Southern African Linguistics and Applied Language Studies 2008, 26(2): 235–254
Printed in South Africa — All rights reserved

SOUTHERN AFRICAN LINGUISTICS
AND APPLIED LANGUAGE STUDIES
ISSN 1607–3614 EISSN 1727–9461
DOI: 10.2989/SALALS.2008.26.2.5.569

Phonetics of intonation in South African Bantu languages

Sabine Zerbian[1*] and Etienne Barnard[2]

[1]Department of Linguistics,University of the Witwatersrand, Private Bag 3, Wits, 2050, South Africa

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

e-mail: ebarnard@csir.co.za

** Corresponding author: e-mail: sabine.zerbian@wits.ac.za*

Abstract: Much is already known about the prosodic systems of the indigenous South African languages from descriptions and analyses in the existing literature. All of the existing work has been carried out in the field of African studies or formal linguistics. In order to be able to implement the generalisations obtained into computational models in speech processing, the existing sources and results must be made accessible to researchers in these areas. In the opposite direction, the modelling of Bantu intonation in speech processing shows the need for more quantitative data. The present article presents a review of durational and pitch-related tonal aspects relating to intonation in South African Bantu languages. Its aim is, on the one hand, to make the results of Africanist and linguistic studies accessible to research in speech processing, laying a foundation on which work in the speech-processing branch of Human Language Technologies can be based. On the other hand, by pointing out gaps in our knowledge, it wants to draw attention to the fact that more quantitative research is needed in order to advance our knowledge also in the theoretical linguistic field.

Introduction

The term *intonation* is loosely defined in Crystal (2003), who refers to 'the distinctive use of patterns of pitch', whereas Hirst and Di Cristo (1998) refer to 'the melody of speech' (thus encompassing not only pitch but also patterns of rhythm and loudness). Traditionally, the term *intonation* is used in stress languages in order to refer to meaningful pitch changes at the sentence level. However, also in tone languages, we find meaningful alternations in pitch across the sentence (for Chichewa see Downing *et al*., forthcoming; for question intonation in equatorial African languages see Rialland, forthcoming; and Xu, 1999, for focus in Chinese). Many different aspects feed into intonation. These aspects include the use of intonation and tone to express emotions, to encode information structure and to signal sentence types; the sensitivity of tone and intonation to lexical specification, word classes and syntactic structure; the phonetic aspects of timing, alignment and segmental influences.

Acoustically, intonation is the modulation of fundamental frequency, intensity and duration across an utterance. Formal linguistics and speech processing share a common interest in the investigation of intonation in South African Bantu languages. Both want to provide an account of Bantu tone systems that predicts the realisation of tones so that tonal contours and intonation over utterances can be modelled, either theoretically or practically.

The current article presents a state-of-the-art report on tonal research into South African Bantu languages in formal linguistics. South African Bantu languages are among the better-documented languages of the African continent. With respect to their intonation systems, a number of case studies exist for all nine official languages. Existing work is cited in Table 1 and serves as the basis for the exposition in the rest of the article.

The article is structured as follows: the section entitled 'Duration' presents durational aspects of intonation in southern Bantu languages. It discusses how phonology, syntax and pragmatics influence vowel length. 'Tone and intonation' reviews the tonal aspects of intonation in southern Bantu languages, discussing the tone inventory, acoustics of low and high tones, as well as tones

Table 1: Literature sources for tone systems of southern African languages

Family	**Language**	**Source**
Nguni	isiZulu	Beuchat (1966), Cassimjee and Kisseberth (2001), Clark (1988), Cope (1959; 1970), Downing (2001), Khumalo (1981; 1982; 1987), Laughren (1984), Peterson (1989), Roux (1995b), Rycroft (1963; 1980a)
	isiXhosa	Cassimjee (1998), Cassimjee and Kisseberth (1998), Claughton (1983), Goldsmith *et al.* (1989), Jokweni (1995; 1998), Louw (1968), Roux (1995a; 1995b)
	IsiNdebele	Rycroft (1980a; 1983), Sibanda (2004), Ziervogel (1959)
Sotho-Tswana	Sesotho	Clements (1988), Demuth (1990), Khoali (1991), Kunene (1972), Köhler (1956), Letele (1955)
	Setswana	Chebanne *et al.* (1997), Creissels (1998; 1999; 2000), Kisseberth and Mmusi (1990), Lets'eng (1994), Mmusi (1992), Mosaka (2000), Van der Pas *et al.* (2000)
	SeSotho sa Leboa	Lombard (1976; 1979), Monareng (1992), Trümpelmann (1942), Zerbian (2006a; 2007a)
Tsonga	Xitsonga	Beuchat (1962), Cole-Beuchat (1959), Endemann (1952), Louw (1968, 1983)
Venda	TshiVenda	Cassimjee (1992), Westphal (1962)

within different tonal and segmental contexts. 'F0 versus intensity' addresses the role of intensity in intonation. These three sections taken together thus lay out how duration, fundamental frequency and intensity interact in the intonation of southern Bantu languages. The summary and critical assessment of the available literature on this topic is meant to facilitate access to basic information for use in speech processing. Though many aspects will need to be oversimplified owing to space restrictions, the list of references is fairly extensive and should allow following up specific aspects for specific varieties or languages.

The section entitled 'Implications for speech technology' discusses the importance of the above matters for speech-processing systems. (Text processing and speech processing are the two main branches of human language technologies; the former is not directly impacted by considerations related to intonation.) The last section of the article provides a conclusion and summary.

Duration

Duration is one parameter that feeds into intonation. As a parameter of intonation, the duration of segments is modified owing to changes in phonological boundaries and/or alternations in discourse prominence. However, not all durational alternations in a language relate to intonation. The present section discusses segment duration in southern African Bantu languages. It starts out with durational alternations which are unrelated to intonation, but which are grounded in the sound inventory of the language (subsection 'Phonemic vowel length') or in phonetic universals (subsection 'Phonetic influences on vowel duration'). It then discusses the durational alternations due to syntactic (subsection 'Syntactic factors of vowel length') and pragmatic factors (subsection 'Pragmatic factors'). The last two aspects fall under intonation.

Phonemic vowel length

Depending on the sound inventory of a specific language, length in vowels can be phonemic or not. If length is phonemic, the language has both long and short vowels and the durational difference of the vowel leads to a difference in meaning. In Kinyarwanda, a Bantu language spoken in Rwanda, vowel length is contrastive. An example of a minimal pair differing only in vowel length is given in example 1.

Kinyarwanda: contrastive vowel length (from Myers, 2003)

1(a) [*gutaka*] 'to scream'

1(b) [*guta:ka*] 'to decorate'

In the Bantu languages of southern Africa, however, vowel length is non-contrastive. Only short vowels occur in the phoneme inventories of these languages (Ziervogel *et al*., 1967). Only in very few idiosyncratic exceptions long vowels do occur, for example, Sesotho *maabane* 'yesterday'. Thus all syllables in southern African Bantu languages contain vowels of phonemically equal length. Differences in the surface length of vowels are thus influenced by the factors outlined in the following subsections.

Phonetic influences on vowel duration

It is known that phonetic factors influence vowel length. We will briefly discuss the contribution of inherent segmental properties, vowel coalescence and global durational structure to the duration of vowels in the Bantu languages of southern Africa.

Among the findings relevant for southern Bantu languages is the observation that low vowels tend to be longer than high vowels (Lehiste, 1970), vowels are longer before a nasal-obstruent sequence (for example, Maddieson, 1985), vowels in longer phrases are shorter than those in shorter phrases (see Lindblom *et al*., 1981). Also, vowels in a phrase-final position tend to be lengthened cross-linguistically (Klatt, 1976).[1]

In the Nguni languages, juxtaposed vowel sequences occasionally lead to vowel coalescence which potentially has an impact on vowel length. Generally, elision of vowels takes place when two words or word groups merge to form a new word group. Thereby, either the initial vowel of the second word – see 2(a) – or the final vowel of the first word – see 2(b) – might be deleted. However, vowels can also merge into one vowel, which is of a different vowel quality than the consisting parts (coalescence), as in 2(c). It remains to be investigated whether merged vowels are longer in duration.

Vowel elision of the initial vowel in isiZulu (Doke, 1927: 23)

2(a) *thina abantu* > *thina 'bantu* 'we people'
2(b) *inkosi enkulu* > *inkos'enkulu* 'a big chief'
2(c) *wa-+umuntu* > *womuntu* 'of the person'

As discussed in the previous subsection, the southern African Bantu languages do not differentiate vowel length in their sound inventory. The consistency in the duration of vowels is further enforced by the global durational structure of these languages. Pike (1945) classified languages into one of two rhythmic classes based on the prosodic units that show the same duration. Some languages time the syllable at isochronous intervals (syllable-timed languages), others organise stressed syllables isochronously (stress-timed languages). The classification as one of these two rhythmic types coincides with the phonological properties given in 3 (a) to (c).

