becoming lyrics: how word prosody and musical meter negotiate the rhythmic terms of prominence

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

becoming lyrics: how word prosody and musical meter negotiate the rhythmic terms of prominence

APPROVED BY

SUPERVISING COMMITTEE:

Scott Myers, Supervisor

Katrin Erk, Supervisor

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)

becoming lyrics: how word prosody and musical meter negotiate the rhythmic terms of prominence

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

Supervisors: Scott Myers
Katrin Erk

This paper presents a microprosodic analysis of fine-grained acoustic-phonetic measurements extracted from recordings of traditional Estonian lyrical *a capella* folksongs called ***regilaul*** ($/re.ki.laul/$). The song verses are set in a trochaic tetrameter, which classically calls for alternations of two (long-short) syllables. The Estonian language itself has three contrastive syllable lengths: short, long, and overlong. Previous studies of regilaul found that the absolute durations of nominally isochronous syllable-notes patterned into two, not three categories: short and long, concluding that semantically relevant duration cues were lost in song. However, I dispute this conclusion in light of their findings: if syllables with more segments (CVCC) are fitting into the same note constituent as those with fewer (CVC, CV), the absolute durations of the individual segments would reduce as the number of segments increased. The absolute duration of syllable nuclei falling on

and off the beat is compared in all three syllable quantities and both stressed and unstressed positions. Results show a collaborative interaction between musical meter and word prosody. The ternary quantity contrast is upheld by a gradient decrease in vowel duration. Likewise, stressed and unstressed syllables maintain their linguistic opposition *relative* to their position in the song. Thus, when the syllables and notes entrain to each other syllable-notes in the metrical long-short pattern of the song, temporal signatures of linguistic rhythm are preserved in the relative duration of their segments.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The entrainment of two independent oscillators (in this case, the musical meter and the prosody of the language) are better described as a collaboration than as a hierarchy where one system dominates the other. By the biological principle of economy, singers of their own language would utilize the highly specialized cognitive systems already developed for the meaningful purposes in their language.

In the same way that one cannot ever un-have their L1 accent, and even thinks their L2 with an accent (Park, 2009), fine-grained acoustic-phonetic cues important in the singer's spoken language would remain in the "interlanguage" that is the song.

How do the word-prosodic requirements negotiate with the imposed prosodic hierarchy of music?

The rhythmic organization of song is said to integrate the prosodic structure of the language with musical rhythmic principles (Palmer & Kelly, 1992).

The intuitions of those who have studied the runosong tradition is that the burden of upholding the temporal structure of the song is the result of symbiosis between the musical rhythm and the natural prosodic features of the lyrical text (Ross, 1992; Tampere, 1934): the song's melody a musical abstraction of the natural prosody of spoken runic verse.

1.1 When words become lyrics

To join song and become lyrics, meaningful linguistic material must be modified to fit the strict temporal structure of music. Consequently, both poet and performer must combine two independent rhythmic systems: the rhythm of the language must fit into the song's rhythm, but enough of the language's own rhythm must remain in order for the lyrics to have meaning.

In particular, the rhythmic features of Estonian prosody include a ternary syllable weight contrast that interacts closely with word stress (Lehiste, 1960, 1965, 1978; Eek & Meister, 1998; Asu & Teras, 2009). Duration functions at several levels of Estonian prosody: it is the strongest correlate of both clear speech and stress (Lippus et al., 2014) and is independently contrastive at the segmental level. In primary stressed syllables, this segmental contrast plays a role in the phonetic realization of the ternary weight contrast illustrated in the minimal triads in 1.1.

To examine the effects of musical metrical structure on grammatical rhythmic components of uttered language, I measure temporal features of sung syllable nuclei in recorded performances of traditional Estonian folksongs known as *regilaul*. Regilaul poets and songwriters use both long and overlong stressed syllables in the long position of a trochaic foot, and short-stressed syllables weak positions of the trochaic foot (Lotman & Lotman, 2013).

Q1	sada *'hundred'*	kabi *'hoof'*
Q2	saada *'send'*	kapi *'of the cupboard'*
Q3	saada *'recieve'*	kappi *'into the cupboard'*

Table 1.1: ternary syllable weight contrast

As can be seen in the examples in 1.1, the length of a given segment can indicate lexical contrasts as in sada vs saada ('hundred' vs 'send'), or differentiate case. In *kapi* and *kappi*, marking 'of' versus ''into' the cupboard is indicated by the quantity of the first syllable.

1.1.1 Phonetics of Estonian Word Prosody

Primary word stress in native Estonian words is fixed, falling on word-initial syllables. By virtue of its predictability, it is not lexically contrastive: instead, it is described as identificational, facilitating intelligibility by demarcating prosodic boundaries (Lehiste, 1965, 1978, 1992; Eek & Meister, 1998; Lippus et al., 2014). However, it is only in primary stressed syllables that the ternary quantity contrast falls.

