Lyrics as entrained musical and linguistic oscillations: acoustic evidence in Estonian *regilaul*

by

sally ransom, B.A., Linguistics

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF ARTS

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2022

I think that when linguists discuss and dispute sound length and stress among themselves they would definitely benefit from inviting ethno-musicologists to join them. They could discuss the issues together, and not only based on written records but also sung records.

That what is written is fiction. Only that what is sung is truth.

Tormis (2007)

Lyrics as entrained musical and linguistic oscillations: acoustic evidence in Estonian *regilaul*

sally ransom, M.A.
The University of Texas at Austin, 2022

Supervisors: Scott Myers
Katrin Erk

This paper presents an empirical analysis of acoustic-phonetic measurements of sung Estonian lyrics using an oscillator-entrainment model. Entrainment is the process by which two independently oscillating systems with differing periods interact and assume the same period. Measurements were extracted from *a capella* recordings of traditional Estonian folksongs called *regilaul* $/re.ki.laul/$, which is part of the ancient indigenous oral tradition of Balto-Finnic *runosong*. Verse lines are composed using trochaic tetrameter, or four measures of syllable-note pairs with the downbeat corresponding to the first syllable-note. In nominally isochronous notes, the absolute durations of notes falling off the beat are shorter compared with their counterparts on the beat. Conversely, Estonian demonstrates foot isochrony such that the duration of the second syllable is inversely proportional to that of the first. Additionally, an increase in syllable duration is an acoustic correlate of primary stress in native Estonian words, which is fixed word-initially. This results in syllable-notes grouped across word boundaries in the song. Previous studies of regilaul found the absolute durations of syllable-notes patterned into two categories based on musical grouping rather than word grouping. In light of these results, I further investigate the strength of the entrainment of musical and linguistic rhythms at the sub-syllabic level, the alternation of consonant and sonorant phoneme segments. Results show that word-level stress and unstress is evident in the vowel durations of syllable-notes off the musical beat, but not apparent in syllable-notes that are on the beat. Therefore, when musical trochees cross word boundaries, word-initial syllables are acoustically distinct from pen-initial syllables.

Chapter 1

Methods

1.1 Design

To test this hypothesis, the vowel durations of isochronous (eighth-note) word-initial and pen-initial syllables of polysyllabic feet are measured. This provides a level of control with regard to feet and word boundaries: no monosyllabic words are used, so every primary stressed syllable included has a following unstressed syllable. This also beneficially excludes possible complications due to secondary stress, which is not only weaker but also acoustically less salient Asu & Lippus (2018) compared to the opposition of primary and unstress in the first two syllables.

1.2 Constructing the Corpus

1.2.1 Materials

Songs for this paper were accessed via The Anthology of Estonian Traditional Music (Tampere, 2016), which includes audio recordings with transcriptions of the lyrics.

The regilaul sample in this paper is a subset of the entire anthology. To control for regional differences (**?**), songs were all from Pärnumaa county. In order to account for diachronic variation (**?**) as well as possible differences due to recording equipment used, all songs came from the same 5 year span from 1961-1966 and were recorded on behalf of the Estonian Folklore Archives by Herbert Tampere, Erna Tampere, and Ottilie Kõiva for the Estonian Folklore Archives(Oras & Västrik, 2002; Tampere, 2016).

1.2.2 Annotating the Song Audio

Each song's lyrics are copied from the site and saved as .txt files in Estonian orthography, each line of the file corresponding to one melody line. Audio files of the selected songs are downloaded from the archive in .ogg format, which is the highest resolution of the two lossy formats available from the digital anthology. Each song is then imported into a Logic Pro X (Cousins & Hepworth-Sawyer, 2014) session for preprocessing. Since a main predictor in the hypothesis refers to the notes with relation to the beat, it is necessary to annotate the location of beats for each measure in the song while accomodating the fine-grained variation in absolute duration and tempo inherent in natural meter. Beat onset detection in general works by analyzing the signal's acoustic dimensions with relation to the downbeat (Robertson & Plumbley, 2007). Once the tempo map is aligned with the downbeat of each measure, Musical Instrument Digital Interface (MIDI) is used to synthesize a metronome to indicate beat location and duration with respect to the tempo variation in each measure. This metronome is converted into a PRAAT(Boersna & Weenink, 2022) TextGrid with an interval tier indicating the onset and offset of the metronome's notes, which correspond to syllable-notes falling "on" the beat.

Lyrical transcriptions of each verse line is added to another interval tier, with each line corresponding to four measures indicated by the click tier. Here, the eSpeak forced aligner for Estonian (Duddington et al., 1995) attempts to align the transcription from orthographic verse line to two tiers: one with phonemic transcriptions of words, and one with individual phonemes. Due to its inherent design for speech, the aligner frequently needed manual adjustment of the word boundaries before reliable phonemic alignment could be produced.

A subset of vowel intervals from initial and pen-initial syllables of polysyllabic words is the data in this study.

