AN ACOUSTIC ANALYSIS OF TONE IN SESOTHO

Lehlohonolo Mohasi[a], *Hansjörg Mixdorff*[b] *& Thomas Niesler*[a]

[a]University of Stellenbosch, South Africa; [b]Beuth University of Applied Sciences Berlin, Germany
mohasi@sun.ac.za; mixdorff@beuth-hochschule.de; trn@sun.ac.za

ABSTRACT

This paper presents one of the first acoustic investigations into the realization of tone in Sesotho, a Southern African language, employing a set of recorded minimal pairs, whose F0 contours are analyzed using the Fujisaki model. Analysis of model parameters shows that high tones are associated with tone commands whereas low tones either follow the phrase contour or are marked by creaky voice. In some minimal pairs, partners differed only with respect to vowel quality, with the expected high and low tone associated with closed and open vowels, respectively. Question intonation is marked by raising the F0 contour, which is reflected by higher phrase command magnitudes and a shortening of the penultimate syllable.

Keywords: Sesotho, Fujisaki model, tone languages

1. INTRODUCTION

Sesotho is a Southern Bantu language spoken as an official language in Lesotho and South Africa. The two countries use different orthographies though the pronunciation is the same. For this paper, the Lesotho orthography is adopted. Sesotho is a tonal language with two contrasting tonemes, high (H) and low (L). The tone of a syllable is carried by the vowel, or by the nasal, if the nasal is syllabic. The interrelationship between the tone and general intonation in Sesotho is hitherto poorly understood and technologically not addressed. This is complicated by the fact that tonal information is not indicated in the orthography of Sesotho [4, 10], as well as most other Bantu languages [11].

In order to quantify the tonal alignments and magnitudes of excursions, *F0* contours are parameterized using the Fujisaki model [3]. This model decomposes a given log *F0* contour into a base frequency *Fb*, a phrase component, which captures slower changes in the *F0* contour as associated with intonation phrases, and an accent component that reflects faster changes of *F0* associated with accents and boundary tones. The phrase and accent components can be interpreted as smooth responses of the model to impulse-wise phrase commands and step-wise accent commands. In earlier studies, the model was applied to F0 contours of Asian tone languages such as Chinese and Thai [8, 9]. Since they are associated with syllabic tones, the accent commands are termed "tone commands". All tone languages examined so far required tone commands of negative polarity in order to model low, falling and rising tones. It will therefore be one of the objectives of this study to examine whether this is also the case for Sesotho.

2. SPEECH MATERIAL AND METHOD OF ANALYSIS

A corpus of tonal minimal pairs was created. Due to the small number of examples provided in the literature (in [2] for instance), the first author, a native speaker of Sesotho, augmented these with her own minimal pairs, yielding a total of 14. The corpus was recorded in a professional recording studio using two female and four male native speakers of Sesotho from Lesotho. All subjects have a university education, two at undergraduate, three at postgraduate, and one at post-doctoral level.

The recording was performed using two strategies: reading and repeating. The first strategy involved subjects reading Sesotho text, in random order, from slides presented on a computer screen. Each slide displayed Sesotho text and its English translation which was intended to guide subjects at choosing the right tone for the critical word in the Sesotho text. The repeating strategy involved the same subjects speaking the Sesotho utterances after the first author, that is, the first author's recordings were used as a reference. Since we were also interested in examining the interaction between syllabic tones and the intonation of questions, one of the partners in five of the minimal pairs was elicited in the interrogative mode. All minimal pairs are listed in Table 1. Some are reported to also exhibit vowel differences in the first syllable which are listed in the table, provided we were able to confirm them in our study.

Table 1: List of minimal pairs showing the critical words, their positions in the utterance, their respective English translations, expected tones, observed tones, and vowel differences (if any) in the nucleus of the first syllable. The means and standard deviations of the amplitude and timing of the underlying tone commands are displayed, the latter given relative to the beginning of the first syllable of most critical words (underlined), and, with respect to the second for "lehata" and third syllable for "lehare", "bobatsi", respectively. The (timing) differences that set apart the tonal assignments of the two partners in a pair are set in bold face.