3(a) Stress-timed languages reduce unstressed vowels in quality and quantity.
3(b) Stress-timed languages distinguish between long and short vowels phonemically.
3(c) Syllable-timed languages usually have simple syllable structures, while stress-timed languages allow complex syllable margins and nuclei.

The southern Bantu languages meet all the characteristics of syllable-timed languages. They do not distinguish between long and short vowels – see 3(b), and they do not have complex syllable margins and nuclei – see 3(c). Characteristic 3(a) is difficult to evaluate as the concept of stress is controversial in Bantu tone languages.

In syllable-timed languages, syllables are at roughly equal intervals. With syllables spaced at roughly equal intervals, no reduction and shortening processes are expected as they occur in stress-timed languages such as English and German. Experimental work by Coetzee and Wissing (2007) has shown that this difference in the rhythmic organisation of Setswana is carried over even into second language English.

Syntactic factors

The preceding two subsections have given phonological and phonetic factors in vowel duration. Durational alternations stemming from these two factors are not considered to contribute towards intonation. By contrast, the influence of syntax and pragmatics crucially determines the intonation of a given utterance. These two aspects will be discussed in the following two subsections.

Independently of the phonemic status of length, Bantu languages show interactions in vowel duration and position within a word or phrase. In languages with contrastive vowel length, such as Chimwiini and Kinyarwanda, these interactions are bidirectional. Long vowels might get shortened if they occur in a certain position, or short vowels might be lengthened (Kisseberth, 2000; Kisseberth & Abasheikh, 1974; and Myers, 2005, 2003).

In languages without contrastive vowel length, such as the southern African Bantu languages, the interaction is always unidirectional in that short vowels get lengthened. Penultimate vowels of words occurring at the right edge of a clause boundary are predictably lengthened. This is shown in example 4 for isiXhosa. The verb shows a short vowel when it is followed by an object within the same phonological phrase (marked by parentheses). It has a lengthened penultimate vowel (with the final vowel being deleted afterwards, cf. 2) when it occurs finally in a phonological phrase, being followed by a right-dislocated object (see Zerbian, 2007a, for similar data in seSotho sa Leboa; see Jokweni, 1995, and Zerbian, 2007a, for a formal account of phonological phrase boundaries in these two languages, respectively). Surface high tones are marked by an acute accent.

IsiXhosa (Jokweni, 1995: 51)

4(a) (*bá-zaku-vul)* (*íncwaadí*)
SC-FUT-open book
'They are going to open the book.'

4(b) (*bá-zaku-yi-vúúl'*) (*íncwaadí*)
SC-FUT-OC-open book
'They are going to open it, the book.'

Although penultimate lengthening is very prominent perceptually, few controlled quantitative studies exist. A first exploration of penultimate lengthening in seSotho sa Leboa suggests that penultimate syllables of words at clause boundaries are nearly twice as long as other syllables (see Zerbian, 2007a, for first measurements; see also Myers, 1999: 221, for a similar result in Chichewa).

Zulu has been claimed to have penultimate lengthening also at the word level, that is, the penultimate vowel of every word is lengthened (Doke, 1927). This has also been found in Kinyarwanda (Myers, 2005). For seSotho sa Leboa, however, this claim could not be confirmed (Lombard, 1979; Zerbian, 2007a). Only penultimate vowels in clause-final syllables are significantly lengthened. However, a carefully controlled study on vowel length depending on phrase position (following the example in Myers, 2005) is still missing for southern Bantu languages.

Pragmatic factors

There are at least two pragmatic factors that are commonly reflected in intonation, namely sentence type and information structure. The influence of these aspects on intonation in southern Bantu languages will be reviewed in this subsection.

In the preceding subsection it has been said that in the Bantu languages of southern Africa lengthened vowels occur at the right edge of clause boundaries. Long penultimate vowels, however, do not occur at the right edge of *every* clause. They only occur at the right edge of declarative clause boundaries. Penultimate lengthening is absent from yes/no-questions in isiXhosa (Jones *et al*., 2001b) and seSotho sa Leboa (Zerbian, 2006b). As can be seen from the examples in 5, there are no morphological or syntactic differences in isiXhosa that distinguish between statements 5(a) and questions 5(b). The load for distinction lies solely on the phonetic implementation.

IsiXhosa

5(a) *Ng-um-fana*.
COP-CL1-boy
'It is a boy.'

5(b) *Ng-um-fana*?
COP-CL1-boy
'Is it a boy?'

Jones *et al*. (2001b) report on a controlled acoustic study into the properties of yes/no-questions in isiXhosa. They recorded yes/no-questions involving a copula construction (as in 5) from 11 isiXhosa mother tongue speakers. On the basis of the recorded tokens they measured duration, pitch and intensity (loudness) for statements and corresponding questions. They found that although every syllable in a question is shorter than its corresponding syllable in a statement, the length

of the penultimate syllable was the most significant feature in distinguishing between questions and statements. A follow-up experiment on the perception of questions in isiXhosa (Jones *et al*., 2001a) manipulated the length of the penultimate syllable by systematically changing its length. This was done in order to determine the crucial duration at which listeners perceive question stimuli as statements. Their results show a significant change in perception towards statements in most of the stimuli when the duration of the penultimate syllable of a question was increased. Interestingly, the opposite could not be observed: a decrease of length in the penultimate syllable does not lead to the perception of a question (see in this article: 'Pragmatics: focus and questions' on overall pitch which is an additional acoustic cue in questions in southern Bantu languages).

The second pragmatic aspect that influences intonation is information structure. The term *information structure* refers to notions such as *focus*, *topic* and *givenness*, and can be seen as a phenomenon of information packaging that responds to the communicative needs of interlocutors (cf. Chafe, 1976; see also Krifka, 2006, for a concise overview). In languages such as English and German, constituents which are new in discourse (=focused) are made prosodically prominent. Acoustic correlates of prominence, both at word and sentence levels, are duration, pitch and intensity. Prominent syllables have a longer duration and carry a higher pitch and more intensity.

Recent research into the expression of information structure in seSotho sa Leboa has shown that this language does not use prosodic means to indicate discourse-new constituents in exchanges as the one in example 6 (Zerbian, 2007c) ([ ]$_F$ indicates the discourse-new constituent as elicited by a preceding question, small caps in the translation indicate pitch accent placement in English).

6(a) *Ke tla be ke šoma* [polase-ng]$_F$.
1ST FUT PROGR 1ST work farm-LOC
'I will be working ON THE FARM.'

6(b) *Ke tla be ke* [*šoma*]$_F$ polase-ng.
1ST FUT PROGR 1ST work farm-LOC
'I will be WORKING on the farm.'

In contrast to English, focus considerations do not contribute towards the intonation contour of an utterance, at least not in seSotho sa Leboa. However, it needs to be kept in mind that focus will be expressed morphosyntactically in this language and that syntax will have an effect on intonation (see, for example, right-dislocation of discourse-old constituents and resulting phrase boundary in example 4).

Interim summary

The present section has discussed how the surface duration of vowels is determined by phonological, phonetic, syntactic and pragmatic aspects. However, only syntactic and pragmatic considerations feed into what is referred to as intonation. Although the southern African Bantu languages do not distinguish long and short vowels in the sound inventory, long vowels are predicted to occur at prosodic boundaries which are determined by syntax as well as by sentence type.

Tone and intonation

According to Crystal (2003), the term *tone* refers to distinctive pitch levels of a syllable. Traditionally, *tone* is reserved for the description of languages that specify pitch for every syllable of a word in their lexicon (=tone languages). Tone is a relative notion. A high tone does thus not refer to an absolute pitch target but to high pitch in relation to surrounding pitch. As pitch changes can apply at the word and sentence levels, we can thus expect to find an interaction of changes at the word level (tone) and changes of pitch at the sentence level (intonation).

The current comprehensive section introduces the reader to the basic issues in tone and intonation in southern Bantu languages. Keeping the application of this knowledge in speech processing in mind, the section will point out open questions as they come up. The section is structured as follows: the subsection 'Tonal inventory' introduces the tonal inventory postulated for southern Bantu languages as based on phonological evidence, and provides a review of the acoustic studies that have been carried out on low and high tones in Bantu languages, respectively. The subsection 'Behaviour of active high tones' presents the processes that are associated with the tonally active

high tones in these languages; namely, high tone spread/shift, adjustment processes regarding adjacent high tones and positional restrictions that high tones underlie. The subsection 'Segmental context' reviews the segmental influences that both low and high tones are subjected to. Finally, the subsection 'Tones at the sentence level' presents intonational aspects in southern Bantu tone in looking at declination and the role of focus. The section ends with an interim conclusion.

Tonal inventory

Phonological inventory

Tone languages use changes in fundamental frequency to reveal lexical or grammatical differences. Bantu languages show a two-tone system with a level high (H) and a low (L) tone. Occasionally, a falling tone is reported which is restricted to long syllables (which are the penultimate syllables of clauses, see in this article 'Phonetic influences on vowel duration'). The examples in 7 show minimal pairs which differ in tone only.

