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

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

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. How do Estonian poets and singers negotiate these complex temporal aspects into a meaningful song? The aim of this paper is to analyze the acoustic-phonetic realizations of metrics in Estonian lyrical folksongs known as ***regilaul*** ($/re.ki.laul/$).

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

1.2 Metrics

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.

1.2.1 Phonetics of Estonian Segments and Syllables

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
[ˈøː.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/	/koː.ti/
	koti	*kotti*
	/kot.ti/	/kotː.ti/
	gooti	*kooti*
	/koːt.ti/	/koːt.ti/

Table 1.2: segmental permutations of initial syllable quantity

In the first row of **??**, 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.

1.2.2 Musical Metrics

In music, the smallest prosodic constituent is an individual note event whose relationship to the other notes in the song are indicated by the time signature, i.e., $\frac{3}{4}$ and $\frac{4}{4}$. The denominator corresponds to the number of divisible beats of a "whole" note (𝅝), while the numerator refers to the number of "beats" in a single measure. In $\frac{4}{4}$, a whole note is sustained for the same duration as four quarter notes (♩) in the same measure, and each measure must culminate in enough notes and rests to equal a whole note.

1.2.3 Metrical Principles of Estonian folksong

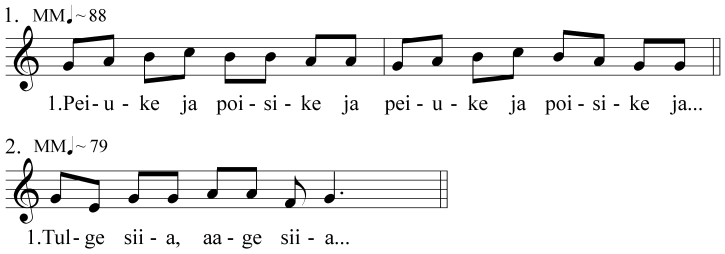

Figure 1.1: "Millal saame sinna maale"

Estonian *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 (𝄾).

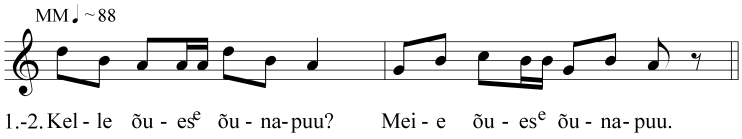

Figure 1.2: notation of “Loomine” performed by Liisu Orik in 1965

1.3 Previous Studies with *regilaul*

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.

1.3.1

Beginning in the early nineties and extending into the millenium, a collaboration between two native speakers of Estonian laid the groundwork for this study. Musicologist Jaan Ross and Phonologist Ilse Lehiste found

common ground investigating the acoustic correlates of regilaul: Ross had publications on one song Ross (1989, 1992), but the inclusion of a linguist in the study of lyrical realization proved crucial. Together they 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 also measuring the duration of segments within the syllable-notes. In their discussion, they mention that some of the durational differences not present at the syllable-note level are present at the segmental level, most visible in heavy syllables with complex and geminate codas (Ross & Lehiste, 2001)

1.4 The present study

To extend the findings of this body of work, the present study examines two acoustic correlates of prominence at the segmental level: namely,

the syllable nucleus. Because the vowel is the most sonorant part of a syllable, it is the most acoustically and perceptually salient portion of the syllable. Measuring the vowel duration will also offer indirect information about the rhyme as a whole: the presence of codas and complex codas should have an effect on the vowel duration in isochronous syllable-note sequences. In the cases where quantity is distinguished by coda consonant length, the syllable nucleus (measured by vowel duration) will be necessarily shorter to accommodate the geminate or complex coda. In cases where quantity is indicated by the length of the vowel, the opposite should be true. This study therefore 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.

In addition to vowel duration, I also include vowel dispersion measured by the euclidean distance from the center of each singer's vowel space

on the (f1, f2) plane. A larger vowel perimeter generally corresponds to hyper-articulation or clear speech, while a smaller perimeter with hypo-articulation or reduction Lindblom (1990); Smiljanić & Bradlow (2005). In natural Estonian speech, reduction is only allowed on unstressed syllables, with /i/ being the most resistant (Eek & Meister, 1998). An early paper by Ross measured formant frequencies f1 and f2, finding a reduced vowel space in song compared to measurements in spoken EstonianRoss (1992). However, upon examination of tokens included in the analysis, it is clear that the sample of sung syllable-notes consists entirely of syllables in non-initial positions of Estonian words: that is, the sample of vowels taken from the song were all unstressed. Conversely, the spoken Estonian used for this comparison contained stressed and unstressed syllables.

This measurement is included for two reasons: one, the acoustic correlates of prominence in both language and music are almost always not a single cue but a convergence of several cues. If duration is less available for word-level contrasts when it is constrained by musical meter, vowel dispersion is a viable candidate for indicating prominence in its stead. The second reason is to follow up on the findings of (Ross, 1989, 1992), this time comparing vowel space in stressed and unstressed syllables within the song.

