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.

?

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 On words becoming lyrics

1.1.1 song lyrics as the music-language interface

In order to join music in song and become lyrics, language elements undergo modification to be fit into the musical structure. At the same time, the interpretability of those linguistic elements which constitute contrasts in semantic meanings has motivation for preservation. If the performer wishes the audience to perceive meaningful lyrical content, these contrasts must be crystallized at some level of salience in the acoustic signal.

This project takes folksong performance as an opportunity to examine the interface of language and music. In particular, I investigate the acoustic manifestations of the word-prosodic features attested in spoken Estonian when they are sung in a temporally restricted domain such as music.

1.1.2 Estonian Phonetics

The role of primary stress in Estonian is described by Ilse Lehiste as *identificational* rather than contrastive (**?**). In other words, there are no stress minimal pairs at the lexical level, so the prominence cue is to indicate the onset of a new word. This is sometimes also called *demarcative* stress.

Three proposed levels of stress: primary stress, unstress, and secondary stress. (**?**)

Duration can be a phonetic correlate of stress, and it can also be independently contrastive at the segmental level (**?**). Primary lexical stress in native Estonian words is fixed, falling on word-initial syllables. **?**

(1) laul-da
[ˈlɑuːl.dɑ]
sing-TR

'singing'

(2) ööbik

[ˈøː.pikː]

nightingale.NOM

'nightingale'

Estonian has three syllable weights, also called degrees or quantities, that are contrastive in primary stress position. The first degree or Q1 is described as short,

This ternary contrast has long been the subject of debate in the phonological literature of metrics: Q3 syllables have been analyzed both as a monosyllabic foot (**?**) and as a trimoraic syllable (**???**).

Acoustically, the difference between a stressed and an unstressed syllable is found not in a single cue but in a convergence of several cues: often

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

duration, increased variability in f0, and increased vowel space, corresponding to a notion of localized hyperarticulation (**???**).

1.1.3 Metrical Principles of Estonian folksong

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., 3/4 or 4/4. The denominator corresponds to the number of divisions of a "whole" note, while the numerator refers to the number of "beats" in a single measure. So the prosodic heirarchy in music begins at the note level, then to each measure as constituent, larger phrases and motifs, and eventually the highest level which is the entire song.

In Eesti, attested segments are organized into syllables, which in turn combine in strong-weak relationships to form prosodic feet, which make up the words of phrases and so on. both regilaul and Eesti have heirarchichal metrical systems by which long and short, stressed and unstressed constituents are arranged. In a regilaul verse line, each syllable corresponds to an eighth note in the measure, with every odd positioned syllable-note coinciding with the "beat" onset in a typical 4/4 measure. Falling on the beat in music generally coincides with increased prominence compared to notes occurring "off" the beat. This is acoustically manifested in increased duration of notes on the beat compared to nominally isochronous notes in other positions. That is, in a measure containing eight eighth notes, the absolute durations of the four eighth notes corresponding to the beat will be longer than the absolute durations of the four eighth notes in even positions, while maintaining the perception of the nominal isochrony.

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 (**?**). However, earlier studies of *regilaul* explored temporal aspects of the songs and found that duration characteristics that would usually indicate important semantic differences lost their distinctions partially or entirely. Assuming that the intention of the singer is for the lyrics to be understood, I hypothesize that if some durational correlates of contrasting word-prosodic constituents are made less distinct in the process of compromising with the song that some other acoustic correlate of the relevant contrast at the word-prosodic level will be present, if not enhanced.

1.1.4 The singing giant: a common folklore epic

the singable song **?**

?

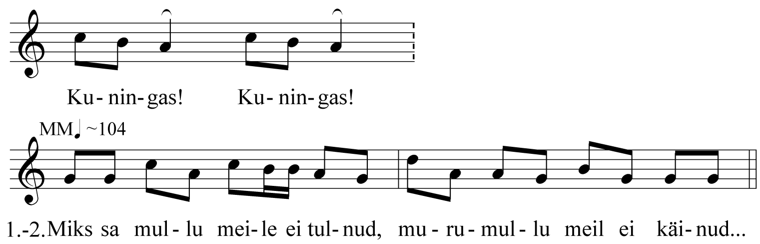

Figure 1.1: music notation of “The King Game” as performed by Liisa Kümmel

1.2 Previous Studies with regilaul

Melodic accents coincide with the stressed syllables, the pitch resolves as the phrase resolves. Said to be a musical abstraction of the natural prosodic intonation of the 8 syllable (spoken) runoverse **?**.