word	position	translation	exp. tone	obs. tone	vowel	A_t	*T1 rel* [*ms*]	*T2* rel [*ms*]
hlola	final	created	LL	HL	[O]	.21/.08	-16/32	194/43
		conquered	HH	HL	[o]	.19/.08	27/56	269/69
seba	final	gossip	HH	HL	[e]	.15/.09	-195/263	**208/115**
		do mischief	LL	LL	[e]	.21/.06	-451/173	**21/65**
pota	medial	coming over	LL	HH	[O]	.18/.08	-268/211	300/67
		talking crap	HH	HH	[o]	.14/.04	-232/163	222/110
tena	final	is getting dressed	HH	HL	[e]	.21/.14	**-134/111**	**56/93**
		is annoying	LL	LL	[e]	.28/.14	**-303/53**	**-1/34**
bolla	medial	was circumcised	HHH	HHH	[o]	.22/.06	-155/212	**309/84**
		decayed	LLL	LLL	[O]	.25/.07	-521/131	**8/107**
hlopa	medial	prepares	LL	HL	[O]	.27/.07	-126/36	273/47
		torments	HH	HL	[o]	.31/.11	-66/50	230/51
bona	initial	them	LL	LH	[o]	.32/.10	**216/61**	**486/208**
		See	HH	HL	[o]	.24/.13	**-22/25**	**175/71**
bopa	medial	sulked	LL	LL	[O]	.20/.07	-515/181	**-177/117**
		molded	HH	HH	[o]	.16/.04	-271/282	**328/221**
lapa	final	patched (it)	HH	HL	[a]	.13/.04	**-155/196**	**261/98**
		became hungry	LL	LL	[a]	.28/.08	**-362/78**	**-73/22**
ts'ela	final	poured	LL	HL	[E]	.19/.07	-17/41	197/58
		crossed	HH	HL	[e]	.18/.06	-12/45	281/71
hloma	medial	acted	HH	HH	[o]	.20/.06	**-13/19**	394/109
		planted	LL	LH	[o]	.29/.02	**229/48**	481/246
lehare	final	razor	LHH	LLL	[e]	**all low tones, no command**		
		middle	LLL	LLH	[e]	**.32/.16**	**-8/60**	**191/33**
bobatsi	final	nettle	LHL	LLL	[o]	**all low tones, no command**		
		width	LLH	LLH	[o]	**.40/.15**	**-2/44**	**231/87**
lehata	medial	liar	LLL	HLL	[e]	.16/.05	**287/72**	745/108
		skull	LHH	HHH	[e]	.16/.05	**-6/54**	624/286

In this paper, high tones will be denoted by acute (ó) and low tones by grave accents (ò) on the vowel.

Initial auditory analysis of the utterances was performed which revealed considerable mismatches between the intended tones and those actually produced. The error rate for reading amounted to 23.6%, whereas for the repeating task it was 7.4%. Therefore only utterances from the repeating task were admitted to the acoustic analysis, together with the first author's sample utterances, yielding a total number of 200 tokens.

F0 values were extracted using the standard method of *Praat* [1] at a step of 10*ms* and inspected for errors. The *F0* tracks were subsequently decomposed applying an automatic method originally developed for German [6]. Initial experiments had shown that the low tones in the critical words could be modeled with sufficient accuracy without employing negative tone commands. As a consequence, only high tones were associated with tone commands. Adopting this rationale, automatically calculated parameters were viewed in the *FujiParaEditor* [7] and corrected if necessary. The utterances were segmented at the word and syllable levels, using the *Praat TextgridEditor* [1].

3. RESULTS OF ANALYSIS

Table 2 lists the subjects whose data were considered for this study, their respective values of base frequency *Fb*, which were kept constant for all utterances by the same speaker, as well as means and standard deviations of syllabic durations, tone command amplitudes *At*, and phrase command magnitudes *Ap*. The latter two parameters reflect the pitch range that subjects employ, with *At* capturing the interval of local tonal transitions, and *Ap* the amount of declination reset at the onset of a phrase. Since

the Fujisaki model is defined in the log *F0* domain, values for male and female subjects are in the same range. Figure 1 shows examples of the sentence "O ile a bolla thabeng" produced by subject MS. Each panel displays from the top to the bottom: the speech waveform, the *F0* contour (extracted and modeled), as well as the underlying phrase and tone commands. The syllable boundaries are indicated by the dotted vertical lines. The top and center panels show "bòllà" and "bóllá", respectively, embedded in statements, and the bottom panel "bóllá" embedded in a question. As can be seen the tone commands extend across several syllables which are hence associated with high tones. The main distinction between "bòllà" and "bóllá" is that the tone command underlying "o ile a" ends before the segmental onset for "bòllà" whereas it extends to the last syllable for "bóllá".