1.4.1 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 by pursuing two possibilities. First, it is possible that apparent neutralization at one level (i.e., syllabic) could still be present at another level (i.e., segmental). Second, if durational differences usually indicative of prominence are indeed neutralized, it is possible that a secondary cue to prominence is enhanced in compensation.

1.4.2 Duration of Syllable Nuclei

Given the findings of (Ross & Lehiste, 2001), isochronous syllable-notes would result in heavier syllables having shorter nuclei to accommodate for the coda and/or complex codas that distinguish the syllable weights from each other. This finding would confirm their earlier results while simultaneously contributing acoustic evidence that prosodic information that is usually suprasegmental is preserved at the segmental level.

- HQ: duration contrasts for syllable quantity will be evident in the vowel duration, with vowel duration *decreasing* as syllable weight increases.

- $HQ_{\emptyset}$: on or off-beat position is best predictor for vowel duration of

syllables in all quantities.

- HA: duration contrasts for word-level stress will be evident at the segmental level after taking into account a syllables beat position in the song.

- $HA_{\emptyset}$: on or off-beat position of the song is best predictor for vowel duration of syllables in both stressed and unstressed syllables.

1.4.3 Dispersion of Perimeter Vowels in F1,F2 Space

An earlier study of vowel quality Ross (1992) found a reduction in the dispersion of vowels in regilaul singing compared to running speech. However, all vowel tokens in that study were taken from syllables that are unstressed at the words level, while the running speech comparison data included all stress positions. Thus we do not know if the vowel space is best predicted by *style* i.e., spoken or sung, or by word stress, or a combination of both. To begin probing this question further, I include measurements of vowel space of stressed and unstressed sung syllables.

- HS: stress/unstress contrasts will be evident in the nucleus in terms of hypo and hyper articulation. For this I measure both nucleus duration and vowel space dispersion.

- $HS_{\emptyset}$: vowel dispersion differences in syllables are random.

Chapter 2

Methods

2.1 Design

To test these hypotheses, vowel duration and location on the F1,F2 plane are taken 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

[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).

inclusion criteria for vowel duration and dispersion measurements.

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. These songs were compiled in part by Herbert Tampere, an early ethnomusicology field work organizer of the EFA, who along with Erna Tampere and Ottilie Kõiva collected these folk songs(Oras & Västrik, 2002; Tampere, 2016).

initial analysis I chose a sample of songs all belonging to the same regional dialect and recording method. 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 and Erna Tampere.

2.3 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. The result is documentationion of the precise moment a given note is realized, the absolute duration of that note, and the fluctuations in tempo from measure to measure. This is beneficial to my purposes in two ways: by accounting for the natural tempo variation of the a capella songs, and by incorporating acoustic parameters to define the beat, which increases the replicability compared to solely using perceptual judgments.

[2] 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]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.

The orthographic text phrases of the song lyrics are then inserted into each verse-phrase interval, which the eSpeak forced aligner for Estonian (Duddington et al., 1995) uses 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.

2.3.1 Adjustment Criteria for Vowel Durations

The beginning of the vowel was broadly aligned according to a combination of acoustic correlates: at the point where 1. intensity was within 2dB of the steady-state medial portion of the vowel with a slope between 0.5 and zero, 2. frequency stabilizing into that syllable-note's pitch category, and 3. the presence of a voicing bar and visible formants f1, f2, and f3. Manner specific criteria: did not include burst in plosives. Boundaries between fricative onsets and vowels was determined by the end of visible high-frequency noise in the spectrum. Coronal fricative /s/ also consistently showed a carat **??** in the frequency track immediately preceding the transition to vowel. For approximants, the additional criteria of steady formants was necessary. Following nasal onsets, vowel intensity *lowered*, but a near-zero slope still reliably coincided with the other acoustic correlates.

The offset boundaries of vowels was set similarly, but instead with slopes less than or equal to -1 in their transition to occlusions. The first

three formants were more variable in transition to codas, so the other cues were relied on more heavily.

Closed syllables with approximant codas /l/were excluded, as neither the forced-aligner nor the phonetician could determine a reliable way to define the boundary between them. In onset position, the boundary between approximant and vowel was more consistently definable by the above criteria (with the additional requirement of steady formants, which generally coincided with the frequency and intensity cues). In cases of vowel adjacency across syllable boundaries, the presence of a visible glottal stop and a similar (though not as strict) pattern to the above 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 or nonse words (verified by native speaker informant), and cases where the transcription indicated epenthesis or severe reduction.

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. For the measurements of F1, F2 space, formants were extracted from the mid point of the vowel.

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

2.3.2 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 corpus: text files of lyrics are connected to the textgrid files corresponding to the audio, and acoustic measurements are extracted from PRAAT using the parselmouth library(Jadoul et al., 2018; Van Rossum & Drake Jr, 1995).

