Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

A model of sonority based on pitch intelligibility

Aviad Albert

Studies in Laboratory Phonology

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

LabPhon-Logo

Studies in Laboratory Phonology

Chief Editor: Martine Grice
Editors: Doris Mücke, Taehong Cho

In this series:

1. Cangemi, Francesco. Prosodic detail in Neapolitan Italian.
2. Drager, Katie. Linguistic variation, identity construction, and cognition.
3. Roettger, Timo B. Tonal placement in Tashlhiyt: How an intonation system accommodates to adverse phonological environments.
4. Mücke, Doris. Dynamische Modellierung von Artikulation und prosodischer Struktur: Eine Einführung in die Artikulatorische Phonologie.
5. Bergmann, Pia. Morphologisch komplexe Wörter im Deutschen: Prosodische Struktur und phonetische Realisierung.
6. Feldhausen, Ingo & Fliessbach, Jan & Maria del Mar Vanrell. Methods in prosody: A Romance language perspective.
7. Tilsen, Sam. Syntax with oscillators and energy levels.
8. Ben Hedia, Sonia. Gemination and degemination in English affixation: Investigating the interplay between morphology, phonology and phonetics.
9. Easterday, Shelece. Highly complex syllable structure: A typological and diachronic study.
10. Roessig, Simon. Categoriality and continuity in prosodic prominence.

ISSN: 2363-5576

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

A model of sonority based on pitch intelligibility

Aviad Albert

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Aviad Albert. 2023. *A model of sonority based on pitch intelligibility* (Studies in Laboratory Phonology). Berlin: Language Science Press.

This title can be downloaded at:
redefine\lsURL
© 2023, Aviad Albert
Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):
http://creativecommons.org/licenses/by/4.0/
ISBN: no digital ISBN
no print ISBNs!
ISSN: 2363-5576
no DOI

Cover and concept of design: Ulrike Harbort
Fonts: Libertinus, Arimo, DejaVu Sans Mono
Typesetting software: XeLaTeX

redefine \publisherstreetaddress
redefine \publisherurl

Storage and cataloguing done by redefine \storageinstitution

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Contents

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Dedication

This book is dedicated to the loving memory of Maurice Sarfati (1947–2018) and Ezra Yitzhak Nawi (1952–2021), two dear friends, inspiring critical thinkers and true Jerusalemite legends, who passed away while I was working on this research.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Acknowledgments

The work presented in this book is based on my doctoral dissertation which was accepted by the Faculty of Arts and Humanities of the University of Cologne in 2022. This work would not have been possible without a doctoral scholarship from the German Academic Exchange Service (DAAD). I am also thankful to the Collaborative Research Centre in Cologne (SFB 1252 *Prominence in Language*) for funding the later stages of this effort. Further thanks should also go to a.r.t.e.s. Graduate School for the Humanities Cologne and the Cologne Center of Language Sciences (CCLS) for their assistance during this project.

I owe a huge debt to my thesis supervisor, Martine Grice, who opened the doors of Cologne's phonetics lab to me. Martine was able to see the potential behind my scattered thoughts and she pushed me forward throughout the path that was required in order to shape them into cohesive proposals that tell a coherent story. Any failure in achieving these goals is mine.

My second thesis supervisor, Doris Mücke, introduced me to Articulatory Phonology, which facilitated critical developments in the way I think about problems in phonology. Doris encouraged me to do my research without losing sight of my creative interest in music, an interest that we both share.

I found personal friendships and academic inspiration in Cologne's phonetics lab. Timo Roettger left a huge mark on my thinking, in terms of both theory and methodology. The same ought to be said about Francesco Cangemi, with whom I also had the pleasure to co-author some of the work that is presented in this book. Francesco's interaction with my projects had a crucial impact on my path.

Many people that I met in Cologne's phonetics lab deserve my gratitude: Simon Wehrle, who reduced the rate of crimes I commit against the English language and remained a steady pillar of sanity; Caterina Ventura, Simona Sbranna and Maria Lialiou, with whom I had the pleasure to co-author papers using the ProPer toolbox; and Anna Bruggeman, who was always willing to assist me if I asked for help, and I always did!

There are many other members, students and friends of the phonetics lab, past and present, to whom I owe my gratitude: Bastian Auris, Stefan Baumann, Mark Ellison, Luke Galea, Katharina Gayler, Harriet Hanekamp, Anne Hermes, Henrik Hess, Alicia Janz, Constantijn Kaland, Janina Kalbertodt, Theo Klinker, Martina

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Krüger, Hauke Lindstädt, Jane Mertens, Eduardo Möking, Lena Pagel, Christine Riek, Simon Roessig, Christine Röhr, Tobias Schröer, Mathias Stöber, Tabea Thies, Drenushë Valera-Kurteshi and Esther Weitz.

While working on this research in Cologne, I had the pleasure to meet and to engage with The following eclectic list of inspiring researchers, to whom I also wish to extend my thanks: Dinah Baer-Henney, Jason Bishop, Ioana Chitoran, Jennifer Cole, Sam Hellmuth, Hae-Sung Jeon, Frank Kügler, Leonardo Lancia, Umesh Patil, Elina Savino, Petra Schumacher, Kevin Tang, Francisco Torreira, Sandra Vella, Kai Vogeley, Nigel Ward, Bodo Winter and Katharina Zahner-Ritter.

A huge gratitude must be reserved for my good friend, Bruno Nicenboim. The fruits of our collaborative effort can be found in the experimental evidence I present herein, but his influence was more far-reaching. I learned a great deal from Bruno's analytical skills, intellectual honesty and technical agility.

I thank Carol Espy-Wilson for being so generous in giving me access to the *APP Detector* code, which allowed this whole project to get off the ground with its first reliable measurements of periodic energy. Many people helped me further in using this code. These include Francesco Cangemi, Yair Lakretz and Doron Veltzer. I also thank Paul Boersma for helping me figure out how to extract periodic energy data using Praat.

I thank Joanna Rączaszek-Leonardi for her kindness and for her role in shaping my understanding of language systems (and I thank Leo for pointing me in that direction). I thank Yoav Beirach for his friendship and the thought-provoking discussions we had about the notions of harmony, sound and time. I thank Eitan Grossman and Noam Faust for their kindness and for the stimulating discussions we had, including invitations to talks in Jerusalem and in Paris.

