ANALYSIS OF SESOTHO TONE USING THE FUJISAKI MODEL

UNIVERSITEIT STELLENBOSCH UNIVERSITY

L. Mohasi [1], H. Mixdorff [2], T. Niesler [1] & S. Zerbian [3]

[1] University of Stellenbosch, South Africa

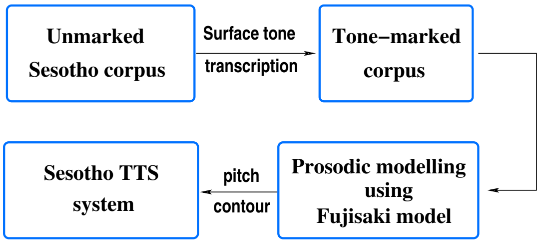
[2] Beuth University of Applied Sciences Berlin, Germany

[3] University of Potsdam, Germany

BEUTH HOCHSCHULE FUR TECHNIK BERLIN

1. Introduction

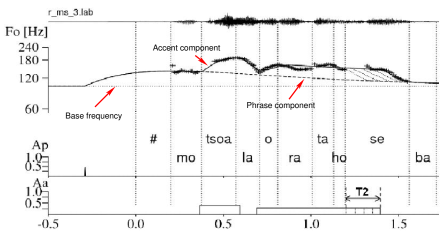
- Sesotho, like most Bantu languages, is a tonal language and uses pitch to distinguish meaning.
- Due to the absence of prosodic marking in the written text, prosodic modelling in Sesotho is a challenge.
- In order for text−to−speech (TTS) systems to produce intelligible and natural−sounding speech, accurate prosodic modelling is crucial.
- We investigate the application of the Fujisaki model to the tonal analysis of Sesotho.
- The tone commands of the Fujisaki model, which represent high tones, are compared with the high surface tones predicted by the tone rules.
- We are particularly interested in how closely related the two predictions are, and how they compare, with the perceived tone.
- The ultimate goal is to develop a technique that is able to predict the Fujisaki tone commands based on the surface tones.



2. The Fujisaki Model

The Fujisaki model is a powerful tool for prosody manipulation which:

- Is reliant on the acoustics of the uttured speech.
- Analyses the F0 contour of a natural utterance and decomposes it into a set of basic components:
  - a base frequency;
  - a phrase component which captures slower changes in the F0 contour as associated with intonation phrases;
  - an accent (tone) command that reflects faster changes in the F0 associated with high tones.



3. Our Corpus

- The corpus was drawn from broadcast speech and constitutes a semantically meaningful passage in a targeted domain (weather forecasting).
- With the goal of TTS in mind, it is efficient to initially focus on a limited domain.

Aspect	Data
Number of sentences	53
Duration	37 minutes
Average length (words)	23
Genre	Weather forecast
Source	Lesotho Meteorological Services
Annotations	Phone, syllable, word
Number of transcribers	2

4. Tone transcriptions

A. Surface tone transcription

- Underlying tones obtained from two tone−marked dictionaries.
- Pronunciation dictionary created from these underlying tones.
- Sesotho tonal rules as per Khoali (1991).
- Surface tones deduced by means of tonal rules and morphological analysis.

B. Perceived tone transcription

- For better analysis of correspondence between the surface tones and the tone commands from the Fujisaki model, each syllable was listened to individually.
- This was followed by a manual and visual inspection of the F0 contour produced by the Fujisaki model.
- The F0 excursions were examined and compared with the surface tone prediction and the perceived tone.

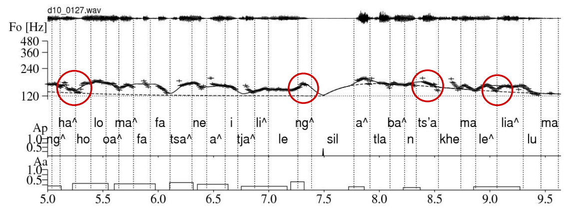
	% Perceived tone high	% Perceived tone low
Surface tone high	79.8	19.1
Surface tone low	18.2	81.3

	% with tone commands	% w/out tone commands
Surface tone high	71.1	27.5
Surface tone low	28.6	69.4

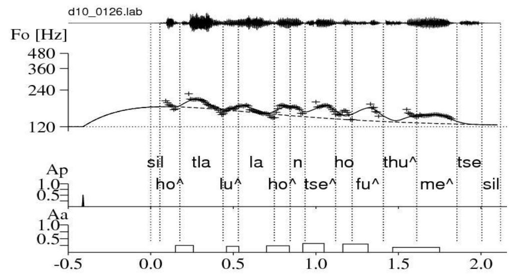
	% with tone commands	% w/out tone commands
Perceived tone high	78.2	21.3
Perceived tone low	29.5	68.7

5. Neighbouring syllables

- Tones of syllables were studied with reference to the influence of their immediate left and right neighbours.
- The relationship between the predicted surface tone, the perceived tone and the output of the Fujisaki model was again considered.
- Partial phrases repeated in the corpus were used to check consistency of both the Fujisaki model parameters and the perceptual tone labels.
- Discrepancies could be accounted for by:
  - F0 extraction errors by the Fujisaki model.
  - Surface tone/perceived tone mismatches.
  - Surface tone prediction rule.



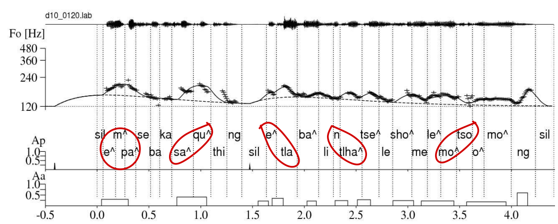
- 1st ellipse − "ha" shows H surface tone yet F0 excursion and perceived tone are L. Suspect incorrect lexical tone pattern leads to wrong surface tone.
- 2nd ellipse − "ng" at phrase final position but shows H F0 excursion and surface tone. This rise is due to continuation rise and the relative verb (itjalileng).
- 3rd ellipse − Microprosody influence on "ts'a" due to "ts" sound.
- 4th ellipse − The dip between "le" and "lia" is due to the d−like sound.



- Tone command associated with H tone can be delayed, e.g. "ho^" should be H but the following "tla" is higher. Also applies to "lu^la", where F0 peak is delayed.

6. Prosodic groupings

- Sequences of consecutive high tones were observed in the Fujisaki analysis.
- In these sequences, the words and syllables link up so closely they appear to be one unit, indicated by long tone commands.



- Most of the groupings (67/125) were due to two or more adjacent H tone syllables, with the last H tone being carried over into the next L−tone syllable.
- Another frequent grouping consisted of alternating tone labels, e.g. HLHLHL.

7. Conclusions

- Since neither surface tone prediction nor perceived tone labels are completely reliable, care must be excercised when interpreting the Fujisaki tone commands. Careful inspection of the Fujisaki analysis in conjunction with listening tests (by different subjects), is currently the only means to achieve consistency.
- Ongoing work is to develop a technique that can predict the Fujisaki tone commands from the orthography and the surface tones. This would allow the Fujisaki model to be integrated into a Sesotho TTS system for the purpose of prosody modelling. The good correspondence between surface/perceived tones and Fujisaki tone commands gives us optimism for the prospects of this approach.

Poster presented at *TAL 2012, 26–29 May 2012 in Nanjing, China.*