Table 2: List of subjects indicating sex, means and s.d. of syllabic durations, *Fb* and means and s.d. of *At* and *Ap*. HL is the first author of the study.

subject	sex	*Fb* [Hz]	syll.dur μ/σ [s]	*At* μ/σ	*Ap* μ/σ
HL	F	140	144/88	.19/.08	.25/.11
MM	F	130	137/61	.26/.11	.34/.11
NR	F	150	162/82	.21/.09	.28/.11
BL	M	85	176/88	.29/.11	.40/.18
SR	M	90	154/75	.23/.10	.31/.13
MS	M	100	149/68	.25/.09	.50/.13
KK	M	95	132/61	.17/.11	.30/.10

The succeeding low tones follow the phrase contour to the end of the utterance. It should be noted that in most instances of utterance-final low tones female subjects produced creaky voice whereas the male subjects exhibit modal vocal fold vibrations.

In the case of the question, the underlying phrase command has a much higher amplitude (.74) compared to that of the statement (.30) which raises the *F0* pattern without changing the tonal assignment for "bóllá". However, the long tone command that was found in the statement is split into two separate commands, one for "o ile a" and one for "bóllá". A comparison with other utterances suggests that this prosodic chunking seems to be an acceptable choice, even for the statement (see Figure 2). The three right-most columns of Table 1 list averaged Fujisaki model parameters for all critical words. In addition to averaged tone command amplitudes *At*, the relative timing is given, expressed as the mean distance between the tone command onset time *T1* and the offset time *T2* and the segmental onset of the syllable marked by underlining in the word in *ms*, henceforth *T1rel* and *T2rel*. Negative timing values therefore indicate tone command onsets or offsets occurring *before* the onset of the syllable, whereas positive values indicate tone command onsets or offsets *after* the syllable onset. The tonal transitions distinguishing the two partners in a minimal pair are indicated in bold face. Since high tones are aligned with tone commands, the onset of a tone command at time *T1* indicates the transition between a low and a high tone, whereas the offset at time *T2* marks the transition from a high to a low tone. In the case of "bolla", *T1rel* is negative, -151*ms* for "bóllá" and -521*ms* for "bòllà", respectively, indicating that the command starts before the beginning of the word. Since, as we described, prosodic chunking might occur, the onset of the command can vary considerably in the case of "bóllá", hence also *T1rel* will vary. This is also reflected by the high standard deviation of 212*ms* in contrast to only 131*ms* for "bòllà". *T2* is the time at which the tone command ends.

Figure 1: Examples of analysis of the sentence "o ile a bolla thabeng", uttered by speaker MS. Panels from the top to the bottom: low tone statement, high tone statement, high tone question.

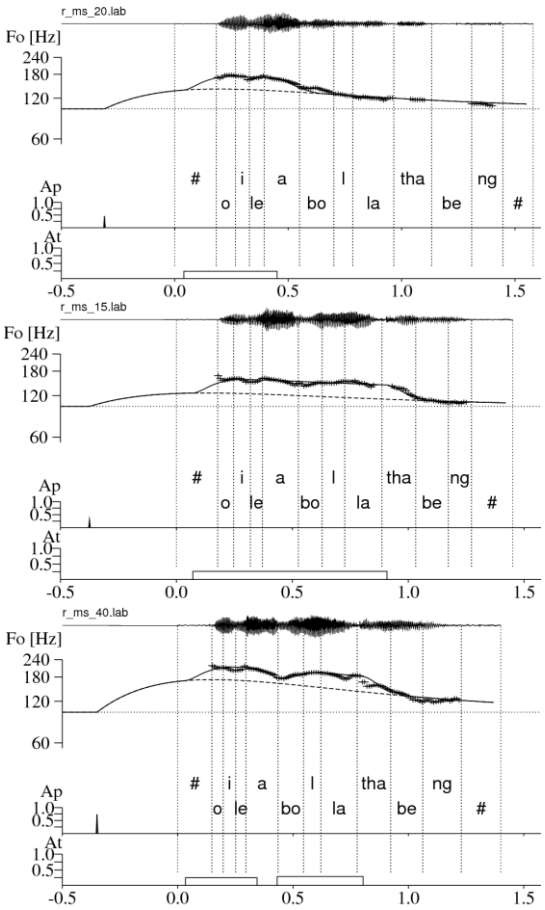
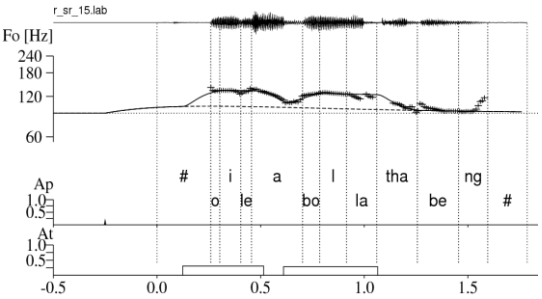


Figure 2: Analysis of the sentence "o ile a bóllá thabeng", uttered by speaker SR uttered as a statement.