2.3.3 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). For models with duration as dependent variable, random intercepts were set for individual song, singer, and word. For HQ, fixed effects were syllable quantity, beat position (ictus), and the interaction of those terms. For HA, the fixed effects were ictus, stress, and the interaction of ictus and stress.

For the model using vowel dispersion (HS) as the predictor, random

intercepts were set for performer, word, and segment quality. Fixed effects were ictus, stress, and the interaction of ictus and stress.

Chapter 3

Results

3.1 Quantity Oppositions in Ictus position

Ternary quantity contrast is only in primary stressed syllables. To analyze vowel duration measurements in all three quantities, a subset of only primary stressed syllables is taken from the dataset. At the song level, the kalevala meter avoids short stressed syllables (Q1) in ictus position, preferring Q2 or Q3 (long and overlong) syllables to fall on the beat. In off-ictus postion, short stressed (Q1) syllables are preferred while Q3 avoided.

3.1 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 A.1 and A.2 in Appendix A. In context of earlier findings that the quantity contrast was “lost” at the syllable level, the decrease in vowel duration as

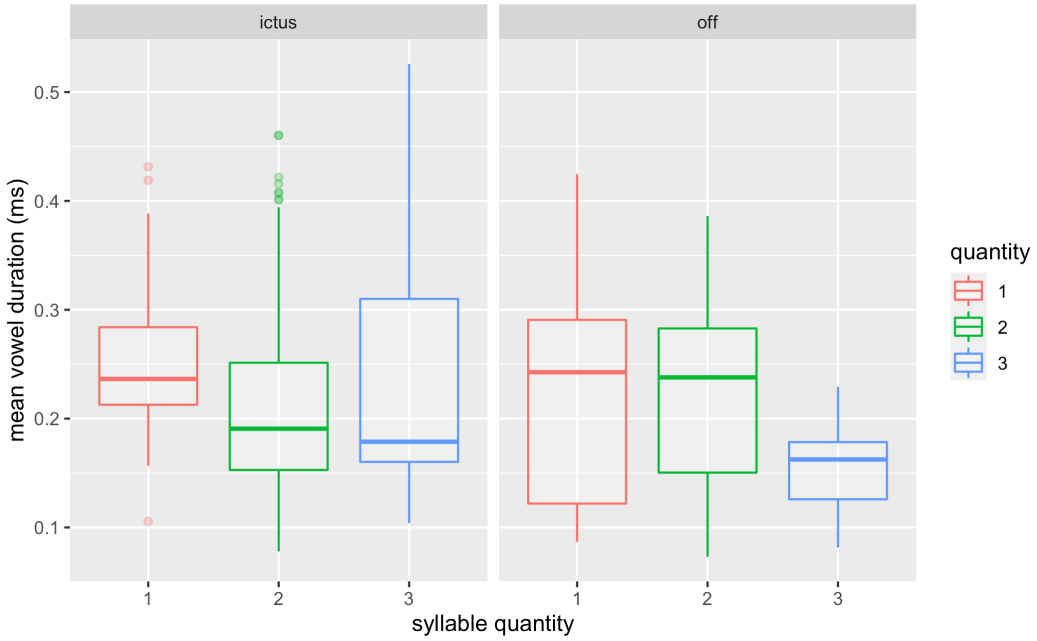

Figure 3.1: density plot of vowel durations in three syllable quantities

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 ($p < 0.001^{***}$), and so I reject the null hypothesis.

These data and analysis support both syllable quantity and ictus as predictors for vowel duration.

