CROSS-LINGUISTIC COMPARISON OF RHYTHMIC AND PHONOTACTIC SIMILARITY

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI‘I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
IN
LINGUISTICS

DECEMBER 2013

By
Diana Stojanović

Dissertation Committee:
Ann M. Peters, Chairperson
Patricia Donegan
Victoria Anderson
Kamil Ud Deen
Kyungim Baek
ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to all who provided inspiration, guidance, help, love, and support during my journey at the Department of Linguistics.

Members of my dissertation committee Professors Ann M. Peters, Patricia J. Donegan, Victoria B. Anderson, Kyungim Baek and Kamil Ud Deen;

All professors at the Department of Linguistics and in particular Byron Bender, Bob Blust, Mike Forman, William O’Grady, Ken Rehg, Albert Shutz, David Stampe, Ben Bergen, Katie Drager, Luca Onnis, Yuko Otsuka, and Amy Schafer;

Department secretaries who made impossible possible: Jen Kanda and Nora Lum;

Classmates and officemates: among many, Kaori Ueki, Yumiko Enyo, Gabriel Correa, Karen Huang, Laura Viana, Tatjana Ilic, Maria Faehndrich, Kathreen Wheeler, and Mie Hiramoto;

East-West Center and in particular Prof. Andrew Mason;

Udacity for teaching me enough Python to support this dissertation;

EWCPA and wonderful neighbors in Hale Kuahine;

Graduate Division, GSO, ISS, and in particular Martha Stuff and Linda Duckworth;

Family away from home Nelda Peterson, Christobel Sanders, Nina, Jo, and Kano;


My family and in particular my grandparents who spoke different languages and instilled the love for language in me; and my parents who supported me unconditionally;

And my dear husband Turro Wongkaren:

THANK YOU
ABSTRACT

Literature on speech rhythm has been focused on three major questions: whether languages have rhythms that can be classified into a small number of types, what the criteria are for the membership in each class, and whether the perceived rhythmic similarity between languages can be quantified based on properties found in the speech signal.

Claims have been made that rhythm metrics – simple functions of the durations of vocalic and consonantal stretches in the speech signal – can be used to quantify rhythmic similarity between languages. Despite wide popularity of the measures, criticisms emerged stating that rhythm metrics reflect differences in syllable structure rather than rhythm.

In this dissertation, I first investigate what kind of similarity is captured via rhythm metrics. Then, I examine the relationship between the assumed rhythm type and the language structural complexity measured by the distributions of 1) consonant-cluster sizes, 2) phonotactic patterns, and 3) word lengths. Materials on which the measures of structural complexity were computed were automatically transcribed from written texts in 21 test languages. The transcriber is implemented in Python using grapheme-to-phoneme rules and simple phonological rules. Complexity measures are calculated using a set of functions, components of the complexity calculator.

Results show that several rhythm metrics are strongly correlated with the phonotactic complexity. In addition, linear relationship found between some metrics suggests that the information they provide is redundant. These results corroborate and extend results in the literature and suggest that rhythmic similarity must be measured differently. Structural similarity in many cases points to historical language grouping. Similarity of word-final clusters arises as a factor that most resembles rhythmic classification, although a large body of independent evidence of rhythmic similarity is necessary in order to establish this correspondence with more certainty.

Based on the results in this dissertation and the literature, a possible model of rhythmic similarity based on feature comparison is discussed, juxtaposing the current model based on rhythm metrics. This new ‘Union of features’ model is argued to better fit the nature of rhythm perception.
TABLE OF CONTENTS

Acknowledgements ...................................................................................................................iii
Abstract .................................................................................................................................iv
List of tables .........................................................................................................................viii
List of figures ........................................................................................................................x

CHAPTER 1: INTRODUCTION ...........................................................................................1
  1.1 Rhythm correlates .........................................................................................................1
  1.2 Rhythm class hypothesis (RCH) ................................................................................2
  1.3 Rhythm metrics ............................................................................................................4
  1.4 Issues present in the current literature .......................................................................8
  1.5 Questions and approaches used to solve them ..........................................................9
  1.6 Contribution of this Dissertation .............................................................................11
  1.7 Outline .......................................................................................................................11

CHAPTER 2: BACKGROUND ...........................................................................................13
  2.1 Durational variability in speech ................................................................................13
  2.2 Phonotactics: Sonority scale and markedness of consonant clusters ......................20

CHAPTER 3: METHODS ....................................................................................................23
  3.1 Model of the transcriber & the phonotactic calculator ................................................23
  3.2 Raw data assembly .......................................................................................................24
  3.3 Creating phonemic corpora .......................................................................................24
  3.3.1 Choice of grapheme-to-phoneme method ...........................................................27
  3.3.2 Implementation of grapheme-to-phoneme method .............................................28
  3.4 Complexity calculator ...............................................................................................29
  3.4.1 Photactic metrics and rhythm metrics ...............................................................29
  3.4.2 Consonant-cluster measures ...............................................................................34
  3.4.3 Word-length measures .......................................................................................35
CHAPTER 4: RESULTS ......................................................................................................37

4.1 Phonotactic component of Rhythm Metrics .................................................................37
   4.1.1 Introduction ........................................................................................................37
   4.1.2 Correlations between the Phonotactic and Rhythm Metrics ................................38
   4.1.3 Classification power of RMs and PMs ..................................................................43
   4.1.4 Language classification based on Phonotactic Metrics .......................................46
   4.1.5 Conclusion ..........................................................................................................45

4.2 Consonant cluster lengths at different positions in the word .....................................50
   4.2.1 Word-initial cluster distributions .......................................................................50
   4.2.2 Word-final cluster distributions .........................................................................52
   4.2.3 Word-medial cluster distributions .....................................................................55
   4.2.3 Summary .............................................................................................................56

4.3 Phonotactic patterns at different positions in the word .............................................58
   4.3.1 Basic sonority (ALT) level .................................................................................59
   4.3.2 Detailed sonority (saltanajc) level .....................................................................67

4.4 Word length distributions .........................................................................................76

4.5 Variability of measures over different materials .......................................................81

CHAPTER 5: GENERAL DISCUSSION AND CONCLUSION .............................................89

5.1 Summary .....................................................................................................................89
5.2 Overview .....................................................................................................................91
5.3 Limitations of the study ............................................................................................92
5.4 Discussion ..................................................................................................................95
   5.4.1 Additional questions .........................................................................................95
   5.4.2 Use of modified speech in addressing questions on rhythmic similarity ..........98
   5.4.3 The nature of rhythm .......................................................................................100
   5.4.4 Proposed model of rhythmic similarity .............................................................101
   5.4.5 Implications/prediction for L2 speech and learning in infants .........................102
5.5 Conclusion .................................................................................................................104
APPENDICES

Appendix 1: Basic properties of the languages from WALS ........................................105
Appendix 2: Texts and transcripts for 21 languages..................................................108
Appendix 3: Values of Rhythm and Phonotactic Metrics .......................................131
Appendix 4: Word-length distributions for 21 languages.........................................134

BIBLIOGRAPHY .........................................................................................................145
# LIST OF TABLES

Table 4.1 Correlation between the consonantal PMs and RM$s$ .....................................38
Table 4.2 Correlation between the vocalic PMs and RM$s$ (long=1)..............................39
Table 4.3 Correlation between the vocalic PMs and RM$s$ (long=2)..............................39
Table 4.4 Distribution of word-initial consonant clusters..............................................51
Table 4.5 Distribution of word-final consonant clusters ...............................................53
Table 4.6 Distribution of word-medial consonant clusters.............................................55
Table 4.7 Distribution of word-medial consonant clusters (re-arranged)......................56
Table 4.8 Language groupings based on word-initial, word-medial, and word-final complexity.............................................................................57
Table 4.9 Word-initial length-0 and length-1 clusters ...................................................60
Table 4.10 Word initial length-2 clusters.........................................................................61
Table 4.11 Word-initial length-3 clusters ........................................................................62
Table 4.12 Word-final length-0 and length-1 clusters .....................................................64
Table 4.13 World-final length-2 clusters.........................................................................64
Table 4.14 Word-final clusters grouped by sonority .......................................................65
Table 4.15 World-final length-3 clusters .........................................................................67
Table 4.16 Saltanajc frequencies in 21 languages ...........................................................68
Table 4.17 Clusters of length-2 in ‘saltanajc’ scale: initial position...............................70
Table 4.18 Clusters of length-2 in ‘saltanajc’ scale: initial position...............................71
Table 4.19 Clusters of length-2 in ‘saltanajc’ scale: final position..................................72
Table 4.20 Clusters of length-2 in ‘saltanajc’ scale: word-final position ........................73
Table 4.21 Clusters of length-2 in ‘saltanajc’ scale: word-medial position.....................74
Table 4.22 Variability of phonotactic metrics over different texts ..................................82
Table 4.23 Variability of consonant cluster complexity in word-initial position.........84
Table 4.24 Variability of consonant cluster complexity in word-final position ..............84
Table 4.25 Variability of consonant cluster complexity in word-final position ..............85
Table 4.26 Variability of word-length distribution: word tokens.................................85
Table 4.27 Variability of word-length distribution: lexical items .................................86
Table A1.1 Phonological properties of test-languages .....................................................106
Table A1.2 Morphological properties of test-languages.................................................107
<p>| Table A3.1 | Rhythm Metrics values .................................................................131 |
| Table A3.2 | Phonotactic Metrics values (long V = 1) ..............................................132 |
| Table A3.3 | Phonotactic Metrics values (long V = 2) ..............................................133 |</p>
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Sonority scale (Vennemann 1988)</td>
<td>20</td>
</tr>
<tr>
<td>2.2</td>
<td>saltanaj sonority scale</td>
<td>20</td>
</tr>
<tr>
<td>2.3</td>
<td>ALT sonority scale</td>
<td>21</td>
</tr>
<tr>
<td>2.4</td>
<td>Preferred initial double clusters (Dziubalska-Kołaczyk 2001)</td>
<td>22</td>
</tr>
<tr>
<td>2.5</td>
<td>Preferred medial double clusters (Dziubalska-Kołaczyk 2001)</td>
<td>22</td>
</tr>
<tr>
<td>2.6</td>
<td>Preferred final double clusters (Dziubalska-Kołaczyk 2001)</td>
<td>22</td>
</tr>
<tr>
<td>3.1</td>
<td>Model of the transcriber</td>
<td>23</td>
</tr>
<tr>
<td>3.2</td>
<td>Model of the complexity calculator</td>
<td>24</td>
</tr>
<tr>
<td>3.3</td>
<td>Example: cluster distribution at ALT level</td>
<td>35</td>
</tr>
<tr>
<td>3.4</td>
<td>Example: cluster distribution at saltanajc level</td>
<td>36</td>
</tr>
<tr>
<td>4.1</td>
<td>Correlation between phonotactic (%) and rhythmic (%) of vocalic intervals</td>
<td>40</td>
</tr>
<tr>
<td>4.2</td>
<td>Correlation between phonotactic (Δ) and rhythmic (Δ) of consonantal intervals</td>
<td>40</td>
</tr>
<tr>
<td>4.3</td>
<td>Correlation between phonotactic (Δ) and rhythmic (Δ) of vocalic intervals</td>
<td>41</td>
</tr>
<tr>
<td>4.4</td>
<td>Correlation between phonotactic (Varco-Δ) and rhythmic (Varco-Δ) of consonantal intervals</td>
<td>41</td>
</tr>
<tr>
<td>4.5</td>
<td>Correlation between phonotactic (Varco-Δ) and rhythmic (Varco-Δ) of vocalic intervals</td>
<td>42</td>
</tr>
<tr>
<td>4.6</td>
<td>Correlation between phonotactic (nPVI-Δ) and rhythmic (nPVI-Δ) of vocalic intervals</td>
<td>42</td>
</tr>
<tr>
<td>4.7</td>
<td>Phonotactic metrics graph (%Δ, ΔCp)</td>
<td>43</td>
</tr>
<tr>
<td>4.8</td>
<td>Rhythm metrics graph (%Δ, ΔCr)</td>
<td>44</td>
</tr>
<tr>
<td>4.9</td>
<td>Rhythm metrics graph (rPVI-Cr, nPVI-Vr)</td>
<td>45</td>
</tr>
<tr>
<td>4.10</td>
<td>Phonotactic metrics graph (rPVI-Cp, nPVI-Vp)</td>
<td>45</td>
</tr>
<tr>
<td>4.11</td>
<td>Grouping of 21 languages based on phonotactic %Δ and ΔCp</td>
<td>47</td>
</tr>
<tr>
<td>4.12</td>
<td>Linear relationship between %Δ and ΔCp</td>
<td>48</td>
</tr>
</tbody>
</table>
Figure 4.13 Grouping of 21 languages based on phonotactic metrics
rPVI-Cp and nPVI-Vp ................................................................. 48
Figure 4.14 Grouping of 21 languages based on phonotactic %Vp and Varco-Vp ........49
Figure 4.15 Distribution of word-final clusters based on sonority .............................. 66
Figure 4.16 Distribution of word-lengths: lexical items ............................................. 78
Figure 4.17 Distribution of word-lengths: word tokens ............................................. 79
Figure 4.18 Average word-length: word tokens ......................................................... 80
Figure 4.19 Average word-length: lexical items ......................................................... 80
Figure 4.20 Stability: Distribution of word lengths (lexical items) ............................ 87
Figure 4.21 Stability: Distribution of word-lengths (word tokens) ............................ 88
Figure 5.1 An example of a characteristic prosodic sequence .................................. 108
Figure A4.1 Distribution of word lengths for Bulgarian ........................................... 134
Figure A4.2 Distribution of word lengths for Catalan ................................................ 134
Figure A4.3 Distribution of word lengths for Czech .................................................. 135
Figure A4.4 Distribution of word lengths for Dutch .................................................. 135
Figure A4.5 Distribution of word lengths for Estonian .............................................. 136
Figure A4.6 Distribution of word lengths for German .............................................. 136
Figure A4.7 Distribution of word lengths for Greek .................................................. 137
Figure A4.8 Distribution of word lengths for Hawaiian ............................................ 137
Figure A4.9 Distribution of word lengths for Hungarian .......................................... 138
Figure A4.10 Distribution of word lengths for Indonesian ........................................ 138
Figure A4.11 Distribution of word lengths for Italian .............................................. 139
Figure A4.12 Distribution of word lengths for Japanese ............................................. 139
Figure A4.13 Distribution of word lengths for Maori ............................................... 140
Figure A4.14 Distribution of word lengths for Polish ................................................ 140
Figure A4.15 Distribution of word lengths for Portuguese ........................................ 141
Figure A4.16 Distribution of word lengths for Russian ............................................. 141
Figure A4.17 Distribution of word lengths for Samoan ............................................. 142
Figure A4.18 Distribution of word lengths for Serbian ............................................. 142
Figure A4.19 Distribution of word lengths for Spanish ............................................. 143
Figure A4.20 Distribution of word lengths for Tongan ............................................. 143
Rhythm of speech has been vigorously discussed during the last century and it continues to
be a topic of research in phonetics, phonology, language acquisition, and the similarity
between speech and music. In linguistics literature, research on rhythm was first seen in
description and analysis of poetry and rhythmic meter. The topic gained controversial status
following Pike’s 1945 observation that English and French differ in their rhythms, describing
the former as more Morse code-like and the latter machine-gun-like. Pike’s statement –
however misunderstood, as he actually does not juxtapose them but says English has possibly
two rhythms, one resembling French (Pike 1945) – has been generally interpreted as a
dichotomy of two types of rhythms, which led to an even stronger claim by Abercrombie
(1967) that all spoken languages must be of one or the other type. At this time, only a small
subset of world languages was considered, so this conjecture was indeed very brave.

Since then, Pike’s and Abercrombie’s statements, later formulated the Rhythm class
hypothesis (RCH), have been studied, refuted, reformulated, and studied again. Approaches
included measuring stress-intervals and syllable durations in the speech signal to test
isochrony, phonological analysis of prosodic shortening and lengthening, phonetic inventory
(long vowels and diphthongs) and phonotactic (consonant clustering) comparisons, and –
having abandoned isochrony formulation – measuring variability in physical speech signal.

1.1 Rhythm correlates
Despite the reference to ‘strong and weak’, which suggests there is more to rhythm than
‘long and short’, and despite the fact that most researchers, even inadvertently, use the term
‘prominent’, they only measure ‘long’; in other words, most quantitative studies are related
to measures of durations.

One repeating theme that we can observe is that quantitative studies of rhythm have
been based mostly on durational analysis, seeking the variability or alternation of long and
short beats. Only a handful of studies on speech rhythm, such as Cumming (2010) and
Grabe&Low (2002), investigated the contribution of other parameters to rhythm, even
though describing prominence often relies on *pitch emphasis* (sentence prominence of ToBI, Barry et al. 2009) or *rate of spectral change* (Kochanski et al. 2005). The dominance of the durational parameter in describing rhythm could be related to the following issues: 1) first description of ‘Morse-code’ (supposedly an alternation of longs (dashes) and shorts (dots)) vs. ‘machine-gun’ (supposedly repetition of equal duration), 2) subsequent formulation of rhythm via *isochrony* (thus related to time) of units, 3) the fact that expression of prominence is not well-defined in terms of physical parameters of the speech signal, and 4) the fact that the speech rhythm itself is not always understood the same way across studies and researchers, or as it was aptly put by Cummins, ‘Much like God, Tidiness, and Having a Good Time™, the concept of rhythm means many things to many people’ (Cummins 2012). The last two tasks, defining speech rhythm and understanding how parameters figuring in that definition are expressed in the speech signal, should be seriously taken in the future research.

1.2 Rhythm class hypothesis (RCH)

In one of the initial formulations of the RCH, Abercrombie (1967) claimed that two posited rhythmic classes differ in the type of isochrony the languages exhibit: isochrony of syllables for syllable-timed languages and isochrony of inter-stress intervals for stress-timed languages. This grouping was based on the perception of salient rhythmic differences between languages such as English or Dutch, called stressed-timed, and languages such as Spanish or French, called syllable-timed. Two basic proposals were made (Abercrombie 1967): (1) that in the stressed-timed languages stressed syllables occur regularly, while in the syllable-timed languages syllables occur regularly, and (2) that languages of the world belong to one or the other class. A third class, *mora-timed*, was later added to accommodate languages such as Japanese that were believed to differ from both of the existing types in that moras occur regularly. Languages included in this group usually have phonological length distinctions.

Because attempts to find evidence of isochrony in the characteristic unit in the speech signal failed (Dauer 1983, Roach 1982), it was proposed that isochrony is a purely perceptual phenomenon (Lehiste 1977) that could not be measured from the acoustic signal. An alternative view of rhythm was put forward by Dauer (1983), who noticed phonological
similarities among the original members of the stress-timed group and the syllable-timed group respectively: languages in the stress-timed group have vowel reduction in unstressed syllables and phonotactics that allows complex syllable structure; syllable-timed languages lack vowel reduction and have simple (C)V(C) syllable structure.\(^1\)

Dauer’s 1983 model of rhythm posits that rhythm emerges from phonological properties such as syllable structure and vowel reduction, as well as duration of stressed syllables, phonemic vowel length distinction, the effect of intonation on stress, the effect of tone on stress, consonantal phonetic inventory, and function of stress.\(^2\)

The more of these properties a language has, the more stress-timed it is proposed to be. Languages thus are said to lie on a continuum between prototypically syllable-timed at one end (Japanese) and prototypically stress-timed (English) on the other end of continuum. This means that various properties combine, possibly with various levels of importance, towards one resultant perception variable: rhythm. Because the listed properties do not always co-occur, in this view two languages can have different properties but be equally syllable- or stress-timed. In this model, called ‘continuous uni-dimensional model of rhythm’ (Ramus 2002), a strict rhythm-class hypothesis is not true. Instead, languages form a rhythm continuum.

The continued interest in the status of the RCH, specifically, the interest in determining whether different rhythmic types exist, and if so how properties of each class can be described and measured, likely comes from two postulates. First, languages are believed to be distinguishable based on rhythmic properties alone.

Results reported in the literature of the perception experiments with infants (Nazzi et al. 1998, Nazzi and Ramus 2003) and adults (Ramus et al. 2003) seem to show that discrimination of languages is possible when segmental and melodic information are hidden, but only if the languages belong to different rhythm types.

---

\(^1\) Mora-timed languages were not discussed in Dauer 1983; based on the properties of Japanese, Yoruba, and Telugu, the mora-timed group has a syllable structure even simpler than that of the syllable-timed group, namely, mostly (C)V. The (C)VC\(_1\) exists but allows only few selective consonants in \(C_1\) position.

\(^2\) Donegan (personal communication) suggests that the following are important: tendency to diphthongize, vowel harmony, geminate consonants, vowel-length distinctions, many vs. few vowel quality distinctions, contour vs. level tone.
Secondly, speech-processing strategy, namely segmentation into words, used by infants during early language acquisition, is suggested to depend on the rhythm class of the first language (Cutler and Noris 1988).

That a processing strategy is depended on the rhythmic characteristics of the language, lies on several assumptions: 1) languages differ rhythmically, 2) there is a finite number of rhythm types or classes, 3) characteristics of the rhythm type can be inferred from speech by young infants who have limited language exposure.

Each of the two postulates represents a strong support for the existence of rhythm types, and thus justifies the search for the ways to quantify characteristics of each class. However, at least one of the two needs to be true, that is, without counter evidence.

The first claim, that languages can be discriminated along the separation of rhythm classes, has recently been challenged. Arvaniti and Ross (2010) fail to replicate language discrimination results for English, German, Greek, Italian, Korean, and Spanish, and suggest that their results ‘cast a doubt in the impressionistic basis of the rhythm class hypothesis.’ However, it is possible that the difficulty of the task involved (different from the original studies) contributed to the lack of consistent results.

I will not discuss the second claim in detail here, but will only note that the burden is on proving that it is exactly the speech rhythm and not some other characteristic along which languages in the reported experiments differed, that are responsible for the emergence of the segmentation strategy. Experiments in support of this theory have so far been conducted on a limited number of languages.

1.3 Rhythm metrics

An alternative approach to quantifying speech rhythm started with a 1999 study by Ramus et al. in which rhythmic differences are seen as differences in durational variability of vocalic and consonantal clusters. Numerous studies were conducted with various levels of success in finding empirical evidence for the existence of rhythm classes.

As mentioned in the previous section, both infants and adults discriminate among languages based on rhythmic properties, which supports the view that languages can be grouped into different types. Because of that, the idea of rhythmic classes has persisted
despite the failure to find measurable evidence of isochrony in the speech signal. Renewed interest in finding evidence for rhythmic differences has occurred with a shift in focus: the distinction of rhythmic classes is not based on isochrony, or lack thereof, among successive units, but is based on a somewhat more relaxed criterion: degree of durational variability among such units. Another difference introduced with the new approach involved a change of unit whose variability is used to characterize rhythm. Instead of syllables and feet (or intervals between two stresses), non-phonological units such as vocalic and consonantal intervals\(^3\) were used. The change of unit was motivated by the results of the studies on infant perception of rhythm. Namely, Nazzi et al. (1998) found that infants, like adults, perceive rhythmic\(^4\) differences between languages. It was shown that infants are able to distinguish speech samples of English from those of French, for instance, but not English from Dutch. Ramus et al. (1999:270) assume that ‘the infant primarily perceives speech as a succession of vowels of variable durations and intensities, alternating with periods of unanalyzed noise (i.e. consonants)’ and suggest that perceiving rhythmic differences must not be based on phonological units such as syllables and feet.

The focus of the new approach was on the formulation of a two-dimensional space in which good exemplars of stress-timed and syllable-timed languages would be separated. Such spaces were defined most of the time by measures that in some way mirror distinguishing phonological properties. Various measures were introduced in the hope of capturing the crucial differences between posited rhythm classes.

The early results were encouraging in that prototypically stress-timed and syllable-timed languages were mapped into opposite corners of the space (Ramus and Mehler 1999, Ramus et al. 1999, Grabe and Low 2002). In subsequent studies, in which larger numbers of speakers per language and new languages were tested with various speech materials and speech styles, it was found that (1) empirical results show more support for Dauer’s continuum hypothesis than for the strict rhythm class hypothesis (Grabe 2002), and (2) various factors compromise successful cross-linguistic classification. Several serious problems of the quantitative approach based on rhythm measures include the following: (1)

\(^3\) In the literature, these are called vocalic and intervocalic intervals.
\(^4\) It was posited that the differences are solely rhythm-based because the samples were filtered to eliminate segmental information
within-language inter-speaker differences may be larger than between-class differences (Benton et al. 2007); (2) speech rate (Dellwo and Wagner 2003) and speech style (Benton et al. 2007) may affect metric values more than the posited rhythm class; (3) different metrics produce contradictory classifications: for instance, $\Delta V$ and $\Delta C$ classify Polish differently (Ramus et al. 1999); (4) different studies obtain contradictory results based on the same rhythm metric (Dellwo 2006, White and Mattys 2007); and (5) rhythm metrics depend on the segmentation rules (Stojanovic 2008).

Others (Dellwo and Wagner 2003) report that discrimination of languages using some rhythm metrics can be explained based on speech rate alone, namely that languages spoken faster are classified as syllable-timed and those spoken slower as stress-timed. This is consistent with the finding that speech rates vary cross linguistically and that those traditionally labeled as syllable-timed have higher average speech rates. Moreover, some measures vary with speech rate more than others, $\% V$ reportedly being more stable across rates than the standard deviations of vocalic and consonantal intervals.

In a subsequent study, Dellwo (2006) shows that another measure, coefficient of variation of consonantal intervals, which is normalized standard deviation, varies more for some languages than for others, in their sample, it varies more for stress-timed English (strangely, at fastest rate it returns to the value given for slow rates) and German than for syllable-timed French.

Szakay (2008), on the other hand reports that two varieties of the same language can be affected differently by the speech rate. Namely, in her study of Maori and Pakeha speakers of English in New Zealand, she finds that Maori English is classified as syllable-timed for all speech rates in the sample (roughly equal as rates for Pakeha English), while Pakeha English changes from being stress-timed at lower speech rates to more-syllable timed at faster rates. Classification in her study is based on pair-wise variability of vocalic intervals, not on perception. Thus, we can say that it is the type of variability that pair-wise index measures that changes, rather than possibly rhythm type. The findings of the three studies (Wagner and Dellwo 2003, Dellwo 2006, and Szakay 2008) are nevertheless important. They suggest that the timing differences are better observed at slower rates, at least for the two
varieties of English examined. Another important piece of information from these studies is that even so-called normalized rhythm metrics may vary across speech rates.

Recent proposals include another view of stress- vs. syllable-timing. Nolan and Asu (2009), for instance, propose that stress-timing and syllable-timing are independent dimensions exemplified by all languages, and so one language can express a certain level of stress-timing and a certain other level of syllable-timing.

There is no consensus in the current literature on whether rhythm can be measured from the acoustic signal, or whether it is different from timing (Arvaniti 2009). In fact, some (Pamies Bertrán 1999) propose that speech is not rhythmic at all. Others view rhythm as coupling between nested prosodic units (Cummins 2002).

Loukina et al. (2011) examined 15 rhythm metrics on a large speech corpus for five test languages: British English, French, Greek, Russian, and Mandarin. They found that no metric was successful overall in separating languages, different metrics being successful for different pairs. If three metrics were used at the time however, it was possible to discriminate all five languages.

I will discuss several important conclusions from their study. First, they note that different metrics were necessary for discrimination of different pairs of languages, suggesting that languages differ rhythmically in different ways. This observation does not seem problematic as rhythm possibly varies along several dimensions, where those dimensions are captured by different metrics.

Next, authors conclude that their results do not support RCH based on three traditional classes as all five languages were discriminated, albeit with use of several metrics. In addition, different metrics produced different groupings. Their results could be interpreted differently, however: not contradicting the existence of three rhythm classes, but taken as evidence that RMs they used are not (all) good classifiers of rhythm, some possibly capturing it better and others not so well.

Authors further mention an important shortcoming of RMs obtained on short speech samples, namely that they differ when calculated for different texts or when same text is
produced by different speakers. To address this issue, they used a large speech corpus and performed automatic speech segmentation into vocalic and consonantal intervals.

The issue of variation of RMs across materials for the same language (including the sample in their corpus) is an important one, however, it is worth noting that their automatic segmenter had varying success in segmenting vowels and consonants in the corpus. Namely, using their best (most consistent with human labelers) algorithm, a large proportion (close to 90%) of voiceless obstruents, but not all, were classified as consonants. In all languages but English, voiced obstruents were equally likely classified as vowels as consonants (in English, 75% of voiced obstruents were classified as consonants). Sonorants were classified as vowels 77-91% of the time, and vowels were classified as vowels 88-92% of the time. Given the discrepancy in labeling basic units, it is unlikely that RM obtained on the automatically labeled set will reproduce the values of RM obtained in studies that used hand-labeled materials – even if the acoustic data were exactly the same. Loukina et al. rationalize that automatic segmentation is a more consistent process across languages than manual segmentation, as no language-specific rules are used in deciding how to label a particular segment; it also captures similarities between sonorant and voiced segments across languages. However, it is also possible that the similarity between these groups of segments is a result of the algorithm relying on voicing and energy, rather than these segments behaving differently with respect to rhythm. Whether they do indeed behave differently is an interesting empirical question.

1.4 Issues present in the current literature

So, can durational patterning successfully capture rhythm? And how large is the ‘noise’ coming from the phonotactic component? My first question addresses these issues.

Next, in order to understand the effect of phonotactics better, I examine phonotactics at different positions in the syllable. This is an interesting question because onsets and codas are reported to affect the duration of nuclei in different ways. Do languages with similar patterns in syllable onset tend to be similar rhythmically? Or is it the structure of the coda that has more effect on rhythmic similarity? To answer these questions, I look at phonotactic
probability\textsuperscript{5}, that is, I characterize onsets and codas by likelihoods of 0) being empty (no coda) (as in ‘no’) or 1) having one consonant (as in ‘mitt’ /mɪt/), or 2) two consonants (as in ‘pest’ /pɛst/) or 3) three consonants (as in ‘angst’ /æŋst/).

While durational patterns in the sample are affected by cluster lengths, similarity between two languages can also be captured by similarity of actual phonotactic sequences that occur. This does not refer to similarity addressed in perception experiments using modified speech, but instead in comparison of unaltered samples, that is, normal speech. Do languages group into similar classes based on phonotactic patterns? If so, are such groupings related to the posited rhythm classes?

Another interesting question to address is the relation between cluster markedness (defined in Chapter 2) and its frequency in the corpora of individual languages. Typologically, less marked clusters are assumed to occur in a larger number of languages. We can check this, although we should be cautious because our language set is possibly too small and too biased towards certain language families. However, we can ask a different question, namely, do less marked clusters occur more frequently than more marked clusters in the corpus of each language? And does the presence of a more marked cluster in a language imply the presence of less marked ones?

Lastly, we are interested in similarity that might stem from word lengths. This parameter is related to durational properties as it has been shown that in most languages, average syllable duration is reduced when the number of syllables increases, or ‘longer words have shorter syllables’. If word length is defined as the number of comprising segments, or, alternatively as number of comprising syllables, is there a relationship between word-length distribution and cluster length in onsets or codas? Word length and phoneme inventory size? Word length and rhythm type?

1.5 Questions and approaches used to solve them

In order to address these questions I first assembled a multi-lingual phonetic corpus from (freely-available electronic) texts converting from orthography by applying rules (see Methods).

\textsuperscript{5} Phonotactic probability refers to the frequency with which phonological segments and sequences of phonological segments occur in words in a given language. (Vitevitch and Luce 1999)
My first research question examines the extent to which the current ways of quantifying rhythmic differences are affected by the phonotactic properties of a language. In particular, here I focus on the length of the phonotactic sequences, that is, number of segments in vocalic and consonantal clusters. Neutralizing any differences in duration that come from a segment’s inherent duration and prosodic emphasis, I calculate a set of metrics based on number of segments instead of real durations and compare them to the durational metrics reported in the literature. I calculate correlations between these phonotactic metrics (PMs) and durational rhythm metrics (RMs); I also observe groupings on the graphs in the literature reported as evidence of language grouping into rhythm types. Possible outcomes are that 1) correlation between RMs and PtMs is very small; therefore the effect of the phonotactic component on the RMs is minimal; i.e., RMs are good correlates of ‘pure rhythm’; or 2) correlation between RMs and PtMs is large; therefore the effect of the phonotactic component on the RMs is large enough to question the use of RMs as a measure of pure rhythm.