I am also thankful for the consistent assistance, engaging feedback and invitations to talks from my ex-teachers in Tel Aviv University, Outi Bat-El and Evan Cohen. My thanks also go to all the friends and colleagues from Tel Aviv University (past and present) who engaged with me and my work in recent years: Daniel Asherov, Si Berebi, Irena Botwinik, Aya Meltzer-Asscher, Avi Mizrachi, Doron Veltzer, Hadas Yeverechyahu and Hadass Zaidenberg. Further gratitude should be extended to Shuly Wintner, Bracha Nir and Ruth Berman for their contributions to my academic transition towards a doctoral endeavor.

I am grateful for my loving family, my beloved mom and my two amazing sisters (and their own beautiful families), for their unconditional love and endless support. Finally, I thank my dear partner and my very best friend – Alma – for wherever she is, I know that we are home.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Part I

Introduction

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

1 General introduction

1.1 The sonority challenge

Sonority is a central notion in phonetics and phonology, with many useful formal applications, yet it has remained vague in too many important respects. The centrality of sonority is primarily derived from the important theoretical weight that it carries in descriptions of syllables in phonology. Sonority is a single hierarchical concept that is most often used to characterize all speech sounds along a single scale in a manner that is pivotal for generalizing preferences and restrictions on syllabic organization.

However, even after many decades in which sonority has played a crucial role in phonology, there is no consensus with regards to its *phonetic* basis in articulation or perception of speech. Many proposals have been debated, but no real consensus has ever been reached. Sonority therefore presents an ongoing phonetic challenge. A good overview of the multiplicity of proposals for the basis of sonority can be found in the various publications by Stephen Parker, from his dissertation (Parker 2002), to subsequent publications (like Parker 2012, 2017, 2018) that meticulously document the vast research related to sonority in the linguistic literature.

The most prevalent models of sonority are based on the *Sonority Sequencing Principle* (SSP) and they are very characteristic of linguistic models from the second half of the twentieth century, whereby the speech signal is represented as a sequence of discrete units, phonological processes are modeled as symbol manipulating rules, and time is accounted for in terms of the non-overlapping linear order of the discrete units in symbolic representations (see Section 3.1).

It may very well be the case that little progress has been made in the theory of sonority since the turn of the century because of the constraining role that SSP-based models play with regards to phonetic dimensions. Specifically, the classic theoretical idea that the speech signal is composed of segments that have a fixed sonority value, which they share with other members of the same category, may have created an impossibility in the traditional theory. This is because the speech signal does not in fact lend itself to such analyses of non-overlapping discrete units with fixed sonority values (see Section 3.1). Such units can only be extracted

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

from a human mind. In other words, the classic theory may have created a formal notion of sonority that simply cannot be found in any phonetic space.

The challenges of sonority are therefore at the intersection of phonetics and phonology. Chiefly, these are issues pertaining to phonetic substance itself and to phonological theory, which accounts for the parts of the system that are linguistically relevant. A better model of sonority would seem to require novelties on both fronts.

1.2 Beyond correlates

Another issue highlighted by the notion of sonority, and directly related to the above, is the lack of explanatory power in many phonological models. A good example of this problem can be gleaned from the common practice of suggesting acoustic *correlates* for various linguistic phenomena, without any related attempt to suggest plausible *causation*. This is very often the case when the suggested acoustic correlates do not seem to have any clear and consistent links to the perception or articulation of speech. In many other cases, the implied causation can be misguided. A case in point is the use of physical acoustic intensity as a correlate of linguistic phenomena – sonority *inter alia* – although the physical amplitude of the entire signal does not consistently relate to perceived loudness, or any other aspect of perception and/or articulation (see details in Section 2.2.3).

Thus, it seems already at the outset of this project that in order to suggest a form of phonetic substance for sonority we need to be able to explain its function in the language system, beyond its ability to exhibit statistical correlations with sonority hierarchies. This cannot be achieved with a few tweaks in traditional discrete and symbolic models, as they are too far removed from cognitive plausibility owing to their classic computer-like architectures (see Chapter 3). The mainstream models of phonology from the second half of the twentieth century are simply not designed to achieve explanatory goals of this kind.

1.3 Goals and motivations of the current endeavor

The challenges that sonority presents are far-reaching, as apparent from the list of problems detailed above. Solving them requires a host of theoretical and methodological novelties that necessitate a relatively large-scale effort. To undertake these challenges, this work aims to determine what sonority *is*, what it *does*, and *how* it does it. To this end, it is also imperative to break away from tra-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

ditional discrete and symbolic models in linguistics by incorporating continuity and dynamics in a perception-based model of phonology.

1.4 A note about terminological choices

Terminology can be confusing. Often there are multiple terms for the same thing, and they are sometimes loaded with differing implications in different scientific and professional circles. The following description of terminological interpretations is intended to reduce confusion for readers of this book.

The terms CONSONANT and VOWEL are used here broadly to denote the phonological entities, which also consider the position within the syllable. In some cases, when it is specifically relevant to only refer to phonetic features (i.e. only to the degree of vocal tract stricture), I use the terms CONTOID and VOCOID (respectively), following Pike (1943).

Successive consonants that follow each other in the phonological description are either referred to as a SEQUENCE or a CLUSTER. The latter has a more specific meaning, implying that *clusters* are syllabified together in a single syllable, thus constituting a tautosyllabic complex onset or coda. The term *sequence* is used when no implication is made about the status of the two successive items, which could also be heterosyllabic.

When writing about auditory perception, it can be useful to keep acoustic and perceptual aspects separate via distinct terminology. The terms INTENSITY, POWER and AMPLITUDE relate to the acoustic signal. They are often interchangeable although whenever the distinction between the raw pressure and the log-transformed dB scale is relevant, I use *power* for the raw pressure and *intensity* for the the log-transformed dB scale. Crucially, the related term for perception of acoustic strength is consistently LOUDNESS. Likewise, F0, or the FUNDAMENTAL FREQUENCY, always refer to the acoustic signal. The related term in perception is PITCH.

Note that the terms TONE and TUNE have specific meanings in linguist contexts. A tone is a pitch target with a communicative function that is either part of the lexical repertoire (e.g. *lexical tone*) or the post-lexical repertoire (e.g. a *pitch accent*/*boundary* tone). A *tune* is often used to describe larger intonation contours (that are commonly analyzed as being composed of tones).