3.2 stress and unstress

3.2.1 duration

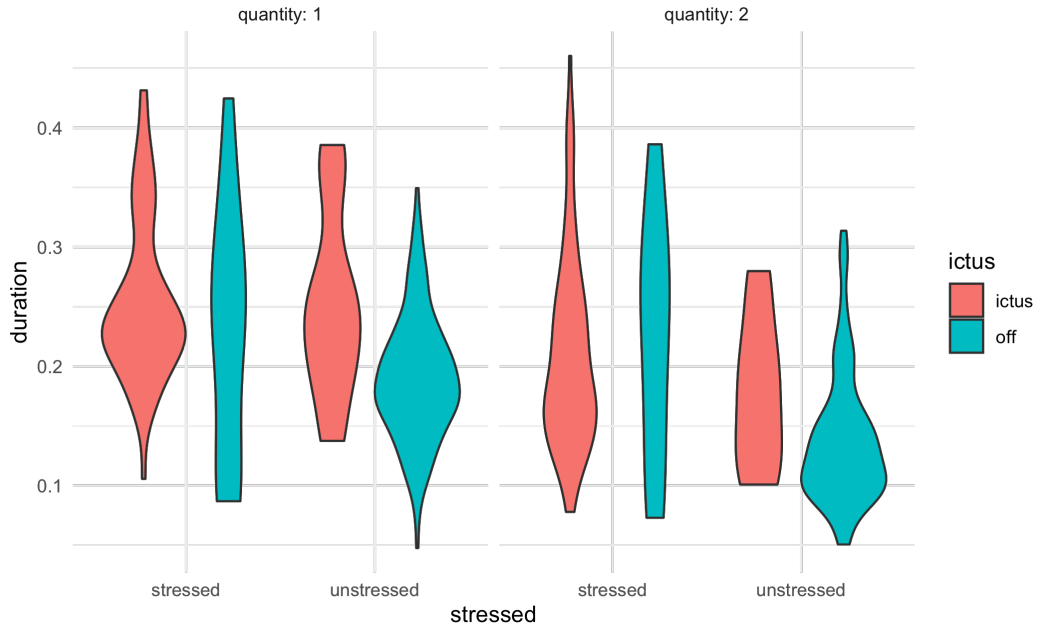

Figure 3.2: vowel durations of stressed and unstressed Q1 and Q2 syllables falling on (ictus) and off the beat

The two graphs in 3.2 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.

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

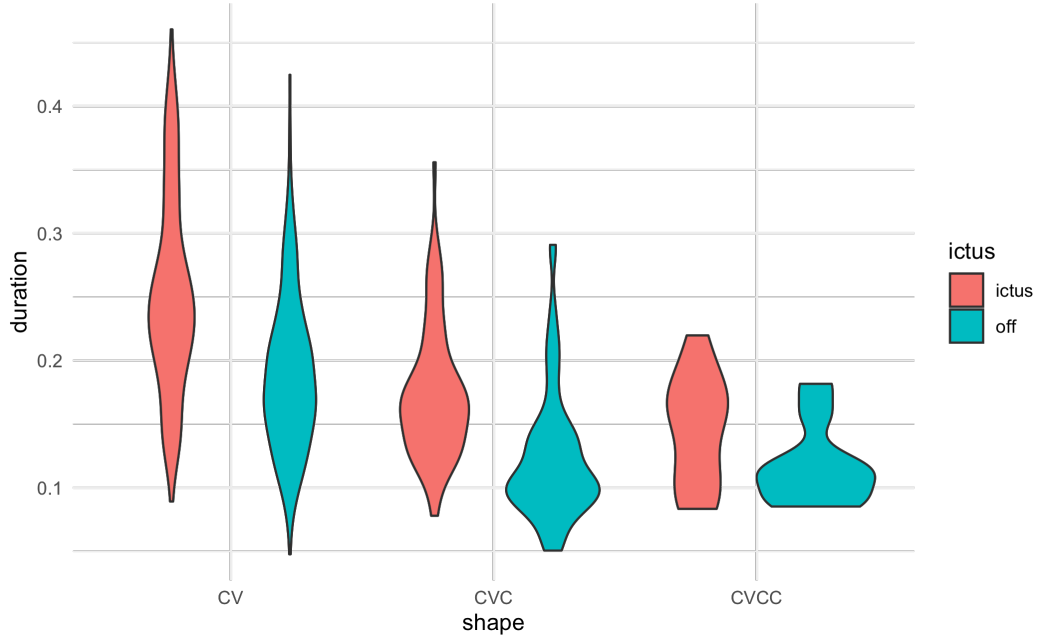

Figure 3.3: vowel dur(s) by beat position and syll. shape

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) as predictors for vowel duration.