SeSotho sa Leboa

7(a) Lexical minimal pairs (Ziervogel & Mokgokong, 1979)

lapá – 'court-yard'	*lapa* – 'become tired'
bóna – '(to) see'	*boná* – 'they'

7(b) Grammatical minimal pairs (Ziervogel *et al*., 1969)

re rútá – 'we teach'	*ré rúta* – 'while we teach'
á rúte – 'he should praise'	*a rúte* – 'he praises habitually'

Southern African Bantu languages, like Bantu languages more generally, show an asymmetry in their tone system. Verb roots can be either high-toned or low-toned, independent of the number of syllables, 8(a). Verbs do not contrast tone on every syllable. Nouns, however, can contrast tone on every syllable, 8(b). Thus nouns exhibit properties of a pure tone system in that pitch can be contrastive on every syllable.

Verb classes in seSotho sa Leboa (Ziervogel & Mokgokong, 1979)

8(a)

go hlaba	'to stab'
go tlogela	'to leave'
go ráta	'to love'
go bóláya	'to kill'

Nouns in seSotho sa Leboa (Ziervogel & Mokgokong, 1979)

8(b)

HL	*púdi*	'goat'
LH	*kgomó*	'cow'
LLH	*morathó*	'younger sibling'
LHL	*mosádi*	'woman'

The verb and noun system is thus asymmetrical with respect to tonal specification. Another asymmetry can be found in the properties of the tones themselves. High tones are phonologically active, that is, they participate in phonological processes such as spread/shift, retraction and deletion. Also, two adjacent high tones are often disfavoured and will lead to repair strategies being implemented in order to resolve these sequences (see in this article: 'Adjacent high tones'). Low tones, on the other hand, are not reported to spread/shift or to be deleted. Neither is adjacency of low tones banned from the tonal grammar.

Whereas high tones are considered to be specified in the underlying form of a word, that is, in the lexicon, low tones are considered inert. They are inserted late in the derivation of the actual tonal surface structure, after specified high tones have triggered tonal rules to apply. Low tones are so-called default tones. Thus, the claim that only high tones are specified underlyingly is mainly based on phonological evidence.

Although it is commonly agreed that Bantu languages have two-tone systems with high tones specified underlyingly, a look at the early literature on these languages reveals that the tone systems are complicated to grasp. Doke (1927) in his study on Zulu posited nine tones for this language. Endemann (1911) found at least three tones to be necessary in order to adequately describe the tonology of seSotho sa Leboa. Thus, the transcription of the perceived tone into H and Ls is not self-evident.

Acoustics of low tones

Myers (1998) investigates the phonetic implementation of tone in the Bantu language Chichewa, spoken in Malawi. He interprets the phonological concept of inertness of low tones (see in this article: 'Phonological inventory') as the absence of an acoustic pitch target. Thus, high tones have an acoustically specified pitch target to be reached during articulation, whereas low-toned syllables lack such a pitch target. The actual pitch value for a low-toned syllable thus depends on the presence and location of preceding high tones. In his experimental study, he shows that this is indeed the case in Chichewa.

In a production experiment that follows methodological standards in the phonetics literature, he investigates the question of tonal inertness in the speech of three Chichewa speakers. He starts out on the premise that high tones present pitch targets, whereas low tones do not have tonal targets. As a result, the pitch value of a low tone should be predictable as a function of the H targets and the interpeak interval. Meyers (1998) analyses data of the following kind:

9(a) *Mlónda* ***á****may****á****mba kunyénya*.
9(b) *Mlónda* ***á****mayen****é****ra kunyénya*.
9(c) *Mlónda* ***á****malepher****é****tsa kunyénya*.
'The watchman begins/must/prevents to goof off.'

The data in 9 show adjacent high tones (given in bold) which are separated by a differing number of low-toned syllables, 1 in 9(a), 2 in 9(b), 3 in 9(c). In his study, Myers measures the F0 minimum of the intervening low tones. The observation is that the F0 dip of the intervening low tone increases when the interpeak interval is longer. Thus, the lowest low tone in 9(c) shows a lower F0 value than the low tone in 9(a). Based on his data, he is able to predict the F0 value for the lowest low tone by means of multiple linear regression analysis. The formula for the frequency of the low tone in Hz is given in 10:

10 Multiple linear regression analysis (Myers, 1998)
$L = 0.89 \times H2 - 102.7 \times T + 21.6$

The value of the low tone L is 89% of the second high tone H2 (in Hz) minus 102.7 Hz per second of the interpeak interval T (in seconds) plus a constant baseline value of 21.6 Hz.

According to Myers, low tones thus lack a clearly defined acoustic pitch target in Chichewa. Rather, their value is defined by the surrounding high tones as well as the number of intervening syllables. This result from Chichewa lends phonetic support to the phonological concept of inertness according to which these syllables do not have a specified tone underlyingly.

Acoustics of high tones

In contrast to low tones, high tones have a phonological, underlying tone specification. Phonological evidence for this claim comes from the 'active' behaviour of high tones (see in this article: 'Behaviour of active high tones'). They participate in spread or shift and they underlie restrictions with respect to position within a word or phrase and with respect to other high tones.

Acoustically, tonal specification translates into an acoustic pitch target. The existence of an acoustic pitch target for high tones is mirrored in Myers's study (1998) by a pitch peak for an underlying high tone. The pitch starts rising on the underlyingly high-toned syllable until it reaches its peak early in the subsequent syllable. Thereafter, the pitch declines again.

Whereas the pitch value for a low tone can be derived from the surrounding high tones (see 10), the specific pitch target of a high tone is much more difficult to predict. Existing literature suggests that a number of factors are influential; most of all, the individual speaker's pitch range, but also the overall length of the utterance, the position of a tone within the utterance and the total number of high tones in the utterance. The absolute pitch of tones is subject to declination (see 'Declination' in this article), which refers to a decrease of pitch over the course of an utterance. Thus, high tones that occur early in the utterance have absolutely higher pitch values than high tones that occur late in the utterance. The number of underlying high tones within an utterance also influences the absolute pitch value owing to high tones being realised at distinct pitch levels from one another. Thus, if there are many high tones in an utterance, the occurring downstep will force the pitch down on each subsequent high tone (see in this article: 'High tone spread/shift').

Perpendicular to position within utterance and total number of tones within an utterance, speakers seem to vary their pitch range according to the overall length of the utterance. In Myers's (1998) study this is reflected in the observation that overall longer utterances are started at a higher pitch than shorter utterances. There is a natural explanation for this observation, as in longer utterances the speaker has to allow for a longer pitch decline. As a result, he/she has to start at a higher pitch because the bottom of his/her pitch range is predetermined.

Whereas studies concerning the specifics of vertical alignment of pitch targets are absent from the literature, the horizontal alignment of pitch targets, namely with respect to the tone-bearing unit, has received some attention in the literature on Bantu tone. In investigating high tone alignment in Chichewa, Myers (1998) finds that the pitch peak is actually not reached in the syllable with which the high tone is associated underlyingly, but only in the following syllable. In a controlled production experiment he investigates the alignment of high tones in the speech of three speakers. He defines a high tone phonetically as a local peak in F0. He finds that F0 starts to rise near the beginning of the syllable from which the high tone originates. However, the pitch peak is only reached at the end of that syllable or at the beginning of the next. This delay in alignment to the beginning of the subsequent syllable is known as *peak delay* and is quite common cross-linguistically.

The fact that the F0 peak occurs at a relatively constant proportion of the duration of a high-toned syllable allows Myers (1998) to model the peak delay by means of multiple linear regression. The formula is given in 11. It should be read as that the peak delay of a high tone can be roughly calculated as the product of the position of the syllable within the sentence (P = 0 for medial position, P = 1 for lengthened penultimate syllable) and the actual syllable duration (S in seconds). As can be seen from the constants in the equation, speaker-dependent variation is taken into consideration.

11 Peak delay in Chichewa for one speaker (regression equation; Myers, 1999: 222)
Peak delay = (–.88 × P + 1.43) × S – 3.89

In medial syllables (P = 0) that are short, the peak occurs as mentioned above, namely in the subsequent syllable. In penultimate syllables which are long in Chichewa just as in southern Bantu languages (see in this article: 'Phonetic influences on vowel duration'); the pitch peak is actually reached within the tone-bearing syllable. Thus, the model in 11 indicates that the F0 peak in medial syllables occurs later, relative to the tone-bearing syllable, than in penultimate position. Thus, syllable position and syllable duration are the two factors that account for peak delay in Chichewa (for other factors in syllable alignment in seSotho sa Leboa, see Zerbian & Barnard, 2008).