This book deals with DISCRETE VS. CONTINUOUS, and with SYMBOLIC VS. DYNAMIC entities or elements. These pairs are interchangeable in the majority of the contexts used here. Likewise, the terms TRAJECTORY and CURVE, as they are used here to denote graphs within plots or, more abstractly, when describing the progression of a certain feature in time, are largely interchangeable.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

1.5 Conventions

Phonemic depictions of speech are presented in this book with mostly standard IPA conventions. A few diversions from the IPA norm include the use of acute accents (´) on the vowel (e.g. é, ó, á) to mark the stressed syllable, and the use of the simple grapheme /c/ to denote the voiceless alveolar affricate, which is transcribed with the complex grapheme /t͡s/ in the standard IPA.

Within phonemic transcriptions, a dash (-) is sometimes used to mark morpheme boundaries, while a dot (.) is used to mark syllabic boundaries. For instance, *dla.t-ót* stands for 'door-PL' in Modern Hebrew, where the suffix *-ot* forms a syllable with the final consonant of the base morpheme (/t/), as shown by the location of the dot.

1.6 Scope of book

The current book is divided into five parts in an attempt to address all the necessary issues mentioned above. The first part, *Introduction*, includes this chapter (Chapter 1) and two more chapters that present the relevant background on sonority (Chapter 2) and linguistic models more generally (Chapter 3).

The second part, *Novel theoretical outlooks*, contains three chapters that present the new theoretical proposals that this work develops: (i) the *PRiORS* framework for modeling auditory perception (Chapter 4); (ii) the proposal for sonority's perceptual cause and acoustic correlate, and the new sonority-based criterion for well-formedness determinations, the *Nucleus Attraction Principle* (Chapter 5); and (iii) the implementation of the Nucleus Attraction Principle in both symbolic and dynamic terms as two separate models (Chapter 6).

The third part, *Evidence in support of the Nucleus Attraction Principle*, consists of two chapters. Chapter 7 presents an experimental study that analyzes data from perception tasks carried out by native speakers of German and Modern Hebrew. Chapter 8 presents a corpus study that looks at some phonologized regularities in Modern Hebrew.

The last chapter in which new data is presented (Chapter 9) constitutes the fifth part of this book, titled *Further contributions of periodic energy to the study of prosody*. This chapter is a showcase for the *ProPer* toolbox, which is an ongoing open-source project building on the assumptions proposed here for sonority in order to develop new acoustic tools for the study of prosody.

This book ends in the *Conclusion* part (Chapter 10), with short discussions on issues that this work can contribute to, such as a more holistic phonotactic

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

division of labor, and the debate regarding the universality of sonority, followed by the closing section, *Directions for future work*.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

2 Sonority: Background

Sonority is a fundamental notion in phonetics and phonology, playing a crucial role in accounts of the syllable in linguistic theory (as a good indication in support of this claim, Parker 2018, cites 2413 titles involving sonority). The topic of sonority can be roughly divided into two related theoretical constructs: (i) *sonority hierarchies* (or *scales*) and; (ii) *sonority principles* (or *generalizations*). Sonority hierarchies locate all speech sounds on a single scale, while sonority principles are universal generalizations about the well-formedness of syllables. Sonority principles require a sonority hierarchy to model the well-formedness of syllables given their underlying sequence of consonants and vowels. Thus, a model of sonority is capable of predicting distributional patterns of consonants and vowels in all human language systems in terms of *phonotactics* – preferences and restrictions with regards to possible combinations of segmental sequences.

Sonority hierarchies and principles are designed to model speech in discrete terms. Thus, sonority models standardly express sonority in terms of integers that are associated to classes of consonants and vowels along an ordinal sonority hierarchy. Syllabic well-formedness is computed from the concatenation of these values in symbolic time (i.e. from linearly ordered non-overlapping symbols). This type of modeling lacks robust cognitive motivations as it assumes – explicitly or implicitly – that linguistic processing is analogous to the workings of a computer despite strong evidence to the contrary (see Chapter 3 and especially Section 3.1).

Although widely used and accepted, the notion of sonority at the same time remains vague and highly contested for various reasons. To date, no real consensus exists with respect to the phonetic basis of sonority in terms of a consistent articulatory or perceptual phenomenon which sonority distinctions could be derived from. This results in a multitude of different sonority hierarchies. Furthermore, the lack of a phonetically useful metric for sonority plagued most sonority models with inherent circularity, as sonority hierarchies are often both determined and confirmed by attested segmental combinations, without recourse to any independently motivated phenomenon (Ohala 1992). Moreover, sonority principles such as the widely used *Sonority Sequencing Principle* (SSP) have been taken as

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

axioms with formal definitions that lack an explicit functional motivation relating to speech articulation or perception. This choice of architecture for the SSP resulted in some of the persistent failures of SSP-based models, such as the unpredicted prevalence of */s/-stop* clusters on the one hand, and the unpredicted rarity of sonorant plateaus on the other (see Section 2.2).

2.1 Hierarchies and principles

2.1.1 Sonority hierarchies

A sonority hierarchy is a single scale on which all consonant and vowel types can be ranked relative to each other.[1] Such hierarchies can be traced back centuries, and concepts akin to sonority hierarchies can be found already in the pioneering works of early Sanskrit grammarians. Donegan (1978) notes that Pāṇini and the Sanskrit grammarians used the term *svara* to imply some kind of harmonic musical quality which applies mainly to vowels. Parker (2002: 58) notes further that the Sanskrit grammarians observed natural classes for speech sounds that are "grouped according to their degree of 'opening' (*vivāra*)". Early versions of current sonority hierarchies are often dated to Sievers (1893), Jespersen (1899), and Whitney (1865), while Ohala (1992) even goes as far back as de Brosses (1765).

While the phonetic basis of sonority hierarchies remains controversial, phonological sonority hierarchies have been primarily based on repeated observations that revealed systematic behaviors of segmental distribution and syllabic organization within and across languages. The general consensus regarding the phonological sonority hierarchy thus stems from attested cross-linguistic phonotactic behaviors of different segmental classes, such as, for instance, the preference for *stop-liquid* sequences in onset positions (e.g. /kl/ in the English word ***cl**ean*) and for the mirror-image *liquid-stop* sequences in coda positions (e.g. /lk/ in the English word *mi**lk***), but not the other way around. See examples in Zwicky (1972), Selkirk (1984), Parker (2002), Jany et al. (2007), and recall Ohala's (1992) related criticism regarding the circularity that results from determining sonority hierarchies according to attested behavior without another independent (phonetic) variable.