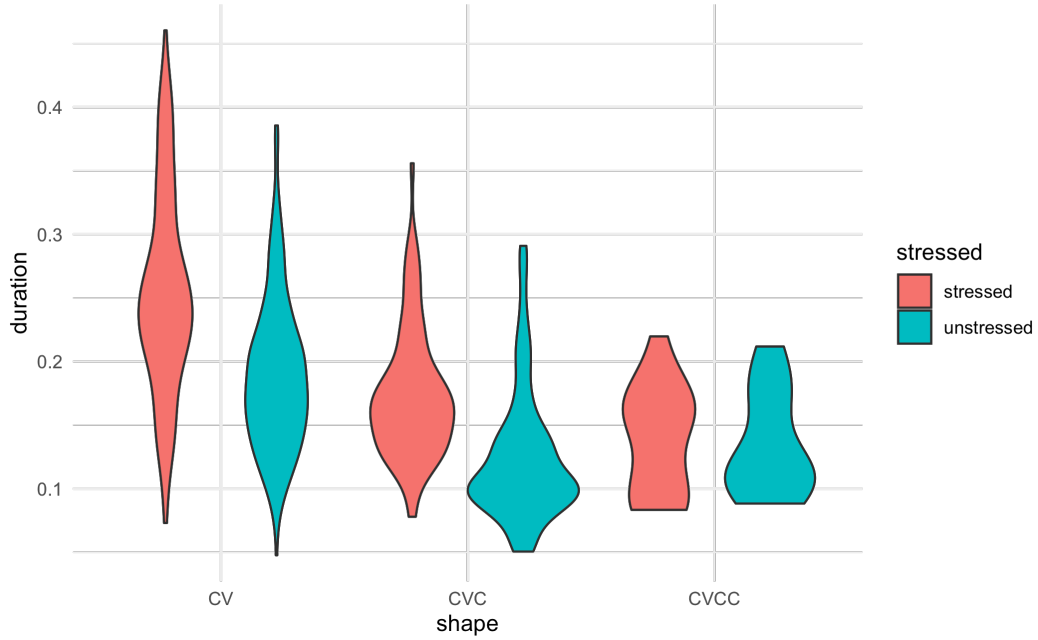

Figure 3.4: vowel dur(s) by word stress and syll. shape

The graph in 3.3 illustrates the distribution of vowel durations in different syllable shapes falling on and off the beat.

A similar pattern can be seen in 3.4, 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.

3.2.2 Vowel Dispersion

A subset of the Q1 and Q2 vowels used for duration measurements above is taken, containing only those five vowel phonemes which occur in

both stressed and unstressed syllables at the word level:

$$a, e, i, o, u$$

. The total number of vowels in this set is **N**. To account for physiological differences between singers, vowel dispersion is calculated as the euclidean distance of each token from the respective singer's vowel center in the (F1, F2) space.

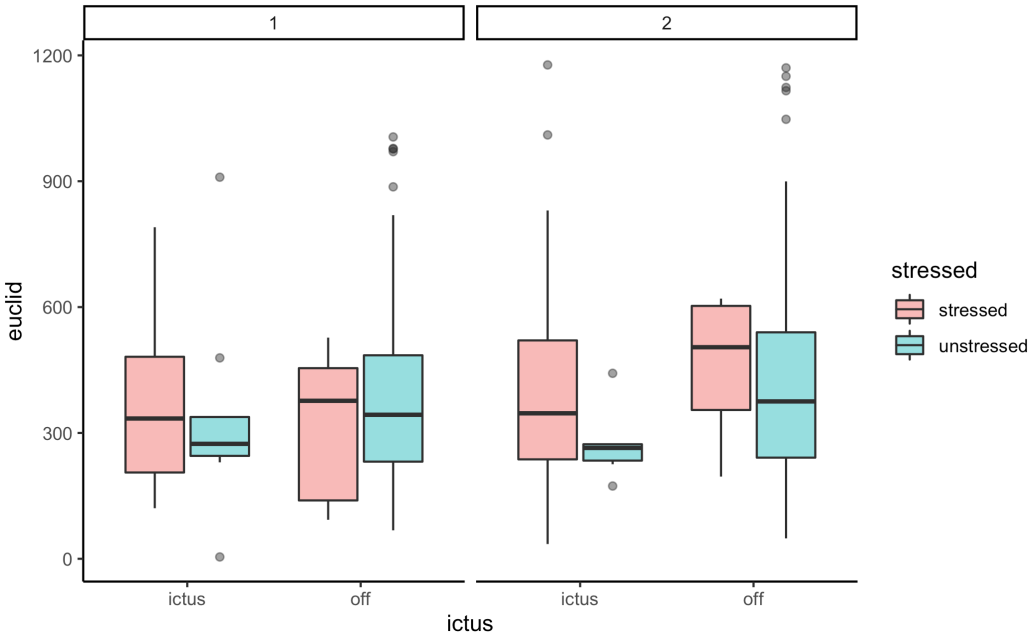

Figure 3.5: euclidean distance of vowels in stress and ictus

A pattern is marginally visible in the two graphs 3.5, faceted by Q1 and Q2. Notice that in off-ictus position, vowel dispersion means of stressed syllables are higher than those of stressed syllables in ictus position. This indicates that stressed syllables falling off the beat are being compensated

for their shortened vowels by way of increased articulation of quality. This pattern, however, doesn't shake out as statistically significant in the model. A linear mixed effects model for vowel dispersion is constructed, but only the uninterpretable intercept shows significance. Comparison with null model is not statistically significant. Thus in the case of vowel dispersion, we fail to reject the null hypothesis. This could be due to the relatively smaller size of the data subset.

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 Vowel Dispersion

Results for the predictive power of ictus and off-ictus, stressed and unstressed syllables and vowel space dispersion were not significant. The patterns that seem to emerge visually in the graphs of distributions suggest that the situation might be ameliorated by increasing the dataset. For this

measurement, there were fewer available tokens than for vowel duration, and therefore less statistical power. Further examination is needed to determine whether or not the observed pattern is simply lacking in power or whether the premises that lead to this hypothesis are missing something entirely.

4.3 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. Increasing the size of the audio corpus would facilitate exploratory analysis in countless dimensions of music and language, and also shed light on the issue of vowel space unresolved here. 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 possi-

bility 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.