Myers interprets the data in that the peak delay in medial/short syllables is due to time constraints of the articulators. Under this view, high tone spread in Chichewa is considered a phonetic carry-over effect owing to the brief duration of the syllables.

Behaviour of active high tones

In Bantu languages high tones are the active tones phonologically. They may spread or shift, they may be deleted and there might be positional restrictions on their occurrence. This subsection presents high tone spread, adjacent high tones and positional restrictions of high tones.

High tone spread/shift

For a high tone that is surrounded by low tones one can observe that its high pitch does not (only) surface on the syllable it is associated with underlyingly but (also) on syllables to the right (for example, Myers, 1998; 1999, for an acoustic study on Chichewa as well as numerous descriptions of other Bantu tone systems). This so-called mobility of high tones is a characteristic feature of Bantu tone systems (Kisseberth & Odden, 2003).

One speaks of high tone spread if the high tones form a plateau from the syllable from which they originate up to the target syllable where the spread ends, as shown in 12(a) for one isiXhosa dialect. In high tone shift, on the other hand, the high tones delink from the sponsor syllable and all intervening syllables and only surface on the so-called target syllable, which is the antepenultimate syllable in 12(b) for another isiXhosa dialect. Underlyingly high-toned syllables are marked by underlining.

Spreading versus shifting dialects of isiXhosa

12(a) *úkúvíngcela*

12(b) *ukuvíngcela*

High tone shift is widespread. However, no acoustic study is available for high tone shift so that one has to rely on transcribed data. It is reported to be characteristic for many Nguni languages (with varying characteristics and restrictions; see Cassimjee & Kisseberth, 2001, for an overview).

High tone spread is just as widespread as high tone shift. The Sotho languages, Xitsonga as well as some Nguni varieties – see 12(a) – are reported to show high tone spread. Two acoustic studies are available that address the phonetic implementation of high tone spread in Chichewa (Myers, 1998; 1999). Myers's (1999) investigation of high tone alignment in Chichewa (see 'Acoustics of high tones' above) is directly relevant for the process of high tone spread. He shows that in short syllables the pitch peak of the high-toned syllable is only reached in the subsequent syllable. We do not find a high tone plateau, that is, a pitch peak on the tone-bearing syllable which is held on to the subsequent syllable. Given this absence of a plateau, Myers concludes that the existence of a phonological process of high tone spread cannot be upheld in Chichewa. Instead, the peak delay to the beginning of the syllable that follows the underlyingly high-toned syllable suggests a phonetic carry-over effect rather than a phonological spreading rule. He concludes that the auditory impression has been misinterpreted by previous researchers. This conclusion is in line with the question raised by Xu (2006) whether the 'spread' of a high tone (that is, the extension of the high pitch phase) on to the immediately adjacent syllable is indeed a phonological process, or whether it is merely a phonetic carry-over effect owing to physiological restrictions.

It is worth noting, however, that the issue deserves further investigation. One relatively uncontroversial diagnostic to distinguish between a phonological and a phonetic process is whether the effect in question is dependent on speech rate and variable in extent (Yip, 2002: 10). If it does depend on speech rate and is variable in extent, it is classified as a phonetic process rather than a phonological one. Myers's study is controlled for speech rate (fast versus slow) as well as for loudness and sentence type. There is no significant difference reported between tokens from different speech rates. Thus, the alleged spread process is independent of speech rate, which would be indicative of a phonological process given the diagnostics by Yip (2002: 10), rather than a phonetic one.

Next to the general parameter of shift versus spread, other language-specific parameter settings determine the surface location of a high tone pitch peak. For example, language varieties differ in the number of intervening syllables in high tone spread and shift. In local (or bound) high tone spread, the high tone spreads only to the immediately adjacent syllable. Unbound high tone spread is normally defined with respect to a prosodic domain boundary up until which a high tone spreads. Similarly, in local high tone shift the high tone is shifted on to the adjacent syllable. This has been reported for Kikuyu (Clements & Ford, 1979), spoken in Kenya, but not for southern African Bantu languages. In unbound shift, the high tone shifts on to a specified syllable at a prosodic domain edge. Normally, this syllable is either the penultimate or the antepenultimate syllable. Final syllables are often excluded as targets for tonal processes (see in this article: 'High tones within the phrase').

Additionally, the target of spread/shift might interact with the number of syllables in the domain in question (minimality considerations; see, for example, Downing, 2006), with the segmental make-up of the target syllables (depressor consonants; see 'Depressor consonants' in this article), or with morphological constituency (Hyman & Mathangwane, 1998, on Ikalanga Shona; Myers, 1987, on Zezuru Shona). Also, our own recent research on seSotho sa Leboa (Zerbian & Barnard, 2008) gives evidence of the importance of morphology in tone alignment. Much more research is needed in order to determine the parameters of variation in high tone spread across Bantu languages. More specifically, no phonetic details are yet available for non-local high tone spread reported in some isiXhosa dialects, nor for high tone shift as reported for other Nguni languages.

Adjacent high tones

A second indicator of the active behaviour of high tones in southern Bantu languages is the so-called *tone sandhi* that can be observed if a high tone occurs together with other high tones. One

has to differentiate between immediately adjacent high tones and co-occurring high tones which are separated by intervening low tones. Both environments will be discussed in turn.

Immediately adjacent high tones

As for immediately adjacent high tones, it is a general observation in phonology that identical adjacent phonetic features are disfavoured. This is called the Obligatory Contour Principle (OCP) (Leben, 1973) and also applies to tone. Adjacent high tones (high tones being the active tones in these languages) are disfavoured. Bantu languages display diverse strategies to resolve the conflict of adjacent high tones (Odden, 1994; Myers, 1997, for a general discussion). Six 'repair strategies' have been reported in the literature, all of which will differ acoustically. They are listed in 13 (a) to (f), following Myers (1997) and Odden (1994).

… H1 H2 …

13(a) H2 is deleted (as in Shona)

13(b) H1 is deleted (as in Rimi)

Strategies 13(a) and 13(b) delete a high tone and thus also delete any high pitch target that is associated with the tone. Deletion of H2 in a sequence of two adjacent high tones is known as Meeussen's rule (Goldsmith, 1984), named after Meeussen who described this rule for the Bantu language Tonga. Acoustically, the output will resemble contexts with only one high tone.

13(c) H1 is retracted away from H2 (Shona)

13(d) H2 is retracted away from H1 (Chichewa)

Strategies 13(c) and 13(d) preserve both high tones and thus both high pitch targets. Retraction of a high tone implies intervening low tones. Languages will differ as to the exact phonological and phonetic implementation of the retraction.

13(e) H1 and H2 are representationally fused into one H (Shona, Kishambaa)

13(f) H2 is downstepped

Strictly speaking, strategies 13(e) and 13(f) do not actually resolve the OCP violation. Strategy 13(e) resolves the unwarranted constellation at the phonological level and represents the two high tones as originating from one underlying high tone. Strategy 13(f) resolves the OCP violation at the phonetic level. Instead of realising two adjacent underlying high tones by a high pitch plateau which stretches over two syllables, the second high tone will be downstepped, that is, it will be realised at a lower pitch level that is significantly lower than the level induced by declination alone.

The resolution of OCP violations in Bantu languages has not yet been studied acoustically. It promises to be a rewarding field of study given the many reported repair strategies (cf. 13, which are themselves subject to further parameters such as morphological constituency (for example, Myers, 1987, on Shona; Zerbian & Barnard, 2008, on seSotho sa Leboa) and phrasal constituency.

High tones within a phonological domain

One of the prominent repair strategies of OCP violations is the downstep, 13(f) (for downstep see also 'Declination' below). Like declination, *downstep* refers to a lowering of subsequent high tones. However, downstep is phonologically determined, most often by an intervening low tone. There is no acoustic study available on downstep in Bantu languages, although several studies describe and analyse this phenomenon and are therefore good sources to establish hypotheses which can subsequently be tested (Clements & Ford, 1981; Kunene, 1972). Next to the empirical task to differentiate downstep from declination or final lowering, one of the central issues is to establish the phonological domain in which downstep occurs in the language under investigation. Some controlled phonetic studies have been carried out on downstep in west African tone languages (Laniran & Clements, 2003; Rialland, 2001; Rialland & Some, 2000). They suggest that there is variation across languages in how far the number of overall high tones in an utterance influences the actual pitch targets.

High tones within the phrase

The last characteristics of active high tone are the positional restrictions that govern their occurrence. The realisation of high tones will also depend on which syllable in a phrase or clause

they originate from. In phonological terms, these positional restrictions on tone are known as *well-formedness conditions*.

