of the frequencies of the first two formants of the vowels represented by the orthographic symbols **i e a o u** are made in several contexts in order to motivate particular phonetic and phonemic representations of these vowels, regardless of the conventional orthography. Our measurements corroborate some of the earlier impressionistic research into Southern Sotho. In particular, the four variants of the vowels orthographically represented by 'e' and 'o' respectively are clearly present, indicating that there are at least seven distinct (phonemic) vowels. These variants are distinguished mainly by vowel height, that is, by F1 differences. On the other hand, our results oppose the established knowledge in some important ways. Most notably, we find that harmonic vowel raising produces a larger change in vowel height than the height difference between the pairs of phonemically distinct middle vowels (/ɔ/ and /o/, /ɛ/ and /e/), respectively. The finding is seemingly reliable as this tendency is constant across all words investigated. An interesting finding is a significant modification of the vowels /u/ and /a/ in certain contexts. These are vowels previously believed not to have any allophones at all. Lastly, a number of important unresolved issues are highlighted.

Introduction

The development of Human Language Technologies (HLTs) depends on the availability of several forms of linguistic information. A state-of-the-art spellchecker, for example, requires not only a sufficiently diverse and accurate lexicon, but also accurate descriptions of morphological processes, syntactic structures and the like. For world languages such as English and Mandarin Chinese, such information is widely available, and the principal challenge of the HLT developer is to represent such knowledge in a way that meets the need of users. For the vast majority of languages, however, the necessary information is much harder to obtain. The relevant expertise is often held by a few (if any) linguists, and is generally not codified in a format that directly supports the development of HLT systems. Since the population of linguists working on a particular language may be small, there is often insufficient independent verification of those findings that have been published, which forces the developers of HLT systems to select from sometimes contradictory accounts of basic linguistic facts (while somehow filling in those portions that are necessary for their systems but not available in the published literature).

For spoken language technologies, such as automatic speech recognition (ASR) systems or text-to-speech (TTS) systems, an exact definition of the phonemic inventory of a language is often the first linguistic foundation required. A complete phoneme set is employed to express all lexical items in terms of a productive set of units; this 'pronunciation dictionary' is then supplemented with relevant information such as the presence of major phonetic variants or suprasegmental phenomena, and used as the basis for the development of the statistical models that are the heart of HLT systems. A reliable description of the phonology (along with major phonetic variants) is therefore of great importance.

In the current contribution, we describe our efforts to establish an exact and reliable description of the vowel system in a less well-studied language, namely Southern Sotho. Although Southern Sotho (along with the closely related languages Setswana and Northern Sotho) has been the subject of active linguistic study for almost a century, there is still considerable debate on the precise nature of its vowel system. In the second section we review this debate, and contrast the issues that remain contentious with those that are considered uncontroversial. The third section proceeds with a description of the experimental methods that were employed in order to address some of the issues raised in the second section. The fourth section contains the results of our investigations, and in the fifth section we conclude with a summary of our main findings as well as a discussion of unresolved issues that require further study.

Background

The 7 phonemes[1] /i e ɛ a ɔ o u/ are generally accepted for Southern Sotho (Tucker, 1929; Doke & Mofokeng, 1957; Guma, 1971; Roux, 1983; Khabanyane, 1991; and Selebeleng 1997) as well as for Setswana (Krüger & Snyman, s.a.) and Northern Sotho (Poulos & Louwrens, 1994). These phonemes are also generally used in pronunciation dictionaries for these languages. They are described in various articulatory terms. Some authors (for example, Tucker, 1929; and Doke & Mofokeng, 1957) make use of the cardinal vowel system introduced by Daniel Jones and also employed by the International Phonetic Association (see Figure 1 for a slightly modified version, where vowel rounding is suppressed since it is not relevant to our discussion). Ignoring rounding, this system indicates 8 positions, 1 cardinal vowel per position. Front as well as back vowels can be close (high),[2] that is, cardinal vowels 1 and 8; and open (or low), that is, cardinal vowels 4 and 5. Cardinal vowels 2, 3 and 6, 7 are all mid, specified by the addition of close or open, yielding the combinations close-mid (cardinal vowels 2 and 7), and open-mid (cardinal vowels 3 and 6). Six of the 7 Southern Sotho phonemes mentioned previously, /i e ɛ ɔ o u/, then correspond to cardinal vowels 1, 2, 3, 6, 7, 8; /a/ takes up a position between cardinal vowels 4 and 5.[3]