My second question addresses the relationship between rhythmic similarity and the length of clusters at different positions in the syllable (onset or coda). However, due to difficulties in syllabification of medial consonant clusters, I have modified the question to examine the relationship between rhythmic similarity and the length of clusters at different positions in the word (word-initial, word-medial, and word-final).

Note that in this question we address only the cluster length, that is, whether a cluster consists of zero, one, two, or more segments; we can call the constraints on how many segments can occur in a cluster at each position in the word durational phonotactics, as it is defined by constraints on sequences but ignores the segmental qualities. Contribution of segmental qualities to cluster durations is not taken into account in this question. We are interested in answering whether any particular position (initial, medial, or final) is particularly important in explaining rhythmic (perception) similarity. In fact, as we compare cluster length distributions, this question addresses probabilistic durational phonotactics to determine similarity across languages.

My third question then focuses on the actual patterns in the clusters and asks to what degree similarity based on the most common cluster patterns is related to perceived similarity.
between languages. Here, however, we are **not** interested in posited rhythmic similarity alone, which is usually tested with segmental qualities hidden (filtered or re-synthesized speech samples). Segmental qualities are now taken into account; this question examines contribution of the segmental component, or **pattern phonotactics**, to the perception of overall similarity between two spoken samples.

My fourth question examines the relationship between the nature of clusters and their frequencies in a particular language. Specifically, we ask whether more marked clusters are less frequent than less marked clusters in each language. Definition of cluster markedness is given in Chapter 2.

My fifth and last question is directed toward the relationship between average word length in the sample and cross-linguistic rhythmic similarity, as well as the relationship between average word length in the sample and the cluster length similarity that was examined in question 2.

### 1.6 Contribution of this Dissertation

This dissertation asks and answers several questions about interplay of rhythm, phonotactics, and perception. The results will contribute to the fields of phonology and speech perception, in particular to the current discussions about quantifying speech rhythm; specifically they will have impact on how the current measures need to be redefined in order to measure rhythmic similarity. The proposed model of rhythmic similarity provides a place for a significant amount of future work, and is a strong alternative to the current geometric model based on rhythm metrics.

The discussion of phonotactic similarity is informative for the field of language typology, especially because it will be related to rhythmic similarity and basic phonological and morphological properties. The use of probabilistic phonotactics allows us to show finer differences in linguistic structure cross-linguistically. It is hoped that the tabulated phonotactic distributions for individual languages, as well as the assembled phonetic corpus, will be used as a reference in other studies.
1.7 Outline

This Dissertation is organized as follows. In Chapter 2 I provide background and introduce necessary concepts and terms that will be used throughout this work. In Chapter 3 I describe my methods: types of speech materials, languages used in the study, as well as the construction of a corpus for 21 languages. In Chapter 4 I present my results, in light of each research question posed above. Chapter 5 contains a brief overview of the results, discusses some issues encountered during this study, and provides place for this work in the current literature of rhythm and phonotactics.
CHAPTER 2

BACKGROUND

In this chapter, I provide background for the two areas related to the dissertation project: 1) durational variability in speech and 2) phonotactics.

2.1 Durational variability in speech

Following the view that rhythm reflects durational patterns in speech, I consider factors that affect duration of vowels, consonants, as well as vocalic and intervocalic intervals. In addition, I discuss the effect of speech rate on change of relative durations of segments within a phrase and highlight the difference between absolute duration measured from the signal and perceived duration experienced by the listeners.

Factors that affect duration

An excellent review of the literature on various factors that affect segmental durations in English is given in Klatt 1976. In addition to listing durational factors and providing references to experimental studies that support them, Klatt also relates these factors to perception studies of duration and to listeners’ ability to use durational differences to help make linguistic decisions. Here, I adapt the factors listed by Klatt to make comparable rules for vowels and consonants. Thus, some of the rules (rule 4 for vowels and rule 5 for consonants) need to be experimentally verified.

Vowels

All else being equal, vowel V1 is longer than vowel V2 if: (1) V1 is inherently longer than V2 (for instance, /ɒ/ in /dɒl/ ‘doll’ is longer than /i/ in /dɪl/ ‘deal’), (2) V1 is phonemically long and V2 is phonemically short (for instance, in Hawaiian, /a:/ in Mānoa /ma:noa/ is longer than /a/ in manu /ma:nu/), (3) V1 is a diphthong and V2 is a monophthong (in English, /ai/ in my is longer than /i/ in me), (4) V1 is a single vowel and V2 is a part of a hiatus (/i/ in /hi nouz/ ‘he knows’ is longer than /i/ in /hi ouz/ ‘he owes’), (5) V1 is in a word with fewer following syllables (in English, /ʌ/ in fun /fʌn/ is longer than /ʌ/ in funny /fʌni/,
Consonants

All else being equal, consonant C1 is longer than consonant C2 if: (1) C1 is inherently longer than C2 (in English, /m/ is longer than /n/ (Umeda 1977), (2) C1 is phonemically long and C2 is phonemically short (for instance, in Italian, /kk/ in *ekko/ ‘here it is’ is longer than /k/ in *eco/ ‘echo’), (3) C1 is a complex consonant, and C2 is a simple consonant (in English, /ʧ/ in /ʧɪp/ ‘chip’ is longer than /ʃ/ in /ʃɪp/ ‘ship’), (4) C1 is a single consonant and C2 is a part of a cluster (/n/ in /ben/ ‘Ben’ is longer than /n/ in /bent/ ‘bent’), (5) C1 is in a word with fewer syllables (/f/ in /fun/ /fʌn/ is longer than /f/ in *funny/ /fʌni/), (6) C1 is in a phrase-final syllable and C2 is not (/n/ in *looks like fun /lʊks laik ˈfʌn/ is longer than /n/ in looks like a fun movie /lʊks laik ə ˈfʌn muvi/), (7) C1 is in a stressed and C2 in unstressed syllable (first /m/ in /mimi/ ‘Mimi’ is longer than the second /m/ in /mimi/ ‘Mimi’), (8) C1 is in a word with sentence prominence and C2 is not (/b/ in *It’s my BOOK, not album is longer than /b/ in It’s in MY book, not hers), (9) there is a language-specific rule that makes C1 longer (in English, /ɛ/ in /sed/ ‘said’ is longer than /ɛ/ in /sett/ ‘set’), and (10) C1 is produced at a slower tempo (speech rate) than C2.

Modeling segmental duration

The factors causing the effects listed in (1–10) are respectively: (1) intrinsic duration, (2) phonemic length, (3) complex segment quality, (4) resource-sharing, (5) word length, (6) prosodic phrasing, (7) lexical stress, (8) prosodic prominence effect, (9) language-specific rule, and (10) speech rate. Some of these 10 factors are universal (1, 6, 10), while others are language-specific and either do not affect duration (5, 7, 9) or are not applicable (2, 3, 4, 8) in
other languages. We can also group the factors based on their nature into structural, prosodic, and pragmatic. Structural factors include intrinsic duration, phonemic length, complex quality, and language-specific rules (1, 2, 3, 4, 5, 9), prosodic factors include word stress, sentence focus, phrasal edge-lengthening (6, 7, 8), and pragmatic factors include speech rate (10).

Next, I consider intervals that consist of more than one vocalic or consonantal phone. In addition to factors (1–10), which affect individual phones, duration of an interval will be higher if its size, measured in number of phones, is larger. Thus /stɹ/ in /stɹɒŋ/ ‘strong’ is longer than /ɹ/ in /ɹɒŋ/ ‘wrong’ and /i i/ in /hiˈɪts/ ‘he eats’ is longer than /i/ in /hiˈstɪs/ ‘he sits’. Typological effects based on syllable structure are briefly discussed.

Definition of intervals

Vocalic intervals. Vocalic intervals consist of more than one vowel only when a (C)V syllable is followed by a V(C) syllable, as in he is /hiˈɪz/ or naïve /naˈɪv/, i.e., where hiatus occurs. Only languages that allow both (C)V (no coda) and V(C) (no onset) syllable types will have hiatuses. In Levelt and van de Vijver 2004, twelve syllable-type inventories are proposed, out of which seven types may have hiatuses. However, some languages have methods to avoid hiatuses through elision (deleting one of the vowels) or syneloepha (merging of consecutive vowels), or consonant insertion.

Ultimately, the average number of vowels in a vocalic interval depends on the distribution of hiatuses in a given sample. No-coda languages that allow a simple V syllable type, such as Cayuva, Mazateko (listed in Levelt and van de Vijver 2004), and Hawaiian, are more likely to have higher variability of vocalic intervals because the maximum size of the vocalic interval and frequency of hiatuses are higher than in other languages. Languages that have both (C)V and V(C) syllable types but also have one or more closed-syllable types, such as Spanish, Finnish, or English, are likely to have hiatuses with much smaller frequency and size. Vocalic interval size is not a significant cause of durational variability in these languages. Finally, in languages in which the obligatory onset principle applies, durational variability will depend only on the factors affecting single vowels.

---

6 More precisely, vocalic interval consists of one or more syllable nuclei. For instance, syllabic /ɾ/ in Serbian can be a part of a vocalic interval.
INTERVOCALIC INTERVALS. The number of consonants in an intervocalic interval is a function of syllable structure, i.e., the occurrence of consonant clusters in onsets or codas, and combinations that occur across word-boundaries (he steals /hi stilz/ as well as his team /hɪz tim/). The first factor is determined by the syllable types in a given language, while the second also depends on the way syllables combine to form words and phrases.

In Levelt and van de Vijver 2004, only two out of twelve types will not allow consonant clusters, that is, languages like Hua and Cayuvava, which allow only open syllables with no complex onsets. Hawaiian also falls in this group. Other types will show the effect of consonant cluster size on the durational variability of intervocalic intervals. The rule of thumb is that the average size of the consonant cluster will be higher in a language with a higher number of closed-syllable types and the prediction is that, for example, English and Dutch will have higher durational variability of intervocalic intervals than Spanish.

Clusters and intervals

The term consonant cluster will be used in this dissertation interchangeably with consonantal intervals. Emphasis will be on the clusters that do not cross word boundaries, in particular, we will consider clusters at the word-initial, word-medial, and word-final position in the word. They will not be discussed in relation to their position in a syllable.

A note to the definition of cluster: while it is sometimes pointed out that a cluster, or a collection, needs at least 2 elements (Vennemann 2012), we will use a more general definition, one which allows a cluster of zero elements or one element. This will greatly simplify characterization of consonant clusters at different positions of the word.

We will call a cluster with no elements a ‘zero-cluster’ and a cluster consisting of one sound/phoneme – a ‘one-cluster’. Clusters of n elements will be called n-clusters. For instance, the word ‘end’ has a zero-cluster at the word-initial (onset) position and a 2-cluster in the word-final (coda) position.

Long vowels will be considered 1-clusters (single vowels) but their phonotactic duration will have value 1 in one analysis and value 2 in a different analysis. In this way, we distinguish long vowels from hiatus sequences of two identical vowels. Short diphthongs are treated as short vowels and heavy/long diphthongs as long vowels.
Re-synthesized and filtered speech and the perception of rhythmic similarity

When judging similarity of two speech samples, a listener can possibly rely on a large number of cues, such as segmental frequencies, inferred phonological structure (phonotactic patterns), frequent grammatical morphemes, and prosodic contour, to name a few. For that reason, in perception experiments that seek to establish the rhythmic similarity or rhythmic difference of two speech samples, we use modified speech in which segmental qualities are masked. Researchers who believe that the melodic component of prosody does not strictly fall under rhythm, mask the melodic information as well.

Examples of modified speech include filtered speech and resynthesized speech. Filtered speech is obtained by removing all frequencies above a certain frequency $F_{\text{filter}}$ from the speech signal. This frequency is usually 400 Hz or 500Hz. It is assumed that the information about segmental qualities is not discernable in filtered samples and thus judging similarity should not rely on segmental identities, their frequencies and their phonotactic grouping.

Two issues arise in relation to this assumption. The first relates to the fact that a single $F_{\text{filter}}$ does not block information on segmental qualities equally in all samples. Namely, depending on the fundamental frequency of the speaker’s voice, more or less information is present in the $[0, F_{\text{filter}}]$ frequency band. Deep voices, those corresponding to low fundamental frequencies and commonly occurring in male speakers, have more information present in $[0, F_{\text{filter}}]$ than higher-pitch voices, that is, those corresponding to high fundamental frequencies and commonly occurring in female speakers and children. Thus, depending on the speaker, some segmental information can remain in the filtered sample.

Another issue arises with respect to phonotactic information. Namely, even when segmental qualities are sufficiently masked, some phonotactic information – more precisely, its durational component – is still available after filtering. This occurs because filtering removes energy contained above $F_{\text{filter}}$, say 400Hz; this procedure turns all the parts of speech segments that correspond to consonants into silences, while parts that correspond to vowels

---

7 Frequency band is frequency interval.
still contain energy. This alternation of tones and silences is assumed to characterize rhythm.\(^8\) Thus, durational variability of silences will be different for languages that allow only one consonant between vowels (such as Maori or Hawaiian) and languages that allow large variation in the number of consonants that can occur between two vowels (such as English or Russian).

Finally, judging what kind of information is present in the filtered sample, I address the question of changing spectral information through filtering, because some experiments point to spectral balance, or intensity level increments in the higher frequency bands, as a correlate of prominence (Sluijter et al. 1997). Since the energy of different vowel qualities will occupy different frequency bands, filtering may change the relative prominence of vowels – those with more energy above \(F_{\text{filter}}\) will result in greater prominence reduction compared to vowels with less energy above \(F_{\text{filter}}\). Thus, information related to energy in the filtered sample is reduced compared to information present in the original sample.

Similar issues arise in relation to re-synthesized speech, in which every consonant is replaced by one representative consonant, usually /s/, and every vowel is replaced by a representative vowel, usually /a/. Durations of each segment, or consonant and vowel sequences as a whole, are preserved. This re-synthesized version is known as ‘sasa’ speech.

We can see that the /s/ portions of ‘sasa’ speech correspond to the silence portions of filtered speech, and are affected by the durational component of phonotactics of a given language. In re-synthesized speech, however, unlike in filtered speech, no segmental quality information corresponding to unmodified speech is present.

There is another reason that filtered speech is considered important in perception experiments. Ramus et al. (1999) claim that newborns perceive speech as an alternation of tones, that correspond to vowels, and unanalyzed noise, that correspond to consonants; the reason for this being that they do not yet recognize segmental qualities. They claim that before birth, only prosodic information can be learned by fetus, as speech in uterus is heard as low-pass filtered variant preserving only rhythmic and melodic information.

---

\(^8\) To my knowledge, this assumption has not been carefully researched in relation to speech rhythm. Whether rhythmic pattern correspond to durational alternation of tones only or include information on durational alternation of silences as well, need to be more thoroughly examined in psycho-acoustic experiments.
As a result, in their studies, consonant intervals are observed as units standing in opposition to vocalic intervals and consequently the measures (rhythm metrics) that they use are defined over vocalic and consonantal intervals as units.

The assumption that newborns do not have any knowledge of segmental qualities is challenged by research reported in Loukina et al. (2011) on the research done by Gerhardt et al. (1990). In this study, authors ‘measured the intrauterine acoustic environment of fetal sheep. They found that high frequencies are somewhat attenuated, but with only a single-pole filter. As a result, enough high frequency information remains so a fetus could potentially discriminate among the consonants or among the vowels.’ As a result, newborns may be able to use segmental information in perception experiments which use unmodified speech, and the results of such experiments should be taken with caution: that is, what seems like rhythmic or durational similarity between two languages, may be in fact based partially on segmental information. However, results of experiments with newborns which use filtered speech can still be taken in support of similarity that does not depend on segmental qualities.

‘sasa’ and ‘saltanaj’ speech

In addition to ‘sasa’ type of re-synthesized speech, researchers have used another type of modified speech, also known as ‘saltanaj’ speech. As opposed to ‘sasa’ speech, in which all consonants are replaced (re-synthesized) as /s/ with the duration of the original consonant, ‘saltanaj’ speech distinguishes consonants by manner of articulation. In ‘saltanaj’ samples all stops are replaced by /t/, fricatives by /s/, nasals by /n/, liquids by /l/, and glides by /j/.\(^9\) In such samples, melodic, rhythmic, and broad phonotactic cues are available while lexical, syntactic, phonetic, and some phonotactic information is removed. The term and the type of stimuli were first reported in Ramus and Mehler 1999. The representative categories were chosen as ‘most universal in their respective categories’, or in other words, most unmarked segments.

In this dissertation, I use modified ‘saltanaj’ form to judge broad phonotactic similarity between languages. The form I use is different in that melodic and durational information are not considered. That is, only broad phonotactic patterns are observed, in

\(^9\) Affricates were not described in the original paper; they were possibly analyzed as a sequence of a stop and a fricative.
particular in consonantal clusters. Another difference includes the addition of /c/ for affricates. Although some researchers analyze affricates as sequences of a stop and a fricative, they can also be understood as single segments. In this dissertation, I consider them single segments in order to compare phonotactic patterns to the preferred phonotactic patterns predicted in (Dziubalska-Kołaczyk, 2001).

Because I analyze similarity based on broad phonotactic patterns including their frequencies in the text, I believe that my result can be informative in interpreting results of the experiments using ‘saltanaj’ speech samples.

2.2 Phonotactics: Sonority scale and markedness of consonant clusters

Phonotactics defines which sequences of segments can comprise words. Phonotactic constraints separate sequences into permissible and non-permissible and are language specific. Traditionally these constraints are formulated based on sonorities of segments in the sequence, or more precisely, based on a function of sonorities (sonority distances) of the neighboring segments.

There are several versions of the sonority scale. The one represented in Figure 2.1 is from Vennemann 1988.

Low V, mid V, high V, central liquids (r-sounds), lateral liquids (l-sounds), …
… nasals, voiced fricatives, v-less fricatives, voiced plosives, v-less plosives

Figure 2.1 Sonority scale (Vennemann 1988)

Finer-grained scales possibly differentiate vowels by frontness, and obstruents and nasals by place of articulation. Less fine-grained scales combine certain groups of sounds. For instance, a possible scale is shown in Figure 2.2:

vowels > glides > liquids > nasals > fricatives > plosives

Figure 2.2 saltanaj sonority scale
This scale corresponds to *saltanaj* (Ramus & Mehler 1999) level of representation discussed in the previous section.

Another more general scale (let us call it ALT scale) defines only 3 categories: vowels (A), sonorants (L), and obstruents (T).

\[
\text{vowels} > \text{sonorants} (\text{glides, liquids, nasals}) > \text{obstruents} (\text{fricatives, plosives})
\]

**Figure 2.3 ALT sonority scale**

Sonority is relative measure; segments in sonority scales are ordered from most to least sonorous; however, the exact values of sonority assigned to individual segments, or classes of segments, are somewhat arbitrary. They usually rank segments in increments of 1. Different values of sonorities are sometimes assigned to classes of segments in order to explain cross-linguistic phonotactic differences.

Phonotactic constraints with respect to syllable structure are defined by the Sonority Sequencing Principle (SSP) (Sievers 1881, Jespersen 1904). SSP states that more sonorous elements should be positioned closer to the nucleus, which forms a sonority peak.

Phonotactic constraints on consonants can be defined with respect to syllable as in SSP; they can also be defined without reference to syllables, by relating to order of consonants in a consonant cluster at different positions in a word: word-initial, word-medial, and word-final.

One such principle, *Optimal Sonority Distance Principle* (OSDP), is presented in Dziubalska-Kołaczyk 2001. Segments are ordered in a sonority scale that can be described as an augmented saltanaj scale; one in which affricates are assigned sonority between fricatives and plosives. Let us call it the *saltanajc* scale. Sonority values in this scale are inversely proportional to their sonority and range from 0, which is assigned to vowels, to 6, which is assigned to plosives. Sonority values thus represent consonantal strength.

Furthermore, phonotactic constraints on consonant clusters in each word-position are defined based on sonority distances between comprising elements and the distance between the vowel (nucleus) and the consonant closest to it. These constraints partition the set of all
possible consonant clusters at the saltanajc level into preferred initials, preferred medials, and preferred finals, with some overlap allowed between medials and finals.

Here, I list preferred double clusters at word-initial, word-medial, and word-final position, adopted from Dziubalska-Kołaczyk 2001. I modify the lists slightly compared to original work in order to represent all clusters using saltanajc level.

\[
tj > cj > sj=tl > nj=cl > lj=sl=tn
\]

Figure 2.4. Preferred initial double clusters (Dziubalska-Kołaczyk 2001)

\[
tt > cc=tc > ss=cs=ct=ts > nn=sn=sc=cn > ll=nl=ns=st > jj=ln=nc > jl=ls=nt \text{(the last three clusters are preferable in final position as well)}
\]

Figure 2.5. Preferred medial double clusters (Dziubalska-Kołaczyk 2001)

\[
jt > jc > js=lt > jn=lc > jl=ls=nt
\]

Figure 2.6. Preferred final double clusters (Dziubalska-Kołaczyk 2001)

Stating constraints as preferences, i.e., in terms of preferred clusters at each position, it is signaled that constraints are sometimes violated. This is well recognized in phonological theory.

To investigate differences in consonant cluster patterns in questions 3 and 4, I will adopt the notion of preferred clusters and the constraints that define them. I will also use the term ‘more marked cluster’ to mean ‘less preferred cluster’ in each position. The notion of cluster markedness will be used to capture cross-linguistic patterns and to test typological tendencies. Namely, I will verify whether more marked (that is, less preferred) clusters are more rare across languages and less frequent within each language.
CHAPTER 3

METHODS

In this Chapter, I describe the process of creating the phonemic corpora from the written materials for the 21 sample languages and the phonotactic calculator for the measures used to answer my research questions.

3.1 Model of the transcriber & the complexity calculator

The model of data collection and transformation is presented in Figure 3.1. We start from texts available in electronic form, and after pre-processing apply two kinds of rules. First, we apply grapheme-to-phoneme rules, which transform written materials into broad phonemic form. Next, some language-specific phonological rules are applied to obtain forms closer to spoken speech.

![Figure 3.1. Model of the transcriber](image)

At the end of process shown in Figure 3.1, we have materials in ‘phoneme level’, that is, words consist of IPA phones and phrases consist of such words. Phrases are assigned based on punctuation marks, as an approximate way to obtain phonological phrases.

Following that, the data is converted into three representational levels:

‘CV’ level, in which all consonantal phonemes are replaced by ‘C’ and all vocalic phonemes, including diphthongs and long vowels are replaced by ‘V’;

ALT level, or ‘obstruent-sonorant-vowel level, in which sonorant consonant phonemes are replaced by ‘L’, obstruent consonant phonemes by ‘T’, and vowels by ‘A’; and
‘saltanajc’ level, or broad sonority level, in which each phoneme is replaced by the corresponding sonority class representative. Sonority class representatives are ‘t’ for stops, ‘s’ for fricatives, ‘c’ for affricates, ‘n’ for nasals, ‘l’ for liquids, ‘j’ for semi-vowels, and ‘a’ for vowels. This particular sonority scale classifies consonants according to their manner of articulation.

This process is accomplished by the ‘Representation converter’, as the first step in the complexity calculator, shown in Figure 3.2.

Data in an appropriate level of representation are passed to the next block in which various measures are calculated. Measures used for answering different research questions are defined in the corresponding chapters. These include rhythm metrics, phonotactic metrics (defined in Chapter 4), distributions of consonant cluster lengths, phonotactic patterns, and word-length distributions.

At the end, the resulting metrics are saved in tables as text files; the raw texts as well as transformed texts in three representation levels are saved as text files in Unicode utf-8 format (http://www.utf-8.com/).

3.2 Raw data assembly

Languages

An important criterion in the choice of languages was that they could be easily phonemicized from texts, or alternatively have large corpora of transcribed speech. The latter is available in
only a few cases, English and French, for instance. For consistency, I opted for the first criterion, that is, creating transcribed speech from the available texts. So, for a language to be useful here, grapheme-to-phoneme mapping should be simple or reasonably easily obtained with the help of phonological transcription rules.

Another criterion was that languages represent several language groups, but at the same time have at least two, and possibly more languages from the same group, so that comparisons based on genetic closeness or distance could be made.

In addition, the languages had to represent all of the putative rhythm classes and preferably have multiple languages in each group. Note that for some languages rhythm type has not been previously assigned in the literature, or a consensus does not exist regarding the type.

Finally, I tried to include languages with different word order, and with various types of syllable complexity and size of phonemic inventory. Some of these criteria are typological correlates, so the criteria might occur in clusters.

English and French were not included because of seemingly complex grapheme-to-phoneme correspondence rules.¹⁰

Based on the mentioned criteria, I selected 21 languages from the following language groups for analysis: Slavic (Bulgarian, Czech, Polish, Russian, and Serbian), Germanic (Dutch and German), Romance (Catalan, Italian, Brazilian Portuguese, and Spanish), Uralic (Estonian and Hungarian), Polynesian (Hawaiian, Maori, Samoan, and Tongan), Other¹¹ (Japanese, Turkish, Greek, and Indonesian).

Slavic and Germanic languages are traditionally considered stress-timed, Romance languages – syllable-timed and Japanese – mora-timed language. I will keep these names for the three posited rhythm types due to readers’ likely familiarity with them. This will make the discussion easier.

¹⁰ As a work in progress, I am trying to address this issue by obtaining already transcribed samples from the spoken databases for these two languages.
¹¹ Naturally, ‘Other’ does not refer to historical language grouping, but only indicates a subset of languages that are single representatives of their language groups: Greek is an Indo-European isolate; Indonesian belongs to the Austronesian group but is not Polynesian; Turkish belongs to Altaic, and Japanese to Japonic language group.
Due to lack of data on perceptual similarity of Polynesian languages to languages in the three rhythm groups, say, to Dutch, Spanish, and Japanese, I assigned their type based on phonological criteria; they have simple syllable structure and phonemic vowel length, so they were assigned to the mora-timed group. Similarly, Turkish, Greek, and Indonesian are assigned to the syllable-timed group. Lastly, Estonian and Hungarian were assigned to the syllable-timed group, although they have both fixed stress (and could be stress-timed) and phonemic vowel length (and could be mora-timed). However, their heavy coda structure and the lack of vowel reduction suggested that they should not be in either the stress-timed or the mora-timed group.

A table summarizing the basic phonological and morphological properties of these languages is provided in Appendix 1. The data is adopted from the World Atlas of Language Structure (WALS); where information was not available from WALS (Dryer and Haspelmath 2011, online), it was filled based on my personal knowledge of the language or based on information from the language grammar books.

Written materials

Ideally, we would like to perform phonotactic analysis of spoken samples. However, because long samples are needed in order to get stable phonotactic frequencies, and because the manual transcription even for a single language would be prohibitively expensive, I decided to start from written texts that can be performed as told stories – fairy-tales and stories from the Bible – and then automatically transcribe them using grapheme-to-phoneme and phonological rules for each language. I analyze several texts for each language in order to test the stability of phonotactic frequencies.

The advantage of the chosen types of chosen texts is that they are less likely to include borrowed words with foreign phonologies and phonotactics. News articles are avoided for this reason, despite their abundant availability.

3.3 Creating phonemic corpora

Creating phonemic representation from a written text is called ‘Grapheme-to-phoneme’ procedure. Text-to-speech systems necessarily use conversion of written text to sequences of
phonemes. Such processes are also a useful basis for making linguistic inquires about phonetic and phonological phenomena for which a large amount of data is needed (or preferred) and where the alternative of recording and then transcribing proves to be too time and effort consuming… or simply impossible\textsuperscript{12}.

Using this sort of corpus generation is a natural choice for phonotactic questions: intonational and durational information are not required but very long texts/materials are needed in order to calculate type frequencies.

3.3.1 Choice of grapheme-to-phoneme method

There are three main types of grapheme-to-phoneme algorithms. Brief characteristics are summarized here from Bisani and Ney (2008).

The first approach is dictionary lookup. It assumes that a dictionary of pairs consisting of a written word and its pronunciations (that is, IPA phonemic representation) is available. To produce such a dictionary is costly and tedious.

The second approach is rule-based conversion. It often incorporates a dictionary or an exception list for the cases where regular rules do not produce correct output. Its drawbacks, according to Bisani and Ney, are that producing rules for a given language can be hard (the interdependence between rules may be complex) and requires specific linguistic skills. Moreover, no matter how carefully designed, such a system is likely to make mistakes with exceptional words (those that use novel spelling, or are recent borrowings with phonological shapes not fitting the host language).

An alternative to dictionary lookup and rule-based conversion, which are two types of knowledge-based methods, is a data driven approach. This approach is based on the principle that ‘given enough examples it should be possible to predict the pronunciation of unseen words purely by analogy’ (Bisani and Ney p.4). According to Bisani and Ney, the benefit of this approach is that it replaces a challenging task (rule making) with an easy one – implementing a learning algorithm and letting the correspondences be found during training.

\textsuperscript{12} How long does it take to transcribe 20 languages times 30,000 words?
Given the available resources and the nature of the questions asked in this dissertation, I decided to construct rule-based conversion modules, one for each language. Rule-based algorithms allow for application of different levels of phonetic details – broad or narrow linguistic transcription – just by regulating which rules to utilize. Also, they allow us to implement a different dialect by making only minor changes to the rules.

As pointed out by Garcia and Gonzalez (2012) who constructed a similar, albeit somewhat more complex, module for Galician, rule-based approaches work very well for languages with transparent spelling. For other languages, however, morphophonological information is needed in order to achieve correct transcription without long exception lists.

Languages I chose include those with transparent to medium-difficulty writing systems. I was able to transcribe materials sufficiently well up to broad sonority, or ‘saltanajc’ level. For languages with the most transparent spelling, the phoneme level was also transcribed reasonably well.

To make the process reasonably simple, I did not include dictionary or exception lists in the grapheme-to-phoneme modules. This limited the success of transcription because part-of-speech and position of accent in the word (where not fixed) were not available. For instance, I was not able to transcribe the positions of reduced vowels in Dutch with very high accuracy. The goal in those cases changed to achieving limited transcription accuracy: 1) correct sequence of CV, or 2) additionally providing vowel qualities (or a class, say \{e,ε\}) where possible, and only minimally a manner of articulation for consonants, and where possible, a place of articulation and voicing.

3.3.2 Implementation of grapheme-to-phoneme method

The transcriber is programmed in Python (http://www.python.org/). Materials are presented in Unicode format and transformed using grapheme-to-phoneme correspondences, correspondences in specific phonetic environments, as well as some phonological rules, such as vowel reduction, vowel lengthening, and place assimilations. Lists of rules used for transforming text to phonemes are listed for each language in Appendix 2.

Several issues arise in this process. One is the assignment of diphthongs. Namely, where two vowel phonemes occur next to each other, they can represent a diphthong or a
hiatus – a sequence of two vowels in different syllables. Since a precise assignment depends on the particular words involved, and the transcribing method did not include access to each language dictionary with word definitions, diphthongs were overcounted. That is, wherever a sequence of two vowels could be interpreted as a diphthong, it was counted as such against the possibility of a hiatus. Hiatus is only assigned in languages that do not have diphthongs or where the sequence is such that it cannot be interpreted as a diphthong in that language. As a result, the percentage of vowels in the text was undercounted in these cases (as one vowel (a diphthong) was counted instead of two).

Most consonants were precisely transcribed, except for the voicing feature. However, although most vowels were assigned correct qualities, in some cases it was difficult to distinguish certain pairs, such as /o/ and /ɔ/ in some languages.

Finally, since the stress location was not known for languages where stress location in words is variable, not all vowel reductions could be implemented correctly. This should not present a problem for the type of questions we are asking.

3.4 The complexity calculator

The complexity calculator produces all the measures required to answer my research questions. Here I summarise definitions of these measures in the order of the questions that require them. They include phonotactic and rhythm metrics, cluster length distributions, cluster pattern distributions, and word length distributions.

3.4.1 Phonotactic metrics and rhythm metrics

To investigate the effect of phonotactics on the rhythm metrics, we define Phonotactic Metrics (PMs). Each of the PMs has an analogous durational measure among the Rhythm Metrics (RMs) used in the literature. PMs are defined on phonotactic durations rather than on temporal durations.

The phonotactic duration of any segment is defined to be 1; thus the phonotactic duration of the consonantal interval /str/ is 3, and that of the vocalic interval /a/ is 1. In other
words, the phonotactic duration of a vocalic or a consonantal interval is equal to the number of segments it comprises.\textsuperscript{13}

Let us now formally define all the metrics we will use. Definitions of RMs are known in the literature and they are repeated here for the readers’ convenience. PMs are introduced here.