4.4 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. Vowel quality patterns were measured but not significant for this dataset.

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

Additional Graphs and Full Statistical Analysis Tables

Table A.1: duration & quantity fixed effects

Predictors	Estimates	Confidence Interval	p
(Intercept)	0.27	0.23 – 0.30	<0.001***
Q2	-0.04	-0.06 – -0.02	<0.001***
Q3	-0.03	-0.06 – -0.00	0.048*
off-ictus	-0.05	-0.08 – -0.02	0.004*
Q2 * off-ictus	0.02	-0.03 – 0.06	0.54
Q3 *off-ictus	-0.02	-0.07 – 0.03	0.44

[1]

[1]Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05

Table A.2: dquantity-duration random effects

Random Effects			
Predictors	**Estimates**		
τ^2	0.0012		
τ_{00} word	0.0034		
τ_{00} song	0.0022		
τ_{00} performer	0.0001		
ICC	0.8227		
N song	9		
N word	298		
N performer	3		
Observations	367		
Marginal R2 / Conditional R2	0.071 / 0.835		

Table A.3: model comparison, duration predicting quantity & ictus

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
null	5	-914.74	-895.21	462.37	-924.74			
design	10	-949.20	-910.15	484.60	-969.20	44.464	5	1.865e-08 ***

Table A.4: anova of model comparison: duration dependent stressed ictus

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
null	5	-1747.0	-1724.4	878.5	-1757.0			
design	10	-1966.8	-1921.6	993.4	-1986.8	229.81	5	< 2.2e-16 ***

Table A.5: anova of design and null lmer models for euclidean distance, stress and ictus

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
null	5	5347.3	5367.3	-2668.7	5337.3			
design	8	5348.7	5380.7	-2666.4	5332.7	4.5855	3	0.2048

Table A.6: duration dependent variable lmer

Predictors	Estimates	Confidence Intervals	p
(Intercept)	0.22	0.18 – 0.25	<0.001***
off-ictus	-0.03	-0.05 – -0.00	0.017**
stressed	0.05	0.03 – 0.07	<0.001***
Q2	-0.05	-0.06 – -0.04	<0.001***
off-ictus* stressed	-0.01	-0.04 – 0.02	0.467
off-ictus* Q2	0.01	-0.01 – 0.03	0.332

Table A.7: random effects duration-stress-ictus model

Random Effects	Confidence Intervals
Predictors	Estimates
σ2	0.0026
□00 word	0.0004
□00 song	0.0018
□00 performer	0.0001
ICC	0.4761
N word	315
N song	9
N performer	3
Observations	676
Marginal R2 / Conditional R2	0.190 / 0.576

Table A.8: euclidean distance dependent fixed effects

Predictors	Estimates	Confidence Interval	p
(Intercept)	351.9	207.53 – 496.26	<0.001***
stressed	48.7	-48.46 – 145.87	0.325
off-ictus	80.8	-14.93 – 176.52	0.098
stressed*off-ictus			

Table A.9: random effects of euclidean distance and stress-ictus lmer

Random Effects	
σ2	32259.12
□00 segment	4058.78
□00 song	5512.35
□00 performer	5432.4
ICC	0.32
N segment	11
N song	9
N performer	3
Observations	401
Marginal R2 / Conditional R2	0.008 / 0.323

