from lexical to lyrical: how linguistic rhythm negotiates with musical metre to join the song

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

from lexical to lyrical: how linguistic rhythm negotiates with musical metre to join the song

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)

from lexical to lyrical: how linguistic rhythm negotiates with musical metre to join the song

Sally Ransom, M.A.
The University of Texas at Austin, 2022

Supervisors: Scott Myers
Katrin Erk

This paper takes the performance of lyrical folksongs as an opportunity for an exploratory corpus phonetics study of syllable prominence using the Estonian language, which has three syllable quantities, a predictable stress pattern at the word level and a robust tradition of folksongs *regilaul* which follow a strict metrical pattern. Using a novel semi-automatic computational annotation method, beat detection algorithms used in live music performance are used to determine the location and duration of the strong beats in a sung performance of a song. Combined with acoustic analysis and natural language processing tools, this paper compares rhythmic prominence in songs with the prominence pattern of the language. Using this method, I analyze the acoustic effect that becoming lyrics has on well established acoustic correlates of rhythmic prominence in Estonian compare the attested primary stress system and ternary quantity contrast doc-

umented in Estonian with the prescribed rhythmic pattern of an ancient folksong known as *regilaul*. Results show that vowel duration is utilized as a prominence cue in both the song and word level. We also examine the long and overlong syllable quantity contrast in the context of the song, and find that the contrast between these two is not preserved in the temporal structure of the song.

The paper first introduces prosodic prominence in general, then delves into the metrical specifics of Estonian and its folksong tradition. After outlining the research questions, the methods section describes the corpus of *regilaul* songs: its construction, annotation, and processing. Investigating whether the established acoustic correlate of stress and quantity in spoken Estonian (duration) are preserved after undergoing subordination to the strict temporal properties of the structure of a song.

Table of Contents

List of Tables

List of Figures

Chapter 1

Background

In order to discuss questions regarding word-level and song-level rhythmic prominence in *regilaul*, it is necessary to understand both the perceptual categories and fine-grained acoustic correlates of both linguistic prominence in Estonian and metrical prominence in music.

I first give an overview of suprasegmental features of Estonian speech, detailing fine-grained acoustic-phonetic features of syllable weight and prominence in spoken Estonian alongside results from several perception experiments in native and non-native Estonian listeners. I give an overview of the perceptual and acoustic correlates of metrical prominence in lyrical and non-lyrical music. The Kalevala meter used in *regilaul* is outlined with examples. I discuss previous research at the intersection of song and word prominence in *regilaul*, synthesizing these findings into the research questions and hypotheses that are the subject of this paper.

1.1 Estonian Word Prosody

INTO PROSE domain of syllabic quantity is evident whre the first, stressed syllable increases in duration with increasing degree of quantity, the fol-

lowing unstressed syllable decreases. Lehiste's duration ratios. Different f0 patterns in Q2 and Q3: early peak, dramatic fall in 3, late peak and no fall in 2 (contradicts that asymmetry paper that still saw falls in all three... that could be related to the Q/A issue! Dang.) Laboratory speech usually confirms temporal and tonal characteristics, but conversational speech shows only duration (ratio) as stable, with Q3 fall often absent.

RQ: can quantities in conversational speech be distinguished by acoustics alone, or are listeners making use of semantic context.

disyllabic words from recorded conversational speech presented without context.

Q3: V1 durations had strongest influence on listener decisions, followed by f0 change within V1. duration ratio had a weaker influence, was only significant for stimuli of single speaker. f0 movement across intervo aclic consonant not significant in any case. similar results for recognition of Q3 when presented in combination.

For combined Q1 and Q2, only duration of V1 significant across speakers, duration ratio only relevant for (same as earlier) speaker.

f0 peak position had no significant effect in the cases where it was present.

for good recognition of all three quantities, duration of first vowel important.

differences in V1 duration robust, even in changing speaking rate.

“characteristic” fall in f0 of Q3 neutralized.

listeners did not recognize the majority of Q3 syllables in the absence of context.

certain minimum duration of V1 needed for high recog rate of Q3,

these were all words that could change in meaning with degree of quantity, **?** doctrine of ternary contrasts, lol a: three contrastive segmental lengths b: segment structure *and* prosody of stressed syllable c: in the foot

domain of prosodic patterns is the foot, but only the stucture of the primary stressed syllable is relevant in determining the Q degree of both syllable and foot. (i.e., there is no Q3 foot that does not contain a Q3 syllable as its initial.)