(1) laul-da
[ˈlɑuːl.dɑ]

sing-TR

'singing'

(2) ööbik
['øː.pikː]
nightingale.NOM

'nightingale'

Q1 and Q2 syllables can be both stressed and unstressed, while Q3 is only present in stressed positions, attracting stress to its (non-initial) syllable in compound and loan words. Pen-initial syllables can only be Q1 or Q2, illustrated in 2.

Q1	Q2	Q3
kodi	*koodi*	*koodi*
/ko.ti/	/koː.ti/	/koːː.ti/
	koti	*kotti*
	/kot.ti/	/kotː.ti/
	gooti	*kooti*
	/koːt.ti/	/koːːt.ti/

Table 1.2: segmental permutations of initial syllable quantity

In the first row of 1.2, we see a minimal triad of the ternary quantity contrast in open 'short' first syllables (Q1). The Q2 'long' and Q3 'overlong' columns demonstrate all the other ways this contrast can be realized using the same segment identities in closed syllables. These three syllable weights can be lexically contrastive, differentiating semantically distinct roots, or morphologically distinguishing varying degrees of grammatical case, direction, aspect, &c.

1.1.2 Metrical Principles of Estonian folksong

I define metrical as the mapping of the pattern on a frame formed of equal time intervals. These patterns have been demonstrated to be more easily replicated by humans (Essens & Povel, 1985), necessary for the transmission of an oral tradition of songs and for the synchronization of human musicians playing together.

Figure 1.1: Song 45: *Millal saame sinna maale*

Regilaul is part of the Finnic runosong tradition shared by several other members of the Finnic language family: Finnish, Karelian, Votic, Ingrian, and Livonian (Ross & Lehiste, 2001). These "singable songs" (Tormis, 1985) follow a metrical pattern such that a given *regilaul* text can be sung to any of the numerous *regilaul* melodies (Ross & Lehiste, 2001). The metrical basis of the tradition is a trochaic tetrameter often referred to as the Kalevala meter (Oras, 2019), which is realized in Estonian 20th century work as syllabic-accentual trochaic tetrameter (Lotman & Lotman, 2013). This pattern opposes metrically strong and weak positions by means of syllable quantity and word stress. Ictus position, which corresponds to 'on the beat' in musical meter, prefers syllables that are both heavy (long and overlong) and stressed, but avoids short stressed syllables. Short stressed syllables can occur off the beat, but this position is avoided by the heaviest (Q3) syllables.

Composed of four of these trochees, each verse line has four beats with eight syllable-note positions (as in 1.1). In $\frac{4}{4}$ measure, the beat (ictus) falls on the note corresponding to the first beat of a measure, and on every other following syllable-note, counted as one (and) two (and) three (and) four (and). The two measures in 1.2 illustrate a melody variation with eight syllables fit into seven positions: in each measure, two weak syllables are halved, with the trisyllabic dactyl having one eighth and two sixteenth syllable-notes. In the first measure, the "extra" position gained from this variation is occupied by the last (heavy) syllable, which is extended to fill a quarter note. In the second measure, the syllable is sung as an eighth note, and the extra eighth position filled with a rest (𝄾).

Figure 1.2: Song 77: *Loomine*

1.2 Previous Studies with *regilaul*

Musicologist Jaan Ross and Phonologist Ilse Lehiste studied the interactions of musical and prosodic meter, comparing nominal transcriptions of syllable-notes to their absolute durations. The findings overwhelmingly were that in nominally isochronous syllable-notes, the absolute durations patterned into two duration categories: short and long. Compared to the inherent durational properties found in Estonian word prosody (primary lexical stress and unstress, ternary syllable quantity), long and short syllable-note category was best predicted by its position relative to the beat: long notes on the beat, short ones off (Ross & Lehiste, 1994, 1996, 1998). They concluded from these findings that semantically relevant duration contrasts ordinarily present in spoken Estonian were "lost" in song, and interpretation of lyrical content must rely heavily on "top-down" processes such as semantic context. Later, they summarize and revisit the question, this time concluding that regilaul folksongs must have originated before Estonian developed its ternary contrast (Ross & Lehiste, 2001). However, a singing tradition is not a time machine. Although it is likely that regilaul developed before the innovation of the ternary quantity, there would be no evidence of that in the songs of singers whose grammar uses the ternary quantity.

1.3 The present study

This study examines the ternary quantity distinction in the context of its syllable shape: no coda (CV), single coda (CVC, CVVC), and complex coda (CVCC, CVCCC) rather than collapsing all syllable shapes according to their quantity. This allows a closer look at the microprosodic features at the segmental level.