Most common phonological sonority hierarchies group segment types into classes that primarily reflect the standard *manners of articulation* in traditional

[1] Note that a related notion of *strength hierarchies* makes similar distinctions, yet in the opposite direction (stronger = less sonorant). Strength hierarchies are mostly evoked in relation to lenition processes rather than syllabic phenomena.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

phonology. The distinct categories commonly used include *stops*, *fricatives*, *nasals*, *liquids*, *glides*, and *vowels*, often with additional distinctions such as voicing and vowel height.[2] Although there are many different proposals for sonority hierarchies (Parker 2002 found more than 100 distinct sonority hierarchies in the literature), a very basic hierarchy that seems to reach a considerable consensus, and is often cited in relation to Clements's (1990) seminal paper is given in (1).

(1) Obstruents < Nasals < Liquids < Glides < Vowels

The ordering of different speech sounds along the sonority hierarchy is assumed to be universal, in line with the common assumption that sonority has a phonetic basis in perception and/or articulation, yet the patterning of segmental classes as distinct groups along the scale is considered to be language-specific, i.e., based on phonological categorization. For example, voiceless stops may be considered universally lower than voiced fricatives on the sonority hierarchy, yet for some languages and analyses they may constitute a single level of *obstruents*. Classes along the sonority hierarchy are most commonly modeled as a series of integers (often referred to as sonority indices) reflecting the ordinal nature of phonological interpretations of the sonority hierarchy.

The main differences that result from variation of the basic hierarchy in (1) concern the class of obstruents, which may contain voiced and voiceless variants of stops and fricatives (to mention just the most prominent distinctions). It is therefore not uncommon to expand the class of obstruents, whereby stops are lower than fricatives and voiceless consonants are lower than voiced ones. Note that vowels are often also commonly divided into subgroups along the sonority hierarchy (see Gordon et al. 2012), but these distinctions will be irrelevant in the context of this work.

The two variants of sonority index values given in Table 2.1 thus reflect two ends of a common sonority hierarchies spectrum. These range from hierarchies that collapse all obstruents together into a single class (resulting in the same sonority index value for all obstruents), to hierarchies that expand the class of obstruents by employing voicing distinctions as well as manner distinctions between stops and fricatives (resulting in multiple sonority index values within the class of obstruents). In what follows I will refer to these two versions of the sonority hierarchy as H_{col} for the *collapsed* sonority hierarchy, and H_{exp} for the *expanded* sonority hierarchy.

[2]The group of *liquids* is the most loosely defined, as it includes both *lateral approximants* (namely /l/) and various types of rhotics such as *trills* (/r,ʀ,ʁ/), *taps* (namely /ɾ/), and alveolar and retroflex *approximants* (/ɹ,ɻ/).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table 2.1: Traditional phonological sonority hierarchies. *Note:* Index values reflect the ordinal ranking of categories in sonority hierarchies. The obstruents in H_{col} are collapsed into one category (bottom four rows = 1), while in H_{exp} they are expanded into four distinct levels.

Sonority index values			
H_{col}	H_{exp}	Segmental classes	Phonemic examples
5	8	Vowels	/u, i, o, e, a/
4	7	Glides	/w, j/
3	6	Liquids	/l, r/
2	5	Nasals	/m, n/
1	**4**	Voiced Fricatives	/v, z/
1	**3**	Voiced Stops	/b, d, g/
1	**2**	Voiceless Fricatives	/f, s/
1	**1**	Voiceless Stops	/p, t, k/

2.1.2 Traditional sequencing principles

Sequencing principles can be understood as a mapping scheme between the ranks of a sonority hierarchy and the linear order of symbolic speech segments. Modern formulations of such principles, which use the ordinal sonority hierarchy to generalize over the phonotactics of consonantal sequences in terms of *sonority slopes* were developed mainly throughout the 1970s and 1980s in seminal works such as Zwicky (1972), Hankamer & Aissen (1974), Hooper (1976), Kiparsky (1979), Lowenstamm (1981), Steriade (1982), Cairns & Feinstein (1982), Selkirk (1984), Harris (1983), Mohanan (1986) and Clements (1990).

Sonority index values, indicating a rank on the sonority hierarchy, can be readily plugged into models that are able to predict distributional patterns of segments vis-à-vis syllabic organization in terms of sonority slopes. Consonants and vowels in a given string are interpreted as a sequence of discrete points in symbolic linear time. The corresponding sonority index values that are associated with these segments are then interpreted in terms of slopes that result from interpolation over the sequence of points. Thus, for instance, going from a low ranking segment to a high one is considered to yield a rising slope, while two adjacent segments that share the same sonority index yield a plateau. Importantly, the notion of the *syllable* is required to define the ranges and types of preferred slopes, which rise from the onset of the syllable to its nucleus and fall from the nucleus of the syllable to its coda. Syllabic well-formedness is therefore defined

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

in terms of universal generalizations over the preferred and dispreferred types of sonority slopes that result from the concatenation of different consonants and vowels, and their grouping into syllables.

The most basic and widely used sonority-based principle that derives phonotactic predictions in terms of syllabic well-formedness is the *Sonority Sequencing Principle* (SSP). The SSP is a simple yet powerful generalization about phonotactics that has been evidently useful in countless theoretical accounts. It identifies three distinct types of slopes – *rises*, *falls*, and *plateaus* – such that sequences of segments should rise in sonority from the consonant(s) in the syllabic onset to the syllable's nucleus (most often a vowel) and fall from the nucleus to the consonant(s) in the syllabic coda. In this project, I focus on syllable-initial onset consonant clusters that precede a vowel, whereby a rising sonority slope (e.g. *plV*) is considered well-formed and a falling sonority slope (e.g. *lpV*) is considered ill-formed (see Figure 2.1). Sonority plateaus (e.g. *pkV*) fare in between, giving way to various interpretations depending on the language and analysis. As such, plateaus may be considered as ill-formed or well-formed (e.g. Blevins 1995, Asherov & Bat-El 2019, Bat-El 1996), although they are generally interpreted as denoting a third, mid-level of well-formedness.