Q1 and Q2 must have at least two syllables, may have three (trochee, dactyl). Having a third syll does not effect quantity. This is in contrast to finnish trisyllables in folksongs, which were 40% longer than disyllables.

Second syllable does not influence the quantity of the first. If the duration of second syllable is predictable from Q of first syllable, this is what phonologists refer to as dependent features. !!!RATIO ARGUMENT

monosyllabic Q3 feet in succession in connected speech *‘khev ‘kõhn ‘poiss ‘läks ‘kepp ‘käes; ‘tõu ‘suur ‘selts ‘kond ‘likkus*

ratio theory initially proposed by Lauri Posti in 1950.

length of the vowel of the second syllable is redundant, dependent, and predictable.

Q3 is monosyllabic foot, making “disyllables” technically trisyllables...

from segments toward long syllables is turning point, segments are a failure (lol) **?**

oot quantity paradigm short-long not separate phonological categories. long monophthongs behave like diphthongs, geminates as consonant clusters

(in english, short and long vowels have different qualities: Wiik 1965

(stressed syllables) Q3 largest area, Q1 most centralized in F1xF2 plane.

when f3 and f4 are removed from spectrum, /i/ is perceived as /ü/, /e/ as /ö/ state that contrast twixt above two vowel pairs on basis of f3. authors say that f3 being close to the strong f2 in round-front is ‘amplified’ the cumulative of f2 and f3.

conclude perceptual param f2 describes well the perceptual phenom governing this contrast.

“effective” f2’ values?? calculate with: Bladon, Fant 1978: 3

long and short vary very little in quality, defining as different phonemes based on length is not justified.

Q3 more “prototypical” or best-contrast version of vowel phonemes in space of stressed syll

in unstress:

Q1 *least* central, unstressed following Q3 init most centralized. V in sylls following Q2 intermediate between others. (ok, they are analyzing these with the "feet" as having the quantity..

/i/ most resistant to centralization

Ross (1990) REASON TO RE-EXAMINE VOWEL FORMANTS/SPACE IN REGILAUL!!! also, data to support "vowel space" as an available acoustic modification (measurable) in singing, evidence favors reduction in word-level-weak syllables AND off-ictus are reduced, but to what extent, and how compared across ictus-stress and ictus-quantity? TLDR are vowels reduced in singing compared to speech, or were those vowels reduced compared with strong positions of song, of word? etc. all these "long" notes are off-ictus: so the shorter notes are corresponding to HEAVIER and STRONGER syllables. singer's formant in untrained female voice very unlikely. measure: LTAS for /a/,/e/, /i/, /u/. SPL of peaks around 3kHz 30-40dB less than that of the first formant in all four vowels.

HOWEVER all vowels selected were in off-ictus– HUAT FIND THIS SONG NOW

READING LIST: Rossing et al 1987- formant SPL in opera and choir singing: singer's formant usually closer to and sometimes converging on f1 SPL.

SAMEAUTHORSAMESONG: no significant timing differences between

performers with respect to note durations (Ross 1989) Ternström and Sundberg 1989: caution against using f3 and f4 standard deviations

standard dev for third and fourth formants less vowel dependent, more "personal," especially compared to standard dev of f2 and f2 f1 f2 which are *fairly* independent of subjects!!!

inverse filter results to confirm T n S 87: f3 lower in singing? saw some similar patterns in with spoken data studies, but overall the size of the variation was small. For the third formant, deviations of the sung vowels from spoken tend to be minor and irregular n = 2

overall f1/f2: sung vowels cluster compared to spoken, specifically: f1 raised in everything but /a/, f2 lower in front, raised in back.

So, gradient modification of vowel quality contrast (less different than in speech).. but I am getting an impression of a larger overall vowel space

1.1.1 syllable prominence

Three proposed levels of stress: primary stress, unstress, and secondary stress. (Lippus et al., 2014)

Primary lexical stress in native Estonian words is fixed, falling on word-initial syllables. Eek (1998)

(1) laul-da
['lɑu:l.dɑ]

a u

sing-TR

‘singing’

(2) ööbik

[ˈøː.pikː]

nightingale.NOM

‘nightingale’

Some loanwords will allow primary stress to fall on a non-initial syllable: *example, example*. Borrowed names occasionally shift stress to the typical Estonian position: for this reason one could find both *Maria* and *Maarja* in the same classroom.

The role of primary stress in Estonian is described by Ilse Lehiste as *identificational* rather than contrastive (Lehiste, 1992). 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.

