Savosavo word stress: a quantitative analysis

Candide Simard[1], *Claudia Wegener*[2], *Albert Lee*[3], *Faith Chiu*[3], *Connor Youngberg*[1]

[1]Department of Linguistics, SOAS, University of London, United Kingdom.
[2] Faculty of Linguistics and Literary Studies, University of Bielefeld, Germany.
[3]Department of Speech, Hearing and Phonetic Sciences, University College London, United Kingdom.

cs75@soas.ac.uk, claudia.wegener@uni-bielefeld.de, kwing.lee.10@ucl.ac.uk, faith.chiu.11@ucl.ac.uk, cy4@soas.ac.uk

Abstract

This paper presents a quantitative analysis of stress in Savosavo (unclassified), an endangered language spoken on Savo Island, (Solomon Islands). Acoustic analyses comprise the measurements of F0, duration, and intensity for each syllable in a dataset carefully selected from elicited speech from one speaker only, aiming to test the effect of increasing morphological complexity on stress realization in a system that displays some variation. Statistically significant variation is found in all correlates between stressed and unstressed syllables, thus fitting with widely attested manifestations of stress cross-linguistically. Findings were further tested with a re-synthesis tool, to confirm our initial hypotheses. Our results demonstrate that the current annotation scheme is a reliable representation of the data, and that the qTA component embedded in PENTAtrainer is effective in modelling F0 contours, even with less controlled data as input. We will argue for the usefulness of instrumental phonetic investigations in describing lesser-known languages, to enhance our understanding of the characterization of the prosodic systems of the world's languages.

Index Terms: prosody, word stress, pitch, duration, Austronesian language, Oceanic language, prosodic typology

1. Introduction

Savosavo is an unclassified, Papuan language spoken on Savo Island (one of the Solomon Islands) by about 2500 people. It does not have any (close) relatives, and has been surrounded by Oceanic languages for at least several hundred years, hence there has been long-standing contact between Savosavo and many of its neighbours. Because of a shift in the younger generations to Pijin and English, Savosavo is an endangered language currently being documented by Wegener [1]. The speech data on which this study is based is part of the documentation of Savosavo, which is thus precious as the sole possible corpus of analysis.

Savosavo is a verb-final language, with postpositions, and adnominal modifiers preceding the head of an NP; it also has a marked-nominative system with case-marking enclitics on syntactic subject noun phrases, but no case-marking on object noun phrases, whereas on verbs only syntactic objects are cross-referenced by means of affixes or stem modification, while syntactic subjects remain unmarked.

In Savosavo, a syllable can either consist of only a vowel nucleus or a vowel nucleus and a consonant onset, i.e. the basic syllable structure is (C)V. Most roots consist of two or three syllables, but roots of four and more syllables also occur. Savosavo has been analysed as a stress language based primarily on auditory impression, with additional qualitative analyses in [1]. Generally, primary stress falls on the penultimate syllable of a root; initial syllables receive a secondary stress if primary stress is on a non-initial syllable. Impressionistically, stressed syllables have been associated with longer duration, clearer pronunciation, higher intensity and sometimes higher pitch. This paper aims to quantify such claims using data extracted from field recordings, however constrained the recording conditions. While lab speech may be desirable and advantageous for such analyses [2], we contend that field-based descriptions are still feasible. The Savosavo data demonstrates a consistency in acoustic correlates sufficient to corroborate its status as a lexical stress language. Its lexical prosody mirrors that of English, a well-studied stress language.

2. Methodology

2.1. Recording

The recordings of Savosavo being analysed are from Wegener (2007-2010), with one male native speaker serving as consultant. Recording took place in a secluded area of the village, though some background noise was unavoidable. The data is extracted from two elicitation sessions lasting about 35 minutes each, in which the subject was prompted by the linguist (using English) to provide a translation for various items, starting from a single word and increasing in morphological complexity e.g. a citation form for a verb, and then various inflections and some negative forms. 207 tokens were analysed in total.

2.2. Annotation and analysis

The initial transcription and annotation was done with ELAN [3] (Sloetjes & Wittenburg, 2008); the 207 selected tokens were converted into Praat [4] textgrid format. The assignment of the stressed syllable was based on the phonological stress assignment rules described above. Syllable boundaries were hand-labeled and glottal pulse data was generated and manually verified for missed or double marked vocal cycles in the wave form. The ProsodyPro script [5] was used to extract all the measurements including mean F0 (based on 10 evenly spaced F0 points from each labelled interval), max F0, duration, and mean intensity, and computed time-normalised F0 contours for each token. The resulting files from ProsodyPro were analysed to investigate the acoustic correlates of lexical prominence.