Let a speech fragment be segmented into vocalic intervals \{Vi, i=1,2,...,Nv\}, and consonantal intervals \{Cj, j=1,2,...,Nc\}, where Nv denotes the number of vocalic intervals and Nc the number of consonantal intervals. Let \(d(x)\) denote the temporal duration of interval \(x\) in ms, and \(L(x)\) its phonotactic duration, or length. Let us subscript phonotactic metrics with the index ‘p’ and durational or rhythm metrics with the index ‘r’.

Next, we present rhythm metrics (RMs) and phonotactic metrics (PMs) side by side. All the rhythm metrics have been extensively used in the literature (Ramus et al. 1999, Grabe and Low 2002, Dellwo 2006, among others).

**PERCENT OF SPEECH OCCUPIED BY VOWELS (\%V)**

This metric evaluates the proportions of vocalic and consonantal material in a speech sample. Let us denote the PM \%V by \%V_p, and the corresponding RM by \%V_r.

Then,

\[
\%V_r = 100 \cdot \frac{\sum_{i=1}^{Nv} d(V_i)}{\sum_{i=1}^{Nv} d(V_i) + \sum_{j=1}^{Nc} d(C_j)}
\]

\[
\%V_p = 100 \cdot \frac{\sum_{i=1}^{Nv} L(V_i)}{\sum_{i=1}^{Nv} L(V_i) + \sum_{j=1}^{Nc} L(C_j)}
\]

These can be transformed to a form in which they are functions of average durations of consonants, \(\overline{d_c}\), and vowels, \(\overline{d_v}\), in the sample (\(\overline{L_c}\) and \(\overline{L_v}\) for the average phonotactic durations).

Note that the sum of durations over all intervals is equal to the product of the number of intervals and the average interval duration. This holds for both vowels and consonants. It

\textsuperscript{13} A special case was considered in some analyses in section 4.1 where long vowels were assigned phonotactic duration equal 2.
is also true that the same sum for vowels (or consonants) can be expressed as a product of the number of all vowels (or consonants) and the average duration of a vowel (or a consonant) in the speech fragment. We then have:

\[
\%V_r = 100 \cdot \frac{1}{1 + \frac{N_c}{N_v} \frac{\overline{d_c}}{d_v}} = 100 \cdot \frac{1}{1 + \frac{n_c}{n_v} \frac{\overline{d_{c1}}}{\overline{d_{v1}}}}
\]

\[
\%V_p = 100 \cdot \frac{1}{1 + \frac{N_c}{N_v} \frac{\overline{L_c}}{\overline{L_v}}} = 100 \cdot \frac{1}{1 + \frac{n_c}{n_v}}
\]

where \(n_c\) stands for number of consonants, \(n_v\) for number of vowels in the sample, and \(\overline{d_{c1}}\) and \(\overline{d_{v1}}\) for average durations of a single consonant and vowel respectively.

We see from the last pair of formulae that the rhythmic \(\%V_r\) will change with tempo depending on the ratio of average consonant duration and average vowel duration.\(^{14}\) On the other hand, phonotactic \(\%V_p\) will be invariable and depend only on the ratio of the number of consonants to the number of vowels in the sample.

**STANDARD DEVIATION OF VOCALIC AND INTERVOCALIC INTERVALS (\(\Delta V\) AND \(\Delta C\))**

These metrics evaluate the variability of interval durations: a standard deviation of 0 corresponds to intervals uniform in duration, while a large standard deviation corresponds to large variation in interval durations. Durational and phonotactic \(\Delta C\) are defined as

\[
\Delta C_r = \sqrt{\frac{\sum_{j=1}^{N_c} (d(C_j) - \overline{d_c})^2}{N_c - 1}} = \overline{d_c} \sqrt{\frac{\sum_{j=1}^{N_c} (\frac{d(C_j)}{\overline{d_c}} - 1)^2}{N_c - 1}}
\]

and

\[
\Delta C_p = \sqrt{\frac{\sum_{j=1}^{N_c} (L(C_j) - \overline{L_c})^2}{N_c - 1}} = \overline{L_c} \sqrt{\frac{\sum_{j=1}^{N_c} (\frac{L(C_j)}{\overline{L_c}} - 1)^2}{N_c - 1}}
\]

\(^{14}\) Both vowels and consonants change duration as tempo increases and decreases, but they change by a different factor.
An often-reported disadvantage of the durational measure, $\Delta C_\rho$ (the standard deviation of consonantal intervals), is that it is proportional to the mean value of the intervals. This means that the durational measures $\Delta C_r$ and $\Delta V_r$ (standard deviations of consonantal and vocalic intervals) will change when the speech rate changes. However, this issue does not exist for phonotactic measures, since the value depends only on the number of segments and not their durations\(^\text{15}\).

The phonotactic metric $\Delta C_p$ measures complexity of syllable structure, or more precisely – complexity of consonant clusters. For languages for which the basic syllable template is (C)V, $\Delta C_p$ will be equal to 0; all consonantal intervals will be the same. Its value will be higher for languages where more complex onsets or codas are allowed, and thus where more complex intervals exist.

The phonotactic metric $\Delta V_p$ will be proportional to the incidence of hiatus (consecutive vowels) in the language. If we allow the phonotactic duration of long vowels to be redefined as 2 (note that in the original version the phonotactic duration of short and long vowels is the same and equal to 1), then $\Delta V_p$ will also be higher for languages, or speech samples, in which the percentage of long vowels is higher\(^\text{16}\). We will calculate metrics for both cases: length of long vowels defined as 1 and defined as 2.

COEFFICIENT OF VARIATION OF VOCALIC AND INTERVALVOCALIC INTERVALS ($\text{VARCOV}, \text{VARCOC}$)

Coefficient of variation is equal to the standard deviation divided by the mean interval duration. RM and PM measures for vocalic intervals are defined as

$$VarcoV_r = \frac{\Delta V_r}{d_v} = \sqrt{\frac{\sum_{i=1}^{N_v} (\frac{d(V_i)}{d_v} - 1)^2}{N_v - 1}}$$

$$VarcoV_p = \frac{\Delta V_p}{d_v} = \sqrt{\frac{\sum_{i=1}^{N_v} (\frac{L(V_i)}{L_v} - 1)^2}{N_v - 1}}$$

Normalizing by the mean value is intended to reduce variations that emerge with change of speech tempo. This normalization method is most successful when all the intervals

\(^{15}\) The only changes happening at faster tempo that will affect $\Delta C_\rho$ and $\Delta V_\rho$ are possible segment deletions

\(^{16}\) We assume there are always more short vowels than long vowels. Otherwise, variability starts decreasing once the number of long vowels surpasses the number of short vowels in the sample.
increase or decrease proportionally. However, speech rate affects speech segments, vowels and consonants, in a different way; stressed vowels do not stretch and expand the same way as unstressed vowels and neither do consonants. Therefore, not all vocalic intervals increase or decrease their durations by the same factor. As a result, Varco Vr still depends on speech tempo but the magnitude of the tempo effect is smaller than for $\Delta C_r$. On the other hand, as with $\Delta C_p$ (and $\Delta V_p$), being a phonotactic metric, VarcoV is independent of speech rate.

**RAW pair-wise variability index (rPVI)**

This metric in its durational variant applies only to consonantal intervals. It measures the variability of intervals with respect to their neighbors. It aims to capture the short-long-short contrast of the prominent (middle) interval\(^\text{17}\).

$$rPVI - C = \frac{1}{N_C - 1} \sum_{k=1}^{N_C-1} |d(C_k) - d(C_{k+1})|$$

The phonotactic metric $rPVI - C_p$ measures variability of length of consonant intervals in a sequence. It is larger when short clusters alternate with longer clusters.

$$rPVI - C_p = \frac{1}{N_C - 1} \sum_{k=1}^{N_C-1} |L(C_k) - L(C_{k+1})|$$

**NORMALIZED pair-wise variability index (nPVI)**

Among rhythm metrics, a normalized version of the pair-wise variability index has generally been applied to vocalic intervals. The motivation, as with the Varco measures, is to achieve invariance with speech rate changes. The definition of the normalized phonotactic pair-wise variability index is given after its durational counterpart:

$$nPVI - V = \frac{100}{N_V - 1} \sum_{k=1}^{N_V} \left( \frac{d(V_k) - d(V_{k+1})}{d(V_k) + d(V_{k+1})} \right) = \frac{200}{N_V - 1} \sum_{k=1}^{N_V} \left( \frac{d(V_k)}{d(V_{k+1})} - 1 \right)$$

\(^{17}\) Or a short-long contrast of a prominent final.
The phonotactic variant measures the variability of vocalic interval lengths, and like $\Delta V_p$, it is affected by the proportion of hiatuses and long vowels (when long vowels are assigned a length of 2). Languages with high $nPV_1 - V_p$ are those with a high percentage of long vowels. Note that only a few terms in the formula will be non-zero – those that apply to an interval that contains a hiatus or a long vowel and an interval that consists of a single short vowel.

3.4.2 Consonant cluster measures

**Cluster-size measures**

Our second question relates to word structure; it asks how many consecutive consonants can occur at the beginning, in the middle, and at the end of a word. In other words, it examines which phonotactic durations (or lengths) of consonant clusters are permitted at these three positions. It also asks how frequently each of these possible lengths occurs in running speech. We represent the frequencies of consonant clusters of different sizes by a discrete distribution.

Remember that in this dissertation a cluster can have length zero, one, two or more. This convention facilitates inclusion in the distribution of words that start with a vowel or with a single consonant to the distribution.

The value of the word-initial (or word-final, or word-medial) distribution for each consonant cluster length $n >= 0$ is the ratio of the number of word-initial (or word-final, or word-medial) clusters of length $n$ to the total number of word-initial (or word-final, or word-medial) clusters in the speech sample. If the sample is large enough, these frequencies correspond to the likelihood of a cluster in that position being of length $n$.

Note that the number of initial clusters (including those of length zero) and the number of final clusters are each equal to the number of words in the sample. The number of
medial clusters can be larger or smaller than the number of words, depending on the proportion of monosyllabic words in the sample. Each disyllabic word contributes one medial cluster, each 3-syllabic word – 2 medial clusters. Monosyllabic words have only word-initial and word-final cluster positions.

Distributions are produced and tabulated for the three positions in the word for all sample languages. To address the issue of stability, namely, whether these distributions vary significantly across speech samples for the same language, I repeated calculations for several samples.

**CLUSTER PATTERN COMPARISON**

Question 3 looks at the similarity across languages in occurrence of length-2 and length-3 clusters. The number of shared highest occurring clusters at each position in the word is used as a measure of similarity.

I compare cluster patterns at two different levels, ALT level and ‘saltanajc’ level, defined in section 3.1. The ALT level is often used in statements of typological universals regarding syllable structure.

The example in Figure 3.3 shows percentages of length-0, length 1, and length-2 initial clusters for Dutch and German at the ALT level:

<table>
<thead>
<tr>
<th>language</th>
<th>#_A</th>
<th>#LA</th>
<th>#TA</th>
<th>#LL</th>
<th>#LT</th>
<th>#TL</th>
<th>#TT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>20.8</td>
<td>21.5</td>
<td>50.5</td>
<td>0.0</td>
<td>n.a.</td>
<td>5.3</td>
<td>1.6</td>
</tr>
<tr>
<td>German</td>
<td>26.9</td>
<td>11.7</td>
<td>53.1</td>
<td>n.a.</td>
<td>n.a.</td>
<td>5.1</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Figure 3.3 Example: cluster distribution at ALT level

The ALT level captures the general phonotactic constraints present in each language. For instance, in this example we see that obstruents are likely single initials than the sonorants, and that obstruent-sonorant is the best double initial in both languages.

At the ‘saltanajc’ level, I will present length-2 clusters in decreasing order of their frequencies in the text. These are compared to the sonorically preferred clusters of length-2 at the same position. The following is an example for initial length-2 clusters in Dutch.
We can see that in Dutch text sonorically preferred initial clusters are not the most frequent. This is an example of violating the sonority constraint or not following the sonority preference. I will also compare languages based on their most frequent clusters, whether they are matching the preferred patterns or not.

I have chosen to observe consonant sequences at these two representational levels because 1) they are more manageable than the exact phoneme levels, given the number of different possible sequences, and 2) I have been able to create these levels for all the languages in this study, while only some have fully correct phonemic levels\(^{18}\).

### 3.4.3 Word-lengths measures

Word lengths are calculated both in terms of number of syllables per word and number of phonemes per word. The number of syllables is defined as the number of vowels in the word; diphthongs and long vowels count as single vowels, just like short vowels.

Languages are compared based on word length distributions. I will present the text distribution (where a given word is counted the number of times it occurs in the text) as well as distribution among lexical items (a given word is only counted ones, irrespective of its frequency in the text).

---

\(^{18}\) Additional attention is needed in transcription process in order to implement correct voicing/devoicing processes in all cases; at present, voicing is incorrect in some places.
CHAPTER 4

RESULTS

In this chapter, I provide empirical results that address research questions 1-5. The chapter is divided into five sections, with section 4.3 addressing both questions 3 and 4. Section 4.5 contains stability analysis. All the results were obtained by analyzing the corpus described in Chapter 3. The corpus contains materials for 21 languages spread over several language families.

Section 4.1 examines the phonotactic component of the durational variation of vocalic and consonantal intervals and its effect on the values of rhythm metrics (RM). Sections 4.2-4.4 examine the structural complexity of words in the test languages and highlights similarity groupings across these factors: consonant cluster lengths (within word), phonotactic patterns, and word lengths. These groupings are then compared to those obtained via rhythm metrics in the literature and to the traditional assignment of rhythm class, as stated in section 3.2.

4.1 Phonotactic component of rhythm metrics

4.1.1 Introduction

In this section, I address the first dissertation question: what is the contribution of phonotactics to the values of various Rhythm Metrics (RMs)? Consequently, how much of the language grouping on the rhythm-metric graphs can be explained by the phonotactic similarities across languages? Lastly, do phonotactic properties alone classify languages more consistently with respect to traditional rhythm type than the rhythm metrics do?

To answer these questions, I calculated Phonotactic Metrics which I defined in Chapter 3. Phonotactic metrics (PMs) are analogous measures to rhythm metrics, based on phonotactic durations, or interval lengths, instead of on temporal durations. They are equal to Rhythm Metrics (RMs) if the contributions of prosody and segmental qualities on durations are ignored.

I measured the effect of phonotactics on Rhythm Metrics by the correlations between the PMs and RMs. Values of the Rhythm Metrics used for comparison were collected from
studies in the literature and are presented in Table A3.1 in Appendix 3. Values of the Phonotactic Metrics were calculated based on equations given in section 3.1 and are presented in Table A3.2 (length of long vowel is 1) and Table A3.3 (length of long vowel is 2) in Appendix 3.

Having established how similar the corresponding phonotactic and rhythmic metrics are, say %Vp and %Vr, I compare the language groupings produced by the Rhythm Metrics to the groupings based on Phonotactic Metrics on a subset19 of 21 languages. In addition, I provide the graphs based on Phonotactic Metrics that show groupings of the complete set of 21 languages and discuss how these groups relate to traditional rhythm type assignment.

In section 4.5 I discuss the stability of PMs when computed over different speech materials and over samples of different lengths.

4.1.2 Correlations between the Phonotactic and the Rhythm Metrics

To calculate the Phonotactic Metrics, I considered two models: in the first one, the phonotactic duration of each consonant and vowel, including long vowels, equals 1. In the second, phonotactic durations of consonants and short vowels (including diphthongs) are defined as 1, and the phonotactic durations of long vowels as 2. Correlations were calculated for a subset of languages for which the values of Rhythm Metrics are available in the literature. Values for RMs are reported in the Table A3.1, and for PMs in Tables A3.2 and A3.3 in Appendix 3.

Correlations for the consonantal metrics are given in Table 4.1. Note that the correlations are calculated on a different subset of languages for each metric, depending on the values of RMs available in the literature; only Rhythm Metrics values obtained in a single study were used in order to avoid discrepancies due to different segmentation methods.

<table>
<thead>
<tr>
<th></th>
<th>rPVI-C</th>
<th>Varco C</th>
<th>ΔC</th>
</tr>
</thead>
<tbody>
<tr>
<td>correlation</td>
<td>-0.01</td>
<td>0.9</td>
<td>0.98*</td>
</tr>
<tr>
<td>p</td>
<td>-</td>
<td>0.1</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

19 Subset consists of those languages for which values of RMs are available in the literature for each metric.
We observe that the standard deviation of consonantal intervals $\Delta Cr$ is significantly correlated ($r=0.98$, $p<0.001$) with its phonotactic counterpart $\Delta Cp$. Rhythmic coefficients of variation of consonantal intervals Varco-Cr and Varco-Cp show tendency towards linear relationship but the number of points for Varco-Cr is too small. Next, we calculate vocalic metrics when the phonotactic duration of long vowel is $\text{longV}=1$.

Table 4.2 Correlation between vocalic PMs and RMs (long=1)

<table>
<thead>
<tr>
<th>correlation</th>
<th>%V</th>
<th>nPVI-V</th>
<th>Varco-V</th>
<th>$\Delta V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>0.87</td>
<td>-0.5</td>
<td>0.21</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.16</td>
<td>0.27</td>
<td>0.92</td>
</tr>
</tbody>
</table>

We see that none of the measures is significantly correlated with the phonotactic component, although percentage vowels shows high coefficient of correlation. Let us then consider model 2 in which long vowels are assigned phonotactic durations of 2. This model accounts for variation caused by the difference between short and long vowels. Results are presented in Table 4.3.

Table 4.3 Correlation between the vocalic PMs and RMs (long=2)

<table>
<thead>
<tr>
<th>correlation</th>
<th>%V</th>
<th>nPVI-V</th>
<th>Varco-V</th>
<th>$\Delta V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>0.93</td>
<td>0.36</td>
<td>0.98</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>0.008</td>
<td>0.38</td>
<td>0.0004</td>
<td>0.15</td>
</tr>
</tbody>
</table>

According to this model, percent vowels $\%V_r$ and the coefficient of correlation of vocalic intervals Varco-$V_r$ are correlated with their phonotactic components at $p<0.001$.

Lack of correlation for the standard deviation of vocalic intervals $\Delta V$ is possibly due to high variability of this measure with speech rate. The pair-wise variability index nPVI-V on the other hand might be influenced significantly by the prosodic factors. This suggests that nPVI-V might be a good rhythm correlate. We return to this point in the next section.

Correlations graphs for each metric, are presented in Figures 4.1-4.6. Rhythm metrics that are highly correlated with their phonotactic counterparts produce close to linear relationship as shown by the regression line. On all graphs, traditionally stress-timed languages are represented by a green triangle, stress-timed languages by a blue circle, and mora-timed languages by a magenta square.
Figure 4.1 Correlation between phonotactic (%Vp) and rhythmic (%Vr) percentage of vocalic intervals

Figure 4.2 Correlation between phonotactic (ΔCp) and rhythmic (ΔCr) standard deviation of consonantal intervals
Figure 4.3 Correlation between phonotactic ($\Delta V_p$) and rhythmic ($\Delta V_r$) standard deviation of vocalic intervals

Figure 4.4 Correlation between phonotactic (Varco-Cp) and rhythmic (Varco-Cr) coefficient of variation of consonantal intervals
Figure 4.5 Correlation between phonotactic (Varco-Vp) and rhythmic (Varco-Vp) coefficient of variation of vocalic intervals

Figure 4.6 Correlation between phonotactic (nPVI-Vp) and rhythmic (nPVI-Vr) normalized pair-wise variability index of vocalic intervals
4.1.3 Classification power of RMs and PMs

My results for phonotactic metrics are presented graphically; the graphs are analogous to those used in the literature to support classification of languages into three rhythm groups. Figure 4.7 graphs the phonotactic metric %Vp against the standard deviation of consonantal intervals ΔCp. This can be compared to the graph of actual rhythm metrics %Vr vs. the standard deviation ΔCr presented in Figure 4.8.

In these graphs, we see languages forming three loose clusters corresponding to the hypothesized rhythm types: stress-timed (green triangle), syllable-timed (blue circle), and mora-timed (magenta square). These domains are demarcated by dashed lines.

We can readily see the similarity between the two graphs even though the values are in a slightly different range for ΔC. We see groupings of languages in the three areas of the graph; these loosely correspond to the notion of stress-timed, syllable-timed, and mora-timed languages.

Figure 4.8 Phonotactic metrics graph (%Vp, ΔCp)
Since phonotactic metrics %Vp and ∆Cp are not based on actual durations, they only reflect the phonotactic but not the prosodic component. Yet, they group languages in the same way as rhythm metrics group them. This suggests that the durational variation in a speech sample that arises from phonotactics is a large component of the overall durational variation. In other words, rhythm metrics (%Vr, ∆Cr) reflect a great deal of phonotactics in addition to some of the rhythmic variations in the sample.

When we compare pair-wise variability graphs however, a different picture emerges. Results for rhythmic and phonotactic pair-wise variability measures are presented in Figures 4.9 and 4.10.

It is interesting to see in Figure 4.9 that (rPVI-Cr, nPVI-Vr) pair of RMs do not clearly separate Japanese from the syllable-timed group, whether we look at each of the metrics separately, or in combination. Additionally, some stress-timed languages, Bulgarian, Russian, and Czech, fall into the region with syllable-timed group and Japanese.
Figure 4.9 Rhythm metrics graph (rPVI-Cr, nPVI-Vr)

Figure 4.10 Phonotactic metrics graph (rPVI-Cp, nPVI-Vp)
The only clear trend in the RM graph (Figure 4.9) is the separation of traditional stress-timed languages (Dutch and German) in the upper half of the graph.

On the PM graph (Figure 4.10), on the contrary, grouping is clear if we assume that Bulgarian, Russian, Catalan, Czech, and Polish all group with German and Dutch. This agrees with the traditional assignments (Slavic and Germanic languages in the same group). This suggests that the lack of correlation between pair-wise phonotactic and rhythm metrics reflects the classification power of PM metric and lack of classification power of the RM metric.

4.1.4 Language classification based on Phonotactic Metrics

Next, let us look at the language groupings of the complete set of 21 sample languages based on Phonotactic Metrics. Consider the graph defined by (%Vp, ∆Cp in Figure 4.11.

Figure 4.11 graphs the phonotactic metric %Vp against the standard deviation of consonantal intervals ∆Cp for all 21 languages. We can now see that there emerges a linear relationship between the two variables, with a strong negative correlation of \(-0.97\). Cassandro et al. (in an unpublished MS referenced in Easterday at al. 2011) postulated a universal linear relationship between the rhythm metrics %Vr and ∆Cr. This correlation may result from the structural properties that phonotactic metrics %Vp and ∆Cp capture. It expresses a tendency for a language with large ratio of vowels to consonants to have a small variation in the size of consonant clusters. This can be explain by observing that percent vowels %Vp can be expressed as 100-%Cp, so large ratio of vowels to consonant means that percentage consonants %C will be small. That further implies that the mean of consonantal duration will be small, and consequently even their standard deviation will be small.
In this figure, we also notice a strong linear dependence between phonotactic measures percentage of vocalic intervals $\%Vp$ and standard deviation of consonantal intervals $\Delta Cp$. Linear trend is marked by a downward-sloped gray line in Figure 4.12.

Phonotactic Metrics $\%Vp$ and $\Delta Cp$ classify languages based on their phonotactic complexity. They provide more detailed description of phonotactic structure than a simple statement on syllable structure; these measures assess the frequencies of vocalic and consonantal clusters in running speech.

Next, we see in Figure 4.13 that the phonotactic pair-wise measures ($rPVI-Cr$, $nPVI-Vr$) can achieve a variety of groupings that agree with the traditional rhythm type, with some minor modifications. For instance, Serbian is positioned in the syllable-timed group on many measures, and Hungarian on these measures seems to group with the stress-timed languages. In Figure 4.12, Hungarian was also bordering stress-timed group.
Figure 4.12 Linear relationship between %Vp and ΔCp

Figure 4.13 Grouping of 21 languages based on phonotactic rPVI-Cp and nPVI-Vp
Finally, in Figure 4.14 we see that phonotactic metrics %Vp and Varco-Vp also group languages according to their traditional type, except that Bulgarian and Serbian pattern with stress-timed languages, while Estonian and Hungarian pattern with the stress-timed.

Figure 4.14 Grouping of 21 languages based on phonotactic %Vp and Varco-Vp

4.1.5 Conclusion

In summary, some phonotactic measures group languages similarly to rhythm metrics. The correlations between phonotactic and rhythmic metrics vary, but they seem positive in most metrics. This suggests that the kind of information provided by the groupings based on RMs is influenced by effects other than prosodic lengthening and vowel shortening, which are normally associated with different rhythmicity; they are at least partially affected by the phonotactics. This leads us to the conclusion that rhythm metrics include durational variability resulting from phonotactics. Procedures for examining rhythmic differences thus need to be based on a different kind of measures, or different kinds of materials in which phonotactics (or average interval complexity) is comparable across languages.
4.2 Consonant cluster lengths at different positions in the word

Having established that the durational aspects of phonotactics affect the measures that are claimed to quantify rhythmic properties, we now turn to examine the relation between phonotactics and rhythm in more detail. In this section I address question 2, in which we investigate the relationship between rhythmic similarity and the length of clusters at different positions in the word (word-initial, word-medial, and word-final). Note that with this question we address only the durational component of the clusters, that is, whether they consist of zero, one, two, or more segments and not the segmental quality. We are interested in investigating to what extent any particular position (initial, medial, or final) can explain rhythmic (perception) similarity.

As described in Chapter 3, I calculated the distributions of consonant-cluster lengths for each language at each of the 3 positions in the word. For each language, I provide the distribution over different lengths of consonant clusters. Values at each point in the distribution (presented in the table as percentages) are calculated over long texts, approximately 10,000 words in length. Exact number of words in each sample is provided in the tables.

4.2.1 Word-initial cluster distributions

Table 4.4 present results for the word-initial cluster distributions. All languages in our samples contain vowel-initial words, although, typologically, some languages must obligatory have at least one consonant at the beginning of the word (i.e., non-empty syllable onset). Since clusters of length-0 and length-1 exist in all languages, initial clusters of length 0 and 1 have been combined into a single category. Two other categories include initial clusters of length-2 and initial clusters of length-3 or higher.

Languages are presented in the increasing order of frequency of the first category (length-0 or length-1).
We can see that languages differ markedly in whether they allow more than one consonant at the beginning of the word, and then in frequency of those length-2 initial clusters. Compare the frequencies in the Slavic languages, Russian, Serbian, Bulgarian, Czech, and Polish (14-30%), with those for the two Germanic languages, German and Dutch (7%). Italian and Catalan have 5-6% clusters of length-2 as well.

Similarly, four of the Slavic languages, Bulgarian, Polish, Russian (the highest), and Czech, have a non-negligible proportion of length-3 clusters; with very few in Serbian and the Germanic languages. It is because of their similarity with respect to initial consonant clusters (and therefore, clusters in general) that Slavic languages have tended to be assigned to the stress-type language group. However, we will see that the similarity does not extend to word-final clusters. And if initial and final clusters have different effects on language
rhythm, then this traditional association of Germanic and Slavic languages might be misguided.

So far, initial cluster distribution correlates well with historic language family. While it appears that all traditionally stress-timed languages appear at the upper end of the table, with high proportion of initial clusters of length-2 or higher, distributions allow us to make finer gradation. Namely, we see that Dutch and German appear more similar in their word-initial cluster distribution to the traditionally syllable-timed Italian and Greek.

Complexity of word-initial clusters also does not correlate with the rhythm type based on phonological property of vowel-reduction. Although Slavic languages have the highest percentage of long initial clusters, they do not – except for Russian – exhibit vowel reduction. Also, while Catalan fits with the Germanic languages by virtue of having vowel reduction, in similarity of word-initial clusters it groups with Italian, which is generally assumed to be syllable-timed and in its standard version does not have vowel reduction.

At the lower end of the table, we find languages from the Polynesian group (Hawaiian, Maori, Samoan, and Tongan) and the Uralic language group (Hungarian and Estonian). They all have simple word-initial clusters.

In sum, Table 4.4 shows a clear grouping by language families, reflecting similarity based on historical relationship. In particular, this grouping classifies languages in the Germanic language group most similar to the language in the Romance group. This means that structure of initial consonant clusters does not correlate well with expected rhythmic similarity. As syllable onsets are sometimes called ‘weightless’ (Hyman 1984), to represent their lack of relevance towards calculation of syllable-weight, and since the syllable-weight is related to stress, it seems appropriate that the initial clusters, which are also syllable onsets, do not correlate with rhythm type.

4.2.2 Word-final cluster distributions

Next, we look at the distribution of word final consonant clusters. Since length-zero cluster represents a meaningful category for the word-final clusters, namely, it defines the canonical final, we will not combine clusters of length-0 and length-1. However, given that percentage of clusters of length-3 or more are rare, we will combine all clusters of length-2 or higher.
Results are presented in Table 4.5. Languages are first ordered by decreasing likelihood of clusters of length-2 or higher; Languages that do not allow such clusters are ordered by decreasing likelihood of length-1 clusters.

Table 4.5 Distribution of word-final clusters

<table>
<thead>
<tr>
<th>Language</th>
<th>length-0</th>
<th>length-1</th>
<th>length-2 or higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>35</td>
<td>48</td>
<td>18</td>
</tr>
<tr>
<td>Hungarian</td>
<td>33</td>
<td>52</td>
<td>15</td>
</tr>
<tr>
<td>Estonian</td>
<td>55</td>
<td>34</td>
<td>11</td>
</tr>
<tr>
<td>Dutch</td>
<td>37</td>
<td>54</td>
<td>9</td>
</tr>
<tr>
<td>Polish</td>
<td>67</td>
<td>31</td>
<td>2</td>
</tr>
<tr>
<td>Russian</td>
<td>62</td>
<td>36</td>
<td>2</td>
</tr>
<tr>
<td>Catalan</td>
<td>56</td>
<td>43</td>
<td>1</td>
</tr>
<tr>
<td>Czech</td>
<td>69</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>Turkish</td>
<td>46</td>
<td>54</td>
<td>1</td>
</tr>
<tr>
<td>Indonesian</td>
<td>46</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>Greek</td>
<td>64</td>
<td>36</td>
<td>0</td>
</tr>
<tr>
<td>Spanish</td>
<td>66</td>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>75</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>Portuguese</td>
<td>79</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>Serbian</td>
<td>82</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>Italian</td>
<td>88</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Japanese</td>
<td>95</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maori</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Samoan</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tongan</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Languages that stand out in this distribution are the Polynesian languages (Maori, Hawaiian, Samoan, and Tongan) plus Japanese, for the low likelihood of non-zero final clusters. Next, we note a group consisting of German, Dutch, Hungarian and Estonian: all of which have a high likelihood of final clusters of length-2 or higher.

While to a certain degree we again see some resemblance to language families, now it is the Uralic group (Estonian and Hungarian) that patterns similarly with the Germanic group, while some of the Slavic languages (Serbian and Bulgarian) look more like the Romance languages in that length-2 final clusters are infrequent. Other Slavic languages
(Russian, Czech, and Polish) appear closer to the Germanic group. Differences in cluster distributions across words are seen in that, while German and Dutch can have heavy clusters on both sides of a word, Slavic languages are only onset-heavy, and Uralic languages are only coda heavy. Because Dutch, German, Catalan and Russian exhibit vowel reduction, it seems at first that this cannot be related to cluster structure\textsuperscript{20}. However, because vowel reduction in Catalan seems to be qualitative and does not result in corresponding shortening (Prieto et al. 2012), it is possible that heavy-finals correspond only to vowel reduction associated with vowel shortening. If that is true, then similarity in the distribution of final consonant cluster could be said to correspond to rhythm type (presumably, stress-timing).

4.2.3 Word-medial cluster distributions

Finally, let us look at the distribution of word-medial clusters. Because medial clusters consist structurally of the coda from a preceding syllable and the onset of the following syllable, medial clusters may combine properties of initials and finals in their influence on rhythm

Since word-medial clusters of length-0 imply that a hiatus occurred, this case can be dealt with when analyzing vowels\textsuperscript{21}. Meaningful categories of word-medial clusters include length-1, length-2, and length-3 or higher. Results are presented in Table 4.6 where languages are ordered by decreasing likelihood of length-3 or higher clusters.

We observe in the table that Dutch, German, Russian, and Catalan have a somewhat higher percentage of length-3 medial clusters (5-9%). The middle group consists of the remaining Slavic, Romance, and Uralic languages, plus Turkish and Indonesian; these language allow complex clusters of length 3 or higher, but their likelihood is small. The third group consists of Japanese and the Polynesian group - languages that do not allow clusters length 3 and higher and have low likelihood of length-2 clusters (the only non-zero value is 8% for Japanese).