Spectral tilt suggested, but no **?**Lippus et al. (2014)

Vowels in this position are rarely reduced, and the full inventory of vowels is allowed. A total of 36 diphthongs are allowed in primary stressed positions. All nine vowels can be the first portion of a diphthong, but only

[ɑ e i o u] appear as the second portion of the diphthong (**?**).

- diphthongs
- example

Unstressed syllables, often have reduced vowels, and are more restricted in inventory: only three diphthongs [ɑi ei ui] are allowed in this position (Lippus et al., 2014).

Examined the phonetic correlates of stress in Estonian, but it was small scale and dealt only with nasal flow, amplitude, and duration, and at multiple levels of prosodic hierarchy (**?**), but with as few as two participants. More recently and with more statistical power, researchers measured mean F0, standard deviation of F0, vowel duration, and spectral emphasis. They found increased vowel duration to be the most important acoustic correlate of primary stress in Estonian words, and that Unstressed syllables in Estonian have been documented to attract creaky voice (Lippus et al., 2014). Estonian facts **?**

A pattern of secondary stress (neutral) has been attested, though phonetic evidence is limited (**?**). Only initial and peninitial syllables are examined in the present study.

studies on the perception of these contrasts to be inserted in prose:

(**???**)

1.1.2 syllable quantity

Primary stress position is where the well-known ternary syllable weight contrast in Estonian appears.

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 (Prince, 1980) and as a trimoraic syllable (Hayes, 1989; **?**; **?**).

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

1.2 Emphasis in Songs

1.2.1 strong beats

1.3 Template for *runosong*

Kalevala and regilaul **?**

The singing of folksongs has long been an important part of the Estonian national heritage, and large scale community gathering to sing traditional

regilaul songs is argued to have aided the preservation of Estonian cultural and linguistic heritage through long periods of occupation. (**?**). The first documented song festival in Estonia was held in the seventeenth century (**?**).

A song called *regilaul* is a type of Estonian folksong, found within the greater songwriting tradition of Balto-Finnic language family: variations on these songs are found in Finnish, Karelian, Votic, and Ingrian traditions (**?**).

Collectively, songs in this tradition are often referred to as "runosongs" or "runic songs."

These runic songs are set in what scholars refer to as The "Kalevala" metre, named for the title of the epic poem in Finnish; Estonia's own version is called "Kalevipoeg" *Kalev's son*.

In studies of metre, a trochee is a disyllabic sequence with a strong first syllable and a weak second. In *regilaul* songs, the strong position of a two note sequence is called "ictus," and the weak position is called "off-ictus"

The basic template of the Kalevala metre is four trochees per line, for a total of eight syllables. There are of course variations, but the basic invariant *regilaul* verse line will contain eight beats evenly divided into a single measure: in a 4/4 song, each of the eight syllables corresponds with one eighth note. In this case each eighth note will correspond to one syllable in the text, the constituent henceforth referred to as the "syllable-note" (**?**).

The late composer and *regilaul* revitalist Veljo Tormis referred to the runic songs as “singable songs” (Tormis, 2007).

“collaborative outcomes” in regilaul, the musical rhythm is directly connected to the verse structure (**?**)

Due to the strict metrical parametres of the form, any runosong text can be sung to any runosong melody. An example of this in English: one can sing *Amazing Grace* to the tune of *House of the Rising Sun,* and vice a versa.[1].

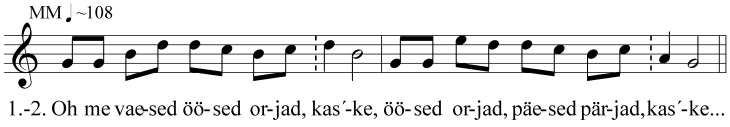

Figure 1.1: Öised orjad, performed by Liisu Orik

In 1.1, we see the musical notation of a typical *regilaul* verse: eight notes evenly divided into a measure (eight eighth notes) and corresponding one syllable per eighth note as indicated by the text. The refrain *kas‘-ke*, which is repeated after every verse line, is part of a separate musical phrase, though it is not a full measure on its own (it contains only three beats). This is indicated by the dashed line following the first measure of the verse. The solid bar after the *kas‘-ke* refrain indicates the onset of a new “regular”

[1]Also try TLC’s “Scrubs” exchanged with America’s “Horse With No Name”

measure, another *regilaul* verse line of eight syllables evenly divided into eighth notes in the measure.