As we saw from Figure 1 this happens before the onset of the critical word for "bòllà" (*T2rel*= 8ms) and afterwards for "bóllá" (*T2rel*=309*ms*).

With respect to "bolla" *T2rel* is the important timing difference. For the sake of conciseness we shall not to discuss other cases in detail, but rather refer to Table 1. The fourth column lists the expected tone for all critical words, and the fifth column the tones that were actually observed. It should be noted that by rule utterance-final high tones on final syllables of verbs are converted to low tones (see, for instance, "hlola" and "lapa"). In the cases of "hlola", "pota", "hlopa" and "ts'ela" only vowel differences could be detected. In all of these minimal pairs, the presumed high tone partner exhibits a closed vowel and the low tone one an open vowel in the first syllable. These perceptual observations were also confirmed by formant measurements carried out with *Praat*. For example for the high "ts'ela" mean *F1/F2* of 364Hz/2174Hz and for the low "ts'ela" mean *F1/F2* of 481Hz/2114Hz were measured. This suggests that tonal and vowel differences might interact in the perceptual assessment of tones.

We compared Fujisaki model parameters for the questions and their corresponding statements. *At* and especially *Ap* for questions were higher for questions (means/s.d.: .24/.10, .42/.16) than for the statements (.23/.09, .32/.15). The literature [6] indicates that questions exhibit compressed penultimate syllables. This was confirmed by our measurements: The mean duration of penultimate syllables was 168*ms* for questions as compared to 254*ms* for the corresponding statements. Finally we examined whether the position of a tonal transition in an utterance had an influence on its interval expressed by *At*. The correlation between the index of the syllable and *At* is highly significant (ρ=-.258; $p < .01$). This indicates that the *F0* range narrows slightly towards the end of the phrase.

4. DISCUSSION AND CONCLUSIONS

The material examined so far only presents a first snapshot on tonal realizations in Sesotho. Since orthography does not reflect tones, the data collection was problematic. The repeating task that yielded a lower error rate can also be questioned as it remains unclear whether subjects identified intended meanings or simply imitated what they heard. The tonal organization of Sesotho is different from that found in Asian tone languages where every syllable is assigned a specific tonal target. In Sesotho only high tone syllables are associated with tone commands and other syllables are either transient or their *F0* follows the phrasal contour. In other words: Low tones could as well be interpreted as the absence of tone since they do not require tone commands. This interpretation has also been suggested in [5], for instance. Furthermore, tone commands may extend over several syllables, facilitating prosodic grouping of utterances. Since perceived tone also appears to depend on the underlying vowel, we will verify as part of future work that minimal pairs which have been found to be distinguished by vowel quality remain distinct once the pitch has been monotonized.

5. REFERENCES

[1] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10), 341-345.

[2] Doke, C.M., Mofokeng, M. 1957. *Textbook of Southern Sotho Grammar.* Cape Town: Longmans, Green and Co.

[3] Fujisaki, H., Hirose, K. 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *J. of the Acoust. Society of Japan* (E) 5(4), 233-241.

[4] Jacottet, E. 1914. *A Practical Method to Learn Sesuto.* Morija Sesuto Book Depot.

[5] Kisseberth, C., Odden, D. 2003. Tone. In Nurse, D., Philippson, G. (eds.), *The Bantu Languages*. London: Routledge, 59-70.

[6] Mixdorff, H. 2000. A novel approach to the fully automatic extraction of Fujisaki model parameters. *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* Istanbul, 3, 1281-1284.

[7] Mixdorff, H. 2009. FujiParaEditor, *http://public.bht-berlin.de/~mixdorff/thesis/fujisaki.html*. Retrieved by 1/10/2009.

[8] Mixdorff, H., Hu, Y., Chen, G. 2003. Towards the automatic extraction of Fujisaki model parameters for Mandarin. *Proceedings of Eurospeech 2003* Geneva.

[9] Mixdorff, H., Luksaneeyanawin, S., Fujisaki, H., Charnvivit P. 2002. Perception of tone and vowel quantity in Thai. *Proc. ICSLP* Denver, USA.

[10] Paroz, R.A. 1946. *Elements of Southern Sotho.* Morija Sesuto Book Depot.

[11] Zerbian, S., Barnard, E. 2010. Word-level prosody in Sotho-Tswana. *Proceedings of Speech Prosody 2010*.