\textsuperscript{20} It is probable that the presence of phonemic vowel length in Serbian, Czech, Estonian, and Hungarian prevents vowel reduction.

\textsuperscript{21} Analysis of vowel clusters is not presented in this work.
Table 4.6 Distribution of word-medial clusters

<table>
<thead>
<tr>
<th>Language</th>
<th>length-1</th>
<th>length-2</th>
<th>length-3 or higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russian</td>
<td>60</td>
<td>31</td>
<td>9</td>
</tr>
<tr>
<td>Dutch</td>
<td>55</td>
<td>38</td>
<td>8</td>
</tr>
<tr>
<td>Catalan</td>
<td>70</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>German</td>
<td>62</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>Czech</td>
<td>70</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>68</td>
<td>29</td>
<td>3</td>
</tr>
<tr>
<td>Spanish</td>
<td>66</td>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>Polish</td>
<td>70</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td>Hungarian</td>
<td>57</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td>Greek</td>
<td>77</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>Serbian</td>
<td>77</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>Italian</td>
<td>56</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td>Estonian</td>
<td>70</td>
<td>29</td>
<td>2</td>
</tr>
<tr>
<td>Portuguese</td>
<td>81</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>Turkish</td>
<td>67</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>Indonesian</td>
<td>75</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>Japanese</td>
<td>92</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Tongan</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Samoan</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maori</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

This grouping allows us to notice stress-timed group in the upper part of the table, and mora-timed group in the lower part. However, the middle group is quite diverse in their distributions. Let us re-arrange the middle group in the increasing likelihood of the length-1 clusters. Results are presented in Table 4.7.

Now we see the languages with simple clusters lower in the table, and those with higher likelihood of length-2 medial clusters – in the upper part, closer to Dutch and German.
Table 4.7 Distribution of word-medial clusters (re-arranged)

<table>
<thead>
<tr>
<th>language</th>
<th>length-1</th>
<th>length-2</th>
<th>length-3 or higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>55</td>
<td>38</td>
<td>8</td>
</tr>
<tr>
<td>Russian</td>
<td>60</td>
<td>31</td>
<td>9</td>
</tr>
<tr>
<td>German</td>
<td>62</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>Catalan</td>
<td>70</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>Italian</td>
<td>56</td>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td>Hungarian</td>
<td>57</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td>Turkish</td>
<td>67</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>Spanish</td>
<td>66</td>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>68</td>
<td>29</td>
<td>3</td>
</tr>
<tr>
<td>Estonian</td>
<td>70</td>
<td>29</td>
<td>2</td>
</tr>
<tr>
<td>Polish</td>
<td>70</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td>Czech</td>
<td>70</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td>Indonesian</td>
<td>75</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>Serbian</td>
<td>77</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>Greek</td>
<td>77</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>Portuguese</td>
<td>81</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>Japanese</td>
<td>92</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Tongan</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Samoan</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maori</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

In this new table, we notice that Serbian and Greek have almost identical distributions, and so do Czech and Polish. These are however not the expected similar pairs. Namely, Polish has penultimate stress, while Czech has word-initial stress. Thus, this grouping also does not fully reflect rhythmic similarity, although it does distinguish stress-timed group (Dutch, German, Russian, and Catalan) from the mora-timed group, with everyone else in the third group.

4.2.4 Summary

Let us now review groupings obtained based on word-initial, word-final, and word-medial cluster distributions. Groupings are shown in Table 4.8.
Table 4.8 Language groupings based on word-initial, word-medial, and word-final complexity

<table>
<thead>
<tr>
<th>INITIAL</th>
<th>MEDIAL</th>
<th>FINAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russian</td>
<td>Russian</td>
<td>German</td>
</tr>
<tr>
<td>Polish</td>
<td>Dutch</td>
<td>Hungarian</td>
</tr>
<tr>
<td>Czech</td>
<td>Catalan</td>
<td>Estonian</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>German</td>
<td>Dutch</td>
</tr>
<tr>
<td>Serbian</td>
<td>Czech</td>
<td>Polish</td>
</tr>
<tr>
<td>Greek</td>
<td>Bulgarian</td>
<td>Russian</td>
</tr>
<tr>
<td>German</td>
<td>Spanish</td>
<td>Catalan</td>
</tr>
<tr>
<td>Dutch</td>
<td>Polish</td>
<td>Czech</td>
</tr>
<tr>
<td>Italian</td>
<td>Hungarian</td>
<td>Turkish</td>
</tr>
<tr>
<td>Catalan</td>
<td>Greek</td>
<td>Indonesian</td>
</tr>
<tr>
<td>Portuguese</td>
<td>Serbian</td>
<td>Greek</td>
</tr>
<tr>
<td>Spanish</td>
<td>Italian</td>
<td>Spanish</td>
</tr>
<tr>
<td>Turkish</td>
<td>Estonian</td>
<td>Bulgarian</td>
</tr>
<tr>
<td>Indonesian</td>
<td>Portuguese</td>
<td>Portuguese</td>
</tr>
<tr>
<td>Japanese</td>
<td>Turkish</td>
<td>Serbian</td>
</tr>
<tr>
<td>Estonian</td>
<td>Indonesian</td>
<td>Italian</td>
</tr>
<tr>
<td>Hungarian</td>
<td>Japanese</td>
<td>Japanese</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>Tongan</td>
<td>Hawaiian</td>
</tr>
<tr>
<td>Maori</td>
<td>Samoan</td>
<td>Maori</td>
</tr>
<tr>
<td>Samoan</td>
<td>Maori</td>
<td>Samoan</td>
</tr>
<tr>
<td>Tongan</td>
<td>Hawaiian</td>
<td>Tongan</td>
</tr>
</tbody>
</table>

We observe that the word-initial cluster complexity reflects mostly historical language groupings. Groupings based on word-medial and word-final cluster distributions both agree with the rhythmic similarity grouping in that prototypically stress-timed and prototypically mora-timed languages are placed on the opposite sides of the table. The Germanic pair (German and Dutch) groups with the vowel-reduction exhibiting Russian and Catalan in the word-medial cluster grouping. They group with Uralic, coda-heavy, Hungarian and Estonian in the word-final cluster grouping. Although Hungarian and Estonian tend to be assigned to syllable-timed group, they both have fixed (word-initial) stress and could be possibly seen as rhythmically similar with the trochaic, stress-timed, German and Dutch.

The placement of the Romance languages, which are traditionally assigned to syllable-timed group, closer together in the word-final distributions (all three are in the lower
part of the middle group) makes this grouping more likely correlate of rhythm than that based on the distributions of the word-medial clusters.

In summary, similarities in word structure, seen through patterning of consonants at the beginning, in the middle, and at the end of the word, produce different groupings of the 21 languages depending on the position. This suggests that the information such as “language allows consonant clusters”, which are a part of certain models of rhythmic similarity (Dauer 1987), may be confounding different kinds of properties – those associated with the word-initial clusters, and those associated with the word-final clusters.

Grouping according to the word-final clusters most closely resembles the rhythmic grouping based on the traditional rhythm classes and some information on the phonological properties (like vowel reduction). To establish true correspondence, however, between the phonotactic and rhythmic similarity, we will need independent evidence of rhythmic similarity for a large number of languages, preferably obtained through the perception experiments.

4.3 Phonotactic patterns

When asked to describe differences between languages, listeners usually pay attention to two aspects: prosodic gestalts and characteristic, frequent segments. In imitations, prosodic properties appear as melodic variations, characteristic stress patterns (for instance, word final stress), and short-long alternations. On the segmental level, listeners often associate Russian with palatal sounds, Polish with clusters involving fricatives, and French with rounded vowels. However, while such characteristic segments are not the absolutely most frequent in a given language, they are usually segments that are more frequent in that language than in other languages.

In section 4.2, we looked at the similarities in consonant-cluster lengths as they occur in word-initial, word-medial, and word-final positions. In this section, we zoom into phonotactic properties by examining the types of segments that make up word-initial, word-medial, and word-final clusters.

We look for similarity of phonotactic sequences across languages at 2 different levels: 1) the ‘saltanaje’ level, which is obtained by replacing each segment by a representative
segment of the same manner of articulation (‘t’ for stops, ‘s’ for fricatives, ‘c’ for affricates, ‘n’ for nasals, ‘l’ for liquids, ‘j’ for semi-vowels, and ‘a’ for vowels); and 2) the ALT (A for vowels, L for sonorants, T for obstruents) level, in which all consonants are divided into 2 classes: sonorants (nasals, liquids, and semi vowels) and obstruents (stops, fricatives, and affricates). We ask two questions related to the phonotactic patterns of consonants cross-linguistically: 1) whether languages that are said to be of the same rhythmic type exhibit similar phonotactic patterns; and 2) whether phonotactic patterns that are defined as ‘better clusters at specific positions in the word’ (according to the sonority principles) are more frequent in each language in the sample.

The first question is asked to elucidate whether languages with similar cluster sizes (lengths in number of segments) differ in rhythm when the composition of those clusters differs; for instance, does ‘having more obstruents’ in certain positions affect language rhythm differently from ‘having more sonorants’ in those positions? Such an effect can come about in two ways: 1) a sonorant behaves differently from an obstruent in that it carries pitch and can be perceived as having duration; and 2) the presence of an obstruent affects the duration of the neighboring vowel.

The second question is asked in attempt to identify unusual patterns that can help distinguish one language or one group of languages from another.

We ask these research questions at two different levels of representation: ALT (vowel-sonorant-obstruent) and saltanajc (manner of articulation representation).

4.3.1 Basic sonority (ALT) level

a) word-initial consonant cluster patterns

Results for length-0 and length-1 initial clusters are given in Table 4.9. Languages are ordered by language family. The most frequent onset for each language is shaded. We see that, judged by frequency in the corpora, obstruents are the ‘best’ initials, except for Samoan which has a large percentage of vowel initial words. The sonority principle requires an initial cluster to manifest an increase of sonority moving towards the nucleus. By virtue of having an obstruent in initial onset position, this sonority rise is maximized.
Table 4.9 Word-initial length-0 and length-1 clusters

<table>
<thead>
<tr>
<th>language</th>
<th>#_A</th>
<th>#LA</th>
<th>#TA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samoan</td>
<td>48.7</td>
<td>29.5</td>
<td>21.8</td>
</tr>
<tr>
<td>Russian</td>
<td>18.2</td>
<td>14.3</td>
<td>34.1</td>
</tr>
<tr>
<td>Czech</td>
<td>14.5</td>
<td>21.1</td>
<td>39.4</td>
</tr>
<tr>
<td>Hungarian</td>
<td>36.9</td>
<td>23.1</td>
<td>39.8</td>
</tr>
<tr>
<td>Serbian</td>
<td>22.2</td>
<td>21.4</td>
<td>41.6</td>
</tr>
<tr>
<td>Polish</td>
<td>11.6</td>
<td>18.9</td>
<td>44.1</td>
</tr>
<tr>
<td>Spanish</td>
<td>33.8</td>
<td>18.9</td>
<td>44.3</td>
</tr>
<tr>
<td>Greek</td>
<td>35.0</td>
<td>8.9</td>
<td>46.0</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>28.7</td>
<td>24.7</td>
<td>46.6</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>17.9</td>
<td>17.9</td>
<td>47.4</td>
</tr>
<tr>
<td>Portuguese</td>
<td>33.5</td>
<td>15.7</td>
<td>47.4</td>
</tr>
<tr>
<td>Italian</td>
<td>24.6</td>
<td>19.0</td>
<td>49.5</td>
</tr>
<tr>
<td>Japanese</td>
<td>22.3</td>
<td>27.4</td>
<td>49.6</td>
</tr>
<tr>
<td>Maori</td>
<td>28.3</td>
<td>22.0</td>
<td>49.7</td>
</tr>
<tr>
<td>Catalan</td>
<td>21.2</td>
<td>22.8</td>
<td>50.2</td>
</tr>
<tr>
<td>Dutch</td>
<td>20.8</td>
<td>21.5</td>
<td>50.5</td>
</tr>
<tr>
<td>Estonian</td>
<td>21.0</td>
<td>26.8</td>
<td>51.5</td>
</tr>
<tr>
<td>German</td>
<td>26.9</td>
<td>11.7</td>
<td>53.1</td>
</tr>
<tr>
<td>Indonesian</td>
<td>16.3</td>
<td>21.9</td>
<td>61.1</td>
</tr>
<tr>
<td>Turkish</td>
<td>23.7</td>
<td>10.9</td>
<td>64.2</td>
</tr>
<tr>
<td>Tongan</td>
<td>15.0</td>
<td>18.8</td>
<td>66.2</td>
</tr>
</tbody>
</table>

Thus, languages do not differ in whether obstruents or sonorants are preferred in initial cluster position, although they do differ in the percentage of obstruents occupying this position. Ordering by the frequency of obstruent in the initial position (in Table 4.9) does not reveal any meaningful language groupings.

Results for the length-2 word-initial clusters are given in Table 4.10, ordered by frequency of #TL (initial obstruent-sonorant). The most frequent type is marked in dark green. In all languages except Polish this is obstruent-sonorant, which is the ideal pattern from the sonority principle point of view (with sonority rising from obstruent to sonorant and from sonorant to the vowel).
Table 4.10 Word initial length-2 clusters

<table>
<thead>
<tr>
<th>language</th>
<th>#TL</th>
<th>#TT</th>
<th>#LL</th>
<th>#LT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russian</td>
<td>14.4</td>
<td>6.2</td>
<td>6.7</td>
<td>0.1</td>
</tr>
<tr>
<td>Czech</td>
<td>12.4</td>
<td>8.6</td>
<td>1.4</td>
<td>0.6</td>
</tr>
<tr>
<td>Polish</td>
<td>9.1</td>
<td>13.2</td>
<td>0.4</td>
<td>-</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>8.8</td>
<td>5</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Serbian</td>
<td>8.6</td>
<td>4.9</td>
<td>0.5</td>
<td>-</td>
</tr>
<tr>
<td>Greek</td>
<td>4.8</td>
<td>4.7</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>German</td>
<td>5.1</td>
<td>2.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dutch</td>
<td>5.3</td>
<td>1.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Italian</td>
<td>4.2</td>
<td>2.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Catalan</td>
<td>3.5</td>
<td>0.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Portuguese</td>
<td>3.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Spanish</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Turkish</td>
<td>1.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Indonesian</td>
<td>0.4</td>
<td>-</td>
<td>-</td>
<td>0.3</td>
</tr>
<tr>
<td>Japanese</td>
<td>0.4</td>
<td>-</td>
<td>0.2</td>
<td>-</td>
</tr>
<tr>
<td>Estonian</td>
<td>0.3</td>
<td>0.2</td>
<td>-</td>
<td>0.1</td>
</tr>
<tr>
<td>Hungarian</td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

In Polish, obstruent-obstruent is most frequent, while obstruent-sonorant is second most frequent. Although at this level we are not able to tell which phonemes contribute towards each pattern, the overall higher presence of obstruents in Polish (2:1 obstruent to sonorants as seen in Table 4.16 in the next section) suggests an overall preference for obstruents in Polish as compared to the other languages in our set.

The second most frequent pattern in all languages in which at least two patterns exist (except Polish which has TL and TT reversed) is TT, and the third most frequent is LL. According to the sonority principle, these two patterns are the second and third most preferred: sonority is constant then rises towards the vowel; in case of TTA the rise in sonority is greater than in ‘LLA’ (A representing a vowel), thus giving TT slight advantage in terms of initial cluster goodness.

Only Czech allows an LT pattern in word-initial position; this pattern comes from the frequent phoneme cluster /js/. This cluster which violates the sonority principle (sonority drops from sonorant to obstruent then rises from obstruent to the vowel) only occurs in one language, and its frequency is only 0.6%. Thus the correlation between ‘preference by
sonority principle’ and frequency in the corpus essentially holds for all our languages. (In Russian the difference between TT and LL is very slight.)

For length-3 clusters, as seen in Table 4.11, most languages have ‘oos’ as most frequent pattern, except Bulgarian for which this is second most frequent, most frequent being ‘oss’. Both of these patterns have rising sonority from the first consonant in C1C2C3V to the nucleus, so it is not possible to decide which one is ‘better’. Requiring a higher slope closer to the nucleus, if such requirement existed, could explain why TTL is more frequent cross-linguistically.

Table 4.11 Word initial length-3 clusters

<table>
<thead>
<tr>
<th>language</th>
<th>#CCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russian</td>
<td>TTL</td>
</tr>
<tr>
<td>Czech</td>
<td>TTL</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>TLL &gt; TTL</td>
</tr>
<tr>
<td>Polish</td>
<td>TTL &gt; TTT</td>
</tr>
<tr>
<td>German</td>
<td>TTL</td>
</tr>
<tr>
<td>Italian</td>
<td>TTL</td>
</tr>
<tr>
<td>Serbian</td>
<td>TTL</td>
</tr>
<tr>
<td>Dutch</td>
<td>TTL</td>
</tr>
</tbody>
</table>

Notice also that the second most frequent pattern in Polish is TTT, again suggesting that obstruents are a preferred segment class in Polish.

We can see that languages obey the universal word (or syllable) building sonority principles with some small exceptions (Polish length-2 clusters), choosing ‘best’ clusters as the most frequent ones. Moreover, the variety of patterns reflects historical language grouping: Slavic languages have three most likely patterns (TL, TT, and LL), Germanic plus Italian allow two patterns, Romance language and Turkish only one pattern, and Japanese, Austronesian languages, and Uralic languages have no initials of length greater than 1.

That the dispreferred pattern LT occurs in Czech, and with some very small frequency in Indonesian, agrees with the view of Dziubalska-Kołascyk (2001) that phonotactic conditions are better understood as preferences rather than as constraints.
We thus conclude based on results in Tables 4.9-11 that – as in our discussion of the
distribution of cluster lengths (section 4.2) – initial clusters better reflect their language group
than their language rhythm classification. In particular, Italian – traditionally assumed to be
of different rhythm type than German and Dutch – has a very similar word-initial cluster
structure to theirs.

b) **word-final** consonant cluster patterns

Next, let us look at the **word-final** cluster patterns. We wonder if at this level of
representation, obstruent-sonorant-vowel, we can gain any information compared to the CV
level examined in section 4.2. Results for length-0 and length-1 final clusters are presented in
Table 4.12.

First, notice that, as opposed to initial position, the most preferred, i.e., most
frequent\(^{22}\), final cluster is length-0 cluster. This finding agrees with the fact that the canonical
syllable (the one present in all languages) is CV. More specifically, across our corpora it is
TA (obstruent+vowel).

If we order languages by the percent of length-2 final clusters, we obtain the same
grouping as in section 4.2 based only on cluster lengths. So instead, we group languages
based on the sum of percentages of finals of length zero (A_#) and finals of length 1 (AL#)
where the final consonant is a sonorant. The sum of these two frequencies is given in the last
column of Table 4.12.

This grouping is based on the often-discussed fact that sonorants in some sense
behave like vowels: they are voiced, can carry pitch, and often have intensity comparable to
that of a vowel. Thus, we have ordered our languages based on the percentage of words that
end in a sonorant segment.

We can observe three salient groups. The first one consists of languages with heavy
finals: Hungarian, Estonian, Dutch, and German; the second consists of languages that
exclusively end in a sonorant segment: Tongan, Samoan, Maori, Hawaiian, and Japanese; the
third group contains intermediate languages, ranging from German-like group towards the
Hawaiian-like group: Russian, Greek, Indonesian, Turkish, Polish, Catalan, Czech,
Portuguese, Spanish, Bulgarian, Serbian, Italian.

\(^{22}\) In Hungarian, length-0 final cluster ties with single obstruent cluster
Table 4.12 Word-final length-0 and length-1 clusters

<table>
<thead>
<tr>
<th>Language</th>
<th>A_#</th>
<th>AL#</th>
<th>AT#</th>
<th>A_#+AL#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hungarian</td>
<td>33</td>
<td>19</td>
<td>33</td>
<td>52</td>
</tr>
<tr>
<td>Estonian</td>
<td>55</td>
<td>9</td>
<td>25</td>
<td>64</td>
</tr>
<tr>
<td>Dutch</td>
<td>37</td>
<td>30</td>
<td>24</td>
<td>66</td>
</tr>
<tr>
<td>German</td>
<td>35</td>
<td>33</td>
<td>23</td>
<td>68</td>
</tr>
<tr>
<td>Russian</td>
<td>62</td>
<td>14</td>
<td>22</td>
<td>77</td>
</tr>
<tr>
<td>Greek</td>
<td>64</td>
<td>14</td>
<td>22</td>
<td>78</td>
</tr>
<tr>
<td>Indonesian</td>
<td>46</td>
<td>33</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td>Turkish</td>
<td>46</td>
<td>34</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>Polish</td>
<td>67</td>
<td>14</td>
<td>17</td>
<td>81</td>
</tr>
<tr>
<td>Catalan</td>
<td>57</td>
<td>25</td>
<td>17</td>
<td>83</td>
</tr>
<tr>
<td>Czech</td>
<td>69</td>
<td>16</td>
<td>14</td>
<td>85</td>
</tr>
<tr>
<td>Portuguese</td>
<td>79</td>
<td>6</td>
<td>15</td>
<td>85</td>
</tr>
<tr>
<td>Spanish</td>
<td>66</td>
<td>20</td>
<td>14</td>
<td>86</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>75</td>
<td>15</td>
<td>11</td>
<td>89</td>
</tr>
<tr>
<td>Serbian</td>
<td>82</td>
<td>8</td>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>Italian</td>
<td>88</td>
<td>11</td>
<td>1</td>
<td>99</td>
</tr>
<tr>
<td>Japanese</td>
<td>95</td>
<td>6</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>Maori</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>Samoan</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>Tongan</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
</tbody>
</table>

Since in this ordering Russian is placed closer to the German-like group, and Serbian and Italian closer to the Hawaiian-like group, with Polish and Catalan – often considered mixed-type languages – somewhere in the middle of the group, it appears that this grouping shows a somewhat better match to perception. Naturally, only results of actual perception tests could confirm or disprove this.

Next, let us look at final clusters of length-2. Results are presented in Table 4.13.

Table 4.13 World-final length-2 clusters

<table>
<thead>
<tr>
<th>Language</th>
<th>LT#</th>
<th>TT#</th>
<th>LL#</th>
<th>TL#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hungarian</td>
<td>7.7</td>
<td>6.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Estonian</td>
<td>3.4</td>
<td>6.1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Dutch</td>
<td>6.1</td>
<td>3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>German</td>
<td>5.8</td>
<td>2.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Polish</td>
<td>-</td>
<td>1.1</td>
<td>-</td>
<td>0.6</td>
</tr>
<tr>
<td>Russian</td>
<td>-</td>
<td>1.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Czech</td>
<td>-</td>
<td>0.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Turkish</td>
<td>0.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Only languages that have length-2 clusters are presented in the table. Frequencies smaller than 0.5% are omitted.

Based on sonority, ‘so’ should be the preferred final cluster, with sonority dropping from vowel to sonorant and from sonorant to obstruent. The close second is TT with a bigger sonority drop from the vowel to the first consonant but with flat sonority between the obstruents. This combination occurs in 7 out of 8 languages that have length-2 final clusters, more than the LT pattern, although the frequencies in the languages that allow both tend to be higher for LT. Let us therefore now order the languages according to length-2 clusters in a fashion similar to that we used to order languages in Table 4.12, namely, by the percentage of words that end in one obstruent: this percentage is a sum of percentages of LT length-2 cluster and T length-1 cluster. Results are shown in Table 4.14.

Table 4.14 Word-final clusters grouped by sonority*

<table>
<thead>
<tr>
<th>language</th>
<th>A_#+AL#</th>
<th>AT#+ALT#</th>
<th>ATT#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hungarian</td>
<td>52</td>
<td>41</td>
<td>7</td>
</tr>
<tr>
<td>Estonian</td>
<td>64</td>
<td>28</td>
<td>6</td>
</tr>
<tr>
<td>Dutch</td>
<td>66</td>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>German</td>
<td>68</td>
<td>29</td>
<td>2</td>
</tr>
<tr>
<td>Russian</td>
<td>77</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>Greek</td>
<td>78</td>
<td>22</td>
<td>-</td>
</tr>
<tr>
<td>Indonesian</td>
<td>79</td>
<td>21</td>
<td>-</td>
</tr>
<tr>
<td>Turkish</td>
<td>80</td>
<td>20</td>
<td>-</td>
</tr>
<tr>
<td>Polish</td>
<td>81</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>Catalan</td>
<td>83</td>
<td>17</td>
<td>-</td>
</tr>
<tr>
<td>Czech</td>
<td>85</td>
<td>14</td>
<td>0.7</td>
</tr>
<tr>
<td>Portuguese</td>
<td>85</td>
<td>15</td>
<td>-</td>
</tr>
<tr>
<td>Spanish</td>
<td>86</td>
<td>14</td>
<td>-</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>89</td>
<td>11</td>
<td>-</td>
</tr>
<tr>
<td>Serbian</td>
<td>90</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>Italian</td>
<td>99</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Japanese</td>
<td>100</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>100</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Maori</td>
<td>100</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Samoan</td>
<td>100</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tongan</td>
<td>100</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*Languages ordered by the frequency of final sonorous segments (L or A)

We now see that the same order describes 3 phenomena across our languages: 1) increase in percentage of words that end in a sonorous element, 2) decrease in percentage of words that
end in one obstruent, and 3) decrease in percentage of words that end in two obstruents (for values larger than 1%). The order of the languages based on this distribution of segments is shown in Figure 4.15.

Figure 4.15 Distribution of word-final clusters based on sonority: cross-linguistic comparison

We see that this ordering corresponds to rhythm grouping to the extent of available perception results. German, Dutch, and Russian are close and so are Japanese and the Polynesian languages. Romance and Slavic languages occupy the middle area, with Catalan and Polish on one side (close to German) and Italian and Serbian on the other side (close to Japanese).

It would be interesting to obtain perception similarity results for all pairs of languages so that the groupings based on phonotactics could be discussed in relation to perceptual (or rhythmic) similarity. At the moment, only some coarse observation can be made regarding this relationship.

Word-final clusters of length-3 are rare across languages and within each language. Results are presented in Table 4.15.
The most frequent length-3 final clusters in each language constitute less than 1% of all final clusters. Most languages that allow length-3 final clusters have LTT as the most frequent type; Hungarian has ‘LLT’ as the most frequent, and LTT as the second most frequent, differing in frequency only by 0.01%.

German and Dutch have the highest frequency of length-3 final clusters and a wider variety of patterns. It would be very interesting to see how English behaves on these measures in relation to German and Dutch. English has not been included in the sample so far because of its less transparent grapheme-to-phoneme rules.

We can thus see that with a careful analysis at the obstruent-sonorant level, it is possible to capture not only some universal tendencies, such as which patterns are the best fit for word-initial or word-final consonant clusters, but also to notice some correlation between patterns in word-final position and rhythm grouping.

### 4.3.2 Detailed sonority (saltanajc) level

At the saltanajc level of representation, each phoneme is replaced by its class (manner of articulation) representative. This will allow us to make more subtle distinctions between phonemes (stops from fricatives, for instance) but at the same time keep the number of cluster combinations manageable.

We again look at the possible consonant clusters at three positions in a word and compare them against preferred clusters in each position as defined by sonority distance (Dziubalska-Kołaczyk 2001). These preference criteria are proposed to be universal, that is, independent of language. We check for each language 1) whether only preferred clusters occur at each position, 2) whether ‘better’ clusters, determined using sonority distance

---

**Table 4.15 Word-final clusters of length-3**

<table>
<thead>
<tr>
<th>language</th>
<th>CCC#</th>
<th>frequency(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>LTT&gt;TTT&gt;LLT</td>
<td>0.59</td>
</tr>
<tr>
<td>Dutch</td>
<td>LTT&gt;TTT&gt;LLT</td>
<td>0.41</td>
</tr>
<tr>
<td>Hungarian</td>
<td>LLT &gt; LTT</td>
<td>0.17</td>
</tr>
<tr>
<td>Estonian</td>
<td>LTT</td>
<td>0.06</td>
</tr>
</tbody>
</table>
principle, occur more frequently in each corpus, and 3) whether ‘better’ clusters occur in a larger number of languages.

We then compare clusters across languages to determine phonotactic similarity. Since the number of possible clusters will affect the absolute frequencies (more combinations, lower frequencies overall), we compare clusters based on their rank. To keep comparisons manageable, we compare the number of allowed clusters in a language and the three most prominent clusters in that language against those in the other languages.

First, let us look at the overall frequencies of saltanajc classes. These are given in Table 4.16. Numbers are rounded to the nearest percent. A value of 0 is given for classes that do not occur, or for those that occur less frequently than 0.5%. Languages are ordered in the increasing value of vowel percentage.

<table>
<thead>
<tr>
<th>vowels</th>
<th>stops</th>
<th>fricatives</th>
<th>affricates</th>
<th>nasals</th>
<th>liquids</th>
<th>semi-vowels</th>
<th>language</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>20</td>
<td>17</td>
<td>0</td>
<td>11</td>
<td>11</td>
<td>3</td>
<td>Dutch</td>
</tr>
<tr>
<td>40</td>
<td>18</td>
<td>21</td>
<td>1</td>
<td>15</td>
<td>4</td>
<td>0</td>
<td>German</td>
</tr>
<tr>
<td>40</td>
<td>20</td>
<td>13</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>Russian</td>
</tr>
<tr>
<td>41</td>
<td>22</td>
<td>12</td>
<td>3</td>
<td>11</td>
<td>9</td>
<td>2</td>
<td>Hungarian</td>
</tr>
<tr>
<td>42</td>
<td>19</td>
<td>15</td>
<td>2</td>
<td>8</td>
<td>10</td>
<td>3</td>
<td>Czech</td>
</tr>
<tr>
<td>42</td>
<td>18</td>
<td>16</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>Polish</td>
</tr>
<tr>
<td>43</td>
<td>21</td>
<td>14</td>
<td>0</td>
<td>9</td>
<td>10</td>
<td>2</td>
<td>Hungarian</td>
</tr>
<tr>
<td>43</td>
<td>17</td>
<td>10</td>
<td>2</td>
<td>11</td>
<td>13</td>
<td>3</td>
<td>Turkish</td>
</tr>
<tr>
<td>44</td>
<td>21</td>
<td>13</td>
<td>2</td>
<td>8</td>
<td>9</td>
<td>3</td>
<td>Bulgarian</td>
</tr>
<tr>
<td>44</td>
<td>17</td>
<td>14</td>
<td>3</td>
<td>10</td>
<td>9</td>
<td>2</td>
<td>Serbian</td>
</tr>
<tr>
<td>45</td>
<td>22</td>
<td>7</td>
<td>2</td>
<td>16</td>
<td>8</td>
<td>1</td>
<td>Indonesian</td>
</tr>
<tr>
<td>45</td>
<td>14</td>
<td>16</td>
<td>0</td>
<td>11</td>
<td>13</td>
<td>0</td>
<td>Catalan</td>
</tr>
<tr>
<td>46</td>
<td>19</td>
<td>9</td>
<td>2</td>
<td>10</td>
<td>13</td>
<td>0</td>
<td>Italian</td>
</tr>
<tr>
<td>46</td>
<td>17</td>
<td>14</td>
<td>0</td>
<td>10</td>
<td>13</td>
<td>0</td>
<td>Spanish</td>
</tr>
<tr>
<td>47</td>
<td>16</td>
<td>20</td>
<td>0</td>
<td>10</td>
<td>7</td>
<td>0</td>
<td>Greek</td>
</tr>
<tr>
<td>50</td>
<td>17</td>
<td>17</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>0</td>
<td>Portuguese</td>
</tr>
<tr>
<td>52</td>
<td>18</td>
<td>9</td>
<td>2</td>
<td>12</td>
<td>4</td>
<td>2</td>
<td>Japanese</td>
</tr>
<tr>
<td>55</td>
<td>20</td>
<td>6</td>
<td>0</td>
<td>11</td>
<td>6</td>
<td>1</td>
<td>Maori</td>
</tr>
<tr>
<td>56</td>
<td>22</td>
<td>10</td>
<td>0</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>Tongan</td>
</tr>
<tr>
<td>57</td>
<td>20</td>
<td>6</td>
<td>0</td>
<td>11</td>
<td>5</td>
<td>2</td>
<td>Hawaiian</td>
</tr>
<tr>
<td>62</td>
<td>9</td>
<td>8</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>Samoan</td>
</tr>
</tbody>
</table>

We see from the table that certain classes of sounds are unexpectedly frequent in some languages. For example, Russian stands out by its high frequency of semivowels; Greek, Portuguese, German, and Dutch by their proportion of fricatives.
These articulatory class frequencies express to a certain extent similarities within a language family. Knowing for instance that Portuguese /rr/ is mapped into a fricative in Brazilian Portuguese (studied here), we see that Romance languages can be characterized by a high frequency of liquids (in particular, rhotics).