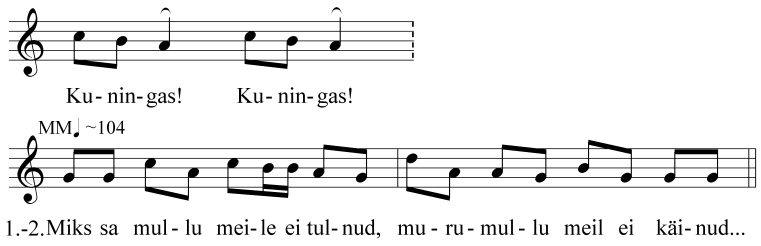

Figure 1.2: 'The King Game" as performed by Liisa Kümmel on track 94 of the anthology

In 1.2 we see an example of extra syllables fit into the Kalevala metre. The first measure in the figure is another refrain. Looking at the verse annotation, we see that there are in fact nine syllables instead of the requisite eight. The measure accomodates the extra syllable by dividing one of the eighth notes into two sixteenths. Thus, the one to one ratio of syllable to note remains, but only seven of the syllable-notes are annotated as isochronous.

Only those syllable-notes which are in evenly divided verse measures are the subject of this study. Non-verse notes, such as those in refrains, as well as notes in verses that occupy a different categorical duration than other notes in its measure are not included.

1.3.1 Previous phonetic studies of Estonian metrics and *regilaul*

(Palmer & Kelly, 1992) hypothesize that prosodic stress and musical metre align in song, but their measurements were only of English vocalists and "western" music. English vocalists were asked to perform songs in which stressed syllables aligned or clashed with the musical meter, finding relative increases in syllable duration in at both word-level and song-level positions of prominence (Palmer & Kelly, 1992), suggesting that the two systems contribute separately but use similar organizational properties.

Ilse Lehiste analyzed spoken poems in Estonian and Finnish, examining the differences in quantity relations under metrical subjugation and found that in strict metrical forms (i.e., the Kalevala metre), the absolute durations of Q2 and Q3 metrical feet lost their contrast, but retained the approximate ratio of durations between the first and second syllable. In freer verse, the differences in absolute durations remained, suggesting that the duration ration between Q2 and Q3 is an important cue for contrast, especially in the contexts of temporal demands of an imposed paralinguistic metre. (Lehiste, 1992) Later, in collaboration with Jaan Ross, three papers examined the phonetics of metrics in old Estonian *regilaul* folksongs. The first (Funeral laments study)(**?**), finding that duration was best predicted by metrical position and not syllable quantity.

They later found that duration was a better predictor of ictus than of word stress: stressed syllables in off-ictus lost their durational contrast with unstressed syllables. (**?**)

In (**?**) Ross and Lehiste measure syllable duration ratios for Q1 initial syllables of disyllabic feet: falling in ictus position and falling in off-ctus position. They found that the duration ratios of Q1 syllables in ictus were greater than those in off-ictus. In ictus position, Q1 syllables were roughly the same length as each other, while those in off-ictus were shortened. They conclude the duration cue for contrastive stress is subordinated to the song. This paper aims to extend the findings of the aforementioned studies (Lehiste, 1992; **?**; **?**; **?**) by annotating a larger corpus of *regilaul*, to compare conflicts of song stress (ictus) and word stress and investigate whether syllable quantity contrast is preserved in song. Vowel duration is an acoustic correlate of both syllable stress and quantity, and so I measure vowel duration in the context of performed *regilaul* songs.

If this question has already been investigated, what motivates the author to pursue it further? First, the aforementioned studies were limited in sample size: each one examined only one word-level interaction with song-level prominence, when both are well documented to influence duration in spoken Estonian, and with each other. Thus, the present study examines all three predictors of prominence: song ictus, word stress, and syllable quantity. I also exponentially increase the *n* of trials for increased robustness, and then analyze the dataset using Bayesian modeling techniques to capture the nuanced covariance of the three predictor categories. In addition to replicating duration measurements undertaken in earlier studies, I also introduce vowel space as a potential predictor for word-level prominence.

- is stress or ictus a better predictor of vowel duration?
- is the ternary syllable quantity contrast preserved in *regilaul* songs?

1.3.2 Design of the Study

Measurements:

vowel duration

vowel dispersion

stressed syllable less reduced, more intelligible,

Fine-grained acoustic-phonetic cues to stress (cross-linguistic) **?????**

1.3.3 On foot isochrony and duration ratios

However, in both cases, the presence of foot isochrony and consistent duration ratios across these three types of feet would not indicate that the foot is the domain of the contrast, nor that the ratio itself is the measurement of importance. Instead, these two findings point to the level of contrast between Q1, Q2, and Q3 syllables at the level of the syllable, and the resulting fluctuations in ratios of first and second syllables as inevitable consequences of preserving the contrast in initial position.