Figure 1.2: default

The question of the acoustic dimensions of Estonian word prosody and the metrical patterns in traditional lyrical folksongs is not a new one. In 1990, an Estonian musicologist measured syllable duration and vowel quality in a single *regilaul* song, publishing a paper on each measurement. same song, timing and quality measured In (**?**), Ross measured formant frequen-

cies f1 and f2 at eighth note positions in (**?**), finding a reduced vowel space in song compared to measurements in spoken Estonian. However, upon examination of the song, it is clear that all the vowel space measurements are from syllable-notes in non-initial positions of Estonian words: that is, the sample of vowels taken from the song were all unstressed, and compared to a mixed sample of spoken Estonian. Thus, the conclusion needs to be evaluated again with comparable samples. (**?**)
In 1992, Ilse Lehiste asked ”Whether there is a correlation between poetic metre and the prosodic structure of a language.(**?**) by means of measuring the acoustic-phonetic realizations of so-called trochaic metrical poetic patterns across several languages including Estonian and Finnish. While these languages share in the more general Balto-Finnic tradition of *runosong* utlizing what is called the *Kalevala* meter, there were significant differences in the phonetic realizations of trochees in each language.

In 1994, their collaboration begins. Ross and Lehiste published several papers examining the temporal dimensions of Estonian word prosody and metrical prominence in *regilaul* folksongs. In (**?**), they conclude that duration differences ordinarily present at the word level (stressed-unstressed) are *"lost"* to the temporal restrictions of the song. In another paper, syllable-notes are again measured, this time examining the role of syllabic quantity in the song. They likewise conclude that the duration of syllable-notes in *regilaul* match more closely with the metrical structure of the songs, that is, the durations of syllable-notes are best predicted by their beat position in the song: on the beat, syllables are longer, off the beat shorter. They extend this finding to conclude that the song ”dominates” the metrical status of the words, claiming that the intelligibility of lyrics is enhanced more strongly by “top-down” processes (i.e., semantic context).

(**?**).

(**?**) Analyzed and concluded that the *regilaul* lyrics are the result of an interaction between word and song prosodic hierarchies. This conclusion relied critically on measuring the durations of syllable-notes, where they found being on or off the beat was the better predictor for duration.

Finally, in

1.3 The present study

I do not dispute the findings that syllable-notes are best predicted by their position in the song. Instead, I argue that instead of an hierarchical interaction between language and music, what occurs is more akin to entrainment. Specifically, bidirectional entrainment of two oscillators.

The syllable-notes, fit into the temporal structure of the music, still have possible acoustic-phonetic dimensions by which to indicate word boundaries (stressed-unstressed) and semantically-relevant quantity contrasts. The durational differences of stress and quantity in spoken Estonian could be indicated either at another level of the prosodic hierarchy, i.e., the segment, or by enhancing another acoustic cue for prominence: i.e., vowel space dispersion.

An EEG study of native Estonian speakers uncovered a perceptual asymmetry: shortening of constituents did not inducing MNN, but lengthening does (**?**). This suggests that stressed syllables, long in spoken Estonian, could be realized as shorter in a song without sounding "off," but that lengthening of shorter elements would be less preferable. So long as syllable-notes maintain the relative durations in the song (i.e., heavy syllables are still longer than light ones), the metrical pattern of the musical phrase can be met.

1.3.1 Hypotheses

Null hypothesis: on or off-beat position is best predictor for vowel duration of syllables in all categories: stressed, unstressed, short, long, and overlong. This would mean that the syllable nucleus is proportional to the total syllable duration, and that the duration contrast is indeed "lost."

H_1: duration contrasts for syllable quantity will be evident in the vowel duration, with vowel duration *decreasing* as syllable weight increases. Given the findings of (**?**), isochronous syllable-notes would result in heavier syllables having shorter nuclei to accomodate for the coda and complex codas that distinguish the syllable weights from each other.

H_2: 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.

Chapter 2

Methods

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.

2.1 Design of the Present Study

autosegmental and segmental tiers

2.1.1 Long Vowels and Diphthongs

CV vs segmental dichotomy of CV phonology, as reflected in the suprasegmental aspects of vowels.

An estonian word game supports the notion of many-to-one phonemes for long vowels, but a one-to-one status for diphthongs.