Previous studies have compared acoustic measurements with the nominal values given in transcriptions made by ethnomusicologists. However, as regilaul is an oral tradition, these annotation represent neither the intuitions of the songs' composers, nor those of the performers. Here I introduce the use of beat-tracking algorithms to define beat location by acoustic means.

1.3.1 Hypotheses

Earlier studies of *regilaul* explored temporal aspects of syllable-notes and found that duration characteristics that would usually indicate important semantic differences lost their distinctions either partially or entirely. The present study wishes to extend these findings. It is likely that apparent neutralization of semantically relevant quantity contrasts at the syllabic level would remain present, and potentially be enhanced, at the segmental level.

The two syllable-note lengths correspond to the metrical pattern of the music, i.e., notes on the beat are predictably longer than notes off.

Lemma 1. *In sung Estonian regilaul, three lexically contrasting syllable lengths fit into nominally isochronous eighth notes have absolute durations patterning as two distinct syllable-note lengths (Ross & Lehiste, 2001).*

The following corollary is a natural consequence of 1:

Corollary 2. *For a syllable with a geminate coda to fit into the same eighth note as one with a single coda consonant, the syllable nucleus (vowel) will have to be shortened to fit the entire complex segment in the note.*

If the findings of (Ross & Lehiste, 2001) are assumed to be true as in 1, the corrolary in **??** should also be true. In either beat position, the segmental durations that are inherent in the three syllable weights will be evident in the duration of the vowel

I thus offer the following hypothesis:

Hypothesis 1. *segmental contrasts inherent to syllabic quantity will be evident in the vowel duration, with vowel duration decreasing in response to the duration of the coda.*

The predictions with word stress are less clear. For one thing, since it is not used for lexical contrast in Estonian, there is a possibility under the hierarchical perspective that this would be a candidate of word-level rhythm to be "lost" to the meter of the song. However, I maintain that the stress pattern of the singers' grammar is inalienable from the grammar as a whole. That is, even if it is not a necessary cue for disambiguating lexemes of different meaning, it would be more difficult for a singer of a language to neutralize the word-level stress pattern than it would be to preserve it. Thus, I hypothesize that the duration of stressed and unstressed syllable nuclei will demonstrate gradient acoustic-phonetic differences in duration within the context of the duration category corresponding to their beat position.

Hypothesis 2. *Stressed syllables in a given beat position will have longer absolute vowel durations than unstressed syllables in the same beat position.*

There would then be nested categories of absolute duration:

$$[long(stressed > unstressed) > short(stressed > unstressed)]$$

Chapter 2

Methods

2.1 Design

To test these hypotheses, vowel duration from the nuclei of first (primary stressed) and unstressed syllables of disyllabic feet. [1]

Within each song, samples of syllable-notes consisted only of those matching the nominal isochrony. For example, in verse lines mostly comprised of syllable-notes divided evenly into eighths, neither quarter notes nor sixteenth notes were taken. Syllables in phrase-final position were generally excluded under this criteria.

2.2 Constructing the Corpus

I first describe the source materials and the selection criteria for the sample corpus of *regilaul* folksongs. Following this, the annotation and measurement procedure is detailed. Then the procedure for assembling the corpus of songs and their text annotations is covered before proceeding to the inclusion criteria for vowel duration and dispersion measurements.

[1] The exclusion of secondary stress is due to recent findings that secondary stress in spoken Estonian does not exhibit significant differences compared to unstressed syllables of polysyllabic words besides the pen-initial Asu & Lippus (2018).

2.2.1 Materials

Songs for this paper were accessed via The Anthology of Estonian Traditional Music (Tampere, 2016). Originally published on four vinyl discs in 1970, the digital version showcases a robust sample of the massive collection of *regilaul* in Estonian Folklore Archives. In addition to audio, the compilation includes photographs, sheet music, and performer demographics of 98 *regilaul* songs and 17 instrumental tunes.

For this paper, I took a subset of the available songs based on shared characteristics such as region, time of field recording, and ethnomusicologists. Once several regions were identified as possible candidates, a native Estonian speaker was consulted on the final selection. The nine songs analyzed in this study were all recorded in Parnümaa county from 1961-1966 by Herbert Tampere, Erna Tampere, and Ottilie Kõiva for the Estonian Folklore Archives(Oras & Västrik, 2002; Tampere, 2016).

2.2.2 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 beat detection, tempo mapping, and trimming. To make the tempo map, the session is set to *flex tempo*. From here a beat onset detection algorithm is given the estimated bpm and time signature as a starting point and then run on the song audio.

From here, a MIDI track is programmed to create a metronome specifically tailored to the song according to the tempo map. Using the aforementioned *flex tempo*, the MIDI track adjusts note and measure length to match the fluctuations documented in the by the beat tracker. Once rendered, the metronome and the song audio file are trimmed to match exactly, and the metronome is converted with a script into a textgrid of beat and measure intervals in PRAAT(Boersna & Weenink, 2022)[2].