In the earlier literature (cf. notably Tucker, 1929; and Doke & Mofokeng, 1957) on the Southern Sotho vowels, two allophones of the open-mid phonemes are acknowledged (that is, front [ɛ̱] and back [ɔ̱]).[4] Generally, such allophones are explained by a process of vowel raising,[5] sometimes also referred to as assimilation or vowel harmony. This process of vowel raising is indicated in Figure 1 by solid arrows, and is predictable on the basis of specific phonetic contexts. A variety of such phonetic contexts is responsible for such vowel raising. The close vowels /i/ and /u/ in following syllables (cited by all of the authors on Southern Sotho quoted previously) are the most important triggers for the raising of vowels in previous syllables. Sometimes close-mid vowels are also identified as having such a raising effect on preceding open-mid vowels. Furthermore, some consonants are mentioned as having similar raising effects too, namely, syllabic nasal consonants /n/, /ɲ/, /ŋ/ (including the locative suffix /-ŋ/) and the alveolar consonants /ts'/, /s/. Vowels in word-final positions are said to be raised as well (especially mentioned by Tucker, 1929; and Doke & Mofokeng, 1957).

In addition, based on the existence of minimal pairs such as /p'ɛp'a/ ('a piece of chuff') and /p'ɛ̱p'a/ ('carry on the back'), Khabanyane (1991) concludes that these vowels are not allophones but phonemes of Southern Sotho, thus to be indicated as /ɛ̱/ and /ɔ̱/, and not as [ɛ̱] and [ɔ̱].[6] Selebeleng (1997) follows her in this assumption.

In more recent work, two more vowels have come to be accepted, in addition to the above-mentioned nine (phonetic) vowels, namely the raised close-mid [e̱] and [o̱] (Guma, 1971; Khabanyane, 1991; Selebeleng 1997; and Roux, 1983, an exception as he feels that a nine-vowel system is adequate).

In summary, then, the existence of seven phonemic vowels and two additional open-mid allophones is currently not contentious. However, the possibility of an additional pair of phonemic vowels (which are collocated with these two allophones) and yet another pair of (close-mid) allophones, is somewhat more controversial.

Note that we began our discussion at the orthographic level because this allows us to indicate unambiguously the tokens under discussion – not because we believe that the five orthographic forms have any special status (as pointed out above, the existence of at least seven vowels in

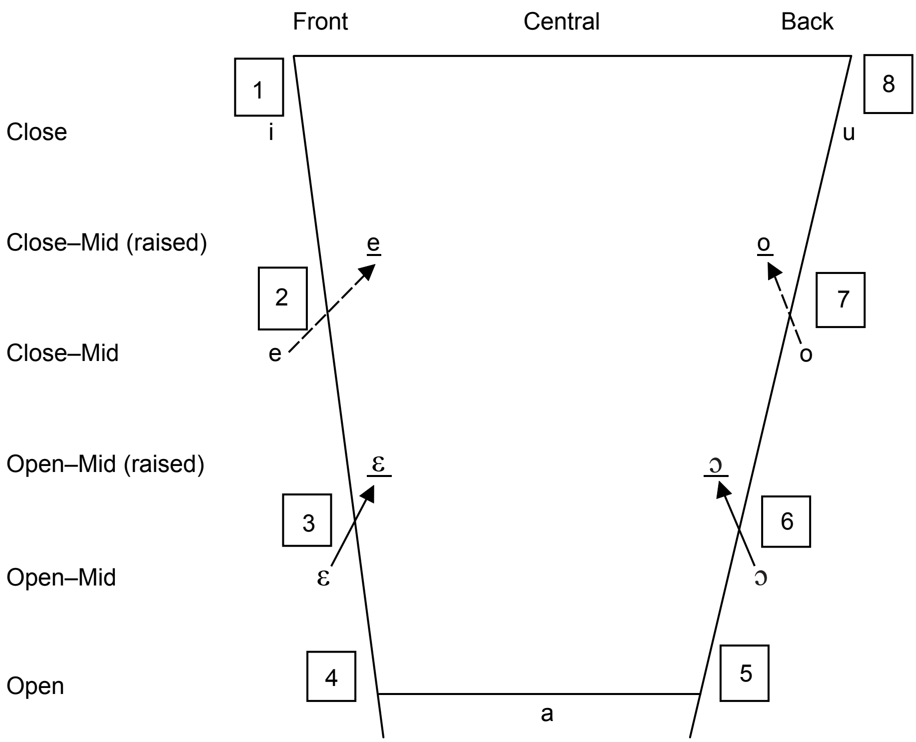

Figure 1: The Southern Sotho phonemes and allophones imposed on the IPA Vowel Chart
Phonemes appear outside the figure, allophones inside (they are underlined)

Southern Sotho is uncontroversial). It is also true that the orthographic form (without diacritics) is usually taken as input by systems for speech recognition and speech synthesis, since these are the forms commonly used in writing.

Methods

Participants

Twelve speakers of Southern Sotho, 6 of each sex, volunteered to read the speech materials (speech stimuli). Of these, 6 were either students or employees of the Qwaqwa campus of the University of the Free State (3 males and 3 females, aged 19–45 years), and 6 were Grade 11 learners at the Lekota Secondary School (also 3 males and 3 females, aged 17–20 years). Both these institutions are situated in the rural region of Phuthaditshaba, Eastern Free State, South Africa.