Another consideration is that the three quantities are not restricted to length contrasts of identical segments, in the case of the geminates seen in minimal triads, but also can be comprised of diphthongs and complex

codas *in addition* to geminates.

contain more phoneme segments. The minimal triads illustrate the segmental length contrasts that provide evidence for the three weights of syllables in primary stress position, but they do not adequately convey that the different quantities are not simply different lengths of segments, but that with an increase syllable quantity comes an increase in the quantity of segments contained within that constituent. To illustrate this, the table in **??** contains representations of available segment combinations in each syllable, and examples of Estonian near-minimal triads in disyllabic words.

Chapter 2

Handout

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

Duration can be a phonetic correlate of stress, and it can also be independently contrastive at the segmental level (Lehiste, 1992).

(Lehiste, 1992) investigated the actual phonetic realization of different metres in different languages, finding evidence for one language's trochee, for example, to be realized in a way that is systematically different from another language's trochee.

Ross found vowel reduction in certain notes, but they were all unstressed, off-ictus for that song(Ross, 1990).

The role of primary stress in Estonian is described by Ilse Lehiste as *identificational* rather than contrastive (Lehiste, 1992). 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. (Lippus et al., 2014)

Primary lexical stress in native Estonian words is fixed, falling on word-initial syllables. Eek (1998)

(3) laul-da
[ˈlɑuːl.dɑ]
sing-TR

‘singing’

(4) 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 (Prince, 1980) and as a trimoraic syllable (Hayes, 1989; **?**; **?**).

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

Table 2.1: ternary syllable weight contrast

Chapter 3

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.

3.1 Measurements and Design

3.1.1 segmentation criteria

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.

Stress, accent, emphasis: not one cue but a convergence of cues.

Song structure enforces perceptual isochrony, allows word prominence as long as melodic prominence is met (in singing/piano study)

Two measurements: vowel duration and vowel space

While f0 is a documented cue of prominence in spoken Estonian, the song melody is strict about frequency and pitch. It is less strict about timbre.

3.1.2 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.

3.1.3 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.

In vocal performance, *timbre* is an element free for the singer to use expressively. If we think of vowel quality as an element of *timbre*, then modulating the size of the vowel space is akin to modifying the size of the filter.

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

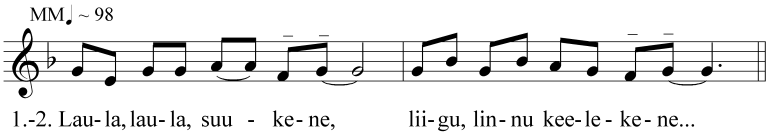

Figure 3.1: Laula! *Sing!*

3.2 Materials

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 **??** is a photograph taken of one of the very field trips to record songs studied in this paper.

Figure 3.2: 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.

3.3 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 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

[1] define lossy

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.

3.4 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: find-

ing 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.

3.5 Study Design

3.5.1 Questions and Hypotheses

3.5.2 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.

3.5.3 Statistical Analysis

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

??

Bibliography

Eek, M. E., Arvo (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.

Hayes, B. (1989). Compensatory Lengthening in Moraic Phonology. *Linguistic Inquiry*, *20*(2), 253–306.

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

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.

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

Prince, A. S. (1980). A Metrical Theory for Estonian Quantity. *Linguistic Inquiry*, *11*(3), 511–562.

Ross, J. (1990). Formant frequencies in estonian folk singing.

Tormis, V. (2007). Some_problems_with_that_regilaul.pdf. In *RING*.

Vita

Sarah Marie Ransom-Laud was born in Sarasota, Florida on 11 July 1990. Her early career was as a songwriter and recording artist, releasing several full-length albums combining digital and analog recording methods over the last decade.

In 2018 she received the Bachelor of Arts degree in Linguistics from the University of Wisconsin, Milwaukee. Upon graduation, she moved to Austin with her partner, Kavi, and taught English as a Second Language (ESL) to adults until her admission into the PhD program in the department of Linguistics at the University of Texas at Austin, where she matriculated in Fall 2019.

Permanent address: 305 E. 23rd Street STOP B5100
Austin, Texas 78712

This thesis was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.