(3) long vowels (Q1,Q2, Q3 contrast)

- sada ‘hundred’ $\rightarrow$ sapida

- saada ‘send’ (2nd sg. imp.) $\rightarrow$ sapiida (*sapiada)(*saapida)
- saada ‘get’ , *-da* infinitive $\rightarrow$ sapii:da (*sapiada)(*saapida)

(4) diphthongs

laulus ‘in the song’ (iness.sg) $\rightarrow$ lapiulus (*laupilus)

In both cases, ‘pi’ carries the prosodic characteristics of the displaced syllable. the stress is moved from first syllable to ’pi,’ and also the length. crucially, the diphthong is split at the segmental level, while the long vowels are not.

long vowels are treated as monophonematic, whereas diphthongs are treated as biphonematic.

??

As previous studies found that the durational properties of quantity contrasts were not found at the level of the syllable or the foot, I investigate the fine-grained acoustic-phonetic properties of these contrasts at the segmental level within the syllable-note of the song.

To investigate the acoustic manifestations of stressed and unstressed syllables in *regilaul*, syllable nuclei from disyllabic feet in Q1 and Q2 are compared, as Q3 never appears in unstressed positions.

Q1 CV CVC Q2 CVV CVVC CVCC

The syllables are further compared by position in the song according to the beat:

To investigate the question of quantity opposition in *regilaul*, disyllabic feet in Q1, Q2, and Q3 are examined according to their relationship to the beat.

2.1.2 segmentation criteria

no syllables in verse-final or musical phrase-final position vowel onset: if sonorant onset, when intensity within 2 dB of steady-state/medial position of the vowel and intensity slope approaching zero or the intensity slope approaching zero (less than or equal to one) for more drastic transitions such as obstruents vowel offset intensity: before slope exceeds one (less than or equal to one) in transition to occlusion

primarily syllables with onsets

adjacent vowels across word boundaries, a glottal stop if present in the acoustic signal, otherwise omitted

lateral approximants in syllable onset usually easier to determine boundary, coda position /l/ no boundary definable and thus analyzed as a diphthong in the case of intersyllabic long /l/, however, as there is no way to determine the end of the coda /l/ and beginning of the onset /l/, are excluded

song 41 contained churchbells cases wherein word boundaries unclear due to transcription of the lyrics/poetic license

nonsense words

epenthesized vowels (i.e., pandi mind paju raiumaie) having mind(e) word-final vowel next to word-initial vowel/difficult or impossible to determine end of preceding and onset of next across word boundaries

vowel onset and offset boundaries were determined using chiefly pitch and intensity contours in addition to the three first formants.

onsets:

plosive: not include burst. three first formants visible. slope of pitch and intensity contours encroaching on the respective steady state, with a slope less than or equal to one.

liquids: formant steady state, steady pitch and steady intensity.

nasals: intensity at least half of level within following vowel, antiformants, steady f0. intensity reduces in vowels.

fricatives: following the end of the visible noise in the spectrum, at the point where intensity and formant contours are both visible and steady. The /s/ have reliable pitch carats immediately preceding vowels. **??**.

codas:

plosives: preceding fall in f0 and intensity allow for more variation in formants of codas, other cues more consistent.

nasals: drop in f0, intensity less than half of the way to the level within-vowel.

liquids: before drop in intensity before formant divergence.

across syllable and word boundaries:

adjacent vowels, when possible, segmented by presence of glotallization pulses and categorical changes in pitch.

In all cases, if the aforementioned cues are unavailable or ambiguous, the token is elided for this analysis.

A total of 757 individual vowel nuclei met the criteria for inclusion in duration measurements. After excluding diphthongs, a subset of $_m$et the criteria for inclusion in vowel space measurements.

2.1.3 Vowel Duration

Why vowels and not entire syllables? The reason for this is twofold. First, measuring only the syllable nuclei affords more accurate automation. Were we to measure entire syllables, we would be limited to those with sonorant onset consonants, or in a restricted set of environments where consonant onset would be definable. By measuring only the syllable nuclei, we can reliably include more of the available vowel instances. Second, using the sonorant portions of the syllables makes way for the use of onset detection algorithms, so a strong beat is defined by a consistent threshold

for each song, relative to the strength of the other beats in the signal.
The previous regilaul studies found evidence of syllable-note isochrony. So, if we see Q2 and Q3 vowels increasing in duration, this would be evidence against those findings. If, however, the long and overlong vowels have less duration than the short ones, this is evidence supporting syllable-note isochrony. Due to the structure of heavy syllables, the vowels must shorten in order to accommodate the additional coda segments within the note.