The finality restriction is one such a well-formedness condition, which states that the final syllable in a certain prosodic domain is extratonal. Phonologically, this means that the domain-final syllable is excluded from certain aspects of the language's tonal system. These representational adjustments are common in Bantu tonology (Kisseberth & Odden, 2003: 64) and also cross-linguistically (Nespor & Vogel, 1986). Extratonality of domain-final syllables can have at least two effects: either domain-final syllables can generally not be associated with (high) tones or domain-final syllables cannot act as a target for tonal rules that spread or shift high tones. As a result, in both cases we will not find high pitch realised on an extratonal syllable acoustically. The Bantu language Chichewa constitutes an example of the first instance, namely the general prohibition of high tones on phrase-final syllables. In Chichewa, underlying high tones retract to the preceding syllable if the syllable with which they are associated occurs in a phrase-final position (Kanerva, 1990).

High tone retraction from final syllable (Kanerva, 1990: 58)

14(a) *mle*^{n}*dó uuyu*	'this visitor'	versus	*mleé*^{n}*do*	'visitor'
14(b) *pezá nyaama*	'find the meat!'	versus	*peéza*	'meat'

SeSotho sa Leboa does not generally prohibit high tones on phrase-final syllables. Underlying high tones are allowed to surface in a phrase-final position (Lombard, 1976; Zerbian, 2007a), as in 15(a). However, seSotho sa Leboa constitutes an example of the second effect of the finality restriction according to which domain-final syllables cannot act as a target for high tone spread. Thus, in 15(b) the verb stem initial high tone spreads on to the following syllable if followed by an object. If the verb appears phrase-finally, the verb stem initial high tone does not spread.

15(a) *kgomó*		'cow'	
15(b) *Ke thúšá mosá:di*.	versus	*Ke a thú:ša*.	
'I help the woman.'		'I am helping.'	

Segmental context

Not only the tonal context is crucial for the determination of acoustic pitch targets in the realisation of tone, the segmental context also exerts an influence. Three aspects are of importance. First, vowels differ in their inherent fundamental frequency; second, voicing in consonants affects the initial fundamental frequency of a following vowel; third, there is a specific group of consonants in the Nguni languages, the so-called depressor consonants, which have a dramatic lowering effect on high tones. All aspects will be discussed in turn in this subsection.

Segment-induced pitch changes

Just as we find interactions between segments and duration ('Phonetic influences on vowel duration'), we also find interaction between segments and pitch. As for vowels, it is long known that the closed vowels /i, u/ are produced with higher intrinsic pitch than the open vowel /a/ (Meyer, 1896/97). There are competing theories as to the reasons for these differences as either physiological, thus unintentional (Whalen & Levitt, 1995), or intentional in order to improve the linguistic contrast (Kingston, 1992), or a combination of both (Honda & Fujimura, 1991).

As for consonants, voiced obstruents are known to lower the pitch of the following vowel, whereas voiceless obstruents have the reverse effect – they raise the pitch of the following vowel. Hombert (1978) studied the pitch on vowels following aspirated English plosives produced by 5 speakers and found that the pitch still significantly differs as a result of the voicing of the preceding stop up to 100 ms into the vowel. The difference was on average around 18 Hz at vowel onset and 4 Hz after 100 ms. A follow-up study of the same effect in the three-tone language Yoruba showed that the voicing of plosives affects the tones in this language (having the greatest effect on high, then low, then mid tone). Furthermore, the effect of voicing on pitch is considerably shorter in Yoruba than it is in English (around 50 to 70 ms into the vowel in Yoruba). Traill *et al*. (1987) found evidence of this same effect in isiZulu. On average, the voiced plosives /b g/ lead to a lowered pitch of 10–15 Hz with greater influence on high-toned vowels. (The acoustic behaviour of /b/ with respect to segment-induced pitch lowering actually led Traill *et al*. (1987) to refuse their phonetic classification

as implosives.) Again, there are competing explanations in the literature as to the intentionality of the pitch lowering effect of voicing.

Depressor consonants

Beyond the probably universal segmental influence of voiced constituents, a more dramatic pitch lowering has been reported in studies on Nguni languages. A phonetically heterogeneous class of obstruents lowers the pitch of the following vowel remarkably (for a comprehensive overview see Traill *et al*., 1987; Rycroft, 1980b; and Downing & Gick, 2005). There are also lexically and grammatically conditioned cases of depression in SiSwati, that is, depression without depressor consonants (Rycroft, 1980b). The so-called depressor consonants have been associated with low tones. There is disagreement in the literature as to which consonants fall into the class of depressor consonants (which is also language-specific) as well as the phonetic properties and phonological classification of this process. With respect to the acoustics of depression, a very conclusive study is that of Traill *et al*. (1987) (see Downing & Gick, 2005, for an acoustic study in a non-Nguni language). Their results as they pertain to pitch will be presented here.

Traill *et al*. (1987) recorded different consonant segments in conjunction with both following high and low tone in order to test, firstly, which consonants have a depressor effect on the following tone, secondly, if both high and low tones are affected and, thirdly, if so, in which way. By recording around 20 to 40 tokens for each condition of 3 speakers of isiZulu they found that depressor consonants have an effect on both high and low tones, though with a regularly greater effect on a following high tone. The relative differences in pitch between a 'normal' low tone and a low tone following a depressor consonant are as high as the differences between a 'normal' high and a 'normal' low tone. However, the difference between a 'normal' high tone and a high tone following a depressor consonant are on average 2 times greater. Thus, depressor consonants have a significant effect on pitch, both on low and high tones but much more prominent on high tones.

Depressor consonants expand the pitch register of a speaker in that both depressed low and high tones start on a pitch that is far lower than low. Traill *et al*. (1987) speculate that the onset pitch following a depressor is close to the lowest limit of a speaker's voice. Furthermore, pitch tracks of depression with voiced fricatives show that the 'maximal depression gesture' takes place even before the tone-bearing vowel begins.

Traill *et al*. (1987) show that depressor consonants constitute a heterogeneous group phonetically. They include voiceless unaspirated stops (<bh, dh, gh> in isiZulu orthography), voiced stops in nasal compounds (<mb, nd, ng>), voiceless fricatives (<f, s, hl>), voiced fricatives (<v, z, dl>) and depressor versions of <w, l, h>. Ejected stops, aspirated stops, voiced plosives and the non-depressor variants of <w, l, h> do not depress the following tone. The depressing effect of voiceless consonants is especially interesting, as it contradicts an approach that considers depression just an exaggerated version of universal segment-induced pitch lowering. Universal segment-induced pitch lowering only occurs with voiced consonants (see in this article: 'Segment-induced pitch changes').

Tones at the sentence level

At the sentence level, the realisation of tones can be influenced by a variety of factors as well, including syntactic and pragmatic factors. It has already been alluded to the fact that the active behaviour of high tones (see subsection 'Behaviour of active high tones') is also affected by the prosodic structure above the word level. This prosodic structure in turn is determined by syntactic factors, a relation that lies beyond the scope of this paper (but see, for example, Nespor & Vogel, 1986). The current subsection will present declination and focus as two factors that influence tone at a sentence level.

Declination

Declination has been mentioned repeatedly in the preceding subsection as one factor that determines the pitch target of high tones. However, declination potentially affects the actual pitch values of all phonetic tones in an utterance. The term *declination* refers to the continuous lowering of fundamental

frequency over the course of an utterance. Its main explication is physiological in that it is attributed to decreasing subglottal pressure which is said to result in decreasing overall fundamental frequency. However, the problem with this explanation is that studies have shown that the subglottal pressure across an utterance is actually kept constant (see Yip, 2002, for a summary).

Connell and Ladd (1990) interpret Hombert's (1978) treatise of Shona in that Shona shows considerable declination in sequences of all high-toned syllables. Within 5 allegedly high-toned syllables the pitch drops from 140 Hz to 80 Hz (Hombert, 1978: 176).[2] The data from Welmers (1973; cited in Connell & Ladd, 1990) on isiXhosa suggest that there is no global declination across the utterances in isiXhosa, but rather smaller domains of downstep and reset. The studies by Myers (1998) and Downing *et al*. (forthcoming) show that in Chichewa high tones are realised at decreasing pitch heights. However, in these studies low tones intervene between the high tones, a phonological environment which is associated with the terms *downdrift* or *downstep*.

Controlled acoustic studies on declination in Bantu languages are missing, which would allow establishing what the parameters of declination are in these languages. Variation is expected with respect to the presence/absence of declination in general, the tones affected (only high tones or both high and low tones), the degree of the F0 decrease across the utterance and the domain in which declination occurs, to name just a few. Furthermore, in a study on declination in southern Bantu languages one should be aware of the interacting, though arguably independent process of *final lowering*. In their study on aspects of pitch realisation, Yoruba, Connell and Ladd (1990) found that low tones, but not mid or high tones, show a significant lowering effect over the last one to three syllables.