Note that while more likely combinations of segments according to sonority principles are expected to guide the presence of cluster types in a language corpus, the occurrence of a cluster containing a certain segment will also depend on the frequency of that segment in the language. So, if affricates are rare in a language, say they occur with low frequency in clusters of length-1 (individually), then the frequency of a ‘lc’ cluster can be expected to be low as well.

Results for length-2 clusters converted to ‘saltanajc’ representation level can be viewed in Tables 4.17, 4.19, and 4.21 for the initial, medial, and final position respectively. The sonority preferred clusters in each position are listed at the bottom of each table in order of predicted preference at that position.

a) word-initial consonant cluster patterns

Table 4.17 presents clusters in word-initial position for languages in which length-2 initial clusters occur. There are 16 such languages in our set but in only 12 do length-2 clusters of a certain type occur more than 0.5% of the time. Cluster listed in parentheses have frequencies less than 1%.

Sonority-preferred clusters in initial positions are listed at the bottom of the table; they are presented in lighter orange (gray) if they do not occur in the corpus. Preferred clusters are highlighted in orange for each language in which they occur. Clusters that do occur, but are not predicted (not considered preferred in initial positions), are given in blue (dark).

The first thing to notice is that both among the higher frequency clusters and lower frequency clusters (those provided in parentheses) there is a mixture of clusters preferred in the initial position and clusters that are preferred in other positions.

Next, we see that the only predicted preferred cluster in initial position that has not been found in the data is ‘cl’, i.e., an affricate followed by a liquid. One possible explanation
for the lack of this cluster is the inherent complexity of affricates, which are thought by some linguists to be sequences of two segments. Indeed, initial clusters with affricates that occur more frequently than 0.5% of the time in our sample are limited to a single pattern in Polish (‘sc’) and Russian (‘ct’). While ‘cl’ is attested in Serbian, as in words ‘član’ member, ‘članak’ article, ‘članarina’ membership fee, its frequency even in the Serbian corpus is low and does not place it in Table 4.17.

Table 4.17   Clusters of length-2 in ‘saltanaje’ scale: word-initial position

<table>
<thead>
<tr>
<th>language</th>
<th>Initial CC clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>tl &gt; sl &gt; st &gt; (tj &gt; ss &gt; sn)</td>
</tr>
<tr>
<td>German</td>
<td>ts &gt; st &gt; ss &gt; tl &gt; sl &gt; (sn &gt; cs &gt; tn)</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>tl &gt; st &gt; sl &gt; ss &gt; tj &gt; sn &gt; nj &gt; (ts &gt; nn &gt; sj &gt; cj)</td>
</tr>
<tr>
<td>Czech</td>
<td>tl &gt; sl &gt; tt &gt; ss &gt; st &gt; nn &gt; ts &gt; sn &gt; tj &gt; (js &gt; tn &gt; nl &gt; sj &gt; sc &gt; tc &gt; tj)</td>
</tr>
<tr>
<td>Polish</td>
<td>ts &gt; tl &gt; st &gt; ss &gt; sn &gt; tt &gt; sj &gt; sc &gt; sl &gt; tj &gt; (tn &gt; cs &gt; nl &gt; nn)</td>
</tr>
<tr>
<td>Russian</td>
<td>tl &gt; nj &gt; tj &gt; st &gt; sj &gt; ss &gt; ct &gt; sl &gt; ts &gt; lj &gt; sn &gt; tt &gt; (lm &gt; ls)</td>
</tr>
<tr>
<td>Serbian</td>
<td>tl &gt; st &gt; ss &gt; sl &gt; sn &gt; (ts &gt; nl &gt; tt &gt; nn &gt; tn)</td>
</tr>
<tr>
<td>Catalan</td>
<td>tl &gt; sl &gt; (st &gt; nt &gt; ls &gt; ns)</td>
</tr>
<tr>
<td>Italian</td>
<td>tl &gt; st &gt; (sl)</td>
</tr>
<tr>
<td>Portuguese</td>
<td>tl &gt; sl</td>
</tr>
<tr>
<td>Spanish</td>
<td>tl &gt; (sl &gt; tj)</td>
</tr>
<tr>
<td>Estonian</td>
<td>(tl)</td>
</tr>
<tr>
<td>Hungarian</td>
<td>(tl)</td>
</tr>
<tr>
<td>Greek</td>
<td>tl &gt; st &gt; ss &gt; sl &gt; ts &gt; (sn &gt; ns &gt; tt &gt; nt)</td>
</tr>
<tr>
<td>Indonesian</td>
<td>(sj &gt; nt &gt; tl)</td>
</tr>
<tr>
<td>Japanese</td>
<td>(sj &gt; tj &gt; lj &gt; nj)</td>
</tr>
<tr>
<td>PREFERRED</td>
<td>tj &gt; cj &gt; sj = tl &gt; nj = cl &gt; lj = sl = tn</td>
</tr>
</tbody>
</table>

Due to the large number of clusters in initial position, let us consider a simplified table from which low frequency clusters have been omitted. This is Table 4.18 in which languages are ordered to reflect the increasing number of patterns.

Notice that cross-linguistically ‘tl’ and ‘sl’ are the most common among the preferred clusters: they occur in all languages that allow length-2 initial clusters, even though in Italian this pattern is in the low frequency group. Thus, by a frequency account, these are the best length-2 initials.
Also, the cluster ‘st’ – which is not OSSP-preferred in initial position – actually occurs in all but two languages, Spanish and Portuguese, with Catalan having it among the low frequency clusters. Another frequent pattern includes ‘ss’; this pattern occurs in languages other than Romance, being in the low frequency group in Dutch.

Table 4.18  Clusters of length-2 in ‘saltanajc’ scale: word-initial position (re-arranged)

<table>
<thead>
<tr>
<th>language</th>
<th>Initial CC clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>tl</td>
</tr>
<tr>
<td>Catalan</td>
<td>tl &gt; sl</td>
</tr>
<tr>
<td>Portuguese</td>
<td>tl &gt; sl</td>
</tr>
<tr>
<td>Italian</td>
<td>tl &gt; st</td>
</tr>
<tr>
<td>Dutch</td>
<td>tl &gt; sl &gt; st</td>
</tr>
<tr>
<td>German</td>
<td>ts &gt; st &gt; ss &gt; tl &gt; sl</td>
</tr>
<tr>
<td>Greek</td>
<td>tl &gt; st &gt; ss &gt; sl &gt; ts</td>
</tr>
<tr>
<td>Serbian</td>
<td>tl &gt; st &gt; ss &gt; sl &gt; sn</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>tl &gt; st &gt; sl &gt; ss &gt; tj &gt; sn &gt; nj</td>
</tr>
<tr>
<td>Czech</td>
<td>tl &gt; sl &gt; tt &gt; ss &gt; st &gt; nn &gt; ts &gt; sn &gt; tj</td>
</tr>
<tr>
<td>Polish</td>
<td>ts &gt; tl &gt; st &gt; ss &gt; sn &gt; tt &gt; sj &gt; sc &gt; sl &gt; tj</td>
</tr>
<tr>
<td>Russian</td>
<td>tl &gt; nj &gt; tj &gt; st &gt; sj &gt; ss &gt; ct &gt; sl &gt; ts &gt; lj &gt; sn &gt; tt</td>
</tr>
<tr>
<td>PREFERRED</td>
<td>tj &gt; cj &gt; sj = tl &gt; nj = cl &gt; lj = sl = tn</td>
</tr>
</tbody>
</table>

It is not surprising that the languages within each historical group, Romance, Germanic, or Slavic, pattern together. Five Slavic languages occupy the higher complexity end of the table with many different patterns allowed. This is consistent with having a high frequency of length-2 initial clusters – the results seen in section 4.2. Many of the patterns are shared although the order of preference (frequency) is different.

Differences in cluster pattern complexity within historical groupings can be noted too. Spanish is the least complex in the Romance group and Serbian – in the Slavic group. Dutch is less complex, i.e., has fewer patterns than German.

In terms of distinguishing patterns, Slavic languages are the only ones that employ the ‘sn’ initial pattern, and German and Polish are similar because of the high frequency (top in both languages) of ‘ts’, stop followed by a fricative, pattern, which is assumed not to be preferred in initial position. Also, these are the only languages (Serbian excluded) that employ patterns with semivowels. However, although such patterns are predicted to be most favorable in initial position, they occur at the lower frequency end in each language.
Since German, Dutch, and Russian, which are considered rhythmically similar, do not have highly similar cluster patterns, phonotactic patterning in length-2 initial clusters does not seem to be related to rhythmic property, although it reveals a different kind of similarity. Such similarity can be used to distinguish or group languages based on their unaltered samples.

b) word-final consonant cluster patterns

Table 4.19 presents clusters in word-final position for languages in which length-2 final clusters occur. There are only 12 such languages and in only 6 do length-2 clusters of a certain type occur more than 0.5% of the time. Clusters listed in parentheses have frequencies less than 0.5%.

Table 4.19  Clusters of length-2 in ‘saltanajc’ scale: word-final position

<table>
<thead>
<tr>
<th>language</th>
<th>Final CC clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>nt &gt; st &gt; lt &gt; ls &gt; tt</td>
</tr>
<tr>
<td>German</td>
<td>nt &gt; st &gt; ls &gt; ns</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>(lt &gt; st &gt; ls)</td>
</tr>
<tr>
<td>Czech</td>
<td>st &gt; (nc)</td>
</tr>
<tr>
<td>Polish</td>
<td>(sc &gt; tj &gt; st)</td>
</tr>
<tr>
<td>Russian</td>
<td>st &gt;(lt &gt; sn)</td>
</tr>
<tr>
<td>Serbian</td>
<td>(st &gt; nc &gt; tl)</td>
</tr>
<tr>
<td>Catalan</td>
<td>(nt &gt;lt &gt;st)</td>
</tr>
<tr>
<td>Estonian</td>
<td>st &gt; ts &gt; nt &gt; lt &gt; ll &gt; (tt &gt; ln)</td>
</tr>
<tr>
<td>Hungarian</td>
<td>tt &gt; lt &gt; nt &gt; st &gt; jt &gt; (ll &gt;nc &gt;tj)</td>
</tr>
<tr>
<td>Greek</td>
<td>(ts)</td>
</tr>
<tr>
<td>Turkish</td>
<td>(lt &gt; st &gt;ns &gt; nc)</td>
</tr>
</tbody>
</table>

Preferred clusters in final positions are listed at the bottom of the table; they are presented in lighter orange (gray) if they do not occur in the corpus. Preferred clusters are highlighted in orange for each language in which they occur. Clusters that occur but are not predicted (not considered preferred at final positions) are shown in blue (dark).

First, we can notice the absence of clusters with semivowels, apart from ‘jt’ in Hungarian, although such clusters are predicted to be most favorable in word-final position.
This can be explained partly by the absence of semivowels from the phoneme inventories (I included semivowels as a part of a diphthong in some languages). However, even in Russian, where the proportion of semivowels is high (10%), such clusters do not occur.

The sonority distance criterion for word-final length-2 clusters states that the first consonant $C_1$ should be closer in sonority to the vowel $V$ than to the second consonant $C_2$ in $VC_1C_2$ word end; it is possible that semivowels are easily attracted to form a diphthong with the vowel in the nucleus.

Another predicted cluster that does not occur is ‘lc’. This cannot be explained by the low frequency of affricates because the clusters ‘sc’ and ‘nc’ do occur, albeit with low frequencies.

Clusters that do occur include ‘lt’, ‘nt’, and ‘ls’. In addition, we find clusters that are considered preferred in positions other than final, in particular ‘st’ which occurs with high frequency in most languages.

Similarity seems to run within language family lines. Similar pairs include German and Dutch; Estonian and Hungarian; and Czech and Russian. Languages traditionally labeled syllable-timed group by virtue of not having length-2 final clusters.

A simplified table for word-final clusters, analogous to Table 4.18 for initial clusters is presented as Table 4.20. In this table, Dutch, German, Hungarian, and Estonian appear to form a grouping. It is interesting to notice that Dutch, German, Hungarian, and Estonian do have something in common phonologically - a salient words stress, albeit in different positions: word-initially for Estonian and Hungarian, and somewhat more variably in German and Dutch.

**Table 4.20 Clusters of length-2 in ‘saltanajc’ scale: word-final position (re-arranged)**

<table>
<thead>
<tr>
<th>Language</th>
<th>Final CC clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>nt &gt; st &gt; lt &gt; ls &gt; tt</td>
</tr>
<tr>
<td>German</td>
<td>nt &gt; st &gt; ls &gt; ns</td>
</tr>
<tr>
<td>Estonian</td>
<td>st &gt; ts &gt; nt &gt; lt &gt; ll</td>
</tr>
<tr>
<td>Hungarian</td>
<td>tt &gt; lt &gt; nt &gt; st &gt; jt</td>
</tr>
<tr>
<td>Czech</td>
<td>st</td>
</tr>
<tr>
<td>Russian</td>
<td>st</td>
</tr>
<tr>
<td><strong>PREFERRED</strong></td>
<td>jt &gt; jc &gt; js = lt &gt; jn = lc &gt; jl = ls = nt</td>
</tr>
</tbody>
</table>
What distinguishes German and Dutch from Estonian and Hungarian is that the most frequent final pattern in German and Dutch is the sonorant-obstruent cluster ‘nt’ that is predicted as preferred final cluster, while in Estonian and Hungarian the most frequent patterns are obstruent-obstruent ‘st’ and ‘tt’ respectively.

c) **word-medial consonant cluster patterns**

Word-medial consonant patterns of length-2 are presented in Table 4.21. Languages are ordered in decreasing order of number of patterns. The frequency of complex medial clusters (length 2 or greater) is given in the last column. Preferred clusters in medial position are given in the bottom row. Those that do not occur in the text are shown in lighter color. Note that there is an overlap between word-medial and word-final preferred clusters.

Table 4.21 Clusters of length-2 in ‘saltanajc’ scale: word-medial position

<table>
<thead>
<tr>
<th>Language</th>
<th>Word-medial clusters of length-2</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>lt &gt; nt &gt; st &gt; ls &gt; ns &gt; sl &gt; nn &gt; lj &gt; sj &gt; ss &gt; ts &gt; tt &gt; ll &gt; tl</td>
<td>44</td>
</tr>
<tr>
<td>Hungarian</td>
<td>tt &gt; lt &gt; nt &gt; st &gt; ss &gt; ln &gt; ts &gt; ls &gt; li &gt; nn &gt; js &gt; tj</td>
<td>42</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>st &gt; tn &gt; tl &gt; lt &gt; sn &gt; nj &gt; ts &gt; ls &gt; sl &gt; lj &gt; ln</td>
<td>32</td>
</tr>
<tr>
<td>Turkish</td>
<td>lt &gt; nt &gt; sl &gt; tl &gt; st &gt; tt &gt; ll &gt; nl &gt; tn &gt; ls &gt; ns</td>
<td>33</td>
</tr>
<tr>
<td>Russian</td>
<td>st &gt; tj &gt; sj &gt; tl &gt; nj &gt; lj &gt; tn &gt; tt &gt; sn &gt; lt</td>
<td>39</td>
</tr>
<tr>
<td>Italian</td>
<td>nt &gt; tt &gt; ll &gt; lt &gt; ss &gt; st &gt; cc &gt; tl &gt; ln &gt; nn &gt; ln</td>
<td>42</td>
</tr>
<tr>
<td>Estonian</td>
<td>st &gt; ll &gt; ts &gt; lt &gt; tt &gt; nt &gt; nn &gt; ln &gt; tl &gt; lj</td>
<td>30</td>
</tr>
<tr>
<td>Polish</td>
<td>st &gt; sn &gt; tl &gt; tn &gt; ts &gt; sc &gt; lt &gt; tt &gt; nt</td>
<td>28</td>
</tr>
<tr>
<td>German</td>
<td>nt &gt; st &gt; lt &gt; tt &gt; ts &gt; ns &gt; nn &gt; nc</td>
<td>34</td>
</tr>
<tr>
<td>Greek</td>
<td>st &gt; ss &gt; ls &gt; ts &gt; sl &gt; sn &gt; tl &gt; tt</td>
<td>21</td>
</tr>
<tr>
<td>Czech</td>
<td>st &gt; tl &gt; sn &gt; sl &gt; tn &gt; tt &gt; lt</td>
<td>30</td>
</tr>
<tr>
<td>Serbian</td>
<td>st &gt; tl &gt; sn &gt; tn &gt; sl</td>
<td>22</td>
</tr>
<tr>
<td>Catalan</td>
<td>nt &gt; st &gt; lt &gt; ns &gt; sl</td>
<td>27</td>
</tr>
<tr>
<td>Spanish</td>
<td>nt &gt; st &gt; lt &gt; tl &gt; ns</td>
<td>33</td>
</tr>
<tr>
<td>Indonesian</td>
<td>nt &gt; lt &gt; nc &gt; ls &gt; ln</td>
<td>22</td>
</tr>
<tr>
<td>Portuguese</td>
<td>st &gt; tl &gt; sc &gt; sn</td>
<td>17</td>
</tr>
<tr>
<td>Japanese</td>
<td>tt &gt; nt</td>
<td>8</td>
</tr>
</tbody>
</table>

**Preferred**

\[ tt > cc = tc > ss = cs = ct = ts > nn = sn = sc = cn > ll = nl = ns = st > jj = ln = nc > jj = ls = nt \ (the last three are also preferred in final position) \]

74
As before, preferred clusters that occur in the corpora are marked in orange. Clusters that are not predicted as preferred medially but nevertheless occur in the text are presented in black. All reported clusters occur more than 1% of the time.

The most frequent clusters that are predicted in positions other than medial are initial-preferred ‘tl’ and final-preferred ‘lt’, followed by initial-preferred ‘tn’ and ‘sl’. By comparing frequencies within each corpus and occurrence across languages, we notice that while ‘tl’ is a very good initial, it is also a reasonably good medial. Similarly, ‘lt’ is both a good final (as predicted) and a very good medial.

On the other hand, ‘tn’ which is predicted to be preferred in initial position is not a very good initial according to the frequency argument, while it seems to be a reasonably good medial. This suggests that some adjustments of the criteria for preferred position might be needed.

Among the predicted best medials, we find ‘tt’, ‘cc’, and ‘tc’. Indeed, ‘tt’ is one of the most frequent patterns, together with ‘ls’, ‘nt’ ‘st’, ‘ts’, and ‘nn’. However, ‘tc’ is not attested in the corpora at higher than 1% frequency, while ‘cc’ occurs only in Italian, and only as geminate. This could be related to the previously mentioned concern that affricates should be considered as complex segments and thus ‘tc’ or ‘cc’ should be considered as clusters of length 3 and 4. Alternatively, affricates, which constitute only a small percentage of segments have proportionally small joint occurrences in clusters. Finally, the high articulatory effort of producing an affricate next to another consonant could be considered as a reason for the low occurrence of clusters that include affricates. Clusters ‘cs’, ‘ct’, and ‘cn’ are also not attested in the corpora.

Finally, although there is a slight tendency of Romance (and therefore ‘syllable-timed’) languages to have a smaller number of phonotactic patterns among length-2 medials, while the traditionally called mora-timed languages have none or very few length-2 patterns, there is no clear way to associate rhythm type with number of phonotactic patterns for the remaining languages. In particular, German and Dutch are quite far apart on the scale of number of patterns, with Italian, Bulgarian and several other languages in between.

There is also no striking similarity in terms of number of possible patterns within language families other than Polynesian and Romance.
In conclusion, looking at phonotactic patterns reveals several properties of cluster and word complexity across corpora. In addition, the obstruent-sonorant-vowel level of representation captures rhythmic similarity better than the more detailed saltanajc level; saltanajc representation reveals segmental similarities that are correlated with similarities within the language families.

If vowels and sonorant consonants are grouped together and then the patterns of final clusters compared across corpora, we do find similarities with rhythm type grouping.

Also, the frequency of length-1 and length-0 clusters confirms that the best initial is a single obstruent and the best final is length-0 cluster, which corresponds to an open syllable. These findings support the fact that CV is the most frequent syllable type cross-linguistically, and in particular, argue for TA (obstruent-vowel) as the best type.

At the more detailed level of sonority representation, we find similarities that reflect language family grouping. Two criteria for similarity can be used: distinguishing patterns, and number of different patterns. This holds for cluster patterns in word-initial and word-final positions while the patterning is somewhat more complex for the word medial clusters.

Goodness of clusters at the coarser obstruent-sonorant-vowel level is verified, with better clusters according to sonority being more frequent in individual corpora and across corpora.

At the more detailed, saltanajc, level, predicted preferred clusters in each position of the word are loosely supported by the data. Many of the predicted clusters occur at the specified position, however, clusters occur even in non-preferred positions and sometimes with very high frequencies. This suggests that, while criteria based on sonority distance gives a good start for a measure of goodness of clusters at certain positions, some modifications are needed in order to account for frequencies in language corpora.

4.4 Word length distributions

In this section we study differences across languages in word length distributions across languages. Then we examine whether there is a relation between rhythm type and word length, and between consonant clustering type and word length.
There are two reasons to study word length. One relates to segmentation: if a language consists only of words of a given length, segmentation into words would be fairly easy, it would require only finding that one length. Variation in word length makes the uniform segmentation impossible. However, speech rate cues might be helpful, as explained later in this section.

Another reason to study word lengths relates to a possible interesting relation with language rhythm. It has been proposed that languages which regulate stress intervals tend to keep word duration consistent by shortening the durations of syllables in longer words, while languages that regulate syllable intervals tend to let the word length increase proportionally to the number of syllables. However, it has been found that speakers of all languages tend to accommodate long words by uttering them faster. Thus, it is possible that variation in word length might contribute to the overall duration variability in speech when words of different lengths (expressed in number of syllables) alternate in a sentence.

Word length, however, is not the only factor expected to influence duration, whether expressed as the number of syllables or the number of phonemes. Any inherent language tendency to downplay (reduce) all but one syllable in a word may also play a role. And so will the internal organization of phonemes into syllables.

Thus, as proposed in Chapter 1, we ask the following two questions: 1) Do languages that are considered similar rhythmically have similar word length distributions? 2) Does word-length distribution correlate with cluster length distribution?

In fluent speech we are exposed to words in succession, thus, like in previous chapters we will look at the set of all token words when determining the length distribution. To make comparison, we will also look at the distributions of word lengths among lexical items (no word counted more than once).

Because the duration of a word depends both on the number of phonemes comprising the word and the number of syllables into which it can putatively be divided, I present distributions for word-types as well as for word-tokens in terms of both independent variables: number of phonemes and number of syllables.
Overall, we have 4 different variables for each language sample: Frequency of words of $N$ syllables in running text, Frequency of lexical items of $N$ syllables, Frequency of words of $N$ phonemes in running text, and Frequency of lexical items of $N$ phonemes. These have been calculated on the corpus of 21 languages. We are most interested in the frequency in the running text because it corresponds to what we hear in speech and what we use to decide which languages sound more similar.

I present results for the 4 variables for each of the 21 languages in Figures A4.1-A4.21 in Appendix 4. Here, I show the summary figures that group languages with similar behavior. Distribution of word-lengths across lexical items is presented in 4 graphs in Figure 4.16, and the distribution of all words in the text, again in 4 graphs, corresponding to 4 different groups, in Figure 4.17.

![Figure 4.16 Distribution of word-lengths: lexical items](figure)

The first group includes German, Dutch, Catalan, and Maori. Languages in this group have high frequency of monosyllabic lexical items while the most frequent words are disyllabic. Group 4 (bottom right), which includes Estonian, Czech, and Hawaiian, also has disyllables as most frequent, but the monosyllables are not as frequent as in the first group.
In the other two groups (top right and bottom left), most frequent lexical items are trisyllabic. In the third group (bottom left), the ratio of disyllabic to trisyllabic words is closer to one than in the second group (upper right).

While the most frequent word-lengths among different lexical items are 2 or 3 syllables long, in many languages monosyllabic words are so frequent that a small number of them dominates the distribution of word tokens (represented in Figure 4.17). Monosyllables are dominant in running speech (text) for 15 out of 21 languages in our sample. For Italian, Polish, and Russian texts, monosyllables are equally frequent as disyllables. Only in three languages, Estonian, Turkish, and Indonesian (represented in bottom left graph), are the monosyllables less frequent than disyllables.

![Figure 4.17 Distribution of word-lengths: word tokens](image)

Dutch, German, Maori, and Catalan have high frequency of monosyllabic words among lexical items and among word-tokens (upper left graph on both Figure 4.16 and Figure 4.17. This suggests possible correlation between having short words in the lexicon and being stress-timed.\(^{23}\)

\(^{23}\) Large percentage of monosyllables in Maori is possibly a result of over-counting diphthongs.
To assess the differences in word-length distributions, we can also compare the average word length across languages. Results for word tokens are presented in Figures 4.18 and for lexical items in Figure 4.19.

Figure 4.18 Average word-length: word tokens

Figure 4.19 Average word-length: lexical items
We notice that in Figure 4.20, where languages are ordered by the average length of word tokens (that is, the average of all words in running text), a potential grouping somewhat resembles traditional rhythm types. Namely, the rightmost five (Japanese and the Polynesian languages, Hawaiian, Samoan, and Tongan), have the longest words; while the leftmost three (Catalan, Dutch, and German) have the shortest words.

However, this should not be interpreted without taking into account other factors. On one hand, long words can be explained if there is a small number of phonemes in a language (phoneme inventory) along with simple syllable structure. A small phoneme inventory can explain the existence of long words in Polynesian languages, but not in Japanese. Relatively simple syllable structure can.

Dutch, German, and Catalan all have relatively large phoneme inventories and moderate to complex syllable structures (see Table A1.1 in Appendix 1). In particular, they are distinguished by their large inventories of different vowel qualities (which allows for a large number of nuclei). A large number of consonant qualities in conjunction with complex syllable structure allows for a large number of possible onsets and codas. Thus, meanings can be expressed by words with smaller numbers of syllables.

In addition, agglutinative languages like Turkish, Hungarian, and Estonian, have long words due to their morphology.

Word lengths then seem to relate to very basic phonological and morphological properties. If word length also groups languages by their rhythm, as Figure 4.19 suggests, then morphology and basic phonological structure together can be said to influence prosodic properties. Some relation between rhythm type and morphological properties and word order are discussed for Munda and Mon-Khmer languages in Donegan and Stampe (2004) and relation between word order and syllable complexity in Tokizaki and Yasumoto (2012).

4.5 Variability of measures over different materials

One of the frequently addressed issues about rhythm metrics is their variation over different speech materials (see Arvaniti 2012, for instance) among many other factors. Authors have used different kinds of texts, trying to maximally differentiate syllable complexity over different test items (texts) and then report values of rhythm metrics. Arvaniti finds that
variation over different texts can be larger than variation over different languages, thus
rendering the value of RMs for language classification minimal.

It is then important to examine whether phonotactic metrics, as well as the other
measures used for classification in this dissertation vary over types of materials as well.
Measures we test include phonotactic metrics, word complexity expressed as distributions of
word-initial, word-final, and word-medial clusters, and distribution of word lengths.

I chose four representative languages: German, Italian, Hawaiian, and Serbian.
German is traditionally considered stress-timed and Italian – syllable-timed; Hawaiian is
assumed here to be mora-timed due to it simple structure, and vowel length distinction;
Serbian is traditionally assumed to be stress-timed, but was on many measures examined in
this dissertation was closer to traditional syllable-timed languages.

In the following, I present tables in which the variables are calculated over 3 different
texts: one is the 10,000 word text for which results are reported in Chapter 4. In addition, I
used one long and one short text for each language. I present results for each of the three
texts, followed by the average over them.

The first table, Table 4.22 reports values of phonotactic metrics.

<table>
<thead>
<tr>
<th>language</th>
<th>Nwords</th>
<th>%V</th>
<th>VarcoV</th>
<th>nPVI-V</th>
<th>stdevC</th>
<th>rPVI-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>10000</td>
<td>38.9</td>
<td>49.8</td>
<td>105.8</td>
<td>147.9</td>
<td>74.1</td>
</tr>
<tr>
<td>German</td>
<td>16043</td>
<td>39.2</td>
<td>48.7</td>
<td>105.6</td>
<td>145.7</td>
<td>73.5</td>
</tr>
<tr>
<td>German</td>
<td>108</td>
<td>37.7</td>
<td>29.4</td>
<td>101.5</td>
<td>153.4</td>
<td>72.2</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>38.6</td>
<td>42.7</td>
<td>104.3</td>
<td>149.0</td>
<td>73.2</td>
</tr>
<tr>
<td>Serbian</td>
<td>10000</td>
<td>46.3</td>
<td>52.1</td>
<td>106.9</td>
<td>97.3</td>
<td>44.6</td>
</tr>
<tr>
<td>Serbian</td>
<td>29456</td>
<td>46.3</td>
<td>49.2</td>
<td>106.0</td>
<td>93.8</td>
<td>42.8</td>
</tr>
<tr>
<td>Serbian</td>
<td>92</td>
<td>46.3</td>
<td>50.1</td>
<td>106.5</td>
<td>89.0</td>
<td>40.7</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>46.3</td>
<td>50.4</td>
<td>106.4</td>
<td>93.3</td>
<td>42.7</td>
</tr>
<tr>
<td>Italian</td>
<td>10000</td>
<td>46.3</td>
<td>55.7</td>
<td>108.6</td>
<td>105.1</td>
<td>50.2</td>
</tr>
<tr>
<td>Italian</td>
<td>55104</td>
<td>46.2</td>
<td>55.6</td>
<td>108.6</td>
<td>105.4</td>
<td>50.6</td>
</tr>
<tr>
<td>Italian</td>
<td>1633</td>
<td>46.3</td>
<td>58.1</td>
<td>109.4</td>
<td>106.8</td>
<td>53.5</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>46.3</td>
<td>56.5</td>
<td>108.8</td>
<td>105.8</td>
<td>51.4</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>18476</td>
<td>56.6</td>
<td>77.4</td>
<td>116.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>3052</td>
<td>56.1</td>
<td>72.2</td>
<td>116.4</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>104</td>
<td>56.6</td>
<td>72.0</td>
<td>117.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>56.4</td>
<td>73.9</td>
<td>116.8</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

*Results are presented for four test languages: German, Hawaiian, Italian, and Serbian.
We see from this table that the values of each metric in each text for a given language are very close to the average value over three texts, except for the Varco-V value of one of the German texts. One the one hand, this is a text of very short length, only 108 words. It is possible that this is an outlier; it is also possible that Varco-V needs sufficiently long text in order to be stable.

It is interesting, however, that we do not see such discrepancy between other metrics or for the Varco-V in other languages. This is particularly interesting because rhythm metrics were reported as varying a lot over different materials (usually short, comparable to the shortest texts reported in Table 4.22). If phonotactic metrics are stable (or similar), and rhythmic metrics are variable over different materials of short length (say, about 100 words), then two scenarios are possible: 1) syllable (or word) complexity in the materials used to show variation does not represent the usual word complexity for that language, that is, it artificially stretches from very simple to very complex; or 2) it is something other than syllable/word complexity that causes variation. The second scenario would suggest that durational variation contributed by prosody varies over these different materials and that longer materials need to be used in such studies – not only because of varying syllable complexity (or not at all because of that), but also because of varying prosodic structure in a single language. This second scenario agrees with the view that rhythmicity of a language cannot be captured in only a few sentences, even though the few sentences may be characteristic of a particular language (or a group of languages).