2.1.4 Vowel Space Area and Dispersion

As a second measure of prominence, we include vowel space area and dispersion. Studies in English have shown that stress can be thought of as localized hyperarticulation or clear speech. **?** Vowel space area and dispersion are well documented acoustic correlates of clear speech in English **?**, and has also been confirmed in cross-linguistic studies with Croatian **?** and others.

While it has not yet been documented as an acoustic correlate of stress in Estonian, I have reason to believe that it will be an available cue for a singer to use, especially in the context of a song. While duration may or may not be an available prominence cue at the word level, vowel space area and dispersion are prominence cues that would not conflict with the prosodic hierarchy of the song.

often, resonators or filters of musical instruments are solid or fixed.

This is not the case with the vocal instrument.

then so long as the word-level categories of quality are preserved, the

2.2 Materials

Figure 2.1: Laula! *Sing!*

The data for this study was sourced at the Estonian Folklore Archives (EFA). (**?**).

Until 1948, songs were collected on wax cylinders, then played on a phonograph and transcribed. shellac discs 1936-38, 746 recordings, analogue is the biggest collection in the archive, with over 80,000 individual recordings. Open-reel tape, cassette recordings since the 1970s. both wax and disc were re-recorded onto open-reel tape in 78-79. Presently, the sound engineer Jaan Tamm has been working on preserving the earlier tape recordings in digital form for the EFA. WAV files are stored on CD-Rom at the EFA in Tartu, Estonia, while .mp3 and .ogg lossy formats are uploaded to the internet database.

Songs for this paper were accessed via The Anthology of Estonian Traditional Music (**?**). 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. Pictured in 2.2 is a photograph taken of one of the very field trips to record songs studied in this paper.

Figure 2.2: Herbert Tampere on a field trip

While the ultimate goal is to continue annotation of the entire available corpus of *regilaul*, for the 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[1] formats available from the digital anthology. Each song is then imported into a Logic Pro X (**?**) session for beat detection, tempo mapping, and trimming. To make the tempo map, the session must be set to *flex tempo*. From here a beat onset detection algorithm (**?**) is given the transcribed bpm and time signature from the archived song data and run on the imported audio file. The result is an annotation of intervals in time, and the bpm for each measure is annotated according to the performance of the song. The tempo map allows us to document when *exactly* in time the particular singer performed a given note, the duration of the sung note, and the acoustic threshold by which the note is defined as "strong" relative to surrounding notes. The process is informed by the transcribed bpm and time signature included in the anthology. This is beneficial to my purposes in two ways: by accounting for the natural tempo variation in live performance, and by using a consistent

[1] define lossy

metric to determine beat strength acoustically rather than just perceptually.

From here, a MIDI track is programmed to create a metronome that is the length of a single syllable-note in the song. In most of these, a 4/4 measure contains eight eighth notes, so the metronome track contains four eighth notes indicating the “ictus” beats. In flex tempo mode, the MIDI track adjusts note and measure length to match the fluctuations in tempo as documented in the map for the song. The metronome and the song audio file are trimmed to match exactly, and the metronome is converted into a textgrid in PRAAT(**?**), where the annotation process continues.

The orthographic text phrases of the song lyrics are then inserted into each phrase interval with a script, and then eSpeak forced aligner for Estonian (**?**) is run on each phrase to the word and phonemic level. Because this forced aligner is trained on spoken, not sung Estonian, the aligner sometimes tries to align words into the signal before they are uttered. In these cases, the word level tier is manually realigned so that it contains all and only the transcribed word, and then the forced aligner is re-run on this word to the segmental level. In the case of a vowel interval containing an obvious silent portion or occlusion, the boundary is manually adjusted to only include sonorant portions of the signal.

At this point, the audio recording of each song has tiers annotated for tempo and strong beat, verse line phrases, two interval tiers force-aligned to word and phoneme levels, and a separate tier with intervals of the individual vowel segments of interest copied from the phoneme tier.

2.4 Assembling the corpus and annotations

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` (**?**). Among other things, the toolkit has a robust dictionary of Estonian grammar, including phonetic transcription of syllables with quantity and stress data.

Thus the data structure of this corpus offers two independent metrics of rhythmic prominence in these songs. From the audio recording and the beat detection, we have an annotation of strong beats based on replicable acoustic measurements, and from the dictionary in the natural language toolkit, we have native speaker intuitions about the lexical weight and prominence in the words of the text. While the stress system is generally predictable, the syllable quantity is not always apparent from the orthography, and not always detectible by a non-native listener. Linguistic descriptions of the Estonian language date back as early as the seventeenth century, but the ternary quantity contrast was not documented until native Estonian linguists contributed their intuitions. The non-native linguists had only described lexical stress (**?**).