The texts of verse lines typeset in Estonian orthography are inserted into each verse-phrase interval (already annotated thanks to the metronome). On each verse line, the eSpeak forced aligner for Estonian (Duddington et al., 1995) to reverse-synthesize from the phrase input to the word and phonemic level. Because this forced aligner is trained on spoken, not sung Estonian, manual verification and realignment of the vowel segments used in this dataset was often necessary.

[2]Using onset detection algorithms such as these (Robertson & Plumbley, 2007) in phonetics research, especially in the interdisciplinary field of linguistics and musicology, will be particularly beneficial to answering questions about rhythm: finding a way to bring our intuitions and impressions about "the beat" together with the acoustic phenomenon. By automating the annotation and measurement process using open source tools, the author hopes to share these machines with those who have similar research interests, and also to invite contributors to the data of this corpus of text data time-aligned to queryable audio signal data.

2.2.3 Adjustment Criteria for Vowel Durations

The beginning of the vowel was aligned according to a combination of acoustic correlates:

- vowel considerations
  - **intensity(dB) contours**: intensity within 2dB of the steady-state medial portion of the vowel, with a positive slope less than one.
  - **fundamental frequency(Hz)** stabilizing into that syllable-note's pitch category
  - **resonant frequencies F1, F2, and F3**: visible and approaching steady-state, combined wtih the presence of a voicing bar.
- onset manner particulars:
  - following **plosive onsets**, vowel boundary is after a burst
  - following **fricative onsets**, the end of visible high-frequency noise in the spectrum was reliable. The phoneme /s/ also consistently showed a carat in the frequency track immediately preceding the transition to vowel.
  - boundaries between **approximant onsets** and vowel nuclei were determined chiefly by the steady state of formants.
  - following **nasal onsets**, vowel intensity actually lowered, but a negative slope approaching zero reliably followed the offset of visible anti-formants.

- The offsets of vowels into codas:
  - negative slopes approaching zero for intensity; positive preceding nasal.
  - Formants were allowed more variation in transitions to codas
  - approximant codas [l, j, w] were excluded from the dataset, as neither the forced-aligner nor the phonetician could determine a reliable way to define the boundary between them.

In cases of vowel adjacency across syllable boundaries, the presence of a visible glottal stop and support from other criteria would qualify both for inclusion. In the absence of these cues, both nuclei were excluded from the measurements. Other exclusions were due to ambient noise (i.e., churchbells in song 41), ambiguity of word boundaries due to wordplay (consulted with native speaker), and cases of adjacent identical vowels across word boundaries.

In all cases, if the aforementioned cues were unavailable, ambiguous, or misaligned, the token was elided for this analysis. From all nine songs in the corpus, a total of 757 vowel nuclei met the criteria for inclusion in duration measurements.

At this point, the audio recording of each song has corresponding textgrid tiers annotated for beat, verse lines, and force-aligned word and phoneme levels.

2.2.4 Connecting acoustic measurements to text corpus

The last step in preprocessing is to integrate the annotation of the song audio with the lexical content of the song. This study accomplishes the task using an open-source natural language toolkit in python called estinltk `https://github.com/estnltk` (Laur et al., 2020). Among other things, the toolkit has a robust dictionary of Estonian grammar, including phonetic transcription of syllables with corresponding quantity and stress data. This is especially important data for testing hypotheses about the ternary quantity contrast, which is not always apparent from the orthography.

Once the annotations are complete, python scripts are used aggregate the audio and text that form this microcorpus: text files of lyrics are connected to the textgrid files corresponding to the audio, and corresponding duration measurements are extracted from PRAAT using the parselmouth library(Jadoul et al., 2018; Van Rossum & Drake Jr, 1995). The process of aggregating the acoustical and textual data for this corpus is available in the two jupyter notebooks in folder preprocess$_j$upyter at `https://github.com/sally-ran-some/eesti_regilaul` and also viewable in the appendices.