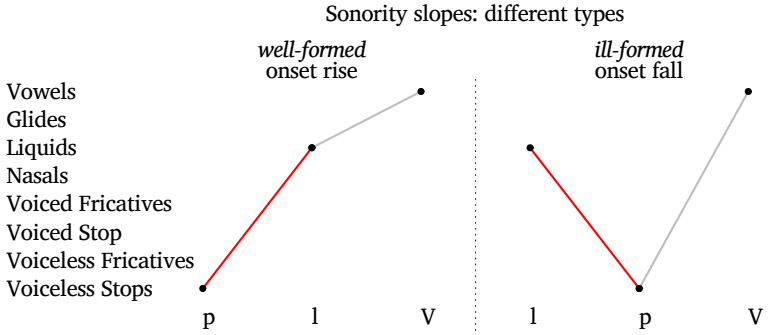

Figure 2.1: Schematic depiction of the sonority slopes of two onset clusters, *plV* and *lpV*. The red line denotes the sonority slope of the onset cluster (i.e. the two onset consonants), while the grey line denotes the slope between the second consonant and the vowel at the nucleus position (always a rise in these cases). The angle of the red lines reflects the well-formed rising sonority slope of the onset cluster in *plV* and the ill-formed falling sonority slope of the onset cluster in *lpv*.

The *Minimum Sonority Distance* (MSD; Steriade 1982, Selkirk 1984) is a well-known elaboration on the preferred angle of sonority slopes compared to basic applications of the SSP, given that the SSP makes no distinction between different angles of rising or falling slopes. The MSD was designed to prefer onset rises

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

with steep slopes over onset rises with shallow slopes, under the assumption that consonantal sequences in the onset are preferred with a larger sonority distance between them. For instance, *plV* has a steeper rise compared to *bnV* and it is therefore better-formed according to the MSD (see Figure 2.2).

The *Sonority Dispersion Principle* (SDP; Clements 1990, 1992) is a slightly different yet related principle that prefers onset rises with a large distance and an equal dispersion of sonority index values across the consonantal sequence and the following vowel. The results of the SDP are highly contingent on the given sonority hierarchy and it is not very clear how to apply the SDP formula with onset sonority falls (among other problems listed in Parker 2002: 22–24). The SDP is therefore not comparable as a model that can generate the full set of well-formedness predictions for onset clusters. Indeed, the SDP is mostly invoked in relation to other generalizations that it makes about the status of the onset versus the coda (not directly related to consonantal clusters), by assuming that onsets prefer to maximize sonority distance from the following nucleus while codas prefer to minimize it.

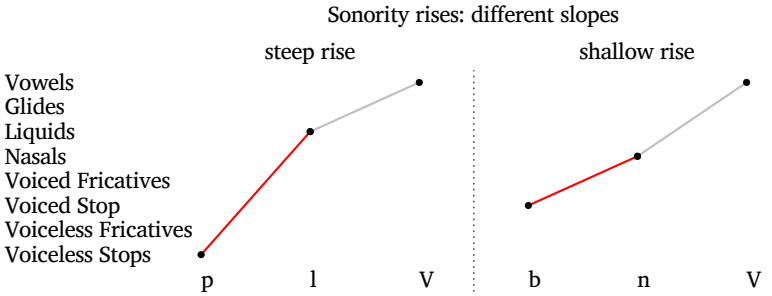

Figure 2.2: Schematic depiction of the sonority slopes of two onset clusters, *plV* and *bnV* (the red solid line denotes the sonority slope of the onset clusters). The angle of the red lines reflects a steeper rise for *plV* (left) compared with *bnV* (right), due to the larger sonority distance between the consonants in *plV*.

2.2 Problems with standard sonority theory

2.2.1 Slippery sonority slopes

The widely accepted use of sonority slopes in order to explain and predict phonotactic behaviors has been adopted by many researchers with only few changes such as the above-mentioned elaborations on the angle of sonority slopes (e.g. the MSD). This is a strong testament to the simplicity and power of the concept of

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

sonority slopes. However, given that the role of slopes is essentially formal, with no explicit functional motivation from articulation, perception or cognition, they remain open for interpretation. In other words, since sonority slopes are not tied to functional aspects such as degree of jaw opening, or the degree of perceived loudness, they pose no inherent limit on what type of phonotactic behavior they can be used to explain and predict. Indeed, sonority slopes have been used in attempts to explain practically all types of phonotactic phenomena, regardless of their different potential sources. This over-application of sonority slopes has resulted in various contradictions in the sonority literature (such as the case of */s/-stop* clusters, see Section 2.2.2), which were highlighted in some prominent objections to a notion of sonority that is not phonetically motivated and appears to act like a cover term for various functionally-different processes (e.g. Ohala & Kawasaki 1984, Henke et al. 2012, Laks 1995, Ohala 1992, Steriade 1999, Wright 2004).

Traditional sonority accounts formalize sonority principles in terms of slopes that are obtained from the sonority index values of members of a consonantal cluster, where only the difference, or distance, between segments in a sequence is taken into account. This suffices to characterize the rough angle of the sonority slope, but not its underlying power, which could potentially differentiate between a low sonority sequence and a high sonority sequence that have the exact same type of sonority slope (see Figures 2.3 and 2.5). To cover this aspect of the sequence, it suffices to obtain the sonority index value of the most marginal member of a sequence alongside the information about the slope of that sequence. For onset sequences, the most marginal member is the first segment that reflects the sonority *intercept* of the onset sequence, which is informative with regards to the overall sonority level of the slope. In the context of the current study, which focuses on complex onset clusters consisting of biconsonantal sequences, the slope is the difference between members of a sequence and the intercept is the sonority index value of the first consonant.

Intercepts play no role in the characterization of traditional sonority profiles although they are informative with respect to the amount of underlying sonority that a certain slope carries. This is a curious fact given that sonority-based accounts stem from the assumption that sonority quantities have an effect on the observed phenomena. Clements' 1990 SDP was actually designed to prefer less sonorant onsets, which could account for this aspect of sonority profiles, but, as mentioned above in Section 2.1.2, the SDP has a host of problems that prevent it from becoming a full model (e.g. it is not designed to account for falling consonantal slopes in onsets). Crucially, the SSP and MSD do not account for this aspect at all, as they only look at sonority slopes. Presumably, it is the propensity

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

for simple and elegant rather than functional generalizations in many theoretical linguistic traditions (see Chomsky 2021) that cemented the formal architecture of sonority principles with a maximally reduced conception of sonority slopes.

Taken together, the over-application of traditional sonority principles that employ a highly reduced conception of slopes leads to consistent cases of misinterpretation of sonority principles, wherein superficially similar qualities are treated as similar regardless of their underlying differences in quantity. For example, the rising slopes in *mlV* and *psV* (see Figure 2.3), are treated similarly in hierarchies such as H_{exp} regardless of their underlying differences in quantity, which are reflected by their different intercepts: the cluster /ml/ has a higher underlying sonority level than the comparable sonority rise in the cluster /ps/. A similar generalization holds for the two plateaus in Figure 2.5 below, in Section 2.2.2.