2.3. Analysis by synthesis

In order to verify the reliability of the stress assignment annotation scheme on which the above analysis is based, the

10.21437/SpeechProsody.2014-89

measurements were re-analysed with PENTAtrainer2 [6], a semiautomatic software package for the analysis and synthesis of speech melody, built upon the Parallel Encoding and Target Approximation model [7] also running on PRAAT [8]. In a preliminary phase, intervals were labeled according to the communicative function tested. The program then extracted the optimal parametric values for the tested communicative function through analysis by synthesis controlled by simulated annealing [6]. The tool assumes three model parameters controlling the F0 movement of each interval (here a syllable), including target slope (*m*), target height (*b*), and strength of target approximation (λ), where *m* and *b* specify the form of the pitch target and λ indicates how rapidly a pitch target is approached.

These results were used to generate F0 contours and compare them with the individual real utterances in the corpus.

In the current annotation scheme, two communicative functions were included, namely Stress condition (stressed (S) vs. unstressed syllable (U)) and Syllable location (Left, Medial, and Right edge of the utterance) as illustrated in Figure 1.

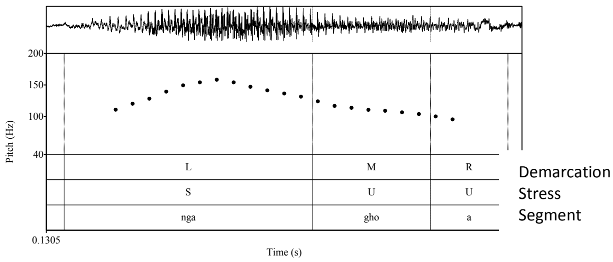

Figure 1. Example of functional annotation of recording for PENTAtrainer2, the top tier indicates the Syllable Location, the 2nd tier its stress condition and the 3rd tier the segmental content.

3. Results

3.1. Acoustic correlates of stress

The acoustic correlates of stress considered in this study are durational (ms.), intensity (dB) and mean F0 which is taken, for our purpose, as the direct correlate of pitch. The results shown in Table 1 indicate that stressed syllables have higher pitch with mean values of 112.06Hz to 96.36 Hz respectively; longer duration, with 212.91 ms to 176.09ms; and greater intensity, with 68.60dB to 64.29dB, than their counterparts. These results are statistically significant for all 3 correlates: F0 ($F=162.299$ (1, 1674) $p<0.001$; intensity ($F=48.933$ (1, 1696) $p<0.001$; duration ($F=146.349$ (1,1696) $p<0.001$.

Table 1. *Acoustic correlates of stress in Savosavo*.

Stress		Mean F0 (Hz)	Duration (ms)	Mean Intensity (dB)
S	Mean	112.06	212.91	68.60
	N	484	484	484
	Std. Deviation	23.50	56.39	9.91
U	Mean	96.36	176.09	64.29
	N	1192	1214	1214
	Std. Deviation	22.61	56.71	12.05

3.2. Analysis by synthesis

Functionally annotated data were fed into PENTAtrainer2, which computed the average parametric value of each interaction of the two functions (i.e. stress condition and syllable location). The results are shown in table 2. The learned values are compatible with our expectations that stressed syllables have a falling underlying target (shown by the negative *m* values) wherever they are located (left, medial or right periphery); these targets are higher than their unstressed counterparts (evident from the greater *b* values) when they are not word-initial. The anomalous Strength value for unstressed syllables at the right edge (100) should be disregarded, as it can be interpreted as a result of missing glottal pulses in the final syllable of some utterances, an artefact of field data commonly observed.

Table 2 *Mean target parameters of the Savosavo data learned through qTA modeling*.

Stress condition	Syllable Location (demarcation)	*Slope m*	*Height b*	*Strength λ*
S	Left	-42.32	1.74	20.15
	Medial	-52.09	-1.86	29.55
	Right	-29.56	4.89	5.34
U	Left	-0.5	2.39	36.86
	Medial	-5.8	-1.66	8.79
	Right	-26.84	-7.27	100

Subsequently, these averaged values were used to resynthesize the F0 contour of each utterance in the corpus. Mean RMSE (Root Mean Square Error is a frequently used measure of the difference between values predicted by a model and the values actually observed from the environment that is being modelled, since the RMSE is a good measure of accuracy, it is ideal if it is small) and Pearson's *r* (Correlation – often measured as a correlation coefficient – indicates the strength and direction of a linear relationship between two variables) of synthesis accuracy were respectively 2.48 and 0.87, these results are comparable to a study of English stress also using PENTAtainer2 [9] who report values of 3.97 ± 0.29 0.478 ± 0.028 for RMSE and R, respectively (for the encoding of the stress function only). The satisfactory accuracy of synthesis, even with only two functions annotated, demonstrates that the current annotation scheme is a reliable representation of the data, and that the qTA component embedded in PENTAtrainer is effective in modelling F0 contours.