Next, Tables 4.23-25 report results for the variability over different texts of consonant cluster complexity in word-initial, word-medial, and word-final positions respectively.
Table 4.23 Variability of consonant cluster complexity in word-initial position

<table>
<thead>
<tr>
<th>language</th>
<th>Nwords</th>
<th># V</th>
<th>#CV</th>
<th>#CCV</th>
<th>#CCCV</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>10000</td>
<td>26.9</td>
<td>64.8</td>
<td>7.3</td>
<td>1.0</td>
</tr>
<tr>
<td>German</td>
<td>16043</td>
<td>26.1</td>
<td>65.5</td>
<td>8.2</td>
<td>0.2</td>
</tr>
<tr>
<td>German</td>
<td>108</td>
<td>20.4</td>
<td>71.3</td>
<td>6.5</td>
<td>1.9</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>24.5</td>
<td>67.2</td>
<td>7.3</td>
<td>1.0</td>
</tr>
<tr>
<td>Serbian</td>
<td>10000</td>
<td>22.2</td>
<td>63.1</td>
<td>13.9</td>
<td>0.8</td>
</tr>
<tr>
<td>Serbian</td>
<td>29456</td>
<td>19.6</td>
<td>65.1</td>
<td>14.7</td>
<td>0.6</td>
</tr>
<tr>
<td>Serbian</td>
<td>92</td>
<td>16.7</td>
<td>64.8</td>
<td>18.5</td>
<td>0.0</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>19.5</td>
<td>64.3</td>
<td>15.7</td>
<td>0.5</td>
</tr>
<tr>
<td>Italian</td>
<td>10000</td>
<td>24.6</td>
<td>68.5</td>
<td>6.3</td>
<td>0.7</td>
</tr>
<tr>
<td>Italian</td>
<td>55104</td>
<td>25.4</td>
<td>69.3</td>
<td>5.0</td>
<td>0.4</td>
</tr>
<tr>
<td>Italian</td>
<td>1633</td>
<td>24.4</td>
<td>69.8</td>
<td>5.4</td>
<td>0.4</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>24.8</td>
<td>69.2</td>
<td>5.6</td>
<td>0.5</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>18476</td>
<td>27.4</td>
<td>72.6</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>3052</td>
<td>27.2</td>
<td>72.8</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>104</td>
<td>28.9</td>
<td>71.2</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>27.8</td>
<td>72.2</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

*Results are presented for four test languages: German, Hawaiian, Italian, and Serbian

All the cluster distributions values look stable, minimally varying from the average, irrespective of the length of the sample. This suggest that phonotactics, or more precisely, the durational phonotactic component defined in Chapter 2, is stable and thus can be easily learned by exposure to a language.

Table 4.24 Variability of consonant cluster complexity in word-medial position

<table>
<thead>
<tr>
<th>language</th>
<th>Nwords</th>
<th>V V</th>
<th>VCV</th>
<th>VCCV</th>
<th>VCCCV</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>10000</td>
<td>10.5</td>
<td>55.7</td>
<td>28.9</td>
<td>4.7</td>
</tr>
<tr>
<td>German</td>
<td>16043</td>
<td>8.5</td>
<td>57.1</td>
<td>28.5</td>
<td>5.2</td>
</tr>
<tr>
<td>German</td>
<td>108</td>
<td>2.0</td>
<td>62.0</td>
<td>30.0</td>
<td>6.0</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>7.0</td>
<td>58.3</td>
<td>29.1</td>
<td>5.3</td>
</tr>
<tr>
<td>Serbian</td>
<td>10000</td>
<td>3.9</td>
<td>74.1</td>
<td>20.1</td>
<td>1.8</td>
</tr>
<tr>
<td>Serbian</td>
<td>29456</td>
<td>4.3</td>
<td>75.8</td>
<td>18.8</td>
<td>1.0</td>
</tr>
<tr>
<td>Serbian</td>
<td>92</td>
<td>13.0</td>
<td>69.6</td>
<td>17.4</td>
<td>0.0</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>7.1</td>
<td>73.2</td>
<td>18.7</td>
<td>0.9</td>
</tr>
<tr>
<td>Italian</td>
<td>10000</td>
<td>6.5</td>
<td>52.0</td>
<td>39.8</td>
<td>1.7</td>
</tr>
<tr>
<td>Italian</td>
<td>55104</td>
<td>6.1</td>
<td>53.0</td>
<td>38.6</td>
<td>2.3</td>
</tr>
<tr>
<td>Italian</td>
<td>1633</td>
<td>7.4</td>
<td>52.7</td>
<td>36.4</td>
<td>3.5</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>6.7</td>
<td>52.6</td>
<td>38.3</td>
<td>2.5</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>18476</td>
<td>18.4</td>
<td>81.6</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>3052</td>
<td>16.1</td>
<td>83.9</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>104</td>
<td>17.8</td>
<td>82.2</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>17.5</td>
<td>82.6</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Table 4.25 Variability of consonant cluster complexity in word-final position*

<table>
<thead>
<tr>
<th>language</th>
<th>Nwords</th>
<th>V #</th>
<th>VC#</th>
<th>VCC#</th>
<th>VCCC#</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>10000</td>
<td>34.6</td>
<td>47.5</td>
<td>15.1</td>
<td>2.7</td>
</tr>
<tr>
<td>German</td>
<td>16043</td>
<td>35.6</td>
<td>46.1</td>
<td>15.5</td>
<td>2.7</td>
</tr>
<tr>
<td>German</td>
<td>108</td>
<td>38.0</td>
<td>41.7</td>
<td>15.7</td>
<td>4.6</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>36.0</td>
<td>45.1</td>
<td>15.4</td>
<td>3.3</td>
</tr>
<tr>
<td>Serbian</td>
<td>10000</td>
<td>81.7</td>
<td>17.9</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>Serbian</td>
<td>29456</td>
<td>81.4</td>
<td>18.3</td>
<td>0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>Serbian</td>
<td>92</td>
<td>77.8</td>
<td>22.2</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>80.3</td>
<td>19.5</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Italian</td>
<td>10000</td>
<td>88.2</td>
<td>11.8</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Italian</td>
<td>55104</td>
<td>87.1</td>
<td>12.8</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Italian</td>
<td>1633</td>
<td>87.6</td>
<td>12.4</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>87.6</td>
<td>12.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>18476</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>3052</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>104</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

*Results are presented for four test languages: German, Hawaiian, Italian, and Serbian

Finally, Tables 4.26 and 4.27 report on variability of word-length distributions, over different texts, where word-length is defined as the number of syllables in the word.

Table 4.26 Variability of word-length distribution: word tokens*

<table>
<thead>
<tr>
<th>language</th>
<th>Nwords</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>10000</td>
<td>66.2</td>
<td>27.8</td>
<td>4.6</td>
<td>1.3</td>
<td>0.1</td>
</tr>
<tr>
<td>German</td>
<td>16043</td>
<td>56.6</td>
<td>31.7</td>
<td>8.6</td>
<td>2.7</td>
<td>0.3</td>
</tr>
<tr>
<td>German</td>
<td>108</td>
<td>65.7</td>
<td>23.2</td>
<td>10.2</td>
<td>0.9</td>
<td>0.0</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>62.9</td>
<td>27.5</td>
<td>7.8</td>
<td>1.6</td>
<td>0.1</td>
</tr>
<tr>
<td>Serbian</td>
<td>10000</td>
<td>36.4</td>
<td>32.2</td>
<td>20.6</td>
<td>9.0</td>
<td>1.7</td>
</tr>
<tr>
<td>Serbian</td>
<td>29456</td>
<td>38.0</td>
<td>34.9</td>
<td>19.5</td>
<td>6.6</td>
<td>1.0</td>
</tr>
<tr>
<td>Serbian</td>
<td>92</td>
<td>40.2</td>
<td>31.5</td>
<td>20.7</td>
<td>5.4</td>
<td>2.2</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>38.2</td>
<td>32.9</td>
<td>20.2</td>
<td>7.0</td>
<td>1.6</td>
</tr>
<tr>
<td>Italian</td>
<td>10000</td>
<td>36.4</td>
<td>36.0</td>
<td>19.4</td>
<td>6.1</td>
<td>1.9</td>
</tr>
<tr>
<td>Italian</td>
<td>55104</td>
<td>34.1</td>
<td>32.2</td>
<td>22.5</td>
<td>8.6</td>
<td>2.2</td>
</tr>
<tr>
<td>Italian</td>
<td>1633</td>
<td>35.4</td>
<td>35.4</td>
<td>19.2</td>
<td>7.2</td>
<td>2.6</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>35.3</td>
<td>34.5</td>
<td>20.4</td>
<td>7.3</td>
<td>2.3</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>18476</td>
<td>48.3</td>
<td>34.2</td>
<td>10.4</td>
<td>4.8</td>
<td>1.4</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>3052</td>
<td>46.5</td>
<td>28.1</td>
<td>11.4</td>
<td>4.5</td>
<td>3.6</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>104</td>
<td>51.9</td>
<td>26.0</td>
<td>9.6</td>
<td>2.9</td>
<td>4.8</td>
</tr>
<tr>
<td>average</td>
<td></td>
<td>48.9</td>
<td>29.4</td>
<td>10.5</td>
<td>4.0</td>
<td>3.3</td>
</tr>
</tbody>
</table>

*Results are presented for four test languages: German, Hawaiian, Italian, and Serbian
Here, we see that the distributions of word-lengths (or, frequencies of words of a specific length) are quite stable when all words in the text are considered, but somewhat less stable when the distribution is calculated over lexical items.

Word-length as a function of the number of syllables over lexical items is presented in Figure 4.20. Four panels present data for German, Serbian, Italian and Hawaiian. In each panel, three distributions are presented; each corresponds to one text from Table 4.26. The line that represents distribution over the shortest text for each language is blue.

Note that there are fewer distinct lexical items than total words, so in a text of 100 words, the number of lexical items indeed is too small to capture frequencies. Thus, the blue line deviates from the two red lines that correspond to long texts. However, even for a very short text, around 100 words, they still follow the same shape.
Finally, the distributions of word tokes over the same texts for the same four languages are given in four panels of Figure 4.21. Even the short text now has the distribution that closely follows the distributions over longer texts.
Figure 4.21. Stability: Distribution of word-lengths (word tokens)

This concludes our brief examination of stability of the measures used. A more thorough analysis can be performed, which I believe will continue to support stability of measures in all languages, given that a certain minimum number of words (or phrases) is taken into account.
CHAPTER 5

GENERAL DISCUSSION AND CONCLUSION

In this chapter, I first give an overview of the results chapter by chapter, with additional discussion of related issues. Next, I summarize theoretical and technical issues that accompanied this work and how improvements can be made in the future. Following that, I reflect on the inter-relations between phonotactics and rhythm and how these two aspects of speech have been interpreted in the literature. Finally, I offer some additional questions that can enrich the present scope of this work.

5.1 Summary

In this dissertation, I examined the relation between the structural properties of words and language (speech) rhythm. Structural properties that I considered include complexity of phonotactic sequences and word lengths, both evaluated based on the word frequency in the test materials. Speech rhythm of each language in the sample is assigned based on a) traditional rhythm type (Pike 1945, Abercrombie, 1967), or b) a set of phonological properties (Gill 1986, Dauer 1987, Auer 1993, Pamies-Bertrán 1999). In addition, I examined the relation between the measures that assess phonotactic complexity, phonotactic metrics (PMs), and measures that have been used to assess rhythm type, rhythm metrics (RMs).

To facilitate this study, I created (close to) phonemic corpora of 21 different languages by designing an automatic transcriber that operates on written materials. I have implemented this transcriber in the Python programming language (http://www.python.org/). In addition, I created a word complexity calculator – a set of tools to calculate the distributions of different classes of segments (CV, ALT, ‘saltanajc’), as well as the related statistical and rhythm related measures).

I then performed a detailed cross-linguistic comparison of word structure on a set of 21 languages, by examining both the distributions of consonant clusters – their lengths and patterns – at different positions in a word, and the distributions of word-lengths. I found that

---

24 I intend to make these tools available for use in the future, as a post-dissertation project.
the similarity among languages differs depending on the criteria/measures used. Some of the measures classify languages in agreement with the rhythm class hypothesis; other measures correlate better with the language family or certain phonological properties.

To make stronger claims on the relation between these measures and rhythmic similarity, we need more comprehensive data on rhythmic similarity across languages: both in terms of the number of languages for which such similarity is evaluated and in terms of validity of the rhythmic similarity data, preferably obtained via an independent criterion.

I have also shown that several measures that have been used to assess language rhythm type are dominated by the effect of segmental phonotactic structure of words, seen through high correlation between the phonotactic and the rhythm metrics. Some of the rhythm metrics that are not correlated with the corresponding phonotactic metrics fit less well to the classification based on traditional types. In addition, the analysis of the interdependence of various rhythm metrics supports the posited linear relationship between the proportion of vocalic intervals (%V) and the standard deviation of consonantal intervals (∆C) proposed by Cassandro et al. (2003).

In Chapter 1, I gave a brief overview of the history of research on speech rhythm and introduced the issues in quantifying rhythmic similarity related to phonotactics. In relation to those issues, I presented my research questions, outlined the methods used, and listed the contributions that my work makes to the field of linguistics. A review of the literature on rhythm and phonotactics, and the definitions of concepts I use in the dissertation are given in Chapter 2; a detailed methodological description of the creation of phonemicized language corpora is presented in Chapter 3.

I organized my study around five questions. These questions and the corresponding summaries of the results are repeated here in order to prepare the reader for the general discussion.
5.2 Overview

First, I examined how the structural complexity of words affects rhythm metrics (RMs) - measures that are claimed to evaluate rhythmic similarity between languages. To do that, I defined phonotactic metrics based on lengths of vocalic and consonantal clusters in the sample. I established that phonotactic %V and ∆C are highly correlated with the rhythmic equivalents and conclude that rhythm measures largely reflect phonotactic, or word structural similarities.

I have also shown in the process that the linear relationship between %V and ∆C proposed by Cassandro et al. (2003) holds for the phonotactic metrics on the set of 21 languages, as witnessed in Figure 4.12 where we see a clear decreasing linear trend. This relationship can also be seen as a positive linear correlation between %C and ∆C (since %C = 100-%V) which is then easier to explain. Namely, %C is higher in languages with more complex consonant clusters; in those languages, the expected variation in the intervals, measured by ∆C, will also be higher. Similarly, where %C is small, syllable structure is simple and there is almost no variation reflected in ∆C.

Furthermore, I established that phonotactic pair-wise measures classify languages better than the rhythmic measures. I argued that this is possibly due to a negative correlation of phonotactics with vocalic variability, causing the phonotactic and rhythmic components to pull the metrics in different directions. While rhythmic grouping of languages using phonotactic metrics is more precise, it is based on a different property – phonotactic complexity of words – rather than on rhythm.

In section 4.2, I examined how similarity in the structural complexity of consonant clusters correlates with perception of two languages as rhythmically similar. I compared consonant cluster length distributions in three positions in the word: initial, medial, and final. The results of this corpus analysis appear in Tables 4.2.1-3. They reveal significant similarities across languages but that similarities based on just one position do not necessarily reflect similarities based on the other positions of the cluster. None of the groupings based on cluster similarity seems to reflect posited rhythm classes exactly, although there is a large overlap between, rhythm-type grouping and grouping observed
based on word-final clusters. More of independent evidence of rhythmic similarity is needed for comparison.

Questions 3 and 4 examined the similarity of phonotactic patterns and relationship between the sonority-based goodness of a cluster and its frequency in the sample. It was shown that at the coarsest sonority level, ALT, the ‘best clusters’ are usually the most frequent, or one of the two most frequent. Frequency results from analysis at the more detailed ‘saltanajc’ level did not agree closely with the preferred clusters proposed by Dziubalska-Kołaczyk, although the more frequent clusters were similar across languages. This suggested that there are factors other than sonority that regulate markedness of clusters.

Finally, Question 5 examined similarity across languages based on word-lengths. Distributions of word lengths were examined for token and type. It was established that the traditional rhythm grouping has some similarities with the distributions of monosyllabic words, as well as with the frequency of longer words.

Finally, let me review the method used to obtain phonemic corpora for 21 languages. As described in Chapter 3, I constructed a two-stage processing of the stories in electronic form. Two modules are responsible for processing: Transcriber and Complexity Calculator.

Transcriber for each language first employed some text processing techniques to eliminate punctuation and separate materials into phrases. Next, a set of phonological rules was applied in order to obtain IPA or close to IPA transcriptions. Following that, the material in IPA form was processed by the Complexity Calculator to obtain several levels of representation, CV, ALT, and ‘saltanjac’. Finally, various measures were computed, including Phonotactic Metrics, statistics of word-initial, word-medial, and word-final clusters, and word-length statistics.

This approach of ‘creating data’ from written texts has not been extensively employed in phonological studies, however it appears to receive some interest in the past few years (Garcia and González, 2012). This dissertation supports efforts in promoting this method.

5.3 Limitations of the study

One limitation of this study is the lack of direct comparison of phonotactic and rhythmic metrics, that is, their comparison on exactly the same phonetic material. Direct comparison
was not possible on longer materials, because lengthy transcribed and segmented spoken samples are not available. The approach taken in this work – transcribing written materials – works well for the examination of phonemic and phonotactic structure, which was the main goal of the dissertation. However, it would be interesting to evaluate rhythm metrics calculated on the same samples, however long, especially when establishing correlations between PMs and RMs. It is possible that additional correlations between corresponding rhythm and phonotactic metrics would be found.

Furthermore, while the achieved level of transcription was deemed sufficient for calculation of the broad phonotactic levels (‘saltanajc’ and ALT), it did not allow for comparison of phonotactic sequences/clusters at the phonemic level in all languages. This will be attempted in the continuation of the project.

Next, the rhythm metrics I evaluate in terms of how they reflect phonotactics are all interval based. In the literature, certain authors (for instance, Nolan and Asu 2009) have suggested that a syllable or a foot might be a better unit for rhythm quantification. While this is out of the scope of the present work, it would be an easy extension to calculate phonotactic metrics on syllables for languages for which syllabification is trivial, although it would be somewhat problematic for languages where it is not. One of my concerns in choosing the interval as a base unit, as well as calculation of cluster distributions as intervals rather than as onsets and codas, was exactly the problem with syllabification. The syllabification problem led me to study interval, or word-structure. However, given that onsets and codas also span zero, one, or larger length clusters, and given that their distribution is different across languages, it is likely that syllable-based phonotactic metrics would group languages based on their similarity in syllabic structure, in the same way that interval-based phonotactic metrics group languages based on similarity in overall cluster structure.

Also, in this dissertation, I focused on similarity based on the structure and patterns of consonant clusters. I did not report on structure of vocalic intervals leaving this for further analysis. The only place in which vocalic interval structure was addressed was in the treatment of long vowels with respect to Phonotactic Metrics computation. A single vocalic phoneme was considered to have phonotactic duration equal to one or equal to two. In these
comparisons, languages in which phonemic vowel length is not marked in writing had calculated values of percentage and the variability of vocalic intervals lower than actual.

Finally, all vowel sequences that possibly occur as a diphthong were counted as diphthongs in all cases. Some instances of vowel hiatus were thus misrepresented, which led to the lower percentage of vocalic phonotactic durations and smaller number of syllable.

In the future, both modules can be improved on. I briefly describe how.

*Grapheme-to-phoneme-transcriber*

As mentioned previously, in some languages there are discrepancies between the materials produced by my Transcriber and the IPA transcriptions of natural spoken speech. One of my future tasks will be to narrow these discrepancies by increasing the number of phonological processes applied during automatic transcription.

I would like to implement my transcriber as a web application so that it can be shared. At present, I have encountered several such applications for a subset of the languages I considered. Some of these applications were more successful than others in transcribing test passages I tried. My main goals are 1) to be able to transcribe a large amount of data from large text files, instead of short passages that need to be pasted into a web-page form and 2) to have multiple languages transcribed using the same general principles in order to gain consistency of level of transcription. Improved transcription accuracy will be the first step towards more accurate calculation of the complexities of clusters and words.

*Complexity calculator*

In this dissertation, I have presented results for analyses at the CV, ALT and ‘saltanajc’ levels. At these levels, we can successfully analyze the structure of consonant clusters, and produce typological generalizations or cross-linguistic comparisons at each of the three levels.

In addition, I implemented but did not report on several other variables that can be useful in structural comparison across languages, such as the average number of syllables and phonemes per word, as well as an analysis of vocalic intervals and nuclei as a distribution of
short monophthongs, short diphthongs, long monophthongs, long diphthongs, and syllabic consonants. This set of functions will be added to the complexity calculator.

5.4 Discussion

Let us now look at the dissertation questions within the larger area of research on speech rhythm. I will discuss related issues in perception, some methodological deficiencies, and propose a model based on what is known in the literature and what we learned in this dissertation.

5.4.1 Additional questions

I will first discuss several issues related to the perception of rhythm that were not directly addressed by the research questions but are important for the understanding of the overall phenomenon and emerge as possible extensions of the current research.

Segmental and phonotactic contributions to rhythm perception

In section 4.1, I established that some Rhythm Metrics are significantly affected by phonotactics and that some Phonotactic Metrics classify languages according to their traditional rhythm type. These results agree with the proposals, such as Dauer’s (1987), that speech rhythm arises from phonological – and in particular some phonotactic – properties of a language. This suggests that the further understanding of speech rhythm must be informed by how phonotactic differences are perceived.

A detailed perception study of which part of the speech signal contributes to perceived rhythmic similarity is needed. Several questions are informative: 1) Are speech sequences with complex word-initial clusters perceived as different from otherwise identical speech sequences with simple word-initial clusters? 2) Are speech sequences with complex word-final clusters perceived as different from otherwise identical speech sequences with simple word-final clusters? 3) Is the answer to question 1) the same when we consider complexity of syllable onsets instead of word-initial consonant clusters? 4) Is the answer to question 2) the same when we consider complexity of syllable codas instead of word-final consonant clusters?
Answering these questions will allow us to make predictions based on phonotactic complexity of a language towards its similarity to other languages. It will also point to the types of measures that can be devised to assess the rhythmic similarity between a pair of languages.

The second important set of questions addresses how different phoneme qualities affect rhythm perception. In sections 4.2–4.3, we have seen that the groupings based on similarities of word-initial (or word-final) clusters are different when we look at the distributions that assume dichotomy between vowels and consonants and when we look at the distributions that assume dichotomy between obstruents on the one hand, and vowels and sonorants on the other hand.

Studies in the literature on Rhythm Metrics are mostly based on the dichotomy of vocalic and consonantal intervals, although some authors examined dichotomy between voiced and voiceless intervals (Dellwo et al. 2007) or sonorant and non-sonorant intervals (Galves et al. 2002). However, these computational studies did not examine how different classes of segments contribute to rhythm perception. It was only suggested that such natural classes (voiced, voiceless, sonorant, and obstruent) are easy for listeners to perceive. Sonorant-obstruent dichotomy, to some degree, may correspond to the perception of filtered speech, which is used in many perception studies as a type of stimuli.

Loukina et al. report in their 2011 paper that their machine learning algorithms often classified sonorant consonants as vowels. This can be explained by the similarity of sonorant consonants and vowels in terms of energy or intensity. However, machine recognition often operates in ways that are different from human perception. The question, then, is: to what extent do human listeners base their perception of rhythmic sequences on energy? Additional uncertainty in interpreting their results arises from the fact that sonorant consonants were classified as vowels some times, but not always.

In sum, it is important to know how each consonant and vowel quality contributes to the creation of perceived rhythmic sequence.
Questions that addressed roles of different segmental qualities in forming a rhythmic sequence rely only on durational properties to define rhythm. Another interesting and very important aspect to understand more clearly is the relationship between different prosodic dimensions and rhythm. As mentioned in the introduction, only a handful of studies have examined the effect of pitch (Cumming 2010, Arvaniti and Rodríguez 2013), spectral balance (Sluijter et al. 1997), and loudness (Kochanski et al. 2005) in relation to the perception of prominence and perception of rhythm.

There are two types of questions to answer. The first is how pitch and loudness (or energy) affect duration. Some studies in the literature address such questions. Low et al. (2000) considered pair-wise variability index for loudness and showed that it was different for British and Singaporean English, similarly to the durational indices. Lehiste (1976) and Cumming (2010) discussed the relationship between fundamental frequency and duration. Kochanski et al. (2005) investigated the role of loudness and fundamental frequency in prediction of prominence.

Another, more interesting question is how do patterns of long/short (durational dimension), strong/weak (prominence dimension), and high/low (pitch dimension) create an overall rhythmic pattern and how are such sequences considered in the judgments of similarity.

Two things need to be foreseen and addressed in such research. The first is that there might be cross-linguistic differences in perception based on listeners’ first language, or a combination of the languages they speak; i.e., prior exposure to language(s) might affect which acoustic dimensions are more salient to the listener.

The second is that languages may differ in the number of dimensions they use to create rhythm, as well as in relation of prosodic dimensions to one another. Long may always be associated with prominent in some languages, but may occur independently in others. While comparing two speech samples, different cues might be used depending on the type of the stimuli – perception may utilize dimensions that maximally differentiate the two samples.

They found that, in English, loudness is more important factor.
and thus vary from comparison to comparison. I will return to the discussion of this issue in section 5.4.4.

**Contribution of phonological processes to rhythm perception**

One of the processes proposed to distinguish stress-timed from syllable-timed languages is vowel reduction. Many models, including the rhythm metrics approach, consider it associated with the stress-timed languages, but not with syllable- or mora-timed languages. However, Easterday et al. (2011) report that vowel reduction did not significantly affect Rhythm Metrics they considered (percent of vocalic durations, and standard deviations of vocalic and consonantal intervals).

It is important to know whether this process affects perception of rhythmic similarity and the Rhythm Metrics do not adequately capture it, or whether the process does not affect perception. This question can be reinterpreted as whether durational difference is perceived in non-prominent positions.

5.4.2 Use of modified speech in addressing questions on rhythmic similarity

To assess the rhythm without segmental and phonotactic influence, studies in the literature often employed perception experiments with filtered, synthesized, or re-iterant speech as stimuli. Two things that are often assumed in such perception experiments are: 1) that the judgments of similarity are based on rhythmic properties, and 2) that unmodified and modified samples are rhythmically equivalent.

The second assumption is justified by the following claims: 1) filtering masks segmental qualities and leaves only prosodic cues and broad phonotactic properties (Ramus and Mehler 1999); 2) all consonants are perceived in the same way with respect to rhythm and thus are all re-synthesized as the same quality, for instance /s/; and 3) imitations using re-iterant speech are rhythmically equivalent to the originals. Here is why this might be problematic.
Low-pass filtering

A low-pass filter eliminates bandwidth above the filtering frequency $F_{LP}$\textsuperscript{26}. Depending on where the energy of an individual segment is situated, more or less of that segment will be eliminated. Most fricatives will be completely filtered out as their energy is above the filtering frequency, which is usually 400Hz. Vowel energy will reduce in comparison to sonorant consonants whose energy is situated lower in the frequency domain. As a result, the energy ratio of segments will change after filtering. If perception of prominence, and therefore rhythm, is based on energy, then filtering may change the rhythmic sequence. Tests of rhythmic equivalence between original and filtered stimuli should always accompany such studies.

Re-synthesized speech

In re-synthesized speech, all consonants are synthesized as one quality, usually /s/, and all vowels as a single vowel quality, usually /a/, resulting in ‘sasa’ speech. This means that in the modified form all the consonants from the original speech sequence will contribute to rhythm in the same way, resulting in CV representation. The resulting rhythmic sequence may differ from the one that low-pass filtering produces, which is more similar to ‘saltanajc’ or ALT representation.

Re-iterant speech

Re-iterant speech consists of /da/ imitations of every syllable, while the prosody is assumed to be preserved. This method attempts to tap into speaker/listeners ability to produce rhythmic imitations of the original stimuli. This type of modified speech stimuli is likely the most similar to the original unmodified speech sequence; however, the ability of the imitators might vary and some tempo adjustments are likely to occur.

In conclusion, the rhythmic equivalence of modified and unmodified stimuli used in the perception experiments should be tested rather than assumed.

\textsuperscript{26} Real filters cannot eliminate all frequency above $F_{LP}$ and keep all frequency below it; elimination is instead gradual; the frequencies just below $F_{LP}$ are somewhat attenuated and those just above $F_{LP}$ still present and wane gradually.
5.4.3 The nature of rhythm

*Is rhythm difficult to define?*

One of the challenges for the research on speech rhythm is the lack of a clear understanding or agreement of what speech rhythm is. It is difficult to compare two approaches or results of different studies if rhythm is defined using different criteria. It is also difficult to judge whether quantitative measures capture rhythmic differences if we do not agree on what makes two rhythms perceptually similar.

Rhythm in speech is often broadly described as being similar to rhythm in music. This entails repetition of alternating qualities, short and longs, strong and weak. Yet, even in music, there was not much attention to models of rhythm until very recently (Hofmann-Engl 2002) and the models do not abound.

Historically, rhythm in speech has been associated with different types of isochrony (Pike 1945, Abercrombie 1967), with durational variability of syllables, feet, or vocalic and inter-vocalic intervals (Ramus et al. 1999, Grabe and Low 2002), or a perceived variability based on phonological properties (Dauer 1983, 1987; Auer 1993; Pamies Bertrám 1999), or underlying mechanisms that arise from regulating, at two levels, the durational properties of syllables and feet (Nolan and Asu 2009), or a word (Schiering et al. 2012) to mention a few approaches.

*Is rhythm one-dimensional or multi-dimensional*

Rhythm has been described in the literature both as a one-dimensional property that ranges between two extremes of stress-timed quality (Dauer 1987) and as a multi-dimensional property where properties along each dimension can vary independently (for example, see Nolan and Asu 2009).

*Is rhythm of speech perceived as small number of classes or as a continuum?*

Humans perception often categorizes continuous percepts/stimuli. In discussions of rhythm, researchers debate on whether rhythm of language is discrete – that is, whether there are small number of classes as posited by the RCH – or whether languages belong to a rhythm continuum.
These questions lead us to a very important discussion point – can we construct a model of rhythmic similarity?

5.4.4 Proposed model of rhythmic similarity

Here I propose a model of speech rhythm similarity based on the earlier discussed issues of the effect of phonotactics, prosody, and phonological processes on perception of rhythmic similarity.

I based the model on Tversky’s 1977 set-theoretical model of similarity, which defines object as a set of features. These features are then compared to obtain similarity value. Tversky opposes this model to the geometrical models that use distance metrics as estimates of (dis)similarity. He points out that while geometrical models apply better to phenomena like tones or colors, feature-model is more appropriate for judging similarity between faces or countries.

I believe that our perception views rhythm more like a face, rather than a color, in that the similarity is based on patterns. Let us call this rhythm perception model ‘Union of features’ model. Different features could be proposed such as existence of long unaccented elements (phonemic vowel length), high tone or low tone prominence, and simple low complexity characteristic prosodic patterns. Characteristic prosodic sequence is shown in Figure 5.1.

\[(\text{high, short, stressed}), (\text{low, long, unstressed}), (\text{low, short, unstressed})\]

Figure 5.1. An example of a characteristic prosodic sequence

To make a good model, however, the three dimensions, pitch height, duration, and the level of prominence must be allowed to vary independently.

This model represents a slight departure from Tverski’s model as it allows characteristic prosodic sequences to act as features. The features in Tverski’s model “may correspond to components such as eyes or mouth; they may represent concrete properties such as size or color; and they may reflect abstract attributes such as quality or complexity.”
However, in the present model, the feature may be sequential in nature, as represented in Figure 5.1.

Possible features in the model can include phonotactic (presence of long vowels), phonological (vowel reduction), or prosodic (high tone prominence, phrase/sequence final lengthening) properties.

In Tversky’s definition, Rhythm Metrics model represent a geometric model with a distance metric – distance on two-dimensional graphs – as a measure of dissimilarity. While it can also be understood as a simplified model, in which several dimensions are projected onto one, RM model does not allow comparison of different dimensions/properties of rhythm.

In fact, even when only durations are observed, RM model conflates two important properties: phrase final lengthening and nuclear lengthening. Edwards and Beckman (1988) argue that these are two distinct processes with different articulatory timing. We may also notice that they differ in location, one associated with the end of a sequence and the other variable.

‘Union of features’ model represents these two properties as two features and is able to compare two samples on both. This gives it an advantage over the Rhythm Metric model.

Another feature that is conflated with phrase-final lengthening and the nuclear stress in the simple durational long-short alternation model like that of Rhythm Metrics, is the existence of long vowels in unstressed positions – one of the properties proposed in Dauer 1987 to give rise to rhythm.

In sum, rhythmicity in the ‘Union of Features’ model can be seen as a repetition of features, whether simple (scalar) or sequential (vector), with no required occurrence at equidistant time points. Because features repeat, they can also be characterized by their frequencies.

5.4.5 Implications/prediction for L2 speech and learning in infants

‘Union of features’ model of rhythmic similarity has interesting predictions for the perception of L2 speakers. Each feature (say, durational patterning) must be operational for
the speaker/listener. If they are not able to judge certain feature, say, phonemic vowel length, differences might not be perceived; two samples with different rhythms might be perceived by these listeners as same. Thus, speakers of different languages may parse the rhythmic space in a different way based on features available to them.