Speech stimuli

We focused on the mid vowels of Southern Sotho, which include the open-mid and close-mid vowels. In light of the uncertain status of some of these (in particular, the uncertainty surrounding the number of distinct phones as described in the second section), we take no *a priori* position as to the precise quality of the vowels of the words in the specific constructions we chose. For this reason we simply refer to these vowels as either ‘e’ or ‘o’. By careful measurement of the formant frequencies of these vowels, we hope to motivate particular phonetic and phonemic representations of these vowels, regardless of the conventional orthography. In this way, we hope to provide an empirical basis for the transcriptions provided in dictionaries and standard texts.

We recorded each speaker 3 times; only the second and third readings were taken into account – the first recording was used merely for acquainting the experimental subjects with the task. We measured, per speaker and per reading, 84 vowel tokens of the 5 orthographically distinct types ('i e a o u'), combined in 36 word tokens. Not all speakers successfully produced all repetitions of all tokens and we had access to the following totals: **i** (97), **e** (1 995), **a** (1 288), **o** (1 229), and **u** (130).

We made use of 2 types of stimuli: sentences and word lists. For the analysis below, we used the data obtained from the sentences only, and only that data will be described. Sentences were derived from the list of basic sentences in the appendix. Readers were presented with the positive forms which they first had to read and then say in the negative.[7] We then measured both stems (that is, the vowels -e-, -o-, -u-), as well as the suffixes -a# and -e#.

We selected 8 minimal pairs (that is, words that are (mostly) written identically but have different meanings), 3 with 'e' as vowel, 4 with 'o', and one with 'u'. Words are bisyllabic in all cases except for those in the past tense, which are trisyllabic (see the appendix). The final pair, with -u- as stem, was included for purposes other than the present ones. That is, we were interested in the tonal behaviour of -u- in minimal pairs such as *Ke **bua** ditaba* ('I speak the news'), and *Ke **bua** kgomo* ('I flay the cow').

The meanings of each word were made clear by the specific contrast in the pair of sentences in which they were embedded. That is, participants were presented with both sentences together and were then asked to distinguish between the 2 intended meanings by choosing the meaning that matched each sentence most closely. In a perception test in which 40 mother tongue speakers of Southern Sotho took part, we confirmed that this was done successfully. The complete sentences with their glosses are presented in the appendix. We therefore had 4 variants of every orthographic word – present indicative positive and present indicative negative versions of both items in a minimal pair. In addition, we also made recordings of different temporal types (future as well as past tenses); these are not included in our main analysis below, except for 2 data points in Figure 4.

Finally, we measured the vowels of the function words, such as subject concords and pronouns. These are the words: *ba*, *di, ha*, *ho*, *ke*, *se*, *tlo*.

Recordings

Readings were done at the campuses of the two institutions mentioned under 'Participants' above. Care was taken to avoid external noises and excessive reverberation. We used professional equipment, and all recordings were stored digitally for future processing.

Speech processing

Oscillograms and spectrograms were produced in Praat (Boersma & Weenink, 2007). During the process of segmentation speech signals were inspected auditorily and visually for determining vowel boundaries. Conventional segmentation criteria were followed (Grabe & Low, 2002). We set out to collect high-quality broadband speech. Speech was sampled at standard 16 kHz, encoded in 16-bit linear PCM and stored directly on the hard disk of a computer in *wav* format.

All relevant vowels were annotated in Praat, in such a way that Vowelyse[8] (Van der Walt & Wissing, 2003) could do appropriate extraction and calculation of acoustic information relevant for the analysis of the parameters in question. Spurious formants and pitch values are automatically detected and flagged for later inspection and evaluation. Most of these were removed from the raw data. The acoustic results were then analysed in Statistica, or with custom-written software.

Statistical analysis

For the purposes of the present study we were interested in vowel quality and possible differences between different vowel classes. Therefore we concentrated on mean values and standard deviations of formant frequencies – we report on the first and second formant frequencies, which we label F1 and F2, respectively. These formant frequencies are widely known to correlate well with listeners' perceptions of vowel height and fronting, respectively. In particular, small values of F1 correspond to the perception of a vowel as 'close' (or high), whereas vowels with small values of F2 are perceived as back vowels. Furthermore, we made use of analyses of variance and some post hoc comparisons.

Since we wish to combine measurements from several speakers, we need to compensate for speaker differences in vowel production: if one were simply to pool the values for all speakers when computing statistical measures, a large fraction of the observed variance would arise because of speaker differences rather than the phenomena that we are trying to investigate. This is an issue that has been researched actively for many years (Lobanov, 1971; Neary, 1989; and Adank *et al*., 2004). No universally satisfactory compensation method has been found (Watt & Fabricius, 2002), and the methods favoured in the literature cannot be applied straightforwardly to our data since it is not clear which units should be grouped together when calculating normalising constants such as variances and extremal formant values. We have therefore chosen a fairly simple approach to normalisation, which allows us to avoid these complications at the cost of less accurate compensation for speaker variability.