4. Discussion

Some limitations due to the field-based nature of our data need be mentioned. First, the data is from only one speaker, and we are aware that in the context of a lab-based study, data from a lone speaker would not be deemed sufficient for a quantitative analysis; in our context it may indeed be a source of confounding factors such as speaker specific speaking style. Second, the male consultant in question uses a low mean F0 (100.7 Hz average across all utterances analysed), often close to his own pitch floor. As a result, some syllables were produced in non-modal phonation, hence did not contain glottal pulses. Third, where glottal pulses are not present, the duration of the syllable in question is determined by the annotator based on auditory

impression. The aforementioned limitations thus need to be taken into account when interpreting our results; nonetheless, we maintain that the acoustical correlates reported in the present paper serve as strong evidence to establish that Savosavo has stress, comparable to English, and unlike other languages, like Urdu, where lexical stress is marked by a lower F0 instead [10]. Further research will also investigate the acoustic difference between primary and secondary stress in Savosavo. Existing literature [1] has postulated the existence of secondary stress within the language, despite the claim of 'no or little [auditory] difference in realisation' between syllables carrying primary versus those with secondary stress. A possible avenue for future research would be to annotate for perceive primary and secondary stress and conduct descriptive statistics between these subgroups. An obvious limitation to this suggestion might be that the learning accuracy of the re-synthesis tool would be biased towards a system that distinguishes between three variables (Stressed, Unstressed and Secondarily Stressed) rather than one which uses two (as in this paper), rendering it incomparable with our above presented annotation scheme.
Finally, expanding from the word domain, continuing analyses will test how the encoding of lexical stress may interplay with sentential prosodic functions, such as modality and focus.

5. Conclusion

In this paper, we have provided quantitative evidence for the marking of lexical prominence in Savosavo through stress to complement previous qualitative analyses. We demonstrate that, even under adverse recording conditions, it is possible to carry out analyses using tools usually reserved for lab speech.
It is important to restate that this study forms the basis of further descriptive work on the prosodic system of Savosavo, so that after establishing the foundation for stress, we can define with more accuracy the nature of its syllable by analysing phonological processes such as syllable fusion. This description, in turn, will contribute to the ongoing areal research aiming to establish whether a historical change took place from a moraic to a syllabic system [11]. These findings will also be the basis for further investigations that will examine, for example, the interplay of word stress and sentence stress.
This paper argues for the usefulness of instrumental phonetic investigations in describing lesser-known languages. Our results demonstrate that the current annotation scheme is a reliable representation of the data, and that the qTA component embedded in PENTAtrainer2 is effective in modelling F0 contours. Finally, this paper also illustrates the usefulness of tools such as ProsodyPro for acoustic measurements, and PentaTrainer2 for hypothesis testing, tools that will eventually make a contribution towards our greater understanding of the characterization of the prosodic systems of the world's languages.

6. Acknowledgements

The authors would like to thank the Savosavo speakers for their continuing efforts to document their language and to make the recordings available for linguistic research. The present study is funded by DoBeS (Volkswagen Foundation) as part of the project 'Discourse and prosody across language family boundaries: two corpus based case studies on contact induced syntactic and prosodic convergence in the encoding of information structure'.

7. References

[1] Wegener, Claudia. A Grammar of Savosavo. Mouton Grammar Library, volume 61. 2012.
[2] Xu, Yi. In defense of lab speech. *Journal of Phonetics* 38: 329-336. 2000.
[3] Sloetjes, H., & Wittenburg, P. Annotation by category ELAN and ISO DCR. In Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008) (pp. 816–820). Marrakech, Morocco. 2008.
[4] Boersma, P. P. G., & Weenink, D. J. M. Praat: Doing phonetics by computer. Retrieved from http://www.praat.org/ on 2013/11/30. 2013
[5] Xu, Y. ProsodyPro: A tool for largescale systematic prosody analysis. In Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013) (pp. O1–1). Aixen Provence, France. 2013.
[6] Xu, Y. and Prom-on, S. (2014). Toward invariant functional representations of variable surface fundamental frequency contours: Synthesizing speech melody via model-based stochastic learning. *Speech Communication* 57, 181-208. 2014.
[7] Xu, Y. Speech melody as articulatorily implemented communicative functions. Speech Communication, 46(34), 220–251. 2005.
[8] Boersma, Paul & Weenink, David (2014). Praat: doing phonetics by computer [Computer program]. Version 5.3.66, retrieved 9 March 2014 from http://www.praat.org/.
[9] Liu, F., Xu, Y., Promon, S., & Yu, A. C. L. Morpheme like prosodic functions: Evidence from acoustic analysis and computational modeling. Journal of Speech Sciences, 3(1), 85–140. 2013.
[10] Hussain, S. Phonetic correlates of lexical stress in Urdu. PhD Thesis. Northwestern University, Evanston, IL. 1997.
[11] Palmer, Bill. Shifting stress. Shifting stress: synchronic variation as a manifestation of diachronic change in Kokota. (Oceanic) prosody LAGB spring 2003.