Using onset detection algorithms such as these (**?**) 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.

2.5 Study Design

2.5.1 Questions and Hypotheses

2.5.2 Inclusion Criteria for Vowel Measurements

Once the annotations are complete, the corresponding text files are aggregated and, the corresponding measurements from PRAAT are concatenated via python using the parselmouth library python interface to PRAAT (**??**). I extracted vowel intervals which met the following inclusion criteria. For this study, we are interested in the durations of vowels in initial and pen-initial syllable-notes that are transcribed as isochronous in the melodic transcription.

2.5.3 Statistical Analysis

Linear mixed-effects model using lme4 in R **?**.

For stress and ictus, only q1 and q2 syllables (Q3 never unstressed). for quantity and ictus, only word-level stressed syllables

Design-based formula Hierarchical Linear Model with group-specific terms

??

Chapter 3

Results

3.1 word and song prosodic prominence

3.1.1 Duration of Vowels

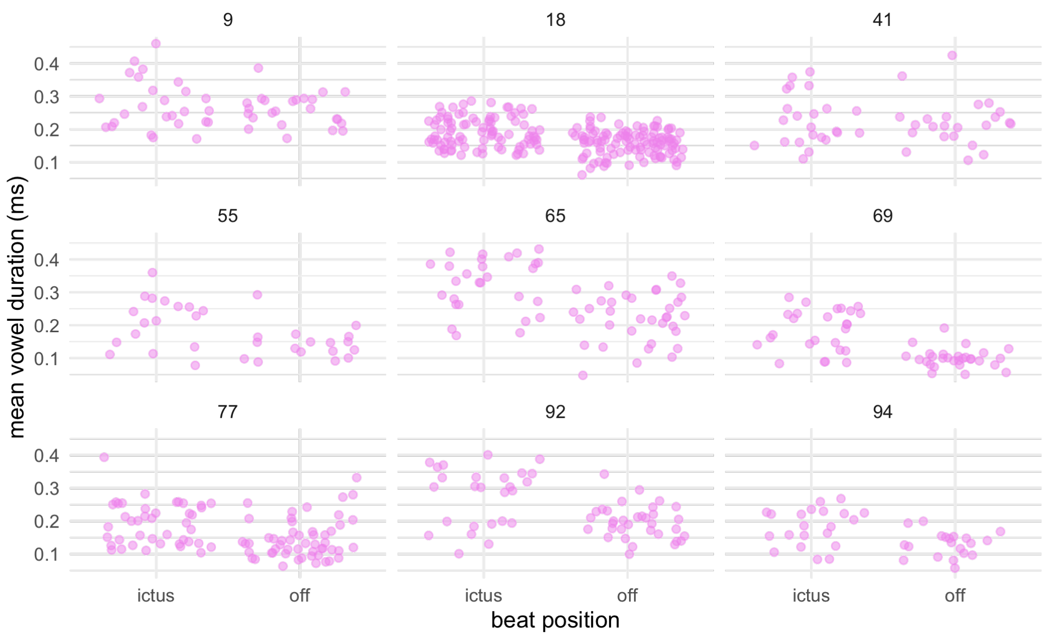

Figure 3.1: within-song vowel durations by beat position

In 3.1, vowel durations of syllable-notes that fall on the beat trend longer than those that fall off the beat. with observations grouped by song 3.2 shows durations of syllable nuclei in two word-stress positions:

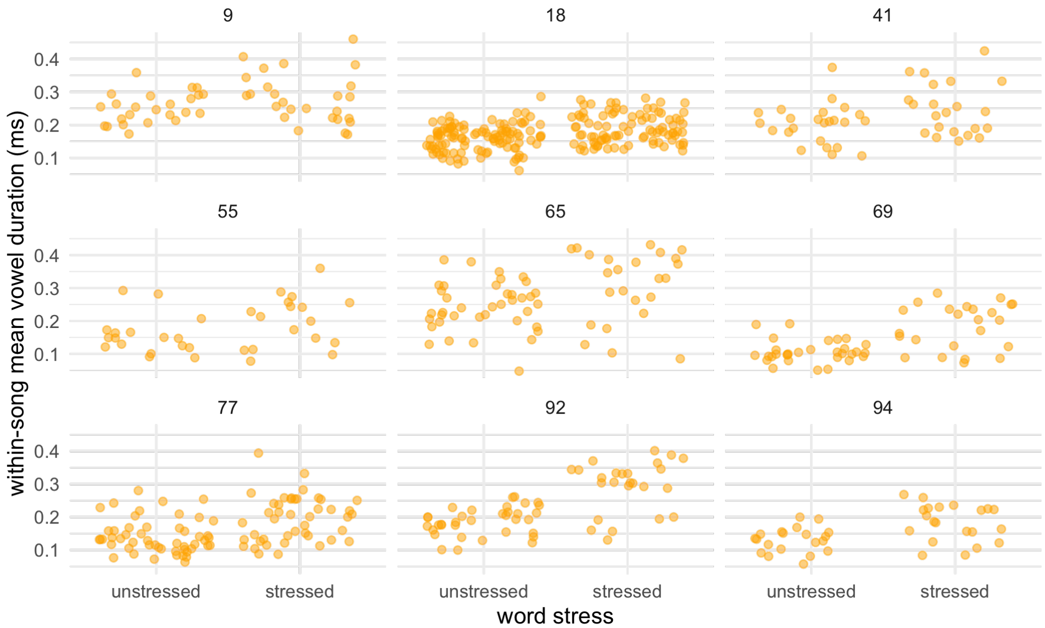

Figure 3.2: within-song vowel durations and word-stress position

stressed and unstressed. Here it is clear that syllable-notes that are ordinarily stressed in spoken Estonian maintain the longer vowel duration even in nominally isochronous syllable-notes.

Taken together, 3.1 and 3.2 show that within each song, vowel durations are distributed in such a way that preserves the word level duration cue of prominence (stressed vowels are longer overall) while simultaneously satisfying the expected distributions of the musical meter: syllable-notes falling on the beat have longer vowel durations than those that fall off the beat.

A similar trend is visible in when the datum are grouped by subject:

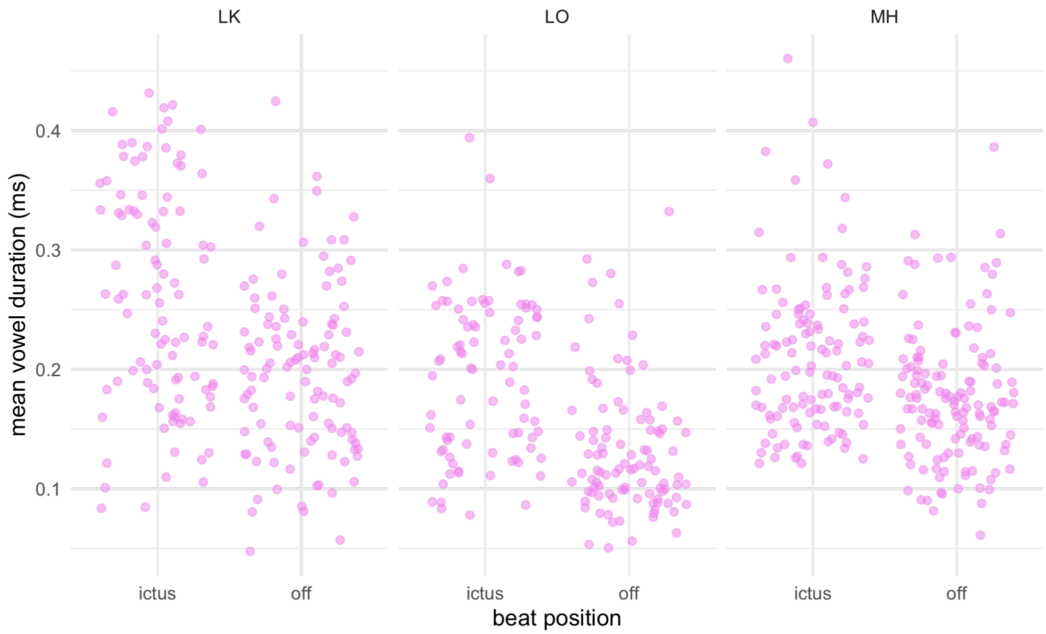

Figure 3.3: within-subject vowel durations by beat position

each performer utlilizes syllable nucleus lengthening in both song 3.3 and word 3.4 positions of prominence.

3.6 shows violin density plots of vowel durations falling on and off the beat with stressed and unstressed syllables at the word level. Here each graph is one of the two quantities that occur in both stressed and unstressed positions in natural spoken Estonian.

3.1.2 Dispersion

Before analyzing the measurements, I look at the overall occurrence patterns of conflicting and concordant lexical and lyrical metrical positions.