The mean F1 and F2 values of each vowel of interest in a particular word (as represented orthographically) for each speaker were computed. The variations of these word-specific vowels were investigated by pooling the offsets relative to these mean values. (Note that we are treating semantically distinct homographs as separate words, and therefore *not* pooling such homographs together in this step.) For example, to determine whether the quality of 'o' changed significantly between two different contexts *Ca* and *Cb*, we first calculated the mean F1 and F2 for each speaker across all occurrences of 'o'; let this mean value for speaker *s* be denoted by the (two-component) vector ***o****s*. The difference between this mean value and the formant frequencies of a particular token *i* produced by speaker *s* in context *Ca* is written as ***d****ai* = ***o****ai* – ***o****s*, and these differences are averaged across all tokens and speakers to form ***d****a* (and similarly ***d****b*). If and only if ***d****a* and ***d****b* differ by a statistically significant amount, do we conclude that these contexts are expressed with different vowel qualities.

Our computations have a limited ability to compensate for speaker differences, since they are equivalent to the assumption that these differences amount to nothing but an offset added to formant frequencies – clearly a significant oversimplification. We have nevertheless found the procedure to be useful towards its stated goal. Graphical representations of the results were created by adding these offset vectors to the mean formant frequencies (averaged across all speakers, types and tokens); standard errors on these graphs represent the variance of the offsets such as ***d****a*.

Results

In the subsection 'Vowel raising' below, we first establish some basic facts about the two forms of vowel raising mentioned in the second section. We then show that similar phenomena also occur for some of the other vowels of Southern Sotho, and that other unexpected phonetic processes have a comparable acoustic impact to the well-documented harmonic vowel raising described in this subsection.

Vowel raising

As discussed in the second section, we needed to verify which of the potential minimal pairs (with respect to vowel quality) indeed displayed the differences in vowel quality described as vowel raising. Figures 2 (a) and (b) show the mean first and second formant frequencies (F1 and F2) obtained for the 4 sets of potential minimal pairs containing the vowels orthographically represented as *e* and *o*, respectively. For each orthographic base form (that is, a word, such as *beha* or *roka*), 4 data points are shown. These represent the 2 semantically distinct forms (that is, the 2 members of a minimal pair) in both the basic and harmonic-raising contexts (that is, when used in the present indicative positive and present indicative negative, respectively). These figures are somewhat complex, and to understand them it is best to first focus on a particular (orthographic) word – for example, *hlola*, which is indicated by the triangular symbols ∆ in Figure 2 (b). The 4 cross hairs containing such triangles, from bottom to top, correspond to: the lower phoneme /ɔ/, without harmonic raising; the higher phoneme /o/, without harmonic raising; the lower phoneme, with harmonic raising; and the higher phoneme, with harmonic vowel raising. The interpretation of Figure 2 (a) is exactly analogous, with /ɛ/ replacing /ɔ/ and /e/ replacing /o/.