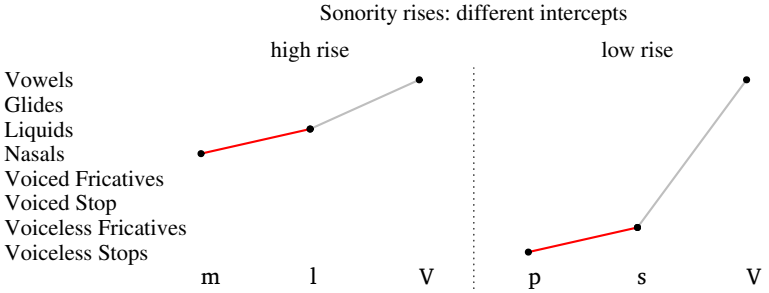

Figure 2.3: Schematic depiction of the sonority slopes of two sonority onset rises, *mlV* and *psV*, with comparable angles but different intercepts. The red solid line which denotes the sonority slope of the onset clusters is higher for *mlV* (left) than for *psV* (right) due to the higher intercept of /m/ compared to /p/.

2.2.2 Inherent failures of traditional sonority principles

2.2.2.1 /s/-stop clusters are well-formed

One rather well-known and well-studied consistent flaw in the empirical coverage of all traditional sonority principles concerns sequences that are often termed */s/-stop clusters*, referring to cases where a sibilant fricative – most often /s/ – precedes a stop consonant, like in the English words ***st****op*, ***sk****y* and ***sp****ort* (see, e.g., Fudge 1969, Goad 2011, Kenstowicz 1994, Olender 2013, Vaux & Wolfe 2009, Wright 2004, Yavaş et al. 2008). The sonority slope of */s/-stop* clusters is either an onset fall or an onset plateau, depending on the given sonority hierarchy (see Figure 2.4). Thus, although */s/-stop* clusters are relatively common in languages that

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

tolerate sequences and should therefore be considered as relatively well-formed (Morelli 2003, Steriade 1999), such clusters are predicted to be rare, or even extremely rare, due to their ill-formed sonority slopes.

As can be seen in the sketches of the syllable *spV*, illustrated here in Figure 2.4 with two different sonority hierarchies, the sonority slopes of the consonantal sequence (red solid line) is either a fall or a plateau depending on the given sonority hierarchy (H_{col} vs. H_{exp} in Table 2.1). The very low intercept of the clusters may serve as an indication that the effect of these ill-formed slopes may be somewhat diminished due to the low amount of underlying sonority. This would make it a case of misinterpretation (i.e. */s/-stop* clusters do not violate sonority principles) due to over-application of sonority slopes, implying that sonority has a limited explanatory contribution to the phonotactics of */s/-stop* clusters.

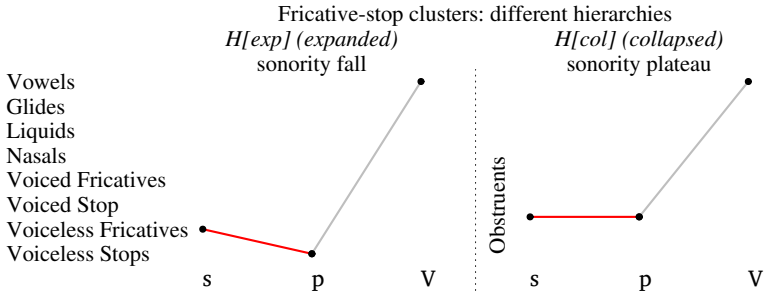

Figure 2.4: Schematic depiction of the two potential sonority slopes of the */s/-stop* cluster *spV*. The red solid line that denotes the sonority slope of the consonantal clusters is falling when applied with the expanded sonority hierarchy H_{exp} (left), and it is a plateau when applied with the collapsed sonority hierarchy H_{col} (right).

Rather than redefining sonority principles to be able to account for the phenomenon of */s/-stop* clusters, more successful attempts to solve this problem in the phonological literature redefined deviant marginal sibilants as exceptional, keeping the traditional sonority principles unaffected by their consistent failure to predict the attested relative well-formedness of */s/-stop* clusters. The main type of exception that is used to explain sibilant-initial clusters is based on tweaking symbolic representations by removing the symbol of the marginal sibilant segment outside of the syllable that contains the following consonant such that – in theory – there is no tautosyllabic complex cluster to trigger sonority restrictions in the first place (see, e.g., Steriade 1982, Kaye 1992, Rialland 1994, Vaux & Wolfe 2009). A slightly different theoretical solution with similar results is to assert that */s/-stop* clusters are, in fact, a single complex segment (see, e.g., Fudge 1969,

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

van de Weijer 1996) such that, again, there simply is no cluster to account for (for an overview, see Goad 2016).

Those theoretical tweaks are not without merit as they follow a strong intuition that marginal sibilants are somehow "outside" the scope of syllabic processes. This intuition is supported by evidence of some unique behaviors of marginal sibilants, such as the kinematic data presented in Hermes et al. (2013), which finds unique coordination patterns in the articulation of sibilant-initial consonantal gestures in Italian. That said, it is important to remember that the problem with */s/-stop* clusters is not that they are common, and, at the same time, unique. The problem with */s/-stop* clusters is that traditional sonority principles fail to account for them in consistent manners, without resorting to exceptions.

I return to this point in Section 10.1 in proposing a more holistic account of */s/-stop* clusters that illustrates the division of labor between sonority and other phonotactic principles.

2.2.2.2 Not all plateaus are equal

A second problem that has received far less attention in the literature (but see Baroni 2014) is the general failure of traditional sonority principles to correctly account for sonority plateaus. This case is perhaps less prominent as it is the *absence* of some plateau types that serves as the main evidence. Sonority plateaus can result from any combination of consonants of the same class, regardless of which class. Thus, a voiceless fricative plateau such as *sfV*, like in the English word ***sph****ere*, should be exactly as ill-/ well-formed as a nasal plateau, such as *nmV* (see Figure 2.5), which is, in fact, a much less common (more *marked*) cluster among the languages of the world (Greenberg 1978, Kreitman 2008, Lindblom 1983). This problem can be, again, attributed to the lack of an intercept in traditional sonority models.

Different sonority plateaus have the same flat line in terms of sonority slopes, yet they differ in their apparent distribution. This difference seems to be linked to the different intercepts of the plateaus A plateau with a low-sonority intercept like *sfV* is better-formed, given that it is cross-linguistically more common, than a plateau with a higher sonority intercept like *nmV*.