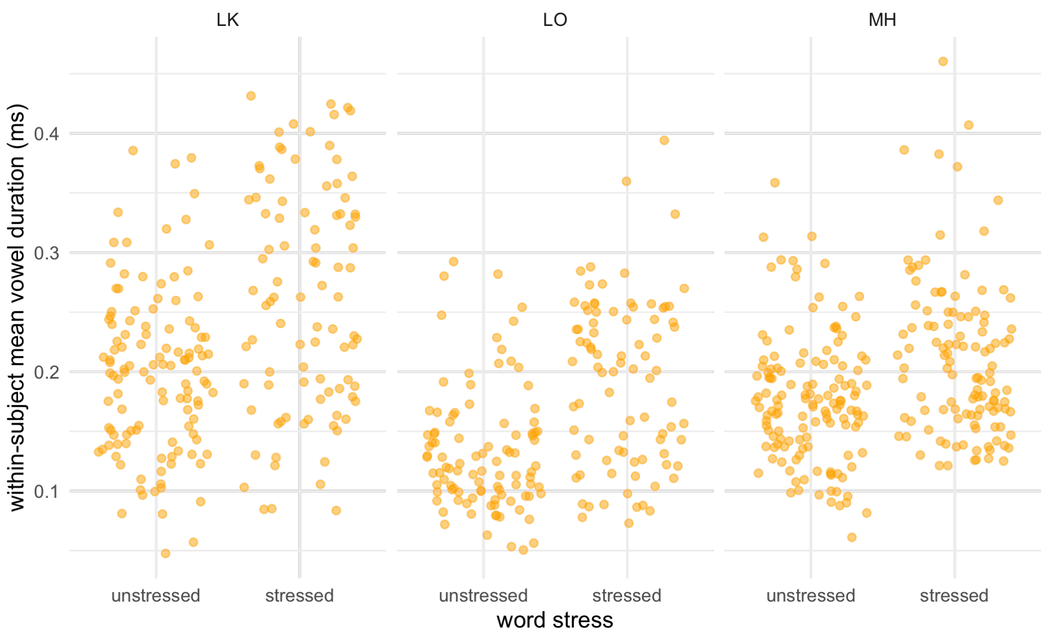

Figure 3.4: within-subject vowel durations by word-stress

That is, do stressed syllables have a requirement or tendency to fall "on the beat" in the song, or are they freely distributed? Likewise, do unstressed syllables tend to align with unaccented notes "off the beat?"

	un	stressed
on	107	129
off	164	210

Table 3.1: counts of conflict(*) and concord strong-weak combinations

The contingency table in **??** illustrates the counts of stressed and unstressed Estonian syllables as they fall on (ictus) and off the musical beats in the song corpus. On the beat, in ictus position, both stressed and unstressed