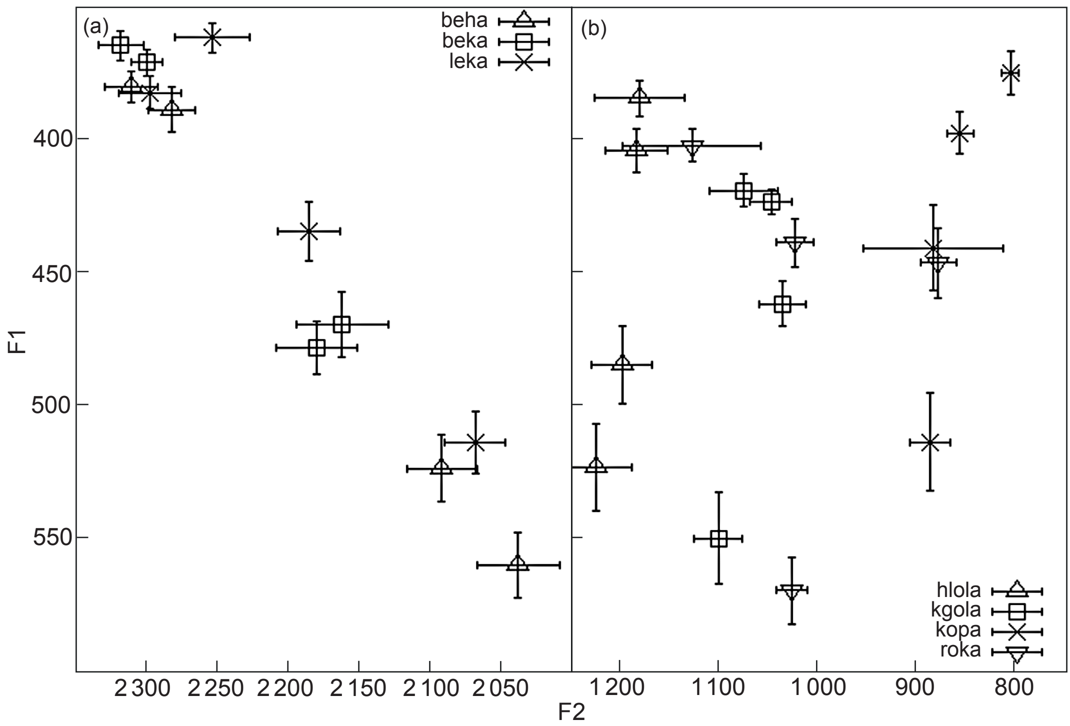

Figure 2: Vowel charts of the initial vowel in several words containing the vowels orthographically represented by e and o, respectively
(The error bars indicate one standard error, when averaged over several speakers and repetitions.)

For the vowels represented by *e*, only *leka* unambiguously displays all 4 variants in F1-F2 space (that is, there are no overlaps in the standard errors along either the F1 or the F2 dimension between all 4 variants). Similarly, *hlola*, *kopa* and *roka* present all 4 variants orthographically represented by *o*. We therefore base our analysis on these 4 orthographic words. (It is not surprising that several tokens do not show the full range of expected variation. On the one hand, it could be that the majority of speakers in our sample group do not use vowel height to distinguish between the different semantic forms of a word – this seems to be the case with *beka*. On the other hand, it could be that limitations on our statistical process, or other linguistic influences, obscure differences which may in fact be present under certain circumstances, as seems to be true of *beha* and *kgola.*)

As expected (Khabanyane, 1991), when 4 variants are present, they differ primarily in their vowel height (that is, in F1). As stated above, the same ordering of vowel heights is observed in all these cases: the lower and higher of the 2 semantically distinct cases (that is, positive forms of members of a minimal pair) are lowest and second-lowest, followed by the harmonically raised variants of the vowels (that is, those appearing in the corresponding negative forms). In all instances, the same semantic case (within a particular minimal pair) which has the higher frequency in the basic form also has the higher frequency when raised by vowel harmony.

Other phonetic effects

Figure 3 shows an F1-F2 plot containing the following vowels:

- -a- in *ha*, occurring in a present indicative negative construct as well as a future negative construct;
- -i- in *di*, occurring in 2 semantically distinct contexts; and
- -u- in *bua*, occurring in the same four variations as the vowels in the previous subsection, and also in the future negative.