1.2.3 Aggregating Acoustic and Text data

The last step in preprocessing is to integrate the audio annotation with the lexical content of the song lyrics. This task was accomplished using an open-source natural language toolkit in python called estinltk `https://github.com/estnltk` (Laur et al., 2020). For each vowel interval included in the study, the vowel's word context from the corresponding TextGrid tier is input into estnltk vabamorf syllabifier, which returns the word broken into individual syllables with corresponding prominence data for each syllable. Vowel duration is measured by the difference between the offset and the onset of the interval on the vowel's TextGrid tier. The estnltk data and

PRAAT measurements are aggregated using the parselmouth library(Jadoul et al., 2018; Van Rossum & Drake Jr, 1995). Jupyter notebooks illustrating each step are available at `https://github.com/sally-ran-some/eesti_regilaul`.

1.3 Statistical Analysis

For each hypothesis, linear mixed-effects models were fit using the lme4 package in R (Bates et al., 2015; R Core Team, 2022). All the models, random slopes and intercepts, comparisons, etc.

Anova comparison of the maximal design model with a null model is also statistically significant ($p < 0.001$***).

Chapter 2

Results

The presence of a coda in comparison to an open syllable-note necessitates a shorter vowel to fit the consonant into the note duration.

2.1 Stress and Unstress

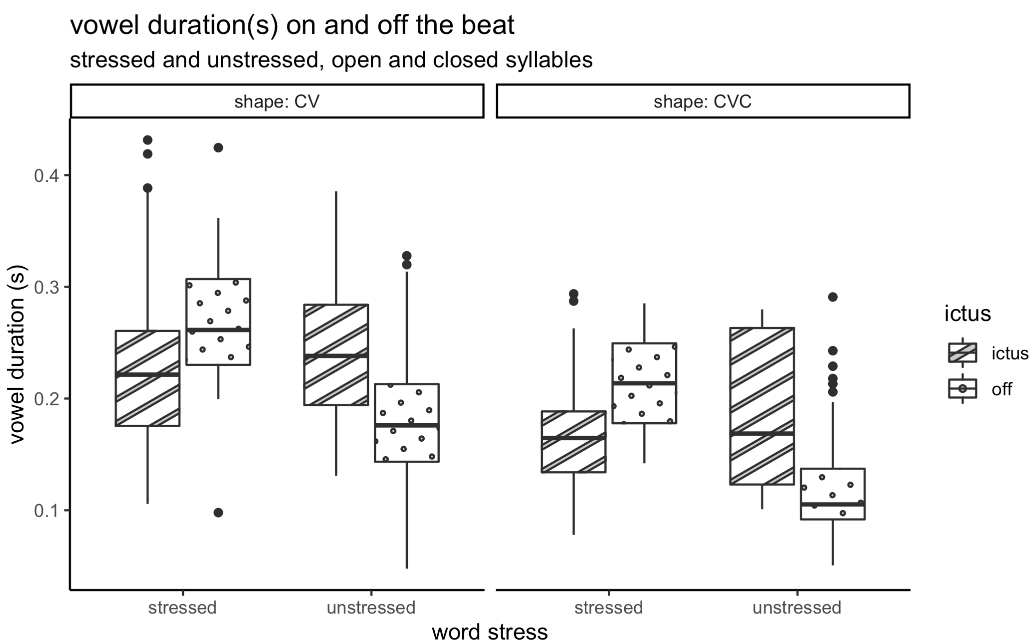

Figure 2.1: vowel durations of CV and CVC syllables in stressed and unstressed positions falling on and off the beat

The two graphs in 2.1 illustrate the overall pattern of interaction between musical meter and word stress. When syllables are on the beat, the vowel duration normally assoficated with differences in word stress are not apparent. Off the beat, however, stress and unstress remain distinct. This pattern is the same in both open (CV) and closed (CVC) syllable shapes, with overall vowel duration being shorter in closed syllables.

	duration		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.22	0.19 – 0.26	**<0.001**
ictus [off]	-0.04	-0.05 – -0.02	**<0.001**
stressed [stressed]	0.01	-0.01 – 0.03	0.247
shape [CVC]	-0.05	-0.06 – -0.04	**<0.001**
ictus [off] * stressed [stressed]	0.04	0.01 – 0.07	**0.014**
Random Effects			
σ^2	0.0010		
$\tau_{00\ \mathrm{word}}$	0.0007		
$\tau_{00\ \mathrm{syll}}$	0.0006		
$\tau_{00\ \mathrm{song}}$	0.0014		
$\tau_{00\ \mathrm{performer}}$	0.0004		
ICC	0.7429		
N_{song}	9		
N_{word}	289		
N_{syll}	157		
$N_{\mathrm{performer}}$	3		
Observations	512		
Marginal R^2 / Conditional R^2	0.187 / 0.791		

Figure 2.2: stress and ictus duration model output

Bibliography

Asu, E. L., & Lippus, P. (2018). Acoustic correlates of secondary stress in Estonian. In *Speech Prosody 2018*, (pp. 602–606). ISCA.