2.3 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). Due to the asymmetry of distribution of the three quantities in Estonian (Q3 is never unstressed), the analysis of vowel duration is broken into two. The first analysis concerns the semantically-relevant lexical contrasts of the three Estonian quantities entrained to the binary long-short pattern of the musical meter. A subset of only primary stressed syllables is used to build this model.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

To examine the effects of word stress on vowel duration, a subset of data excluding Q3 syllables was taken. While retaining Q1 and Q2, the interaction of syllable quantity and word stress will be apparent, as both Q1 and Q2 fall in stressed and unstressed positions.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Chapter 3

Results

	duration		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.30	0.25 – 0.36	**<0.001**
quantity [2]	-0.09	-0.12 – -0.07	**<0.001**
quantity [3]	-0.08	-0.11 – -0.05	**<0.001**
ictus [off]	-0.05	-0.08 – -0.03	**<0.001**
quantity [2] * ictus [off]	0.03	-0.00 – 0.07	0.084
quantity [3] * ictus [off]	-0.01	-0.05 – 0.03	0.631
Random Effects			
σ^2	0.0011		
$\tau_{00\ word}$	0.0008		
$\tau_{00\ syll}$	0.0004		
$\tau_{00\ shape}$	0.0035		
$\tau_{00\ song}$	0.0019		
$\tau_{00\ performer}$	0.0005		
ICC	0.8689		
N_{song}	9		
N_{syll}	191		
N_{shape}	13		
N_{word}	298		
$N_{performer}$	3		
Observations	367		
Marginal R^2 / Conditional R^2	0.162 / 0.890		

Figure 3.1: quantity model

The presence of a coda in comparison to an open syllable-note necessitates a shorter vowel to fit the consonant into the note duration. Complex and geminate codas shorten the nucleus even further. Likewise, primary stress and unstressed syllables maintain their durational differences within the context of their position relative to the beat. That is, while on the beat (long position) vowel durations are longer in general than off the beat (short positions), a stressed syllable on the beat has a longer vowel duration than an unstressed syllable in its same position, and in short positions, stressed syllables have longer vowels than unstressed. Thus, when the syllables and notes entrain to each other syllable-notes in the metrical long-short pattern of the song, the segmental contrasts are preserved for both semantically relevant (quantity) and prosodic boundary (word stress) levels.

3.1 shows the full output of the linear mixed effects model.

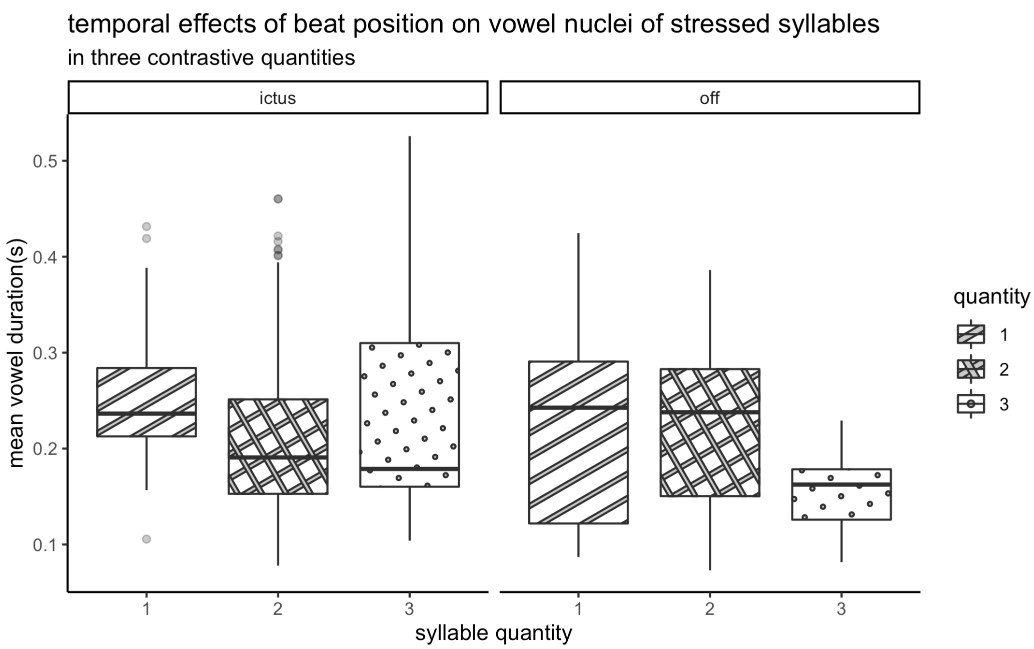

Figure 3.2: density plot of vowel durations in three syllable quantities

3.2 shows the vowel durations of all three syllable quantities, grouped by ictus and off-ictus positions in the song. In ictus position, median vowel duration descends as quantity increases, with the greatest difference between Q1 and Q2. Off the beat, a similar descending pattern is evident: however, in this case the largest difference is between Q2 and Q3 syllables. The intercept is set at ictus position, Q1. Findings are significant results for Q2($p<0.001$), Q3, and off-ictus positions ($p< 0.05$). Comparison with null model was statistically significant ($p<0.001$). For full model output see **??** and **??** in Appendix A. In context of earlier findings that the quantity contrast was "lost" at the syllable level, the decrease in vowel duration as syllable weight increases supports the notion that the contrast is preserved at the segmental level. That is, rather than the full syllable lengthening in duration, the song-level isochrony of syllable-notes results in vowel nuclei shortening to accommodate codas in Q2, and further for geminates and complex codas in Q3.