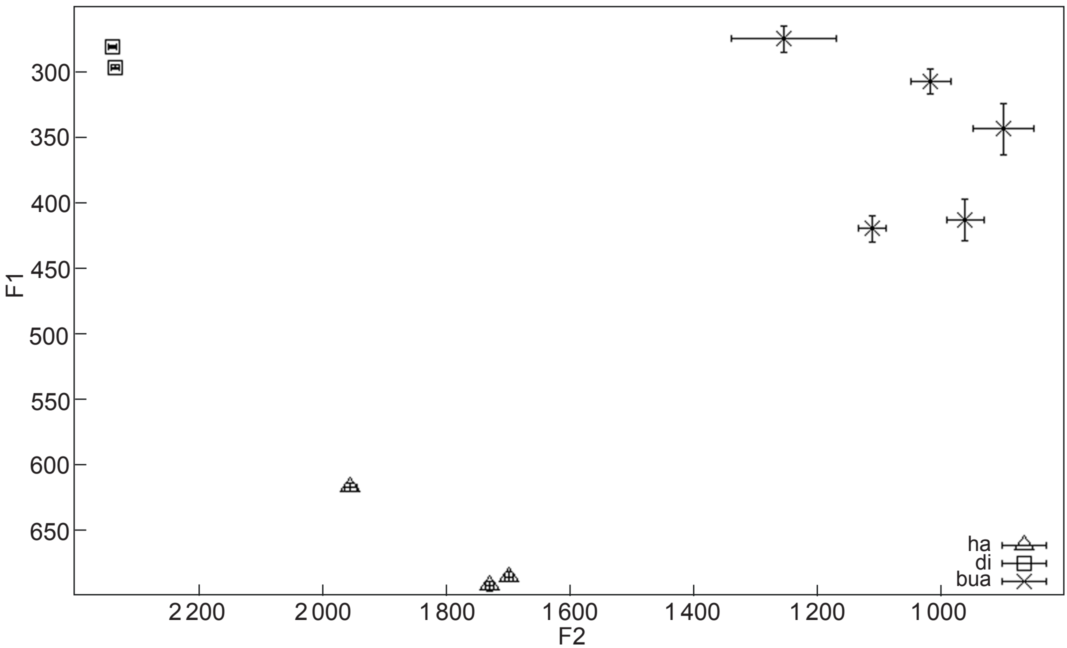

Figure 3: Vowel chart showing variations of /a/, /i/ and /u/

We see that the semantic context does not modify the vowel quality of -i-; that no such variation occurs would be predicted based on earlier descriptions (Doke & Mofokeng, 1957). The formants of -u- in *bua*, however, are seen to vary significantly across the 5 contexts, and details of the association between syntactic/semantic construct and vowel location are shown in Figure 4. Since /u/ is not described as having 2 phonemic variants in the literature referred to above, the semantic distinction between the 2 cases of *bua* is traditionally ascribed to tonal differences. Here we see, however, that a significant change in vowel quality is at least associated with this distinction (and may even be its primary indicator).

Finally, we see that both -a- in *ha* and -u- in *bua* undergo significant acoustic changes when used in the future negative. Both vowels are raised significantly in comparison with their basic version (though, in the case of *bua*, not as much as the raising that occurs in the present indicative negative). However, the second formants of the 2 vowels change in opposite directions: F2 of -u- is reduced, indicating a stronger back vowel, whereas F2 of -a- is increased. In order to see all the relationships between all vowels studied in this research, we also show the vowels in Figures 2 and 3 on the same coordinate system in Figure 5.

Discussion and conclusions

We have seen that detailed measurements of the formant frequencies of core vowels in Southern Sotho confirm several of the facts that have been reported in the literature and educational materials. In particular, the four variants of the vowels orthographically represented by *e* and *o*, respectively, are clearly present and distinguished primarily by vowel height. For *e*, we unambiguously see all four variants in the orthographic word *leka*, whereas *hlola*, *kopa* and *roka* display the analogous variability of *o*.

Our measurements do, however, also contradict the received wisdom in some ways. Most significantly, we find that harmonic vowel raising produces a larger change in vowel height than the height difference between the pairs of phonemically distinct middle vowels (/ɔ/ and /o/, /ɛ/ and /e/),

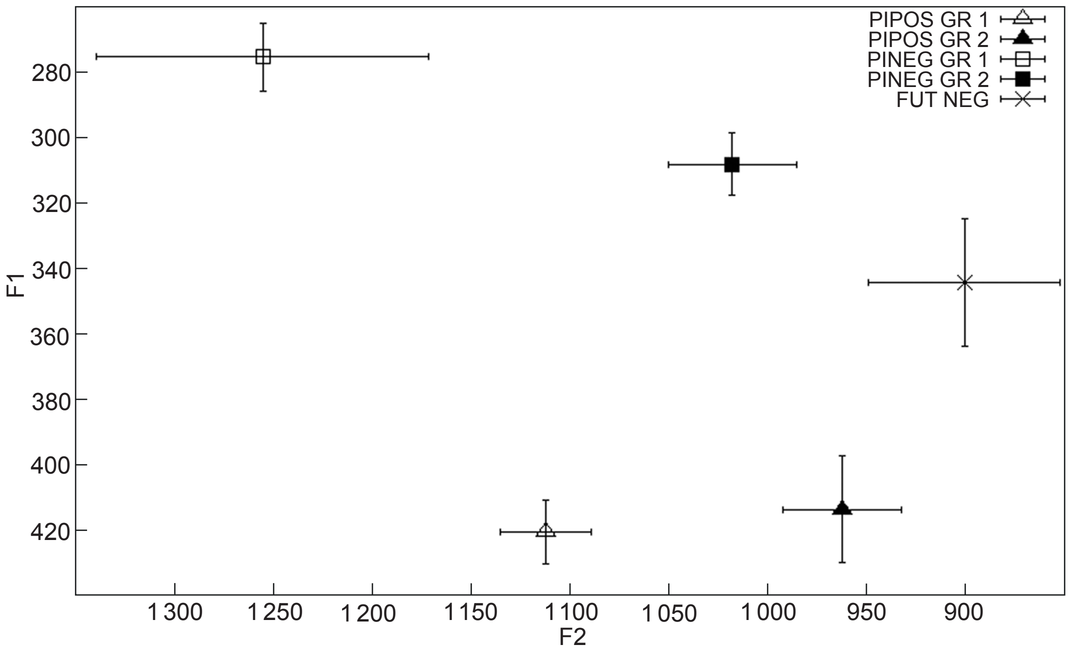

Figure 4: Detail of vowel chart showing the 5 variations of -u- in bua
(PIPOS and PINEG indicate the present indicative positive and negative, respectively, whereas FUT NEG9 indicates the future negative. GR 1 and GR 2 distinguish between the 2 different semantic cases which are traditionally described as tonal variants.)

respectively. That is, we find that the middle vowels are ordered from low to high as {/ɛ/, /e/, [ɛ̱], [e̱]} and {/ɔ/, /o/, [ɔ̱], [o̱]}, where the ordering would be {/ɛ/, [ɛ̱], /e/, [e̱]} and {/ɔ/, [ɔ̱], /o/, [o̱]} if the height difference between phonemically distinct vowels were larger than the change owing to vowel raising. Since this trend is consistent across all words that we have investigated, we believe that this is a reliable contrast.