If true, this would have important predictions about the infant rhythm perception, in relation to their experience or lack of experience with the language specific features. If, like segmental distinctions, features of rhythmic similarity are available at birth and later retained or lost based on the need for processing first language (L1), then infants should be successful in differentiating any two language samples that are differing in at least one feature. This may not be true for adults native speakers of a language that does not exemplifies the differentiating feature and for whom such feature may be lost.

If, on the other hand, the features are not universal, but learned from exposure, then infants would need time to learn it and, at least at very early age, lack the ability to perceive some distinctions that the adult listeners can perceive. Which prosodic features are acquired and which ones are present at birth is an important question. Perception experiments on both infant and adult rhythmic perception in many languages can be formulated to answer this question.

5.5 Conclusion

This dissertation asks and answers several questions about the interplay of rhythm, phonotactics, and perception. It points out where the problems in current quantitative approaches lie, but also explains how a particular factor – phonotactics – affects the methodology. The results will thus have an impact on methodology in quantitative studies of rhythm, contributing to the fields of phonetics and phonology, in particular to the ongoing discussions about quantifying speech rhythm. Specifically they will have impact on how the current measures need to be redefined in order to measure rhythmic similarity.

This dissertation also examines in detail the relation between rhythmic grouping and specific structural properties: 1) consonant cluster complexity and 2) word-length. These results add to the efforts to understand the relation between prosody and phonotactics, and to some degree between prosody and morphology.
The discussion of (broad) phonotactic similarity – similarity of cluster lengths and cluster patterns – will also be informative for the field of language typology. Approach I took in using the probabilistic phonotactics, that is, comparing cluster lengths and patterns across languages in conjunction with their frequencies, is in line with the views that our perception is gradient. The Phonotactic Metrics can be in fact used as a measure of word complexity that is more refined than a simple average cluster length, or general form of a syllabic shell. More refined comparison of word complexity can thus be achieved.

The cross-linguistic phontactic similarity is especially interesting because it is related to rhythmic similarity and basic phonological and morphological properties. It is hoped that the tables of cluster length distributions and most frequent clusters for individual languages, as well as the created phonetic corpora for 21 languages, will possibly be of use to others.

Deciding to use automatically transcribed orthographic materials as a base of the study contributes to the efforts to make this approach more widely known. Despite existence of studies like Kučera and George 1968, which analyzed the similarity between Russian, Czech, and German by analyzing transcribed corpora, it is only recently that the researchers have started to emphasize the use of large corpora for phonetic studies (Loukina et al. 2011) or have gotten interested in ways of producing such corpora (Garcia and González 2012).

Although the computational part of this dissertation is not yet at the level that it can be meaningfully shared, I hope to facilitate the use of these tools by others in the near future. Specifically, I plan to produce a web application that will consist of both a phonemic transcriber and a complexity calculator, which could be used independently of the transcriber for the speech that is already phonetically transcribed.

Finally, based on this study and the current views in the literature, I propose a feature model of rhythm perception. This model makes predictions regarding which languages will be judged as similar, as well as regarding possible differences in perception based on the listener’s first language.

I also provide several questions in the scope of future work – interplay of segments and rhythm, as well as interaction of prosodic dimensions in the perception of rhythm – that will contribute to better understanding of the nature of rhythm and provide details for the proposed rhythm model.
APPENDIX 1

BASIC PROPERTIES OF THE LANGUAGES FROM WALS

Appendix 1 provides two tables adopted from the World Atlas of Language Structures (WALS) (Dryer and Haspelmath 2011). Basic phonological properties are listed in Table A1.1, and basic morphological properties in Table A1.2. Information presented in italic was not present on the website; I added it based on similarity to other languages in the table or based on information available in language textbooks.
<table>
<thead>
<tr>
<th>Language</th>
<th>Consonant Inventories</th>
<th>Vowel Quality Inventories</th>
<th>Consonant-Vowel Ratio</th>
<th>Syllable Structure</th>
<th>Fixed Stress Locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawaiian</td>
<td>Small</td>
<td>Average (5-6)</td>
<td>Low</td>
<td>Simple</td>
<td>Penultimate</td>
</tr>
<tr>
<td>Maori</td>
<td>Small</td>
<td>Average (5-6)</td>
<td>Low</td>
<td>Simple</td>
<td>No fixed stress</td>
</tr>
<tr>
<td>Samoan</td>
<td>Small</td>
<td>Average (5-6)</td>
<td>Low</td>
<td>Simple</td>
<td>No fixed stress</td>
</tr>
<tr>
<td>Tongan</td>
<td>Small</td>
<td>Average (5-6)</td>
<td>Low</td>
<td>Simple</td>
<td>No fixed stress</td>
</tr>
<tr>
<td>Japanese</td>
<td>Moderately small</td>
<td>Average (5-6)</td>
<td>Average</td>
<td>Moderately complex</td>
<td></td>
</tr>
<tr>
<td>Turkish</td>
<td>Average</td>
<td>Large (7-14)</td>
<td>Average</td>
<td>Moderately complex</td>
<td>No fixed stress</td>
</tr>
<tr>
<td>Greek</td>
<td>Average</td>
<td>Average (5-6)</td>
<td>Average</td>
<td>Complex</td>
<td>Antepenultimate</td>
</tr>
<tr>
<td>Italian</td>
<td>Moderately large</td>
<td>Average (5-6)</td>
<td>Average</td>
<td>Moderately complex</td>
<td>No fixed stress</td>
</tr>
<tr>
<td>Spanish</td>
<td>Moderately large</td>
<td>Average (5-6)</td>
<td>Average</td>
<td>Moderately complex</td>
<td>No fixed stress</td>
</tr>
<tr>
<td>Catalan</td>
<td>Average</td>
<td>Large (7-14)</td>
<td>Average</td>
<td>Moderately complex</td>
<td>No fixed stress</td>
</tr>
<tr>
<td>Portuguese</td>
<td>Moderately large</td>
<td>Average (5-6)</td>
<td>Average</td>
<td>Moderately complex</td>
<td>No fixed stress</td>
</tr>
<tr>
<td>German</td>
<td>Average</td>
<td>Large (7-14)</td>
<td>Low</td>
<td>Complex</td>
<td>No fixed stress</td>
</tr>
<tr>
<td>Dutch</td>
<td>Average</td>
<td>Large (7-14)</td>
<td>Low</td>
<td>Complex</td>
<td>No fixed stress</td>
</tr>
<tr>
<td>Russian</td>
<td>Moderately large</td>
<td>Average (5-6)</td>
<td>High</td>
<td>Complex</td>
<td>No fixed stress</td>
</tr>
<tr>
<td>Polish</td>
<td>Moderately large</td>
<td>Average (5-6)</td>
<td>High</td>
<td>Complex</td>
<td>Penultimate</td>
</tr>
<tr>
<td>Czech</td>
<td>Moderately large</td>
<td>Average (5-6)</td>
<td>High</td>
<td>Complex</td>
<td>Initial</td>
</tr>
<tr>
<td>Serbian</td>
<td>Moderately large</td>
<td>Average (5-6)</td>
<td>High</td>
<td>Complex</td>
<td>No fixed stress</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>Large</td>
<td>Average (5-6)</td>
<td>Average</td>
<td>Complex</td>
<td>No fixed stress</td>
</tr>
<tr>
<td>Hungarian</td>
<td>Moderately large</td>
<td>Large (7-14)</td>
<td>Average</td>
<td>Complex</td>
<td>Initial</td>
</tr>
<tr>
<td>Estonian</td>
<td>Moderately large</td>
<td>Average (5-6)</td>
<td>High</td>
<td>Moderately complex</td>
<td>Initial</td>
</tr>
<tr>
<td>Indonesian</td>
<td>Average</td>
<td>Average (5-6)</td>
<td>Average</td>
<td>Complex</td>
<td>Penultimate</td>
</tr>
<tr>
<td>Language</td>
<td>Prefixing vs. Suffixing in Inflectional Morphology</td>
<td>Order of Subject (S), Object (O), and Verb (V)</td>
<td>Order of Adposition and Noun Phrase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bulgarian</td>
<td>Strongly suffixing</td>
<td>SVO</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catalan</td>
<td><em>Strongly suffixing</em></td>
<td>SVO</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czech</td>
<td>Weakly suffixing</td>
<td>SVO</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>Strongly suffixing</td>
<td>No dominant order</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estonian</td>
<td>Strongly suffixing</td>
<td>SVO</td>
<td>post</td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>Strongly suffixing</td>
<td>No dominant order</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greek</td>
<td>Strongly suffixing</td>
<td>No dominant order</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawaiian</td>
<td>Little affixation</td>
<td>VSO</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hungarian</td>
<td>Strongly suffixing</td>
<td>No dominant order</td>
<td>post</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indonesian</td>
<td>Strongly suffixing</td>
<td>SVO</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>Strongly suffixing</td>
<td>SVO</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>Strongly suffixing</td>
<td>SOV</td>
<td>post</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maori</td>
<td>Little affixation</td>
<td>VSO</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polish</td>
<td>Strongly suffixing</td>
<td>SVO</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portuguese</td>
<td>Strongly suffixing</td>
<td>SVO</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian</td>
<td>Strongly suffixing</td>
<td>SVO</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Samoan</td>
<td>Little affixation</td>
<td>No dominant order</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serbian</td>
<td>Strongly suffixing</td>
<td>SVO</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>Strongly suffixing</td>
<td>SVO</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tongan</td>
<td>Little affixation</td>
<td>No dominant order</td>
<td>pre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkish</td>
<td>Strongly suffixing</td>
<td>SOV</td>
<td>post</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX 2:

TEXTS AND TRANSCRIPTS FOR 21 LANGUAGES

Appendix 2 contains the following information for each language:

1) List of rules used in transcription of each language
2) Limitations that exist in transcriptions of each language
3) Transcription illustration: one paragraph in original text followed by its broad transcription equivalent

Languages in the test set include the following 21 languages, organized according to historical grouping (http://www.ethnologue.com/)

**Uralic:**
Estonian and Hungarian

**Slavic:**
Bulgarian, Czech, Polish, Russian, and Serbian

**Germanic:**
Dutch and German

**Romance:**
Catalan, Italian, Brazilian Portuguese, and Spanish

**Polynesian:**
Hawaiian, Maori, Samoan, and Tongan

**Other** (belong to different language groups):
Japanese, Turkish, Greek, and Indonesian

To mark diphthongs and triphthongs in IPA notation, I used dots around vowel sequences that are understood to be in the same syllable.

Columns (/:/) are used to mark phrase ends, while the IPA diacritic for long /ː/ are used to mark long vowel.
ESTONIAN

Assumed rhythm type: not agreed; It has vowel quantity distinction but also rich set of consonant clusters. In this work, I assign it to the syllable-timed group.

Short description: Estonian has long vowels and diphthongs. Although some phonologists posit 3 levels of vowel length, only two were transcribed here. The super-long is considered by some (e.g., Lippus et al. 2009) to be expressed as an interaction of vowel length and the pitch contour. Nasality in vowels was not transcribed.

Rules used in transcription:
- grapheme-tp-phoneme rules
- sequences of vowels were transcribed as diphthongs

Transcription example (text):
Kägu.
Ühes peres kasvas kaunis tütarlaps; sirgus nagu osi salus, oli lahke ja virk ja vaga, nii et ümberkaudu kedagi temataolist ei olnud. Isa ja ema hoidsid tütrekest kui silmatera ja armastasid teda ülväga. Seal tuli aga surm ja viis lahke eidekese ära, enne kui tütar neiualiseks sai. Isa ja tütreke leinasid kaua kadunud eite, ning tema hauaküngas seisis alati leinalille-vanikutega kaetud.

Transcription example (IPA):
kægu : yhes peres kasvas k.au.nis tyutarlaps : sirgus nagu osi salus , oli lahke ja virk ja vaga , niː et ymberk.au.du kedagi temat.ao.list .ei. olnud : isa ja ema h.o.i.dsid tytrekest k.u.i. silmatera ja armastasid teda ylivæga : s.ea.l tuli aɡa surm ja viis lahke .ei.dekeste æra , enne k.u.i. tytar ne.iu. .ea.liseks s.ai. : isa ja tytreke l.ei.nasid k.au.a kadunud .ei.te , ning tema h.au.akyangas s.ei.sis alati l.ei.nalille vanikutega k.ae.tud :
HUNGARIAN

Assumed rhythm type: not agreed; It has vowel quantity distinction but also rich set of consonant clusters. In this work, I assign it to the syllable-timed group.

Short description: Hungarian has long vowels; it does not have diphthongs. Long vowels were transcribed.

Rules used in transcription:
- grapheme-to-phoneme rules
- transcribing long vowels in certain environments

Transcription example (text):
A vak király.
Hol volt, hol nem volt, még az Öperenciás-tengeren is túl volt, volt a világon egy vak király. Mindenféle orvosok-doktorok próbálták meggyógyítani, de mindhiába, egyik sem ment semmire. Maga a király tudott volna ugyan egy orvosságot szemének, de azt senkinek sem mondtá meg, hogy mi; akárki kérdezte tőle, csak azt felelte, hogy mihásna mondja meg, mikor úgyse tudják megszerezni.

Transcription example (IPA):
a vak kiraj : hol volt , hol nem volt , meːg az oːpetɛnsiaːʃ tɛŋɡɛɾɛn iʃ tuːl volt , volt a vilaːɡon eːj vak kiraj : mindenfeːle orvoʃok doktorok proːbaːltɑːk meːɡoːʒiːtɑni , de mindhiaːba , ejik ʃɛm ment ʃɛmmire : maga a kiraj tudott volna uʃan eːj orvoʃʃaːgot semenɛk , de azt şɛŋkinek ʃɛm mondtɑ meg , hoʃ mi : akaar ki keːrdɛztɛ tposé , ʃʃak azt felelte , hoʃ mihasna mondjɑ meg , mikor uʃʃe tudjaːk megsɛrɛzni :
Assumed rhythm type: stress-timed as a member of the Slavic group. See Dimitrova 1997 for some discussion.

Short description:
First step in transcribing involved translation of Cyrillic letters to IPA phonemes. Since the stress does not occur in a fixed position, and I did not have access to stress location within word, I transcribed all back vowels as if they were stressed. Consonants are devoiced word-finally and before a voiceless consonant. They are voiced before a voiced consonant, except before /v/. Only word-final devoicing is implemented. In accordance with H_IPA (2000), no palatal consonants are assumed. Instead, they are analyzed as consonant+/j/ before back vowels and not phonemic before front vowels and /j/.

Rules used in transcription:
- grapheme to phoneme translation
- word-final consonant devoicing

Discrepancies:
1) voicing of some word-final consonants is incorrect
2) vowel qualities of unstressed vowels /e/ and /o/ are represented as /a/ or /ɤ/, and /ɔ/ or /u/ respectively
Neither of the two affects our current analyses.

Transcription example (text):
Северният вятър и слънцето се скарали кой от двамата е по-силен. Видели човек, който си вървял по пътя, облечен в дебело палто. Те решили да разрешат спора си, като видят кой пръв ще го накара да си съблече палтото. Първо опитал вятърът, но колкото по-силно духал, толкова по-плътно се увивал пътникът в палтото си. Тогава се появило слънцето и започнало да грее. Скоро човекът усетил топлината му и свалил палтото си. Така вятърът бил принуден да признае, че слънцето е по-силно.

Transcription example (IPA):
severnijat vjatɤr i sɬɤnʦɛtɔ se skarali kɔj ɔt dvamata ɛ pɔsilɛn : vidɛli ʧɔvɛk , kɔjtɔ si vɤrvjaɬ ɬɤpɤtja , ɔbleʃɛnɛbeɬɔ pɑɬtɔ : te refili da razreʃat spɔra si , katɔ vidjaɬ kɔj prɤf şte ɡo nakara da si sɬɤbliɬɛ paɬtɔ : pjɤrvɔ æpiɬal vjatɤrɬγ , no kɔɬkɔtɔ pɔsilɛn duxɬ , tɔlkɔva ɬɤpɤtɬɛ ɡa uwiɬal pɤtnikɤtpaɬtɔ si : tɔɡava se pojaviɬ sɬɤntstɛtɔ i zapaʃnaɬa da gree : skɔɾo ʧɔvɛkat uʃeti tɔplinata mu i svalil paɬtɔtɔ si : taka vjatɤt biɬ prinuden da priznaɬ , ʧe sɬɤntstɛt ɛ pɔsilɛn :
CZECH

Assumed rhythm type: stress-timed

Short description: Possible diphthongs are transcribed as diphthongs; ‘ou’ represented as a single vowel /ˌou/. (dots used to represent that a sequence is a diphthong). Long vowels are transcribed as long.

Rules used in transcription:
- graphemes-to-phonemes
- possible diphthongs are marked as diphthongs
- long vowels (marked in Czech texts) are marked as long
- alveolar nasal is changed to palatal nasal before palatal stops
- syllabic /r,l,m/ are marked as nuclei

Issues:
word ‘z’ accidently erased; supposed to be attached to the next word

Transcription example (text):
O Červené Karkulce.

Byla jednou jedna sladká dívenka, kterou musel milovat každý, jen ji uviděl, ale nejvíce ji milovala její babička, která by jí snesla i modré z nebe. Jednou ji darovala čepeček karkulku z červeného sametu a ten se vnučce tak líbil, že nic jiného nechtěla nosit, a tak ji začali říkat Červená Karkulka.

Transcription example (IPA):
o ʧɛrvɛnɛː karkulʦɛ : bɪla jɛdn.ou. jɛdna sladkaː diːvɛŋka , ktɛr.ou. musɛl mɪlovat kaʒdiː , jɛn jɪ uvyɛel , alɛ nejviːʦɛ jɪ mɪlovala jɛjiː babɪʧka , ktɛr.ou: bɪ jɪ: snesla ɪ modrɛː nebe : jɛdn.ou. jɪ: dɛɾɔvala ďɛpeʃɛk karkulku tʃɛrvɛnɛːʃo samɛtu a tɛn se vnuʧʦɛ tak liːbɪl , ʒɛ ɲuts jnɛːʃo nextjɛla nosɪt , a tɑː jɪ: zaʧalɪ řiːkat tʃɛrvɛnaː karkulka :

/ː/ marks long vowel, while /ː/ marks sentence break
Assumed rhythm type: stress-timed

Rules used in transcription:

No diphthongs or long vowels are assumed; palatalized /k/ and /g/ are single phones
- transVowels
- nasal vowel + /p/ transformed to vowel + /mp/
- grapheme(s) to phoneme(s)
- /n/ place assimilation to palatals stops
- ‘nn’ translated into single /n/; ‘kk’ translated into single /k/
- palatalize before /i/+vowel
- denasalize before /l,w,m,n/
- join independent /z/ and /v/ to the next word

Transcription example (text):

CZERWONY KAPTUREK.
Była sobie kiedyś mała, prześliczna dziewczynka. Jej buzia była tak słodka i radosna, że każdy, kto tylko raz na nią spojrzał, od razu musiał ją pokochać. Dziewczynka wraz z rodzicami mieszkała nieopodal lasu. Często odwiedzała babcię, która gotowa była jej przyczyścić nieba. Babcia mieszkała w niewielkim domku otoczonym przez zielony las. Pewnego dnia jej wnuczka otrzymała od niej prezent - czerwony aksamitny kapturek, który dziewczynka polubiła tak bardzo, że za nic nie chciała się z nim rozstawić i wszędzie nosiła go na swojej cudnej główce! Przez to zaczęto ją nazywać „Czerwonym Kapturkiem”.

Transcription example (IPA):

ʧɛrvɔnɨ kapturɛk : biwa sobie kɛdiɛ mawa , pʃɛɕliʃna jɛvtfiŋka : jes buza biwa tak swɔdka i radɔsna , ʒe kajdi , kto tiɭko raz na ne spojʒaw , od rau muɕaw jo pokɔxats : jɛvtfiŋka vraz zɾdɛitsamı mięʃkawa ɲeopɔdal lasu : tʃɛstɔ ovdiɛzawab babcɛ , ktura gotɔwa biwa je pʃiŋlits ɲeba : babca mięʃkawa ɲnwiɛlki domku otɕʃɔnim pʃez żelɔni las : pɛvneɡ dŋa jej vnutʃka otʃimawa ɔd nej prezent tʃɛrvɔni aksamitni kapturek , kturi jɛvtfiŋka polubiwa tak bardɔ , ʒe za nits ne xɛawa ɭe znim rɔzstavats i vʃɛ̃ɟɛ ɲɔsiwa go na svojej tsudnej gwuvtse : pʃez to zaɭɛtɔ jo nazivats , tʃɛrvɔnim kapturkɛm :
Assumed rhythm type: stress-timed

Short description:

First step in transcribing involved translation of Cyrillic letters to IPA phonemes. No long vowels or diphthongs assumed

Rules used in transcription:

- grapheme-to-phoneme
- palatalize vowel that occurs after another vowel
- word ‘в’ (in) erased by mistake; it was supposed to attach to the next word; only 130 cases in 10,000 word text; considered acceptable

_Transcription example (text):_

Каменный цветок.

_Тоже и в наших заводах, сказывают, это мастерство имели. Та только различка, что наши больше с малахитом вожгались, как его было довольно, и сорт - выше нет. Вот из этого малахиту и выделывали подходящие. Такие, слышь-ко, штучки, что диву даешься: как ему помогло._

_Transcription example (IPA):_

kamjennɨj ʦvjetok : je odni mramorski na slavje bili po kamjennomu to djelu : toʒɛ i naʃix zavodax , skazɨvajut , ɛto mastjerstvo imjeli : ta tolko razliʧka , tʃto naʃi bolʃəmalaxitom voʒgalis , kak jego bɨlo dovolno , i sort , vɨʃɛ njet : vot iz etogo malaxITU i vɨdjelɨvali podxodjaʃe : takije , sliʃ ko , ʃtuʧki , tʃto divu daʃsja : kak jemu pomoglo :
SERBIAN

Assumed rhythm type: stress-timed

Rules used in transcription:
- grapheme to phoneme translation
- place assimilation of /n/ to palatal stops /k/ and /g/ (within word or across word boundary)
- word ‘s’ (shorten form of ‘sa’ with) attached to the next word

_Transcription example (text):
Tri praseta.

_Transcription example (IPA):
    tri praseta : nekada davno na obodu jasenove šume rasla su tri praseta : rojen od iste kemaťe bejaxu braca odrasla u istom oboru : kada stasaše nastupio je tjas da se osamostale i zasnuju svoj dom : kako su bili priliţno vezani jedan za drugoga odluţiše da svoje nove kuću podignu u susedstvu : na taj naţin bi se uvek mogli naci u nevolji jedni drugima :
DUTCH

Assumed rhythm type: stress-timed

Short description: Long vowels and diphthongs were transcribed

Rules used in transcription:
- vowel sequences transcribed as diphthongs
- grapheme-to-phoneme rules
- Transcribing long vowels as a function of environment
- Vowel reduction
- Schwa insertion to break long clusters
- Consonant sequence simplification
- double /d,t,s,k,l/ were transcribed as single consonants

Transcription example (text):
De noordenwind en de zon waren erover aan het redetwisten wie de sterkste was van hun beiden. Juist op dat moment kwam er een reiziger aan, die gehuld was in een warme mantel. Ze kwamen overeen dat degene die het eerst erin zou slagen de reiziger zijn mantel te doen uittrekken de sterkste zou worden geacht.

De noordenwind begon toen uit alle macht te blazen, maar hoe harder ie blies, desto dichter trok de reiziger zijn mantel om zich heen; en ten lange reste gaf de noordenwind het op. Daarna begon de zon krachtig te stralen, en hierop trok de reiziger onmiddellijk zijn mantel uit. De noordenwind moest dus wel bekennen dat de zon van hun beiden de sterkste was.

Transcription example (IPA):
de noːɾdəʋɪnt en de zɔn varə ɛɾɛːvər am fɛt reːdɛtʋɪstə vi de stɛɾəkstə vas fɑn fɪŋ b.ei.do : j.œj.st op dat mɔːmɔt kwɑm ər ə r.ei.zɪχər ən , di ɛʃiːvt vas in ə vɑrmə mantel : zə kwɑmə ɔvɛɾən dat ɪbɛːnə di fiɛt erɛst ɛɾɪn z.lu. slɑχə ə r.ei.zɪχər z.ei.n mantɛl tə dʊn ʊtɾɛːkə də stɛɾəkstə z.lu. wɔrdə ɛɾəχt : de norɛdəvɪnt bɛχɔn tun .œj.t ələ məχ tə blɑzə , mɑɾ fi fiːdər i blis , dɛstə dɾɪʃtər tɾɔk də r.ei.zɪχər z.ei.n mantɛl əm zɪχ fɪɛn : ən tə ɫɑŋə lɛstə ɣaf de norɛdəvɪnt fiɛt əp : dɑɾnə bɛχɔn də zɔn ɪbɛːtəɾ əm ʊtɾɛːkə də r.ei.zɪχər əmɪdɛl.ei.k z.ei.n mantɛl .œj.t : də norɛdəvɪnt must dɨs vɛl bɛkənə də də zɔn fɪŋ b.ei.do də stɛɾəkstə vas :
GERMAN

Assumed rhythm type: stress-timed

Short description:

Diphthongs /ai, oi, au/ transcribed from corresponding letter sequences. Combination of vowels and the following coda /r/ considered diphthongs as well.

Eight long vowels transcribed. Some contrast in quality to the corresponding short vowel.

Rules used in transcription:

- grapheme to phoneme rules
- vowel sequences transcribed as diphthongs (where appropriate diphthong exists)
- vowel + /r/ transcribed as diphthongs except in monosyllables
- final schwa + /n/ transcribed as /n/; this should have been transcribed as syllabic /n/
- double letters r, t, l, f, n, s transcribed as single consonants.
- word initial and intervocalic /s/ transcribed as /z/
- vowel before h is transcribed as long
- vowel before double consonant is transcribed as short
- ‘r’ in VrC or Vr# (# denotes word boundary; V either short or long) becomes vocalic: /ɐ/
- word final ‘mn’ transcribed as /m/
- word final /t/ erased if the preceding sound is a consonant and the next word starts with an obstruent

Issues:

- final schwa + /n/ transcribed as /n/; this should have been transcribed as syllabic /n/
- some vowel sequences possibly transcribed as diphthongs

Transcription example (text):

Einem reichen Manne, dem wurde seine Frau krank, und als sie fühlte, daß ihr Ende herankam, rief sie ihr einziges Töchterlein zu sich ans Bett und sprach "liebes Kind, bleibe fromm und gut, so wird dir der liebe Gott immer beistehen, und ich will vom Himmel auf dich herabblicken, und will um dich sein."

Transcription example (IPA):

.aɪ.nem ʁ.aɪ.χən manə , dem v.ʊɐ.də z.aɪ.nə ʁ.aʊ. kʁaŋk , unt als zɪ fyltə , das iɐ endə heʁaŋkam , ʁɪf zɪ ie .aɪ.nʦɪɡəs tʊχt.ee.lɐ.ɐ n tso zɪɕ ans bet un jʊrɐχ ɪbʊs kɪnt , b.lɐ.ɐ ʃʃʊm un gʊt , zɔ vɪɛt dɪə d.ɐɐ. ɪbʊ ɡʊt ɪmme b.aɪ.stɐn , unt ɪɾ ɪlc vɪm hɪmm .au.f ɪɾ ʃɛkɐbbling , ʊnt ʊɾ ʊm dɪɾ z.aɪ.n :
CATALAN

Assumed rhythm type: mixed

Short description:

Catalan has diphthongs and triphthongs but not long vowels.

Rules used in transcription:

- grapheme to phoneme rules
- diphthongs and triphthongs transcribed from letter sequences (diphthong /ui/ become /wi/ before a consonant and /uj/ after a consonant
- word stress assigned
- vowel reduction applied
- word initial fricative voicing before vowel or voiced consonant
- fricative voicing between vowels
- fricative devoicing before voiceless consonant
- plosive voicing before voiced consonant
- plosive devoicing before vowel and voiceless consonant
- /r/ erased word finally or before a consonant that is not word final
- word final C1C2 transcribed as word final C1
- place assimilation of /n/
- voice plosive spirintization
- degemmination of /m, n, l, lj/
- hiatus resolution
- /j/ and /k/ that are surrounded by word boundaries are joined to the next word

Transcription example (text):
Saltimbanqui.
Era un barri portoriqueny, amb rètols en castellà; els carrers eren amplies, molt oberts, i les cruïlles separaven pendants oposats. Cinc cantonades més avall, ja començava Central Park: sobre el plànol, a Barcelona, m’havia semblat un lloc per estar-m’hi molt cèntric i adequat, però en el trajecte des de l’aeroport amb el taxi, els gratacels només els havia vist quan havíem travessat el pont, intuïts com un resplendor entre una boira que, malgrat que eren les deu de la nit, feia sensació de matinada.

Transcription example (IPA):
səltimbaŋki : eɾə m bari purturikeɲ , am retul æŋ kəstəʎa : el kəɾer erəŋ amblaʃ , mɔl uβəɾt , i les kruiʃəs səɾəɾəbəm pəɾəɾən upuzat : siŋ kəntunədəz mez əβəɔ , .iə. kʊməɾəʃə sentɾəl par : səɾəɾə βlənul , æ βəɾəɫəɾə , m æβ.iə. səmbəʈən ək pə sta m i mɔl sentɾig jəɾək.wa.t , pəɾo n eɾ tɾe(parenthesized)ɛktə də ləəɾəɾəɾə aməɾ təʃi , el γrαtəζɛl numez el æβ.iə. bɪs k.wa.n æβ.iə.m trəbəɾət æɾ pɔɾ , intuɪt kɔm un θɛɾβləndɛ nترا nə βɛɾə kə , məɣɾət kəɾəŋ lez d.ɛw. də lə nɪt , fəə sənəɾəɾə də məɾɪnədə :
ITALIAN

Assumed rhythm type: syllable-timed

Short description: Language has diphthongs and triphthongs but no long vowels. In this version all geminates are represented by two consecutive consonantal phonemes. Alternatively, double affricates and stops can be considered single consonants.

Rules used in transcription:
- Grapheme-to-phoneme rules
- Place assimilation of /n/ before palatal stops
- Voicing of intervocalic /s/ into /z/
- Triphthongs and then diphthongs are transcribed from vowel sequences

Transcription example (text):
I vestiti nuovi dell'imperatore.

Molti anni or sono, viveva un Imperatore, il quale dava tanta importanza alla bellezza ed alla novità dei vestiti, che spendeva per adornarsi la maggior parte de’ suoi danari. Non si curava de’ suoi soldati, non di teatri o di scampagnate, se non in quanto gli servissero di pretesto a far mostra di qualche nuovo vestito. Per ogni ora della giornata, aveva una foggia speciale, e, come degli altri re si dice ordinariamente: è al consiglio, - di lui si diceva sempre: è nello spogliatoio.

Transcription example (IPA):
i vestiti n.uo.vi dellimperatore : molti anni or sono , viveva un imperatore , il k.ua.le dava tanta importantsa alla belleʦʦa ed alla novita dei vestiti , ke spendeva per adornarsi la magʤʤor parte de s.uoi. danari : non si kurava de s.uoi. soldati , non di teatri o di skampaɲate , se non in k.ua nto ἃi servissero di pretesto a far mostra di k.ua.lke n.uo.vo vestito : per ὀνι ora della çhornata , aveva una foʧʧa speʧʧale , e , kome deʎi altri re si ditʧe ordinar.ia.mente : e al konsiʎo , di lui si ditʧeva sempre : e nello spoʎa.to.io.
PORTUGUESE

Assumed rhythm type: syllable-timed (Brazilian Portuguese rules were applied)

Short description: In Portuguese, I implemented diphthongs and nasalized vowels.

Rules used in transcription:
- Grapheme to phoneme rules
- Vowel sequences transcribed as diphthongs
- intervocalic /s/ voiced to /z/
- word-final /l/ is velarized
- syllable final /r/ is transcribed as uvular fricative
- vowel nasalized before syllable final /m/ or /n/

Transcription example (text):
O vento norte e o sol discutiam qual dos dois era o mais forte, quando passo um viajante vestido numa capa.
Ao vê-lo, poem-se de acordo em como aquele que primeiro conseguisse obrigar o viajante a tirar a capa seria considerado o mais forte. O vento norte começou a soprar com muita fúria, mas quanto mais soprava, mais o viajante se aconchegava à sua capa, até que o vento norte desistiu. O sol brilhou então com todo o esplendor, e imediatamente o viajante tirou a capa. O vento norte teve assim de reconhecer a superioridade do sol.