Note that the critique regarding the lack of intercepts in traditional sonority principles is given from within a discrete symbolic framework, where non-overlapping segments and their associated sonority values are interpolated into slopes in symbolic time. This is only the first step towards a more radically different treatment of sonority with continuous entities and dynamic procedures, which I will propose in Chapter 5.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

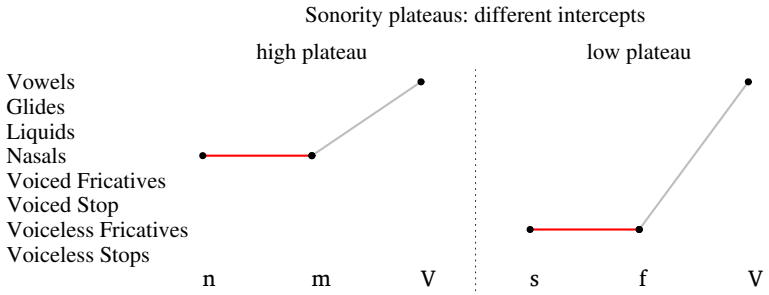

Figure 2.5: Schematic depiction of the sonority slopes of two sonority plateaus, *nmV* and *sfV*. The red solid line which denotes the sonority slope of the onset clusters is higher for *nmV* (left) than for *sfV* (right) due to the higher intercept of /n/ compared to /s/. This difference is not accounted for by traditional sequencing principles.

2.2.3 Sonority "correlusions"

Although no strong consensus has ever been reached with respect to the phonetic basis of sonority, acoustic *intensity* is perhaps the most widely assumed correlate of linguistic sonority. This is evident from the many influential studies on sonority that consider acoustic intensity as its phonetic correlate (e.g., Sievers 1893, Blevins 1995, Clements 1990, Heffner 1969, Ladefoged 1975, Parker 2008, and Gordon et al. 2012, to name just a few prominent examples).[3]

The main problem with intensity-based accounts is related to the distinction between *causation* and *correlation*. It is possible to find acoustic markers that correlate with some linguistic phenomenon. A discovery of this kind is valuable, but to advance our knowledge further we would also need to know if the correlation between the linguistic phenomenon and our acoustic marker of choice can be characterized in terms of causation. Establishing causation from acoustic signals necessitates a theory that can reliably map acoustic markers to consistent operations or processes in sensorimotor speech articulation and/or auditory speech perception. The problem with accounts that are based on acoustic intensity is that the general acoustic intensity of the signal does not consistently map to any aspect of human auditory perception, not even perceived loudness.

[3]In his overview of existing literature, Parker (2002) found close to 100 different proposals for correlates of sonority, and he tested five leading proposals in laboratory conditions: *intensity*, *intraoral air pressure*, *F_1 frequency*, *total air flow*, and *duration*. In his study, the tightest correlations with sonority classes were obtained for acoustic intensity measurements, a conclusion that was repeated and elaborated upon in Parker (2008).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

2.2.3.1 Acoustic intensity ≠ perceived loudness

The acoustic signal has certain physical qualities contributing to its overall power, but they have different effects on the perceptual system of the human hearer. This discrepancy between acoustic intensity and perceived loudness is a well-known problem, playing a role at different dimensions of the mapping between acoustics and perception. The prominent points of departure between acoustic intensity and perceived loudness include the following: (i) loudness perception differs for sine waves with the same intensity level at different frequencies (e.g. Fletcher & Munson 1933, Plack & Carlyon 1995, Suzuki & Takeshima 2004); (ii) loudness perception differs for comparable sounds at different durations (e.g. Turk & Sawusch 1996, Moore 2013: 143, Olsen et al. 2010, Seshadri & Yegnanarayana 2009); and (iii) loudness perception differs for otherwise comparable periodic (harmonic) vs. aperiodic (noise) sounds, and noise, like sine waves, is not uniformly loud across the frequency spectrum (e.g. Hellman 1972, Bao & Panahi 2010, Moore 2013: 140). Acoustic intensity is therefore a physical description of sound waves in space which does not consistently relate to how loud we perceive these sounds, or to any other perceptual phenomenon for that matter.

2.2.3.2 Loudness is not a good candidate for sonority

Note also that the relevance of perceived loudness to syllabic organization requires some sort of functional explanation, which seems to be lacking. The systematic differences in intensity of adjacent speech sounds imply that these differences are neutralized in perception, as it should make sense to assume that the different sounds that compose coherent speech are perceived as having comparable loudness. The literature on perceived loudness supports this assumption given that speech portions with relatively low acoustic intensity, like voiceless fricatives, appear in speech next to portions with relatively high acoustic intensity, like vowels. Our auditory system perceives the aperiodic high-mid frequencies of many obstruents as exceptionally loud compared to the periodic low-mid frequency ranges of vowel sounds, thus compensating in perception for physical differences in acoustic intensity.

Given the above, we should anticipate that perceived loudness will not be a good candidate for the acoustic correlate of sonority hierarchies, as a measure of perceived loudness would bring all speech sounds closer together by diminishing the distinctions provided by acoustic intensity. Indeed, although good approximations of perceived loudness from acoustic signals are available (e.g. Seshadri & Yegnanarayana 2009, ITU-R 2015, Lund & Skovenborg 2014, Skovenborg 2012), I am not aware of any attempts to employ such measures for sonority.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Instead of attempts to map acoustic intensity to perception in terms of perceived loudness, most successful endeavors that use intensity-based measures as correlates of sonority do so by essentially enhancing the intensity-loudness discrepancy. Certain frequency bands are targeted to – roughly speaking – discriminate against energy at the higher frequencies (which are more characteristic of obstruents) in favor of energy at low-mid ranges of the spectrum (which are more characteristic of sonorants and particularly vowels).[4] The relative success of such metrics is not commonly based on perceptual grounds. However, they are often tightly linked to the perceptual quality that is identified with sonority in this work – the capacity to perceive pitch.