This finding is of great importance to developers of HLT systems, who invariably need to group allophones together for reasons of data and computational efficiency. The results of this article indicate that grouping is not feasible for the allophones that arise from vowel harmony – these should be kept separate, since their combination would completely obscure the distinction between the semantically different cases. We also find that /u/ and /a/ are modified significantly by syntactic context (and /u/ by semantic contrasts as well); it may therefore be sensible to expand the set of major allophones to contain these variants as well.

A number of important issues remain unresolved. Most importantly, we have consciously avoided issues of tone, even though tone and vowel quality are known to interact with one another (Hillenbrand *et al.*, 1995). (Since we have recorded our contrasting tokens in contexts where tonal shifts are not expected to occur – except for *bua*, as discussed in the fourth section – we are at least confident that the well known interaction between tone and vowel height does not confound our main conclusions.) This matter requires careful annotation of the lexicon to distinguish between the two forms of vowel raising discussed here, on the one hand, and high tones, on the other, and is currently being investigated. We have also not investigated the possibility that there are three *phonemically* distinct vowels corresponding to each of the orthographic symbols *e* and *o*, respectively – as suggested in Khabanyane (1991). Our current measurements do not allow us to address this issue, and additional experiments should be undertaken. Finally, we have not investigated whether differences of dialect or speaker age interact with any of the findings reported here,

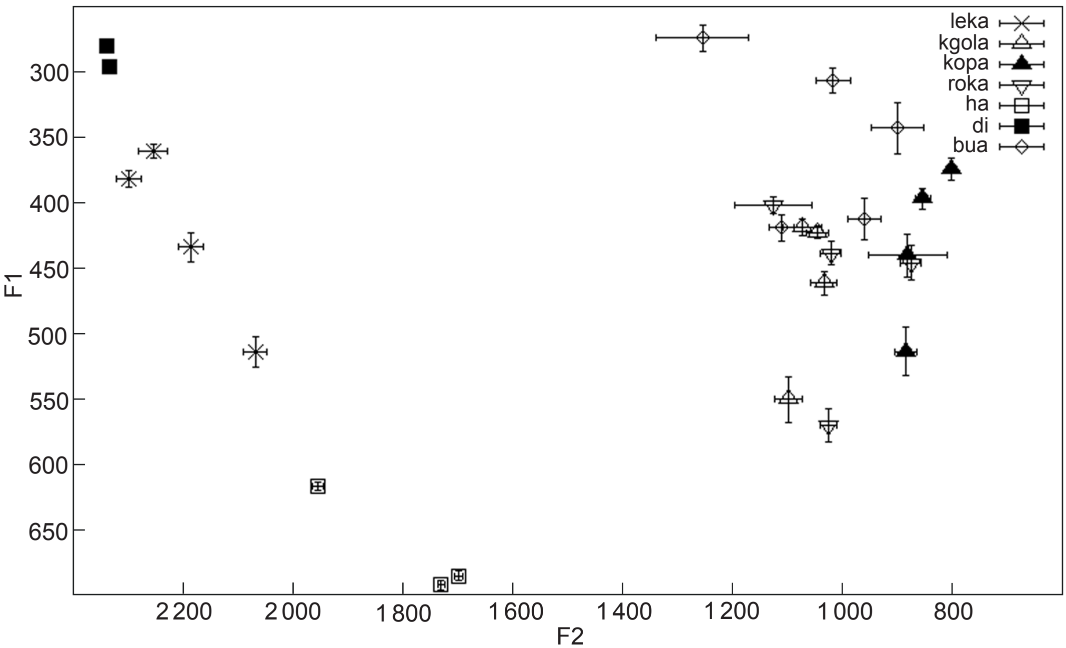

Figure 5: Combined representation of all vowels included in this study

although our preliminary analyses indicate that speaker age may indeed be important in the current context.

We believe that our research solidifies the foundations on which HLT systems, such as pronunciation dictionaries, speech recognisers and speech synthesisers can be built for Southern Sotho. Not only does this enable us to select appropriate phone inventories for the implementation of HLT systems, but it also provides concrete guidance on the contextual effects that need to be considered when developing systems for speech recognition and speech synthesis. Such systems are currently under development in our laboratories, and their success will be an important verification of the current research.

Notes

1 According to Tucker (1929: 23) there is a 'wonderful similarity of the vowel-systems of all the known members of the Suto-Chuana group of Bantu languages'.