Issues: Some vowel qualities are transcribed incorrectly. This does not affect our analysis

Transcription example (IPA):
o vê-to noɾtʃi i o soɬ jiskutʃiá k.ua.ɗ dos d.o.i.s era o m.ai.s foɾtʃi , k.uã.do paso ũ v ia.ʒâtʃi veʃʃido numa kapa : ao vi lo , po.iẽ. si ji akɔ̃kdo .iẽ. komo akeli ki prim.ei.ro kõsegisi obrigaɾ o v ia.ʒâtʃi a tʃiɾaɾ kapa seria kõsiderado o m.ai.s foɾtʃi : o vê-to noɾtʃi komesou a sopraɾ kô m.ui.ta fur.ia. , mas k.uã.to m.ai.s sopraɾa , m.ai.s o v ia.ʒâtʃi si akõʃegava a sua kapa , aɾʃi ki o vê-to noɾtʃi dezístʃ.iu. : o soɬ briʃou ét.ẽu. kô todo o esplêdoɾ , i imed.ia.tamẽtʃi o v ia.ʒâtʃi tʃiɾou a kapa : o vê-to noɾtʃi tevi aɾši ji ɐkõ̃pẽeɾ a superioridaɾi do soɬ :
SPANISH

Assumed rhythm type: syllable-timed
Short description: Spanish has diphthongs; doesn’t have long vowels.

Rules used in transcription:
- grapheme to phoneme rules
- sequences of vowels should be transcribed as diphthongs. Unfortunately, this rule failed to apply, so
- /b/, /d/, and /r/ are spirintized intervocalically

Issues:
- Intervocalic voicing has not been implemented (but will be in a new version). This should not affect analysis.
- ‘y’ before vowels failed to be transcribed as /j/. There are only 85 cases in 10,000 words, so this error should not significantly affect the frequencies or metrics. CV word initials were under counted (<1% total words) in favor of zero-consonant onsets (vowel initials).

Transcription example (text):
Caperucita Roja

Había una vez una niña muy bonita. Su madre le había hecho una capa roja y la muchachita la llevaba tan a menudo que todo el mundo la llamaba Caperucita Roja.
Un día, su madre le pidió que llevara unos pasteles a su abuela que vivía al otro lado del bosque, recomendándole que no se entretuviese por el camino, pues cruzar el bosque era muy peligroso, ya que siempre andaba acechando por allí el lobo.

Transcription example (IPA):
kaperuθita roxa : aβia una beθ una niɲa mui bonita : su madre le aβia eʧo una kapa roxa i la muʧaθita la ʌβaθa tan a menuðo ke toðo el mundo la ʌβaθa kaperuθita roxa : un dia , su madre le piðio ke ʌθase unos pasteles a su abuela ke biβia al otro laðo del boske , rekomendandole ke no se entɾetuβiese por el kamino , pues kruθar el boske era muí peligroso , ia ke siempre andaβa aθεʃando por aʌθ el loβo :
HAWAIIAN

Assumed rhythm type: mora-timed

Short description:
Hawaiian has long vowels and diphthongs, long and short. All were transcribed; diphthongs were possibly overcounted.

Rules used in transcription:
- grapheme-to-phoneme rules
- transcribing sequences of 3 vowels as triphthongs
- transcribing sequences of two vowels as diphthongs
- letters corresponding to non-Hawaiian phonemes were replaced according to this set of correspondences \{b:p, d:k, v:w, s:k, r:l, f:p, t:k, g:k\}

Transcription example (text):
Ka Moʻolelo o nā Kamehameha.
No ka Noho aliʻi ʻana o Liholiho
ma luna o ke Aupuni, a ua Kapa
ʻia ʻo Kamehameha .

1 I ka ʻeiwa a me ka ʻumi paha o nā maka hiki o Keōpūolani, ua holo mai ʻo Kamehameha
I mā i ke kaua ma Maui me Kalanikūpule, ke keiki a Kahekili, a ua hoʻouka ke kaua nui ma
ʻĪao i Wailuku, a ua ʻauheʻe ʻo Kalanikūpule mā me nā aliʻi a pau o Maui iā Kamehameha, a
ʻo Keōpūolani kekahi i ʻauheʻe me Kekuʻiapoiwa, kona makuahine; ma luna o nā pali
kūninihi ka hāʻawe ʻana o kona makuahine, a mai make lāua, a e ʻole ka ikaiaka o ke kahu i
ka hāʻawe, pakele ai ko lāua ola.

Transcription example (IPA):
ka moʔolelo o nɑː kamehameha : no ka noho aliʔi ?ana o liholiho ma luna o ke .au.puni , a ua kapa ?ia ?o kamehameha : i ka ?.ei.wa a me ka ?umi paha o nɑː
maka hiki o keoʔpu:olani , ua holo m.ai . ʔo kamehameha i mɑː i ke k.au.a ma m.au.i
me kalaniku:pule , ke k.ei.ki a kahekili , a ua hoʔ.ou.ka ke k.au.a nui ma ?i:a. o i
w.ai.luku , a ua ?.au.heʔe ʔo kalaniku:pule mɑː me nɑː aliʔi a p.au . o m.au.i .iəː·
kamehameha , a ʔo keoʔpu:olani kekahi i ?.au.heʔe me kekuʔiap.oï.wa , kona
makuahine : ma luna o nɑː pali kūninihi ka hɑːʔawe ?ana o kona makuahine , a
m.ai. make l.a:u.a , a e ʔole ka ik.ai.ka o ke kahu i ka hɑːʔawe , pakele .ai. ko l.a:u.a
ola :
Assumed rhythm type: mora-timed

Short description:
Maori has long vowels and diphthongs, long and short. All were transcribed; diphthongs were possibly overcounted.

Rules used in transcription:
- grapheme-to-phoneme rules
- transcribing sequences of 3 vowels as triphthongs
- transcribing sequences of two vowels as diphthongs

Transcription example (text):
Whaitere – te whai ātahu.

E noho ana a Koro Pat i runga i tāna tūru ātawhao, e mātakitaki ana i ngā tamariki e toru e kanikani ana i te taha o te ahi, e tahutahu ana i ngā ngārehu, ka rere atu ki te pō.
"I kite koe i tērā?" E tohu whakawaho ana a Koro Pat ki te moana. E rua ngā pākau tapatoru i kitea e pakaru mai ana i waho, e heke ana hoki ki te papaki i te wai.
"He whai!" te tioro a ngā tamariki, me te oma atu ki te takutai. Ka tuohu a Kimi ki te tiki kōhatu, ā, e whakareri ana ki te whiu, ka pā te ringa o Koro Pat ki a ia.
"Hoi, ka whiu e koe he kōhatu ki tō māmā?" Ka noho pōnānā a Kimi, ka taka te kōhatu i tōna ringa.

Transcription example (IPA):
f.ai.tere , te f.ai. ā:tahu : e noho ana a koro pat i ruŋa i tɔːna tʊːru tɔː:f.əo.f.əo. , e mɔːtakitaki ana i ŋaː tamariki e toru e kanikani ana i te taha o te ahi , e tahutahu ana i ŋaː ŋaːrehu , ka rere atu ki te poː : i kite k.o. e i tɛːɾaː : e tohu fakawaho ana a koro pat ki te moana : e r.ua. ŋaː pɔː:k.au. tapatoru i kit.ea. e pakaru m.ai. ana i waho , e heke ana hoki ki te papaki i te w.ai. : he f.ai. : te t.io.ro a ŋaː tamariki , me te oma atu ki te takut.ai. : ka t.uo.hu a kimi ki te tiki kōhatu , ɑː , e fakareri ana ki te f.iu. , ka pɔː te riŋa o koro pat ki a .ia. : h.oi. , ka f.iu.a e k.o.e. he kōhatu ki toː mɔːmɔː : ka noho pɔːnɔːnaː a kimi , ka taka te kōhatu i toːna riŋa :
SAMOAN

Assumed rhythm type: mora-timed

Short description: Samoan has phonemic vowel length: 5 short vowels and corresponding 5 long vowels. However, they are not marked in the texts I have so these were all transcribed as short. There are 7 diphthongs.

Rules used in transcription:
- grapheme-to-phoneme rules

Issues:
- Long vowels were not marked in the texts and were thus transcribed as short. They do not affect present analysis, but knowing position of long vowels would improve advanced method for RM calculation.

Transcription example (text):
A’o gasegase pea le Tupu Tafa’ifa o Fonoti i Mulini’u Laloga’afu’a ma Sepolataemo, sa malaga atu le tama o Aputiputiatoloula ma lona tina o Melegalenu’u e asi le fa’atafa o Fonoti le Tupu, ma sa fa’apea lava fo’i le tele o le atunu’u sa gasolo i ai i le taimi lea. Fai mai sa potopoto ai Tumua e lipoi le gasegase o le Tupu. Na i ai fo’i Fuatino le masiofo a le Tupu fa’atasi ai ma lana tama o Muagututui’a le atali’i o Fonoti

Transcription example (IPA):
aʔo ŋaseŋase pea le tupu tafaʔifa o fonoti i mulinuʔu laloŋafuʔafuʔa ma sepolataemo , sa malaŋa atu le tama o aputiputiatol.ou.la ma lona tina o meleŋalenuʔu e asi le faʔatafa o fonoti le tupu , ma sa faʔapea lava foʔi le tele o le atunuʔu sa ŋasolo i .ai. i le t.ai.mi lea : f.ai. m.ai. sa potopoto .ai. tumua e lip.oi. le ŋaseŋase o le tupu : na i .ai. foʔi fuatino le masiofo a le tupu faʔatasi .ai. ma lana tama o muaŋututui’a le ataliʔi o fonoti:
TONGAN

Assumed rhythm type: mora-timed

Short description: Tongan has long vowels. It is said to not have diphthongs, unlike other Polynesian languages I transcribed. This increases %V for Tongan compared to Hawaiian, Samoan, and Maori, as well as slightly larger average word length (expressed in number of syllables).

Rules used in transcription:
- grapheme-to-phoneme rules

Transcription example (text):
Fakafe'iloakí.
Ngaahi Mātu'a, Kau Faiako, Kau 'Etivaisa mo e Kau Taki 'Ofeina 'o e To'u Tupú:

Kuo uiui'i kimoutolu 'e he 'Eikí ke mou tokoni ki hono fakaului 'o e to'u tupú ki he ongoongolelei. Ko ha tāpuaki faka'ofo'ofo mo'oni ia! 'Oku mou ma'u 'a e faïngamālie ke fokotu'u ha fetu'utaki tu'uloa mo e to'u tupu pelepelengesi kuo fakafalala atu 'e he 'Eikí ke mou tokanga'i. 'I he'ene mahino kiate kimoutolu 'a 'enau ngaahi fie ma'u makehe mo e holi 'o honau lotó, te mou lava ai 'o tokoni ke nau a'usia iate kinautolu pē 'a e ngaahi tāpuaki 'o hono ako mo mo'ui 'aki 'o e ongoongolelei 'i he 'aho kotoa pē.

Transcription example (IPA):
Assumed rhythm type: Japanese is said to be mora-type with respect to rhythm. Having phonemic length for both vowels and consonants it belongs to quantity languages.

Short description: Japanese has long vowels; it does not have diphthongs.

To transcribe Japanese texts, I used simple transformation of the syllabaries (hiragana and katakana) into phonemes and combining two identical vowels into a long vowel. A short list of kanji is also interpreted based on translations from http://life.ou.edu/stories/. Example transcription is given below.

Syllable final nasal stop was transcribed differently from /n/ in the onset. They are considered as consonants.

Rules used in transcription:
- kana-to-IPA rules
- transcribing sequence of two identical vowels as long vowel

_Transcription example (text):_

ももたろう。

むかし、むかし、あるところに
おじいさんと おばあさんがいました。
おじいさんがやまで きてきりにいけば、
おばあさんは川（かわ）へせんたくにてかけます。
「おじいさん、はよう もどってきなされ。」
「おばあさんもきをつけてな。」
まいにち やさしく いいあって でかけます。

あるひ、おばあさんが
川でせんたくをしていたら、
つんぶらこ つんぶらこ
ももがながれてきました。ひろって
たべたら、なんとも おいしくて
ほっぺたが おちそうです。おじいさんにも
たべさせて あげたいと おもって、
「うまいもの こっちゃ こい。
にがいもの あっちゃ いけ。」
といったら、
どんぶらこ どんぶらこ
でっかい もも が ながれてきました。
おばあさんは よろこんで、ももを
いえに もって かえりました。
Transcription example (IPA):
momotarouː mukaʃi, mukaʃi, aru tokoro ni ojiːsaN to obaːsaN ga imajitaː ojiːsaN
gu jama he ki o kiri ni ikeba, obaːsaN ha kawa he seNtaku ni dekakemasuː ojiːsaN,
hajou modotte kinasareː obaːsaN mo ki o tsukete naː mainiʃi jasajikuː iː atte
dekakemasuː aruhi, obaːsaN ga kawa de seNtaku o ōite itara, tsuNbuːrako
 tsuNbuːrako momo ga nagarete kimaʃitaː hirotte tabetara, naNtomo oiiʃikuː
hoppeta ga oiiʃiouː ojiːsaN nimo tabesasete agetai to omotte, : umai momo koʃʧa
koiː nigai momo aʃʧa ikeː to ittara, doNbuːrako doNbuːrako dekkai momo ga
nagarete kimaʃitaː obaːsaN ha joroκoNde, momo o ie ni motte kaerimajitaː
Assumed rhythm type: undecided;
Short description: Turkish has diphthongs and long vowels. Both were transcribed.

Rules used in transcription:
- grapheme-to-phoneme rules
- diphthongs transcription
- palatalization of velar stops
- velarization of /l/
- deletion of syllable final ‘soft g’

Transcription example (text):
Pinokyo.

Transcription example (IPA):
pinokjo : biɾ vəɾmuʃ , biɾ jokmuʃ tʃook esci bir zamanda çytʃyk bir kasabada jeppetto adunda ihtıjar bir ojunʤakʧɯ jafarmɯʃ : japtɯɣɯ tahtadan ojunʤakɬarɯ satarak jeʃiʃими saɫarımɯʃ : ihtıjar ojunʤakʧɯnun hajattay yzyldyy tek şe:j. bir tʃoʤuɣunun olmamas.ɯj.mɯʃ : bir tʃoʤuɣunun olmas.ɯ ifiʃ neler veɾmezmiş ci : bir jyn jeni bir ojunʤak jarpak ifiʃin ormana jidip cytyk aramaja başɬamɯʃ : dercen tam aradɯɣu jibi bir cytyyy buɬmuʃ :
GREEK

Assumed rhythm type: syllable-timed

Short description: Texts were given in Greek alphabet. Diphthongs in Greek were implemented as a hiatus resolution of /oi, ao, uo/.

Rules used in transcription:
- grapheme-to-phoneme rules
- /n/ and /s/ liaison after unstressed but before stressed vowel in the next word
- stops are voiced after /n/, within word or over word boundary
- transcribing some unusual letters
- vowel deletion as a resolution of a hiatus other than those cases that result in diphthongs

Issues:
- Voicing of some consonants is possibly incorrect

Transcription example (text):
Ο βοριάς κι ο ήλιος μάλωναν για το ποιος απ’ τους δύο είναι ο δυνατότερος, όταν έτυχε να περάσει από μπροστά τους ένας ταξιδιώτης που φορούσε κάπα. Όταν τον είδαν, ο βοριάς κι ο ήλιος συμφώνησαν ότι όποιος έκανε τον ταξιδιώτη να βγάλει την κάπα του θα θεωρούνταν ο πιο δυνατός. Ο βοριάς άρχισε τότε να φυσάει με μανία, αλλά όσο περισσότερο φυσούσε τόσο περισσότερο τυλιγόταν με την κάπα του το ταξιδιώτης, ώσπου ο βοριάς κουράστηκε και σταμάτησε να φυσάει. Τότε ο ήλιος άρχισε με τη σειρά του να λάμπει δυνατά και γρήγορα ο ταξιδιώτης ξεστάθηκε κι έβγαλε την κάπα του. Έτσι ο βοριάς αναγκάστηκε να παραδεχτεί ότι ο ήλιος είναι πιο δυνατός απ’ αυτόν.

Transcription example (IPA):
o voɾʝas c.oи.ʎos malonan ja to pços ap tus dьo in o dинατотерос , ota netiçe na perasi apo brosta tu senas taksidjotis pu foruse kapa : ota do niðan , o voɾjas c.oи.ло simfonisa noti opço sekane to daksidjoti na vyali ti gapa tu ða theorudan o pçо dìnatos : o voɾjas arçiçe tote na fisai me mania , al.ао.so perisothero fisuse toso perisotero tiliyotan me ti gapa t.uo. taksidjotis , ospu o voɾjas kurastice ce stamatise na fisai : tote .ои.ло sarçiçe me ti sira tu na labi dînata ce γριɣor.ао. taksidjotis zestaðice cevyale ti gapa tu : etsi o voɾjas anagastice na paraðexti oti .ои.ло sine pçо dìnatos ap afton :
Assumed rhythm type: not reported

Short description: Three diphthongs are reported for Indonesian: /ai/, /au/, and /oi/; Rules are made to transcribe these but only /au/ occurred in our 10,000 word text. It occurred 139 times only. This supports the fact that examples of diphthongs are usually given for loan words or a very limited set of lexical items.

In Indonesian, letter ‘e’ can be pronounced as mid-vowel /e/ or a schwa. Since there is no rule determining grapheme-to-phoneme mapping, I transcribed all ‘e’ vowels as schwa.

Rules used in transcription:
- phoneme-to-grapheme rules
- verb that have transitive suffix /i/ were checked to not transcribe the final /ai/ as a diphthong (diphthongs do not occur across morpheme boundary).

Issues: Vowels /e/ and /ə/ are both represented as /ə/. This should not affect the present analysis.

Transcription example (text):
Malin Kundang.

Transcription example (IPA):
malin kundaŋ : pada zaman dahulu kala , ada səoruŋ anak bərnama malin kundaŋ : ia tiŋgal di səbuah dəsa tərpəncil di pəsisir pantaisumatəra barat bərsama ibuɲa : ajah malin kundaŋ pərgi məncari nafkah di nəgeri sabaranʃ daŋan maŋaruŋi lautan jaŋ luas : tapi , əntah kənapa saŋ ajah tidak pərnah kəmbali kə kampuŋ halamanŋa : jadi , ibu malin kundaŋ harus məŋgantikan posisiŋa untuk məncari nafkah :
### APPENDIX 3

VALUES OF RHYTHM AND PHONOTACTIC METRICS

Table A3.1 Rhythm Metrics values

<table>
<thead>
<tr>
<th>Language</th>
<th>rPVI-Cr</th>
<th>Varco-Cr</th>
<th>∆Cr</th>
<th>%Vr</th>
<th>nPVI-Vr</th>
<th>Varco-Vr</th>
<th>∆Vr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulgarian</td>
<td>45.0</td>
<td>-</td>
<td>41.0</td>
<td>46.0</td>
<td>47.0</td>
<td>-</td>
<td>33.5</td>
</tr>
<tr>
<td>Catalan</td>
<td>67.8</td>
<td>53.0</td>
<td>45.2</td>
<td>45.6</td>
<td>44.6</td>
<td>55.0</td>
<td>36.8</td>
</tr>
<tr>
<td>Czech</td>
<td>70.0</td>
<td>61.5</td>
<td>52.0</td>
<td>46.0</td>
<td>45.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dutch</td>
<td>57.4</td>
<td>44.0</td>
<td>53.3</td>
<td>42.3</td>
<td>65.5</td>
<td>65.0</td>
<td>42.3</td>
</tr>
<tr>
<td>Estonian</td>
<td>40.0</td>
<td>0.0</td>
<td>52.0</td>
<td>0.0</td>
<td>45.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>German</td>
<td>55.3</td>
<td>54.0</td>
<td>62.0</td>
<td>39.8</td>
<td>59.7</td>
<td>51.5</td>
<td>-</td>
</tr>
<tr>
<td>Greek</td>
<td>59.6</td>
<td>46.8</td>
<td>41.1</td>
<td>48.2</td>
<td>48.7</td>
<td>57.4</td>
<td>-</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hungarian</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Indonesian</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Italian</td>
<td>49.3</td>
<td>51.7</td>
<td>48.1</td>
<td>45.2</td>
<td>48.5</td>
<td>55.0</td>
<td>40.0</td>
</tr>
<tr>
<td>Japanese</td>
<td>62.5</td>
<td>-</td>
<td>35.6</td>
<td>53.1</td>
<td>40.9</td>
<td>-</td>
<td>40.2</td>
</tr>
<tr>
<td>Maori</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Polish</td>
<td>79.1</td>
<td>-</td>
<td>51.4</td>
<td>41.0</td>
<td>46.6</td>
<td>-</td>
<td>25.1</td>
</tr>
<tr>
<td>Portuguese</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Russian</td>
<td>61.0</td>
<td>-</td>
<td>54.0</td>
<td>-</td>
<td>45.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Samoan</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Serbian</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Spanish</td>
<td>57.7</td>
<td>50.2</td>
<td>47.4</td>
<td>43.8</td>
<td>29.7</td>
<td>53.3</td>
<td>33.2</td>
</tr>
<tr>
<td>Tongan</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Turkish</td>
<td>67.0</td>
<td>-</td>
<td>53.0</td>
<td>-</td>
<td>47.0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table A3.2 Phonotactic Metrics values (long V = 1)

<table>
<thead>
<tr>
<th>Language</th>
<th>rPVI-Cp</th>
<th>Varco-Cp</th>
<th>∆Cp</th>
<th>%Vp</th>
<th>nPVI-Vp</th>
<th>Varco-Vp</th>
<th>∆Vp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulgarian</td>
<td>54.1</td>
<td>80.2</td>
<td>109.7</td>
<td>43.5</td>
<td>104.4</td>
<td>42.3</td>
<td>45.0</td>
</tr>
<tr>
<td>Catalan</td>
<td>51.3</td>
<td>79.0</td>
<td>107.11</td>
<td>43.3</td>
<td>102.9</td>
<td>34.8</td>
<td>36.2</td>
</tr>
<tr>
<td>Czech</td>
<td>58.1</td>
<td>83.1</td>
<td>118.4</td>
<td>42.1</td>
<td>103.2</td>
<td>36.5</td>
<td>38.2</td>
</tr>
<tr>
<td>Dutch</td>
<td>66.6</td>
<td>88.1</td>
<td>139.7</td>
<td>39.0</td>
<td>103.3</td>
<td>37.1</td>
<td>38.9</td>
</tr>
<tr>
<td>Estonian</td>
<td>50.8</td>
<td>79.8</td>
<td>109.1</td>
<td>43.3</td>
<td>104.4</td>
<td>42.3</td>
<td>45.1</td>
</tr>
<tr>
<td>German</td>
<td>66.0</td>
<td>86.9</td>
<td>132.0</td>
<td>40.4</td>
<td>105.4</td>
<td>48.1</td>
<td>52.2</td>
</tr>
<tr>
<td>Greek</td>
<td>44.2</td>
<td>73.5</td>
<td>94.0</td>
<td>46.7</td>
<td>108.2</td>
<td>54.9</td>
<td>61.6</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>56.7</td>
<td>116.7</td>
<td>76.0</td>
<td>95.6</td>
</tr>
<tr>
<td>Hungarian</td>
<td>58.0</td>
<td>82.8</td>
<td>119.8</td>
<td>41.0</td>
<td>103.1</td>
<td>35.7</td>
<td>37.3</td>
</tr>
<tr>
<td>Indonesian</td>
<td>44.6</td>
<td>74.6</td>
<td>98.7</td>
<td>44.8</td>
<td>106.0</td>
<td>49.0</td>
<td>53.4</td>
</tr>
<tr>
<td>Italian</td>
<td>50.2</td>
<td>77.5</td>
<td>105.1</td>
<td>46.3</td>
<td>108.6</td>
<td>55.7</td>
<td>62.7</td>
</tr>
<tr>
<td>Japanese</td>
<td>13.5</td>
<td>44.4</td>
<td>47.7</td>
<td>52.1</td>
<td>109.7</td>
<td>59.3</td>
<td>67.8</td>
</tr>
<tr>
<td>Maori</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>55.4</td>
<td>114.4</td>
<td>70.3</td>
<td>85.4</td>
</tr>
<tr>
<td>Polish</td>
<td>60.1</td>
<td>84.8</td>
<td>122.7</td>
<td>42.3</td>
<td>105.3</td>
<td>45.1</td>
<td>48.6</td>
</tr>
<tr>
<td>Portuguese</td>
<td>31.6</td>
<td>64.4</td>
<td>76.7</td>
<td>50.4</td>
<td>112.9</td>
<td>67.5</td>
<td>80.6</td>
</tr>
<tr>
<td>Russian</td>
<td>71.6</td>
<td>89.7</td>
<td>139.7</td>
<td>39.5</td>
<td>103.1</td>
<td>35.5</td>
<td>37.1</td>
</tr>
<tr>
<td>Samoan</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>61.8</td>
<td>137.7</td>
<td>92.3</td>
<td>141.4</td>
</tr>
<tr>
<td>Serbian</td>
<td>44.6</td>
<td>75.2</td>
<td>97.3</td>
<td>46.3</td>
<td>106.9</td>
<td>52.1</td>
<td>57.6</td>
</tr>
<tr>
<td>Spanish</td>
<td>47.8</td>
<td>76.9</td>
<td>102.4</td>
<td>46.0</td>
<td>108.3</td>
<td>55.0</td>
<td>61.7</td>
</tr>
<tr>
<td>Tongan</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>56.3</td>
<td>120.3</td>
<td>76.6</td>
<td>98.7</td>
</tr>
<tr>
<td>Turkish</td>
<td>45.3</td>
<td>74.5</td>
<td>98.1</td>
<td>42.9</td>
<td>102.5</td>
<td>32.4</td>
<td>33.6</td>
</tr>
<tr>
<td>Language</td>
<td>rPVI-Cp</td>
<td>Varco-Cp</td>
<td>ΔCp</td>
<td>%Vp</td>
<td>nPVI-Vp</td>
<td>Varco-Vp</td>
<td>ΔVp</td>
</tr>
<tr>
<td>---------------</td>
<td>---------</td>
<td>----------</td>
<td>-----</td>
<td>------</td>
<td>---------</td>
<td>----------</td>
<td>-----</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>54.1</td>
<td>80.2</td>
<td>109.7</td>
<td>43.5</td>
<td>104.4</td>
<td>42.3</td>
<td>45.0</td>
</tr>
<tr>
<td>Catalan</td>
<td>51.3</td>
<td>79.0</td>
<td>107.11</td>
<td>43.3</td>
<td>102.9</td>
<td>34.81</td>
<td>36.31</td>
</tr>
<tr>
<td>Czech</td>
<td>58.1</td>
<td>83.1</td>
<td>118.4</td>
<td>45.6</td>
<td>115.0</td>
<td>66.6</td>
<td>80.6</td>
</tr>
<tr>
<td>Dutch</td>
<td>66.6</td>
<td>88.1</td>
<td>139.7</td>
<td>43.0</td>
<td>117.4</td>
<td>69.7</td>
<td>86.6</td>
</tr>
<tr>
<td>Estonian</td>
<td>50.8</td>
<td>79.8</td>
<td>109.1</td>
<td>46.0</td>
<td>112.8</td>
<td>67.2</td>
<td>80.2</td>
</tr>
<tr>
<td>German</td>
<td>66.0</td>
<td>86.9</td>
<td>132.0</td>
<td>40.8</td>
<td>106.5</td>
<td>51.6</td>
<td>56.6</td>
</tr>
<tr>
<td>Greek</td>
<td>44.2</td>
<td>73.5</td>
<td>94.0</td>
<td>46.7</td>
<td>108.2</td>
<td>54.9</td>
<td>61.6</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>58.7</td>
<td>124.5</td>
<td>84.4</td>
<td>114.9</td>
</tr>
<tr>
<td>Hungarian</td>
<td>58.0</td>
<td>82.8</td>
<td>119.8</td>
<td>45.5</td>
<td>118.9</td>
<td>71.4</td>
<td>89.9</td>
</tr>
<tr>
<td>Indonesian</td>
<td>44.6</td>
<td>74.6</td>
<td>98.7</td>
<td>44.8</td>
<td>106.0</td>
<td>49.0</td>
<td>53.4</td>
</tr>
<tr>
<td>Italian</td>
<td>50.2</td>
<td>77.5</td>
<td>105.1</td>
<td>46.3</td>
<td>108.6</td>
<td>55.7</td>
<td>62.7</td>
</tr>
<tr>
<td>Japanese</td>
<td>13.5</td>
<td>44.4</td>
<td>47.7</td>
<td>52.7</td>
<td>111.4</td>
<td>63.2</td>
<td>73.7</td>
</tr>
<tr>
<td>Maori</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>58.6</td>
<td>127.5</td>
<td>83.4</td>
<td>115.3</td>
</tr>
<tr>
<td>Polish</td>
<td>60.1</td>
<td>84.8</td>
<td>122.7</td>
<td>42.3</td>
<td>105.3</td>
<td>45.1</td>
<td>48.6</td>
</tr>
<tr>
<td>Portuguese</td>
<td>31.6</td>
<td>64.4</td>
<td>76.7</td>
<td>50.4</td>
<td>112.9</td>
<td>67.5</td>
<td>80.6</td>
</tr>
<tr>
<td>Russian</td>
<td>71.6</td>
<td>89.7</td>
<td>139.7</td>
<td>39.5</td>
<td>103.1</td>
<td>35.5</td>
<td>37.1</td>
</tr>
<tr>
<td>Samoan</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>61.8</td>
<td>137.7</td>
<td>92.3</td>
<td>141.4</td>
</tr>
<tr>
<td>Serbian</td>
<td>44.6</td>
<td>75.2</td>
<td>97.3</td>
<td>46.3</td>
<td>106.9</td>
<td>52.1</td>
<td>57.6</td>
</tr>
<tr>
<td>Spanish</td>
<td>47.8</td>
<td>76.9</td>
<td>102.4</td>
<td>46.0</td>
<td>108.3</td>
<td>55.0</td>
<td>61.7</td>
</tr>
<tr>
<td>Tongan</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>57.0</td>
<td>123.6</td>
<td>79.3</td>
<td>105.3</td>
</tr>
<tr>
<td>Turkish</td>
<td>45.3</td>
<td>74.5</td>
<td>98.1</td>
<td>43.1</td>
<td>103.1</td>
<td>36.0</td>
<td>37.6</td>
</tr>
</tbody>
</table>
APPENDIX 4: WORD-LENGTH DISTRIBUTIONS FOR 21 LANGUAGES

Figure A4.1 Distribution of word lengths for Bulgarian

Figure A4.2 Distribution of word lengths for Catalan
Figure A4.3 Distribution of word lengths for Czech

Figure A4.4 Distribution of word lengths for Dutch
Figure A4.5 Distribution of word lengths for Estonian

Figure A4.6 Distribution of word lengths for German
Figure A4.7 Distribution of word lengths for Greek

Figure A4.8 Distribution of word lengths for Hawaiian
Figure A4.9 Distribution of word lengths for Hungarian

Figure A4.10 Distribution of word lengths for Indonesian
Figure A4.11 Distribution of word lengths for Italian

Figure A4.12 Distribution of word lengths for Japanese
Figure A4.13 Distribution of word lengths for Maori

Figure A4.14 Distribution of word lengths for Polish
Figure A4.15 Distribution of word lengths for Portuguese

Figure A4.16 Distribution of word lengths for Russian
Figure A4.17 Distribution of word lengths for Samoan

Figure A4.18 Distribution of word lengths for Serbian
Figure A4.19 Distribution of word lengths for Spanish

Figure A4.20 Distribution of word lengths for Tongan

143
Figure A4.21 Distribution of word lengths for Turkish
ARVANITI, AMALIA and RODRIGUEZ, TARA. 2013. The role of rhythm class, speaking rate, and F0 in language discrimination. Laboratory Phonology, 4.


EASTERDAY, SHELECE, TIMM, JASON AND MADDIESON IAN. 2011. The effects of phonological structure on the acoustic correlates of rhythm. ICPhC XVII, Hong Kong


TORGENSEN, EVIND NESSA and SZAKAY, ANITA 2012. An Investigation of speech rhythm in London English Lingua, 122.822-40.


WAGNER, PETRA and DELLWO, VOLKER. 2004. Introducing YARD (Yet Another Rhythm Determination) and re-introducing isochrony to rhythm research. Proc. Speech Prosody