A null model constructed containing only random effects was compared to the design model by two-way ANOVA. Results are significant for the design model ($p < 0.001$***).

3.1 Stress and Unstress

The two graphs in 3.3 illustrate the distribution of vowel durations in stressed and unstressed syllables falling on and off the beat. In Q1 syllables, ictus position predicts longer vowels in both stressed and unstressed syllables, while stressed syllables are longer overall than unstressed. In Q2, we see longer vowel durations for ictus position in stressed syllables, and higher means for ictus position in unstressed, though the distributions overlap much more here.

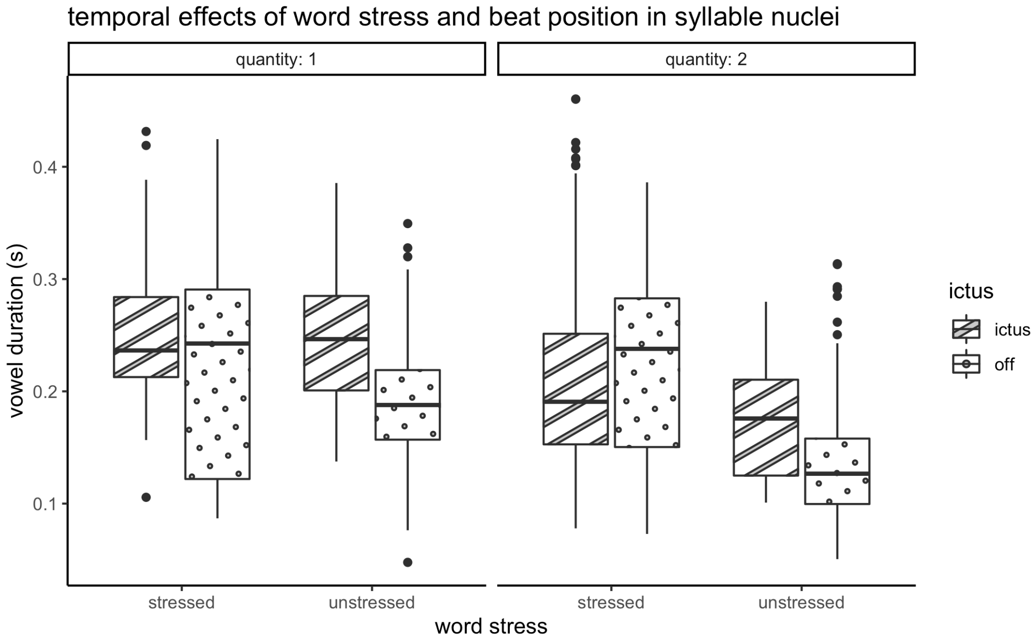

Figure 3.3: vowel durations of stressed and unstressed Q1 and Q2 syllables falling on (ictus) and off the beat

Linear mixed-effects model results are significant for off-ictus ($p < 0.05^{**}$), stressed ($p < 0.001^{***}$), and Q2 ($p < 0.001^{***}$).

Compared to Q1 unstressed syllables in ictus position (the intercept, off-ictus positions have a negative slope and are overall shorter. Stressed syllables have a small positive slope, indicating longer vowel durations. Q2 syllables have a negative slope, highlighting the shortening of syllable nuclei to accomodate the codas of these syllables.

Anova comparison of the maximal design model with a null model is also statistically significant ($p < 0.001^{***}$). I reject the null hypothesis: these results support both word-level stress and beat position in song (ictus)