2 'High' and 'low' are synonyms for 'close' and 'open', respectively, with the latter terms currently preferred by the IPA.

3 The conceptual status of these vowels – whether they correspond exactly to the auditory qualities defined by Jones or only approximately so – is a complex question that we do not address here.

4 A variety of symbols are used in the cited works. This is not the place to debate the issue here. We use underscores to indicate such vowels.

5 Selebeleng (1997: Chapter 2) calls this process 'vowel laxing'.

6 Note the convention of using / / as an indication of phonemes, as opposed to [ ] for (allo)phones.

7 For example, the present indicative sentence *Ke* ***beka*** *dihwapa* was presented in writing. Respondents were asked to create the negative themselves, namely, *Ga ke* ***beke*** *dihwapa*; as discussed above, this construct is known to produce vowel raising.

8 Praat is a comprehensive speech analysis programme, and Vowelyse is a specialised add-on script for the analysis of vowels.

[9] Examples of these are: *Ke* ***bua*** *ditaba* (PIPOS); *Ga ke* ***bue*** *ditaba* (PINEG); *Ha* ke *tlo* ***bua*** *ditaba* (FUT NEG).

Acknowledgement — The current paper benefited greatly from the detailed comments of two anonymous reviewers.

References

Adank P, Smits R & Van Hout R. 2004. A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America* **116**: 3099–3107.

Boersma P & Weenink D. 2007. Praat: doing phonetics by computer. www.praat.org.

Doke CM & Mofokeng SM. 1957. *Textbook of Southern Sotho*. Cape Town: Longman.

Grabe E & Low EL. 2002. Durational variability in speech and the rhythm class hypothesis. In Gussenhoven C & Warner N (eds) *Laboratory Phonology VII*. Berlin: Mouton de Gruyter, pp 515–546.

Guma SM. 1971. *An outline structure of Southern Sotho*. Pietermaritzburg: Shuter and Shooter.

Hillenbrand J, Getty LA, Clark MJ & Wheeler K. 1995. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* **97**: 3099–3111.

Khabanyane KE. 1991. The five phonemic vowel heights of Southern Sotho: an acoustic and phonological analysis. *Working Papers of the Cornell Phonetics Laboratory,* No. 5, pp 1–36.

Krüger CJH & Snyman JW. s.a. *The sound system of Setswana*. Goodwood: Via Africa.

Lobanov BM. 1971. Classification of Russian vowels spoken by different listeners. *Journal of the Acoustical Society of America* **49**: 606–608.

Nearey TM. 1989. Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America* **85**: 2088–2113.

Poulos G & Louwrens LJ. 1994. *A Linguistic Analysis of Northern Sotho*. Goodwood: Via Africa.

Roux JC. 1983. Vokaalkwaliteit, toon en konsonantale invloede in Sotho. Unpublished research report. University of Stellenbosch, Stellenbosch.

Selebeleng KZ. 'Phonetic and phonological aspects of vowel raising in SeSotho'. (Masters thesis, University of Stellenbosch, 1997).

Tucker AN. 1929. *The comparative phonetics of the Suto-Chuana group of Bantu languages*. London: Longman, Green and Co.

Van der Walt AJ & Wissing DP. 2003. *Vowelyse: tools for acoustic analysis and optimising phonetic actions*. http://www.puk.ac.za/fakulteite/lettere/HLT_Resoruces/.

Watt D & Fabricius A. 2002. Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1 ~ F2 plane. *Leeds Working Papers in Linguistics and Phonetics* **9**: 159–173.

Appendix: Complete set of sentences used in the recordings

-e- stem words
Ho ***beha*** *thebe* ('To sway a shield')
Ho ***beha*** *diphahlo* ('To carry the cloth')
Ke ***beka*** *ngwetsi* ('I receive a daughter-in-law')
Ke ***beka*** *dihwapa* ('I cut dried meat')
Di ***leka*** *batho* ('They try the people')
Di ***leka*** *letswai* ('They lick the salt')

-o- stem words
Ba ***kopa*** *dipabi* ('They ask maize')
Ba ***kopa*** *letheka* ('They cover the hips')
Ho ***hlola*** *dintwa* ('To win the battle')
Ho ***hlola*** *bobe* ('To conjure the evil')
Ho ***roka*** *mose* ('To sew a dress')
Ho ***roka*** *morena* ('To praise the king')
Ba ***kgola*** *ditolwana* ('They pluck fruit')
Ba ***kgola*** *meputso* ('They receive payment')

-u- stem words
Ke ***bua*** *ditaba* ('I read the news')
Ke ***bua*** *kgomo* ('I flay the cow')

Speakers were asked to say each of these present-indicative sentences in the positive and negative case. In addition, past and future tenses of both the positive and negative cases were recorded, but these were only analysed in a few cases (see Figure 3).