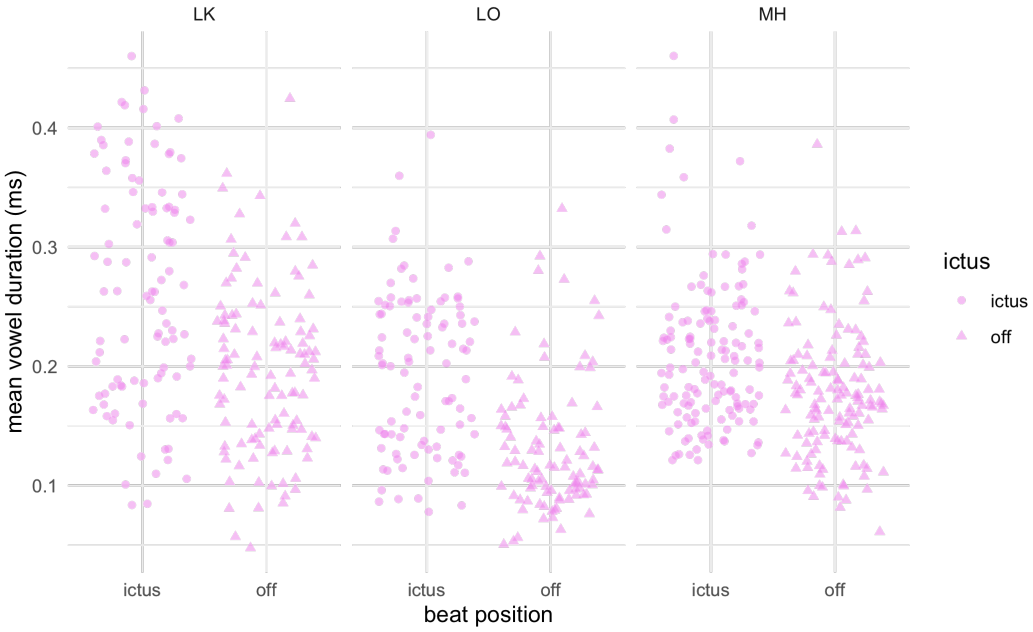

Figure A.1: vowel durations on and off the beat in each performer

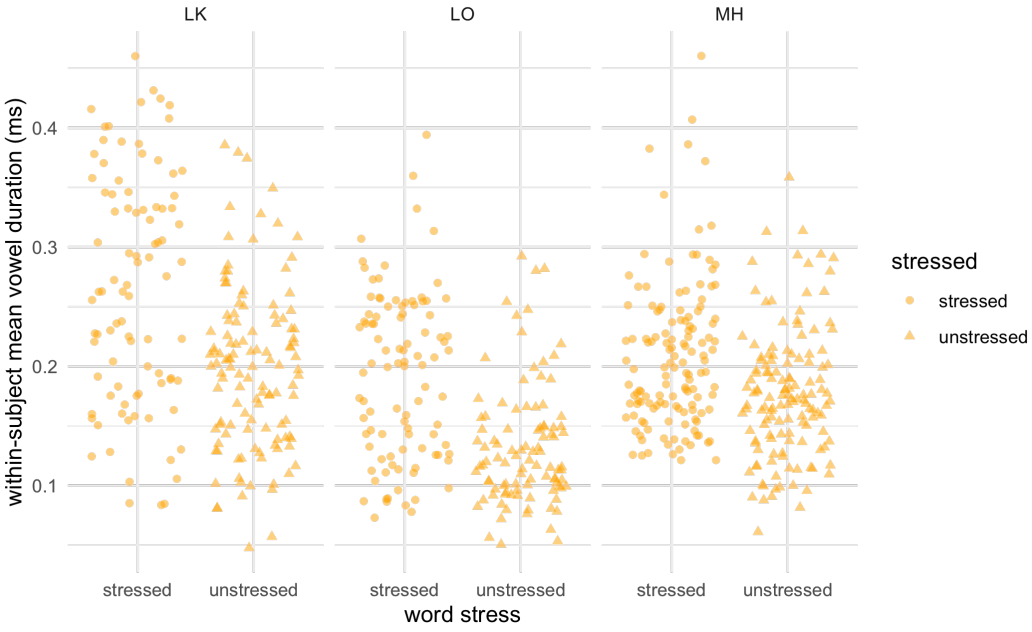

Figure A.2: vowel durations of word-stress in each performer

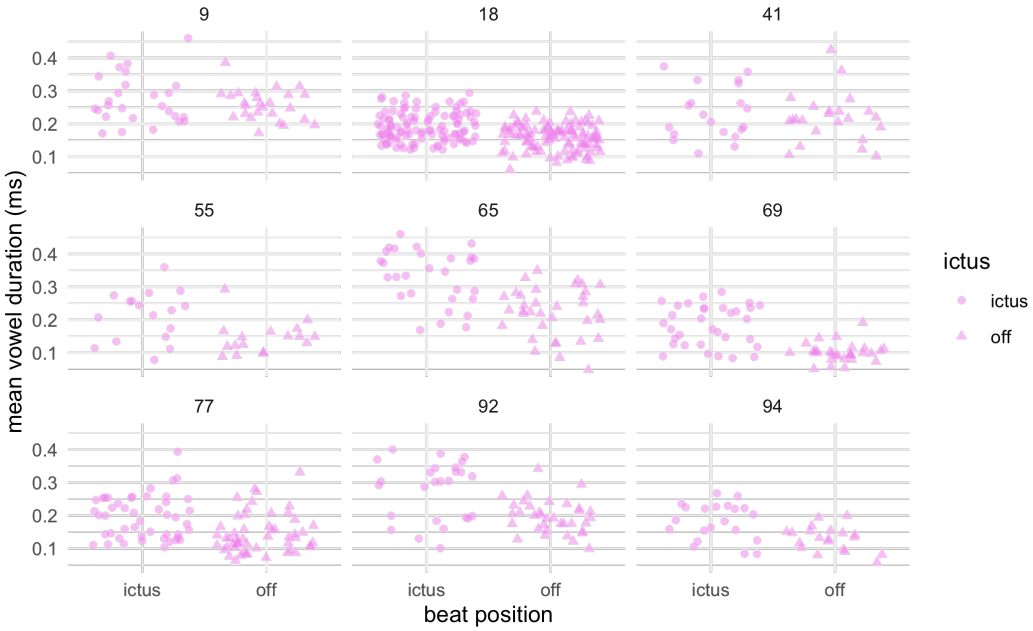

Figure A.3: vowel durations on and off the beat by song

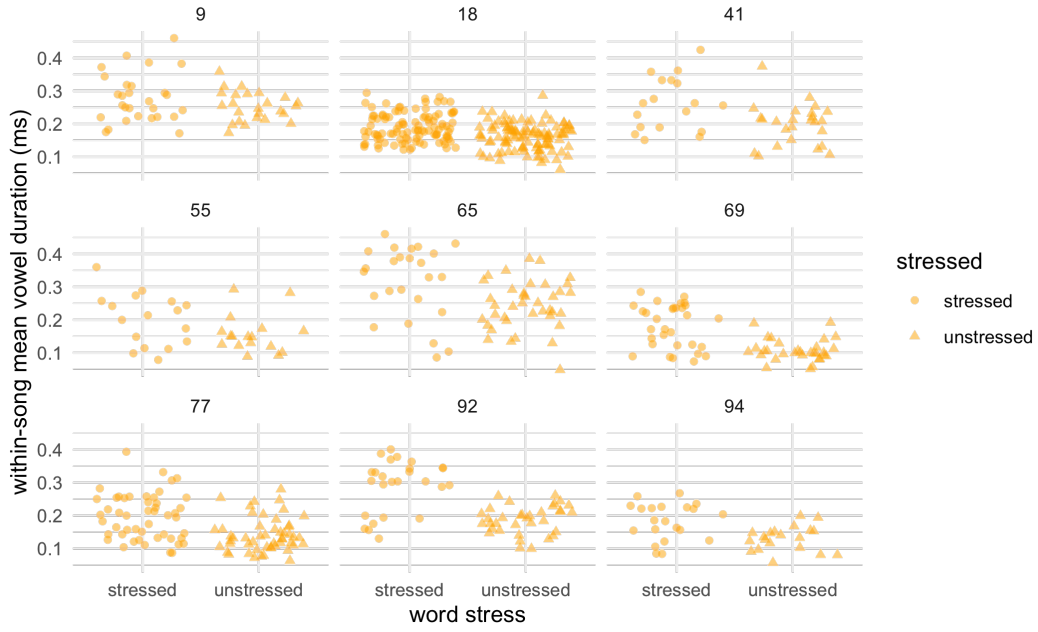