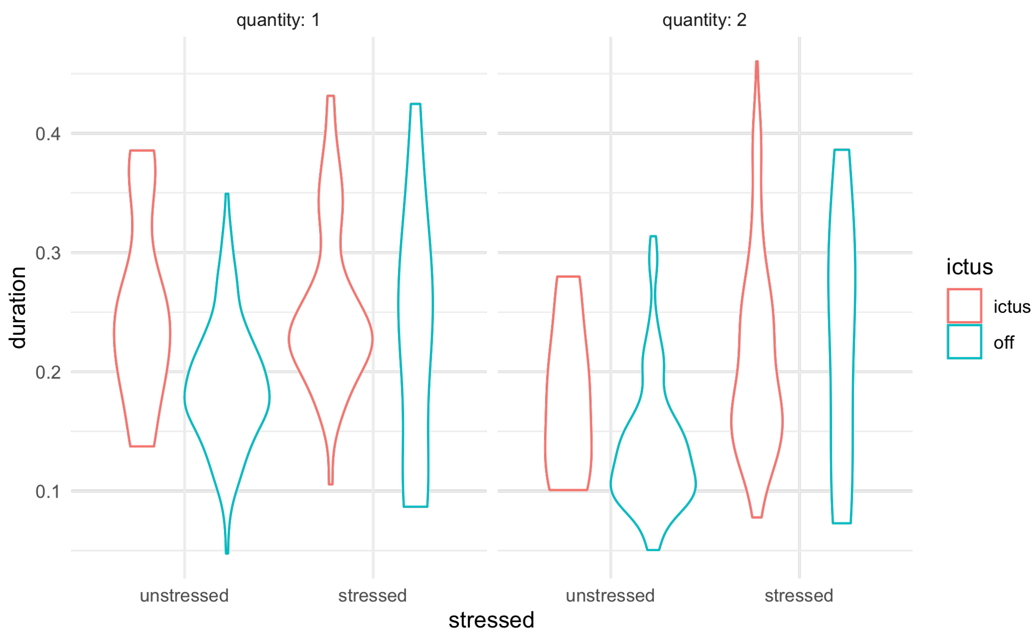

Figure 3.5: violin plots of vowel durations

syllables occur, and Chi-squared test finds no significant difference. Thus, both stressed and unstressed syllables may occupy the ictus ”on beat” note positions in a measure. A similar pattern is seen in note positions that are off the beat, with both stressed an unstressed syllables given equal opportunity to fall in note positions that are off the beat.

This suggests that word-level and song-level rhythmic prominence are able to act independently of each other: that is, one cannot predict from beat position in the song the stressed status of the syllable, or vice a versa. a chi-squared test on the contingency table using the scipy library in

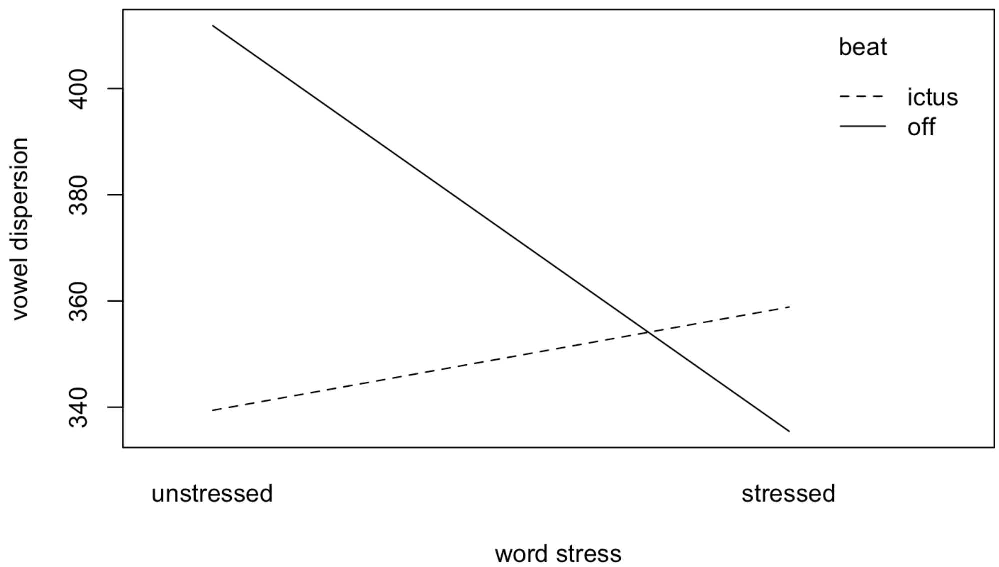

Figure 3.6: vowel dispersion interactions of word stress and beat position

python (**??**)[1]

In 3.2, we see that there is no significant (p = 0.78) tendency for stress to align with ictus or not, and the expected values are very close to the actual value counts in 3.3.

Table 3.2: chi-squared of stress-ictus contingency table

chi-stat	0.077
p-value	0.78

[1](**?**) can be found at `https://github.com/sally-ran-some` (**?**)

htb

Table 3.3: expected values predicted from chi-squared

	un	stressed
on	166	207
off	104	131

3.1.3 Vowel Duration: song & word stress

To examine the overall effects of stress and ictus on vowel duration, I exclude Q3 syllables, as they always fall in stressed position in native Estonian words.[2]

Using the (**?**) Stan extension package for R, I first construct a design-based Bayesian heirarchical regression model following **?**.

$$duration\ stress + ictus + stress * ictus + (1|singer) \tag{3.1.1}$$

However, upon looking at the posterior distributions, it was clear there was a pattern not being explained by the model. The random intercept for singer had been chosen based on the assumption that singers would vary from each other. However, as you can see in **??**, there are distinct groupings that may instead correspond to each song. The singer as random intercept notion had been based on phonetics research in natural speech, however, in this case the song was the best random intercept to include in the model for vowel duration. Performers tend to be consistent compared with each

[2]Vowel durations for Q3 falling on and off the beat are included in appendix C

other in songs, while songs with differing speeds (beats per minute) are better predictors for vowel duration. Thus, the (ad hoc) model I will focus on for this analysis is as follows:

$$duration\ stress + ictus + stress * ictus + (1|song) \tag{3.1.2}$$