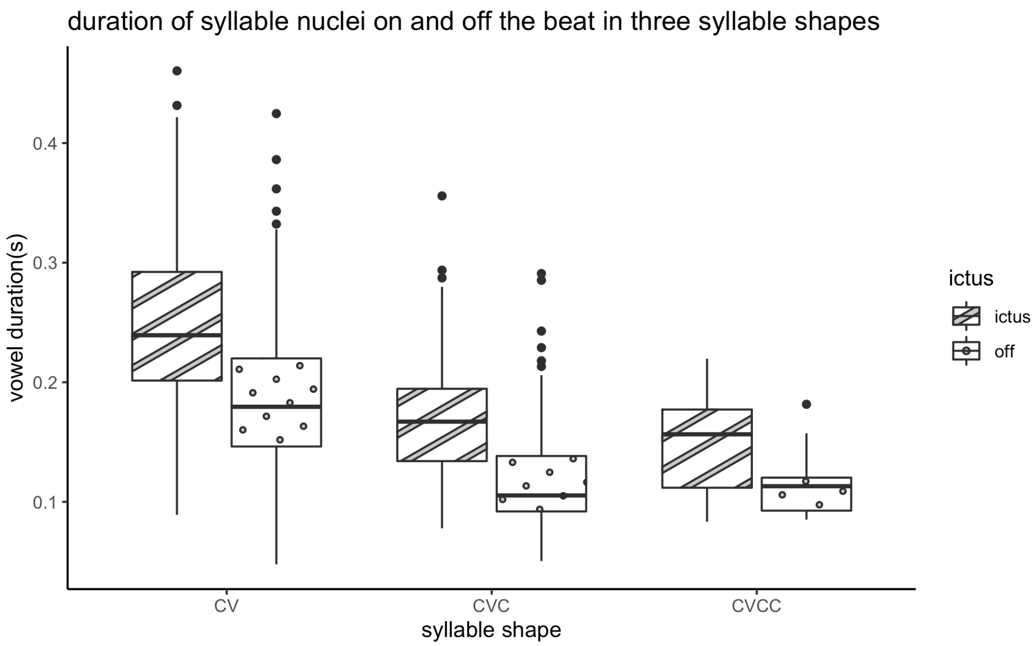

Figure 3.4: vowel dur(s) by beat position and syll. shape

as predictors for vowel duration.

The graph in 3.4 illustrates the distribution of vowel durations in different syllable shapes falling on and off the beat.

A similar pattern can be seen in 3.5, where the stressed or unstressed status is shown instead. At both song and word levels of prominence, CV or Q1 syllables are the longest, gradually decreasing in CVC and CVCC, both of which are Q2 syllables. This further confirms the gradience of the quantity contrast at the segmental level.

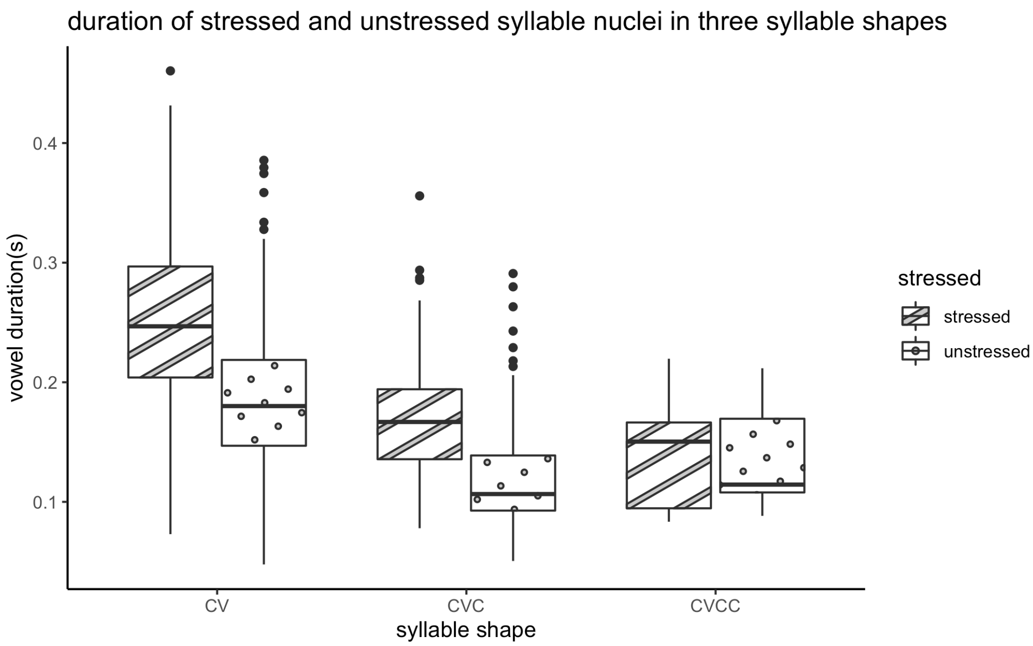

Figure 3.5: vowel dur(s) by word stress and syll. shape

		duration	
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.25	0.20 – 0.30	**<0.001**
ictus [off]	-0.04	-0.06 – -0.02	**<0.001**
stressed [stressed]	0.03	0.01 – 0.05	**<0.001**
quantity [2]	-0.07	-0.09 – -0.06	**<0.001**
ictus [off] * stressed [stressed]	-0.01	-0.04 – 0.01	0.381
ictus [off] * quantity [2]	0.04	0.02 – 0.06	**<0.001**
Random Effects			
σ^2	0.0011		
$\tau_{00\ word}$	0.0005		
$\tau_{00\ syll}$	0.0004		
$\tau_{00\ shape}$	0.0027		
$\tau_{00\ song}$	0.0017		
$\tau_{00\ performer}$	0.0004		
ICC	0.8355		
N_{song}	9		
N_{syll}	259		
N_{shape}	15		
N_{word}	315		
$N_{performer}$	3		
Observations	677		
Marginal R^2 / Conditional R^2	0.109 / 0.853		

Figure 3.6: stress and ictus duration model output

Chapter 4

Discussion

4.1 Temporal Prosodic Features Crystalized at Segmental level in isochronous syllables

These results support the hypothesis that prosodic timing modifications resulting from the synchronization of independent rhythms (in this case, syllables and notes into syllable-notes) do not result in the subordination of one system to another. In this case, when syllables and notes become one timing unit, the temporal acoustic correlates often found at suprasegmental levels (syllable, foot, word) are found at the segmental levels.