Pragmatics: focus and questions

As mentioned in the subsection entitled 'Pragmatic factors', information structure and more specifically, focus, crucially determine intonation in languages such as English. The focus constituent is marked with a specific pitch accent transcribed as HL. Thus, linguistically significant pitch changes can be observed under focus in English. The role of pitch for the marking of information structure in Bantu tone languages is much more unclear, also owing to the fact that pitch changes already indicate grammatical and lexical information. However, research on information structure in Chichewa has shown that focus can be expressed prosodically, though not by the addition of a pitch accent. The pitch changes occur owing to either the occurrence of tone sandhi at focus-induced phrase boundaries (Kanerva, 1990) or the manipulation of downdrift through focus-induced boundaries (Downing *et al*., forthcoming).

Recent research on seSotho sa Leboa shows, however, that focus (at least new information focus) does not directly influence pitch realisation in this language (Zerbian, 2007c). The same is suggested for Xitsonga (Zerbian, 2007b). There is no doubt that pitch changes occur owing to morphosyntactic changes which occur under focus (cf. Zerbian, 2006c; Jokweni, 1995).

Durational changes employed for the grammatical marking of yes/no-questions in southern Bantu languages have been addressed in the subsection 'Pragmatic factors' already. Yes/no-questions in the southern African Bantu languages are pronounced with an overall raised pitch and a shortened penultimate syllable (Jones *et al*., 2001, for isiXhosa; Zerbian, 2006b, for seSotho sa Leboa). In their detailed acoustic and perceptual study, Jones *et al*. (2001) show that the overall pitch range is raised by Hz in yes/no-questions. Furthermore, they found that listeners already decide based on the pitch of the first syllable how to interpret an utterance with respect to the utterance being a question or a statement.

Interim conclusion

This section has presented the basic pitch-related issues of tone and intonation as they pertain to the Bantu languages of southern Africa. Although the tone of these languages can be captured by a two-tone system phonologically, it was shown that the actual phonetic realisation of tone and pitch-related intonation will depend on a number of other phonological and phonetic factors, such as tonal processes, adjacent tones and positional restrictions, as well as segment-inherent pitch and declining pitch over an utterance.

F0 versus intensity

Mostly, fundamental frequency is considered a correlate for intonation and tone, both in studies on tone languages such as the southern Bantu languages and in studies on stress languages such as English. However, the work on stress in Germanic languages has long revealed a more differentiated picture: several acoustic parameters, namely, F0, duration, intensity and the composition of the spectral tilt, contribute to the perceived sentence intonation. From among these, the composition of the spectral tilt and F0 have been found to be the most reliable parameters of intonation in Germanic languages. Similarly then, tone could have several acoustic parameters of which F0 is not necessarily the most reliable one.

In the acoustic studies available on Bantu tone, F0 variations have been taken as the sole indicator for tone. Roux (1995a) suggests that intensity might even be a more suitable acoustic parameter to capture tonal variations in isiXhosa. From an inspection of his illustrative example one can deduce that that might be the case owing to intensity not underlying downstep or declination which would lower a high tone that appears late in an utterance. However, our own research into the correlation of intensity and fundamental frequency in seSotho sa Leboa tone does not support the importance of intensity for tone. Although it is true that higher intensity co-occurs with high-toned syllables, it seems from the timing of F0 and intensity peaks that the intensity rather follows the laryngeal settings for tone.

Implications for speech technology

As information technology becomes increasingly pervasive in society, there is a strong desire to support local languages through technology (Goronzy *et al*., 2006). For speech technology in the southern African Bantu languages, the issues addressed above are of great importance in terms of both the resources and the algorithmic approaches that are employed. Below, we briefly discuss the importance of intonation for pronunciation dictionaries (which are fundamental resources for speech processing) and for two classes of algorithms, namely those employed in speech recognition and speech synthesis.

Pronunciation dictionaries

Pronunciation dictionaries, which specify a mapping between the orthographic representation of a word and its pronunciation, are fundamental to most modern speech processing systems. The details of this mapping depend to some extent on the purpose of the speech processing system – for a limited-vocabulary speech recognition system, a phonemic transcription may suffice, whereas a high-quality system for speech synthesis may require finer phonetic distinctions in the pronunciation dictionary. Since phrase-level intonation cannot be derived from the properties of words in isolation, information on this level is generally not included in the pronunciation dictionary. However, in light of the importance of tone in the languages studied here, the incorporation of tone information in their pronunciation dictionaries is likely to be important – as it is in well-studied tone languages such as Mandarin.

Unfortunately, the literature displays significant variability in how tone is assigned to even common words (Roux, 1995a), which may complicate the development of such tone-enhanced pronunciation dictionaries. In addition, the assignment of underlying tones is thought to be independent of the orthography, which means that predictive algorithms (Davel & Barnard, 2006) are of limited use in this task. It is therefore important that all available constraints should be employed to assist the dictionary developer: for example, the fact that verbs are classified with a single binary distinction – high or low, see 8(a) – or the observation that two adjacent underlying high tones are not favoured in surface tone structure (see ‘Immediately adjacent high tones’). The development of software tools that utilise such constraints to assist in dictionary creation should receive urgent attention, and will greatly assist in the development of speech technology for these languages.

In addition, the important phrasal effects that have been described in our review suggest that there will be a need for sophisticated post-lookup algorithms to account for the interactions between adjacent words as well as the interaction between words and phrase boundaries. Given the ubiquity of these effects in the Bantu languages, it will probably become standard practice to consider the

pronunciation dictionary along with the contextual modification algorithms as an 'extended pronunciation dictionary'.

Speech recognition

The use of suprasegmental cues in speech recognition has a long and largely undistinguished history (Shafran *et al*., 2001). Although humans are highly sensitive to these cues and numerous experimental systems have shown significant benefit from their use (see, for example, Vergyri *et al*., 2003), these benefits have rarely been sufficiently robust to justify the use of suprasegmental information in practical speech-recognition systems (at least for non-tone languages). For tone languages, such as Mandarin, there is more evidence that suprasegmental information can be useful in practice (Siu & Ostendorf, 2005), and we similarly expect that speech recognition in the southern African Bantu languages will eventually benefit from using information derived from pitch, duration and (possibly) intensity contours. However, the relative paucity of minimal pairs distinguished by tone alone suggests that this benefit will only occur when systems of fairly significant complexity are deployed. For less complex systems, it will be more reliable to disambiguate based on context rather than prosody.

In order for speech recognition to make use of intonation, a number of prerequisites need to be established. Firstly, tone-enhanced pronunciation dictionaries, enhanced with post-processing rules to account for contextual effects (as described above), are required in order to represent all lexical units in a manner that supports intonation modelling. In addition, phonological, phonetic and computational models are required to predict how the underlying tones will combine to create the surface realisation and hence the acoustic realisation. Finally, algorithms that allow the speech recogniser to deduce the intonation structure from a recorded signal and combine that with other sources of phonetic information need to be developed. Initial investigations in this regard (Govender *et al*., 2007) suggest that standard pitch-processing algorithms are sufficient for the first level of analysis, but little is known about the additional steps that are required to successfully incorporate tonal information in speech recognition.

Speech synthesis and text-to-speech (TTS) systems

Since suprasegmental information is unavoidably embedded in any spoken utterance – synthesised or natural – intonation models have been present in speech synthesisers from the outset. It has been shown that accurate models of intonation are crucial for natural-sounding speech synthesis (Black & Taylor, 1994), and for tone languages intelligibility clearly depends equally strongly on such models. Thus, the importance of reliable models of tone and intonation stands beyond doubt in this application (in contrast to speech recognition, where basic systems may succeed in the absence of such models, as discussed above).

The development of such models for TTS applications will proceed along similar lines as those for speech recognition, with the following exception: for recognition, the development of algorithms to accurately extract the physical parameters related to tone and intonation is a major challenge. For synthesis, however, the converse process of *generating* those physical parameters is rather straightforward, and the same methods that have been applied successfully for other languages (see, for example, Black & Taylor, 1994) should be directly applicable to this set of languages. To our knowledge, the only published attempt to incorporate an awareness of intonation in a Bantu language TTS system is the fairly crude system for isiZulu described by Louw *et al*. (2005). There it was found that rules that appropriately indicate the presence of phrase boundaries were of great importance, but that listeners were remarkably tolerant to the absence of further tonal processing in the reported system. Despite this surprising finding, we believe that more sophisticated intonation models will greatly enhance TTS systems in these languages.

Conclusion

The review of duration and pitch-related aspects of tone and their relation to intonation in the preceding sections has brought to light that much is already known about the prosodic systems of the indigenous southern African languages. However, in order to be able to implement these

generalisations into a computational model, more quantitative data are needed. Most of the generalisations summarised here have been published as transcribed data, using either H or L as labels (with an additional ! used to indicate downstep in the most recent literature).

Thus it is not surprising that the current article repeatedly pointed out the need for quantitative data. Quantitative acoustic data on intonational processes in southern Bantu languages provide valuable information not only for the modelling of intonation but also for the improvement of theories of intonation.

Notes

[1] An impression, however, suggests that the vowel in final syllables in at least seSotho sa Leboa is often deleted. This is especially prevalent for syllables that have *-i* as the syllable nucleus, such as *kgoši*.