Figure A.4: within-song vowel durations in each word-stress position

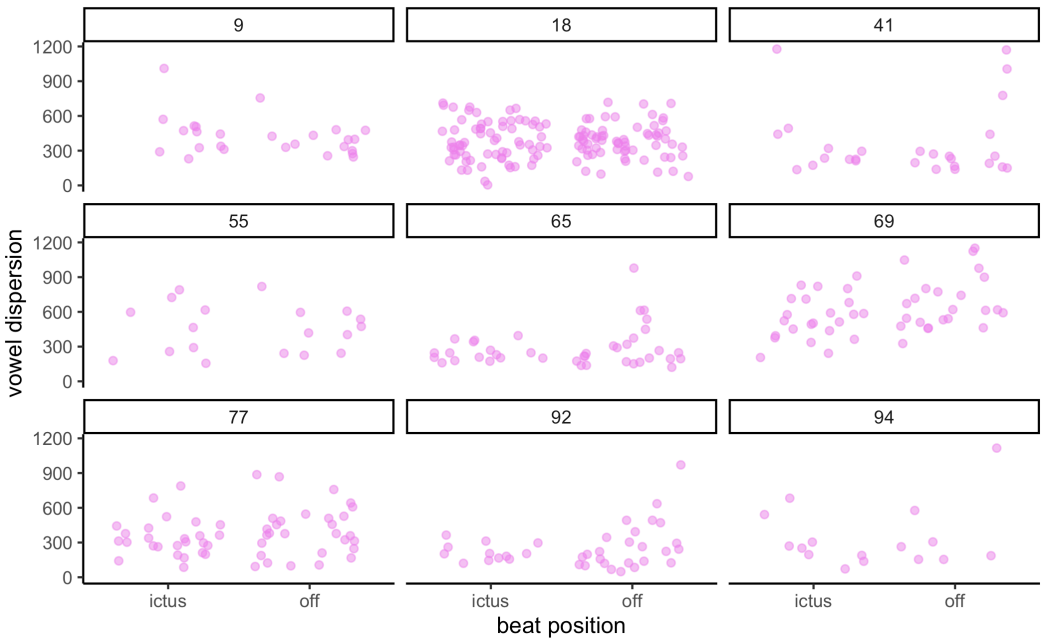

Figure A.5: euclidean distance of vowels on and off the beat by song

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.

URL `https://www.jstatsoft.org/index.php/jss/article/view/v067i01`

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.

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.