Asu, E. L., & Teras, P. (2009). Estonian. *Journal of the International Phonetic Association*, *39*(3), 367–372.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Boersna, P., & Weenink, D. (2022). Praat: Doing Phonetics by Computer.

Cousins, M., & Hepworth-Sawyer, R. (2014). Logic pro X.

Duddington, J., Avison, M., Dunn, R., & Vitolins, V. (1995). eSpeak: Speech Synthesizer.

Eek, A., & Meister, E. (1998). Quality of standard Estonian vowels in stressed and unstressed syllables of the feet in three distinctive Quantity degrees. In *Proceedings of the Finnic Phonetics Symposium,* Linguistica Uralica. Tallinn.

Essens, P., & Povel, D.-J. (1985). Metrical and Nonmetrical representations of temporal patterns". *Perception & Psychophysics*, *37*, 1–7.

Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, *71*, 1–15.

Laur, S., Orasmaa, S., Särg, D., & Tammo, P. (2020). EstNLTK 1.6: Remastered estonian NLP pipeline. In *Proceedings of the 12th Language Resources and Evaluation Conference*, (pp. 7154–7162). Marseille, France: European Language Resources Association.

Lehiste, I. (1960). Segmental and syllabic quantity in Estonian. *American studies in Uralic linguistics*, *1*, 21–82.

Lehiste, I. (1965). The Function of Quantity in Finnish and Estonian. *Language*, *41*(3), 447.

Lehiste, I. (1978). The Syllable as a Structural Unit in Estonian. In *Syllables and Segments*. North-Holland Publishing Company.

Lehiste, I. (1992). The Phonetics of Metrics. *Empirical Studies of the Arts*, *10*(2), 95–120.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle, & A. Marchal (Eds.) *Speech Production and Speech Modelling*, (pp. 403–439). Dordrecht: Springer Netherlands.

Lippus, P., Asu, E. L., & Mari, M.-L. K. (2014). An acoustic study of Estonian word stress. In *Speech Prosody 2014*, (pp. 232–235). ISCA.

Lotman, M.-K., & Lotman, M. (2013). The Quantitative structure of Estonian syllabic-accentual trochaic tetrameter. *TRAMES*, *3*, 243–272.

Oras, J. (2019). Individual rhythmic variation in oral poetry: the runosong performances of Seto singers. *Open Access Linguistics*.

Oras, J., & Västrik, E.-H. (2002). Estonian Folklore Archives of the Estonian Literary Museum. *The World of Music*, *44*(3), 153–156.

Palmer, C., & Kelly, M. H. (1992). Linguistic Prosody and Musical Meter in Song. *Journal of Memory and Language*, *31*(4), 525–542.

Park, H. (2009). Linguistic experience's influence on foreign accent detection in short, slightly accented speech. *The Journal of the Acoustical Society of America*, *125*(4), 2756–2756.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL `https://www.R-project.org/`

Robertson, A., & Plumbley, M. (2007). B-Keeper: A beat-tracker for live performance. In *Proceedings of the 7th International Conference on New Interfaces for Musical Expression - NIME '07*, (p. 234). New York, New York: ACM Press.

Ross, J. (1989). A study of timing in an Estonian runic song. *The Journal of the Acoustical Society of America*, *86*(5), 1671–1677.

Ross, J. (1992). Formant frequencies in Estonian folk singing. *Journal of the Acoustical Society of America*.

Ross, J., & Lehiste, I. (1994). Lost Prosodic Oppositions: A Study of Contrastive Duration in Estonian Funeral Laments. *Language and Speech*, *37*(4), 407–424.

Ross, J., & Lehiste, I. (1996). Trade-off between quantity and stress in Estonian folksong performance? *Folklore: Electronic Journal of Folklore*, *02*, 116–123.

Ross, J., & Lehiste, I. (1998). Timing in Estonian Folk Songs as Interaction between Speech Prosody, Meter, and Musical Rhythm. *Music Perception*, *15*(4), 319–333.

Ross, J., & Lehiste, I. (2001). The temporal structure of Estonian runic songs / by Jaan Ross, Ilse Lehiste. In *The Temporal Structure of Estonian Runic Songs*, Phonology and Phonetics ; 1. Berlin, [Germany] ;: Mouton de Gruyter, reprint 2015 ed.

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, *118*(3 Pt 1), 1677–1688.

Tampere, H. (1934). Mõningaid mõtteid Eesti rahvaviisist ja selle uurimismeetodist. *Eesti Muusika Almanak I*, (pp. 30–38).

Tampere, H. (2016). Anthology of Estonian Traditional Music.

Tormis, V. (1985). Kalevala – the Estonian perspective. *Finnish Music Quarterly*.

Tormis, V. (2007). Some problems with that *regilaul*. In *proceedings of the international RING conference*.

Van Rossum, G., & Drake Jr, F. L. (1995). *Python Reference Manual*. Centrum voor Wiskunde en Informatica Amsterdam.