[2] Hombert's (1978) interpretation of the Shona data is debatable. A sequence of five adjacent high tones is probably the result of high tone spread, given the avoidance of adjacent underlying high tones in the Bantu languages (see in this article: 'Adjacent high tones'). The pitch track that he provides could just as well be interpreted as two to three high toned syllables with a subsequent gradual decline in pitch which corresponds to low tones. The amplitude graph that he provides would support this analysis although the relation between tone and intensity is far from direct (see in this article: 'F0 versus intensity').

References

Beuchat P. 1962. Additional notes on the tonomorphology of the Tsonga noun. *African Studies* **21**(3): 105–122.

Beuchat P. 1966. *The verb in Zulu.* Johannesburg: Witwatersrand University Press.

Black A & Taylor P. 1994. 'Assigning intonation elements and prosodic phrasing for English speech synthesis from high level linguistic input'. Proceedings of the International Conference on Speech and Language Processing, Yokohama, Japan, pp 715–718.

Cassimjee F. 1992. *An autosegmental analysis of Venda tonology.* New York: Garland.

Cassimjee F. 1998. *IsiXhosa tonology: an optimal domains theory analysis.* Munich: LINCOM Europe.

Cassimjee F & Kisseberth CW. 1998. Optimality domains theory and Bantu tonology: a case study from isiXhosa and Shingazidja. In Hyman LM & Kisseberth CW (eds) *Aspects of Bantu tone*. Stanford: CSLI Publications, pp 33–132.

Cassimjee F & Kisseberth CW. 2001. Zulu tonology and its relationship to other Nguni languages. In Kaji S (ed.) Proceedings of the Symposium Cross-Linguistic Studies of Tonal Phenomena, pp 327–359.

Chafe WL. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Li CN (ed.) *Subject and topic*. New York: Academic Press, pp 25–55.

Chebanne AM, Creissels D & Nkhwa, HW. 1997. *Tonal morphology of the Setswana verb.* Munich: LINCOM Europe.

Clark MM. 1988. An accentual analysis of the Zulu noun. In Van der Hulst H & Smith N (eds) *Autosegmental studies on pitch accent.* Dordrecht: Foris Publications, pp 51–79.

Claughton JS. 1983. *The tones of Xhosa inflections.* Grahamstown: Rhodes University.

Clements GN. 1988. Tonology of the SeSotho verb. Manuscript.

Clements GN & Ford KC. 1979. Kikuyu tone shift and its synchronic consequences. *Linguistic Inquiry* **10**(2): 179–210.

Clements GN & Ford KC. 1981. On the phonological status of downstep in Kikuyu. In Goyvaerts DL (ed.) *Phonology in the 1980s*. Ghent: Story-Scientia, pp 309–357.

Coetzee AW & Wissing DP. 2007. Global and local durational properties in three varieties of South African English. *The Linguistic Review* **24**: 263–289.

Cole-Beuchat P. 1959. Tonomorphology of the Tsonga noun. *African Studies* **18**(3): 133–145.

Connell B & Ladd DR. 1990. Aspects of pitch realisation in Yoruba. *Phonology* **7**: 1–29.

Cope AT. 1959. Zulu tonology. *Afrika und Übersee* **43**(3): 190–200.

Cope AT. 1970. Zulu tonal morphology. *Journal of African Linguistics* **9**(3): 111–152.
Creissels D. 1998. High tone domains in Setswana. In Hyman LM & Kisseberth CW (eds) *Theoretical aspects of Bantu tone*. Stanford: CSLI Publications, pp 133–194.
Creissels D. 1999. The role of tone in the conjugation of Setswana. In Blanchon JA & Creissels D (eds) *Issues in Bantu tonology*. Cologne: Rüdiger Köppe, pp 109–152.
Creissels D. 2000. A domain-based approach to Setswana tone. In Wolff E & Gensler O (eds) Proceedings of the 2nd World Congress of African Linguistics, Leipzig 1997. Cologne: Rüdiger Köppe, pp 311–321.
Crystal D. 2003. *A dictionary of linguistics & phonetics. (*5th edition) Malden: MA Blackwell.
Davel M & Barnard E. 2006. Bootstrapping pronunciation models. *South African Journal of Science* **102**(7/8): 322–329.
Demuth K. 1990. Accent, tone and the acquisition of underlying representations. Manuscript.
Doke CM. 1927. *Textbook of Zulu grammar.* Cape Town: Maskew Miller Longman.
Downing LJ. 2001. How ambiguity of analysis motivates stem tone change in Durban Zulu. *UBC Working Papers in Linguistics* **4**: 39–55.
Downing LJ. 2006. *Canonical forms in prosodic morphology* Oxford: Oxford University Press.
Downing LJ & Gick B. 2005. Voiceless tone depressors in Nambya and Botswana Kalang'a. Proceedings of the Berkeley Linguistics Society **27** (2001): 65–80.
Downing LJ, Mtenje A & Pompino-Marschall B. Forthcoming. Focus phrasing and raising in a variety of Chichewa.
Endemann K. 1911. *Wörterbuch der Sotho-Sprache*. Hamburg: L. Friedrichsen.
Endemann TMH. 1952. *Die intonasie van Tsonga ('n Sinkroniese studie van die Tsonga-spreektone)*. Pretoria: University of South Africa.
Goldsmith J. 1984. Tone and accent in Tonga. In Clements GN and Goldsmith J (eds) *Autosegmental studies in Bantu tone*. Dordrecht: Foris, pp 19–53.
Goldsmith J, Peterson K & Drogo J. 1989. Tone and accent in the Xhosa verbal system. In Newman P & Botne R (eds) *Current approaches to African linguistics*. Dordrecht: Foris, pp 157–178.
Goronzy S, Tomokiyo LM, Barnard E & Davel M. 2006. Other challenges: non-native speech, dialects, accents, and local interfaces. In Schultz T & Kirchhoff K (eds) *Multilingual speech processing*. New York: Academic Press, pp 273–317.
Govender N, Barnard E & Davel M. 2007. Pitch modelling for the Nguni languages. *South African Computer Journal* **38**(1): 28–39.
Hirst D & Di Cristo A. 1998. *Intonation systems: a survey of twenty languages.* Cambridge: Cambridge University Press.
Hombert J. 1978. Consonant types, vowel quality, and tone. In Fromkin VA (ed.) *Tone: a linguistic survey.* New York: Academic Press, pp 77–111.
Honda K & Fujimura O. 1991. Intrinsic vowel F0 and phrase-final F0 lowering: phonological vs biological explanations. In Gauffin J & Hammerberg B (eds) *Vocal fold physiology: acoustic, perceptual, and physiological aspects of voice mechanisms.* San Diego: Singular Publishing Group, pp 149–157.
Hyman LM & Mathangwane JT. 1998. Tonal domains and depressor consonants in Ikalanga. In Hyman LM & Kisseberth CW (eds) *Theoretical aspects of Bantu tone*. Stanford: CSLI Publications, pp 195–229.
Jokweni MW. 'Aspects of isiXhosa phrasal phonology' (Doctoral thesis, University of Illinois, 1995).
Jokweni MW. 1998. Parametric phonology and boundary tonology in Xhosa. *South African Journal of African Languages* **18**(2): 29–32.
Jones J, Louw JA & Roux JC. 2001a. Perceptual experiments on Queclaratives in Xhosa. S*AJAL*, supplement **36**: 19–31.
Jones J, Louw JA & Roux JC. 2001b. Queclaratives in Xhosa: an acoustic analysis. *SAJAL*, supplement **36**: 3–18.
Kanerva JM. 1990. *Focus and phrasing in Chichewa phonology.* New York/London: Garland Publishing.