This interaction is then more analagous to entrainment of two independent rhythmic systems than to language being forcibly pigeon-holed into the metrical structure of music.

4.2 Future Studies

This study used a sample of nine songs and three singers, all recorded in the 1960s. Annotation has already begun on the remaining songs that fit into this sample's criteria: In total, there are seventeen songs and seven singers from Parnumaa county recorded in the 1960s. Several of the singers featured in this sample set were also recorded speaking. Annotating their natural speech would provide a valuable contribution to the song corpus, as findings from the songs could be compared with speech of the same person.

Extending the findings of vowel duration and the ternary quantity contrast has several obvious paths: synchronic analysis of with song samples from the same approximate time period but differing according to region, or even language: several other Balto-Finnic families have a trochaic tetrameter folksong tradition. Diachronic analysis with song samples of same singers in the same region at different points in time is another possibility with data from the Estonian Folklore Archives. Both these goals are achievable only with the continued annotation of the corpus of regilaul, which is quite demanding work. As I continue to build this corpus, I am also actively exploring ways in which to automate the process. The inclusion of beat tracking software eliminated much observer subjectivity, and also facilitated the forced-aligner: by automatically grouping verse lines into measures, the aligner was given phrase groupings to synthesize and compare, rather than attempting to align the entire song in a linear fashion. However, as the forced aligner used here was made specifically for speech, one way to improve the accuracy of the forced aligner (decreasing manual adjustment of annotations) would be to train a the aligner on sung material using supervised machine learning. As I plan to continue with the annotations either way, I can use the corrections I make as training material for the algorithm, with the hopeful result that the forced-aligner will eventually reach some threshold of accuracy, voiding the need for manual adjustments.

4.3 Conclusion

This study examined fine-grained acoustic-phonetic features of Estonian prosody in the context of traditional folksongs known as regilaul. The data support the hypothesis that duration contrasts inherent in the ternary quantities of Estonian are still present at the segmental level, even after undergoing modification to fit syllables into isochronous notes of the song. The data also supports that the durational correlates of word-level stress are also crystallized and evident at the segmental level, so while it isn't lexically contrastive, the role of primary stress and unstress to mark word boundaries in spoken Estonian is still present in sung Estonian.

This indicates a relationship more akin to the collaboration of two independent rhythmic systems rather than one rhythmic system dominating the other. Combined with the fact that any regilaul text can be sung to any existing regilaul melody brings into question whether *spoken* metrical verse text is truly independent of the temporal constraints of the musical meter they are made with, or somewhere between language and music.

Appendices

Appendix A

songs

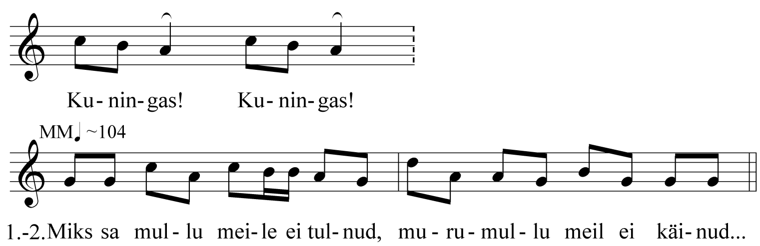

Figure A.1: Kuningamäng “The King Game” performed by Liisa Kümmel

Figure A.2: Põld lindudele performed by Marie Helimets

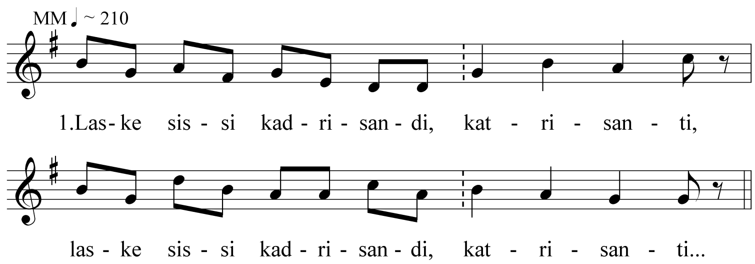

Figure A.3: Kadrilaul performed by Marie Helimets

Figure A.4: *Loomine* performed by Liisu Orik

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.