Khoali B. 'A Sesotho tonal grammar' (Doctoral thesis, University of Illinois, 1991).
Khumalo JSM. 1981. Zulu tonology, Part I. *African Studies* **40**(2): 53–130.
Khumalo JSM. 1982. Zulu tonology, Part II. *African Studies* **41**(1): 3–125.
Khumalo JSM. 'An autosegmental account of Zulu phonology' (Doctoral thesis, University of the Witwatersrand, 1987)
Kingston J. 1992. The phonetics and phonology of perceptually motivated articulatory covariation. *Language & Speech* **35**: 99–113.
Kisseberth CW. 2000. The phonology-syntax interface: Chimwiini revisited. Manuscript.
Kisseberth CW & Abasheikh MI. 1974. Vowel length in Chi-Mwi:ni – a case study of the role of grammar in phonology. In Bruck A (ed.) Papers from the Parasession on Natural Phonology. CLS, pp 193–209.
Kisseberth CW & Mmusi SO. 1990. The tonology of the object prefix in Setswana. *Studies in the Linguistic Sciences* **20**(1): 151–161.
Kisseberth CW & Odden D. 2003. Tone. In Nurse D & Philippson G (eds) *The Bantu languages.* London/New York: Routledge, pp 59–70.
Klatt D. 1976. Linguistic uses of segment duration in English: acoustic and perceptual evidence. *Journal of the Acoustical Society of America* **59**: 1208–1221.
Köhler O. 1956. Das Tonsystem des Verbum im Südsotho. *Mitteilungen des Instituts für Orientforschung* **IV**(3): 435–474.
Krifka M. 2006. Basic notions of information structure. In Fery C, Fanselow G & Krifka M (eds) *Interdisciplinary Studies on Information Structure 06*. Potsdam: University of Potsdam.
Kunene DP. 1972. A preliminary study of downstepping in southern Sotho. *African Studies* **31**(1): 1–24.
Laniran YO & Clements GN. 2003. Downstep and high raising: interacting factors in Yoruba tone production. *Journal of Phonetics* **31**: 203–250.
Laughren M. 1984. Tone in Zulu nouns. In Clements GN & Goldsmith J (eds) *Autosegmental studies in Bantu tone*. Dordrecht: Foris Publications, pp 183–234.
Leben W. 'Suprasegmental phonology' (Doctoral thesis, MIT, 1973).
Lehiste I. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.
Letele GL. 1955. *The role of tone in the southern Sotho language*.
Lets'eng MC. 1994. Les tiroirs verbaux du Sesotho. Manuscript, Grenoble III.
Lindblom B, Lyberg B & Holmgren K. 1981. *Durational patterns of Swedish phonology: do they reflect short-term memory processes?* Bloomington: Indiana University Linguistics Club.
Lombard DP. 'Aspekte van toon in Noord-Sotho' (Doctoral thesis, University of South Africa, 1976).
Lombard DP. 1979. Duur en lengte in Noord-Sotho. *Studies in Bantoetale:* 39–69.
Louw JA. 'The intonation of the sentence and its constituent parts in Xhosa and Tsonga' (Doctoral thesis, University of South Africa, 1979).
Louw JA. 1983. Some tone rules of Tsonga. *Afrika und Übersee* **66**(1): 13–24.
Louw JA, Davel M & Barnard E. 2005. A general-purpose isiZulu speech synthesiser. *South African Journal of African Languages* **25**(2): 92–100.
Maddieson I. 1985. Phonetic cues to syllabification. In Fromkin V (ed.) *Phonetic linguistics*. Orlando: Academic Press, pp 203–222.
Meyer EA. 1896/97. Zur Tonbewegung des Vokals im gesprochenen und gesungenen Einzelwort. *Phonetische Studien* **10**: 1–21.
Mmusi SO. 'Obligatory contour principle effects and violations: the case Setswana verbal tone' (Doctoral thesis, University of Illinois, 1992).
Monareng WM. 'A domain-based approach to Northern Sotho tonology: a Setswapo dialect' (Doctoral thesis, University of Illinois, 1992).
Mosaka NM. 2000. Stress assignment in syllabic structures in Xhosa and Tswana. *South African Journal of African Languages* **20**(2): 177–185.
Myers S. 'Tone and the structure of words in Shona' (Doctoral thesis, MIT, 1987).
Myers S. 1997. OCP effects in optimality theory. *Natural Language and Linguistic Theory* **15**(4): 847–892.

Myers S. 1998. Surface underspecification of tone in Chichewa. *Phonology* **15**: 367–391.
Myers S. 1999. Tone association and F0 timing in Chichewa. *Studies in African Linguistics* **28**(2): 215-239.
Myers S. 2003. F0 timing in Kinyarwanda. *Phonetica* **60**: 71–97.
Myers S. 2005. Vowel duration and neutralization of vowel length contrasts in Kinyarwanda. *Journal of Phonetics* **33**: 427–446.
Nespor M & Vogel I. 1986. *Prosodic phonology.* Dordrecht: Foris Publications.
Odden D. 1994. Adjacency parameters in phonology. *Language* **70**(2): 289–330.
Peterson K. 1989. A comparative look at Nguni verbal tone. In Haik I & Tuller L (eds) *Current approaches to African linguistics.* Dordrecht: Foris, pp 115–137.
Pike KL. 1945. *Tone languages*. Ann Arbor: University of Michigan Press.
Rialland A. 2001. Anticipatory raising in downstep realization: evidence for preplanning in tone production. In Kaji S (ed.) Proceedings of the Symposium Cross-linguistic Studies of Tonal phenomena, pp 301–326.
Rialland A. Forthcoming. The African lax question prosody: its realisation and geographical distribution. *Lingua*.
Rialland A & Some PA. 2000. Dagara downstep: how speakers get started. In Carstens V & Parkinson F (eds). *Advances in African linguistics.* Trenton, NJ: Africa World Press, pp 251–263.
Roux JC. 1995a. On the perception and production of tone in Xhosa. *South African Journal of African Languages* **15**(4): 196–204.
Roux JC. 1995b. Prosodic data and phonological analyses in Zulu and Xhosa. *South African Journal of African Languages* **15**(1): 19–28.
Rycroft DK. 1963. Tone in Zulu nouns. *African Language Studies* **IV**: 43–68.
Rycroft DK. 1980a. Ndebele and Zulu: some phonetic and tonal comparisons. *Zambezia* **III**: 109–128.
Rycroft DK. 1980b. *The depression feature in Nguni Languages and its interaction with tone*. Grahamstown: Rhodes University.
Rycroft DK. 1983. Tone-patterns in Zimbabwean Ndebele. *BSOAS* **46**(1): 77–135.
Shafran I, Ostendorf M & Wright R. 2001. 'Prosody and phonetic variability: lessons learned from acoustic model clustering'. Proceedings of the ISCA Workshop on Prosody in Speech Recognition and Understanding, pp. 127–131.
Sibanda G. 'Verbal phonology and morphology of Ndebele' (Doctoral thesis, University of California, 2004).
Siu MNT & Ostendorf M. 2005. A quantitative assessment of the importance of tone in Mandarin speech recognition. *Signal Processing Letters* **12**(12): 867–870.
Traill A, Khumalo JSM & Fridjhon P. 1987. Depressing facts about Zulu. *African Studies* **46**(2): 255–274.
Trümpelmann HD. 'Intonasie in Sepedi' (Masters thesis, University of Pretoria, 1942).
Van der Pas B, Wissing D & Zonneveld W. 2000. Parameter resetting in metrical phonology: the case of Setswana and English. *South African Journal of Linguistics*, supplement **38**: 55–88.
Vergyri D, Stolcke A, Gadde VRR, Ferrer L & Shriberg E. 2003. **'**Prosodic knowledge sources for automatic speech recognition'. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Hong Kong, pp 208–211.
Welmers W. 1973. *African language structures.* Berkeley: University of California Press.
Westphal EOJ. 1962. Venda: tonal structure and intonation. *African Studies* **21**(2): 49–69.
Whalen D & Levitt A. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* **23**: 349–366.
Xu Y. 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* **27**: 55–105.
Xu Y. 2006. Principles of tone research. International Symposium on Tonal Aspects of Language, La Rochelle.
Yip M. 2002. *Tone.* Cambridge: Cambridge University Press.
Zerbian S. 2006a. Variation in HTS in the Sotho verb. In Mugane J, Huchtinson J, & Worman D

(eds) *Trends in African linguistics*. Somerville, MA: Cascadilla Publishing Project, pp 147–157.
Zerbian S. 2006b. Questions in Northern Sotho. *Linguistische Berichte* **208**: 385–405.
Zerbian S. 'Expression of information structure in Northern Sotho' (Doctoral thesis, Humboldt-University, 2006c).
Zerbian S. 2007a. Phonological phrasing in Northern Sotho. *The Linguistic Review* **24**: 233–262.
Zerbian S. 2007b. A first approach to information structuring in Xitsonga/Xichangana. *Research in African Languages and Linguistics* **7** (2005–2006): 1–22. (Also in *SOAS Working Papers* **15**: 65–78).
Zerbian S. 2007c. Investigating prosodic focus marking in Northern Sotho. In Hartmann K, Aboh E & Zimmermann M (eds) *Focus strategies: evidence from African languages*. Berlin: Mouton de Gruyter, pp 55–79.
Zerbian S & Barnard E. 2008. Phonetic and morpho-phonological factors in the alignment of a single high tone in Sepedi. Manuscript.
Ziervogel D. 1959. *A Grammar of Northern Transvaal Ndebele.* Pretoria: J.L. van Schaik.
Ziervogel D, Lombard DP & Mokgokong PC. 1969. *A handbook of the Northern Sotho language.* Pretoria: J.L. van Schaik.
Ziervogel D, Louw JA, Ferreira JA, Baumbach EJM & Lombard, DP. 1967. *Handbook of the speech sounds and sound changes of the Bantu languages of South Africa.* Pretoria: University of South Africa.
Ziervogel D & Mokgokong PC. 1979. *Klein Noord-Sotho woordeboek.* Pretoria: J.L. van Schaik.