[4]For example, Pfitzinger et al. (1996), Port et al. (1996), Fant et al. (2000), Galves et al. (2002), Wang & Narayanan (2007), Tilsen & Arvaniti (2013), Patha et al. (2016), Nakajima et al. (2017), and Räsänen et al. (2018).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

3 Linguistic models: Between symbolic discreteness and dynamic continuity

The study of the sound system of human languages has been one of the longest-standing intersections of symbol-based categorical analyses on the one hand, and signal-based continuous descriptions on the other. These two different types of analysis stand at the core of the distinction many researchers make between *phonetics* and *phonology*, where the former addresses continuous and measurable aspects of the speech signal (namely, sensorimotor aspects of articulation, acoustic signals and neurological patterns in perception), while the latter addresses categorical aspects of the speech signal using discrete and symbolic units like *consonants*, *vowels* and *phonemes* (see overviews in Harris 2007, Ladd 2011).

3.1 Thesis and antithesis: Problems with symbol-based models

The incompatibility between the continuous and the discrete types of description did not escape early studies. Menzerath & Lacerda (1933), Wickelgren (1969) and Fowler (1980) noted how the reality of co-articulation of segments defies the idealized conception of speech signals as consisting of a sequence of non-overlapping discrete phonemes. Warren (1982: 172–187) also provides an overview of this problem given the limitations of auditory perception.

Morris Halle acknowledged the problem of discrete descriptions already in Halle (1954), which was written in defense of phonemes (p. 198): "It is now necessary for us to show why the discrete picture of language is preferable. Our answer is that it enables us to account for many facts which on the assumption of continuity would be extremely difficult, if not impossible, to explain". Halle wrote this a decade before he published *The sound pattern of English* with Noam Chomsky (Chomsky & Halle 1968), perhaps the most influential work in phonology from the second half of the twentieth century, in which speech sounds are modeled as discrete symbolic units and phonological processes are modeled as rules that manipulate symbols.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Beyond phonology and linguistics, many successful enterprises in the cognitive sciences of the second half of the twentieth century likened cognitive capacities to symbol-processing machines. In that context, John Searle provided some well-known attacks on common notions of *Artificial Intelligence* (Searle 1980) and the computer metaphor of the mind (Searle 1990), responding, among others, to prominent voices like Fodor (1983), Pylyshyn (1985) and Cummins (1985).

The *connectionist* program in cognitive psychology (e.g. Rumelhart et al. 1986a, Rumelhart et al. 1986b, Bates & Elman 2002) was set to change that classic view with the introduction of connectionist models to phonology (e.g. Goldsmith 1992, Joanisse 2000, Laks 1995, Smolensky & Legendre 2006, Tupper & Fry 2012). These models replaced classic symbolic models with *neuromimetic* models (Laks 1995: 52) that attempt to improve the cognitive plausibility of language models with architectures that resemble neurobiological systems. They were faced with fierce opposition from voices like Fodor & Pylyshyn (1988) and Pinker & Prince (1988), who criticized the connectionist models of the time for lacking a symbolic level of representation.

It should be noted, however, that the main focus of connectionist models is not so much on the symbols in the system as on the classic processes of symbol manipulation. Connectionist models present alternatives to the notion of *rules* of symbol manipulation that directly transform symbols (see Harnad 1990). Indeed, as Smolensky & Legendre (2006) point out, (some versions of) connectionist models in phonology are largely compatible with *Harmonic Grammar*, which describes the learning processes in a constraint-based system like *Optimality Theory* (Prince & Smolensky 2004). In that sense, connectionism is like a low-level description for which Optimality Theory provides the high-level description.

The problem with all the models that Smolensky & Legendre (2006) mention is that they still take discrete symbols in their input in order to generate discrete symbols in their output. As Gafos (2006: 57) points out in the context of modeling variation, these models deal with "variation among discrete alternatives", without accounting directly for the continuous aspects of the system. For different reasons, related to the architecture of neural networks in connectionism, Harnad (1990: 337) even suggested that it may be reasonable to consider connectionism as "a special family of symbolic algorithms".

In the context of the present work, connectionist models can be effective in modeling the phonology of speech perception in a top-down model, which represents processes that start and end with discrete symbols. However, I assume here that there is a functional source for linguistic distinctions in perception, which has to be accounted for via the bottom-up route, originating from continuous events in real time.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

3.2 Synthesis: Integrating dynamic and symbolic notions

Port & Van Gelder (1995), Kelso (1997) and Spivey (2007) have presented a comprehensive refutation of the computer metaphor of the mind at the turn of the century, relying primarily on the many advancements achieved with dynamical systems models of cognition. Dynamical systems have been also successfully implemented in phonology, underlying the enterprise known as *Articulatory Phonology* (see, e.g., Browman & Goldstein 1992, Goldstein et al. 2009, Nam et al. 2009). This approach models phonology with continuous motor gestures, whereby coordinative structures can be understood in terms of the coupling and decoupling of oscillations with respect to syllabic organization (see Haken et al. 1985 for early incarnations of these models).

Within dynamical systems models it is possible to integrate continuous aspects of the speech system (articulatory trajectories and the related output on the acoustic surface) with the discrete symbolic categories that linguists postulate. One way to achieve this is via attractor landscape models, where discrete categorical units can be modeled as stable states in a continuous phase space in terms of attractor basins (see Haken 1990). In this type of model, various continuous events can contribute more or less to the (partial) activation of different, often competing, attractors. Convincing examples for the application of such attractor landscapes can be found in Tuller et al. (1994), Case et al. (1995), Rączaszek et al. (1999), Gafos & Benus (2006), Roessig & Mücke (2019) and Roessig et al. (2019).

In fact, using attractor landscape models shows not only that discrete alternatives can be selected by continuous events, but that attractor landscapes can also be advantageous in modeling categories. This is especially true for models that embody a more nuanced understanding of the nature of discrete categories in responding to multiple – often redundant – cues and exhibiting *fuzzy* category boundaries that can be readily accounted for in terms of *noise* in the system (see Roessig et al. 2019: 8–9).

Attractor landscapes are therefore an essential component of models concerned with the interaction between continuous and discrete entities in a language system. However, much like the connectionist models in phonology, attractor landscapes explicate the process by which a discrete alternative can be (partially) activated, but they have little to say about the components of the system otherwise. For example, attractor models cannot explain or predict the shape and behavior of the attractor landscape itself (e.g. universal and idiosyncratic language categories), they cannot address the limitations on dynamic events that the system can reliably detect (e.g. selecting the relevant effects in auditory perception), and they pose no restrictions on what a valid symbol is in a natural

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

language system (what Harnad 1990 called "the symbol grounding problem"). In other words, attractor-based models are good at integrating continuous variables with discrete alternatives, but they are not designed to reveal what drives and limits the dynamic and symbolic modes of the system.

3.3 Dynamics in perceptual phonology

Dynamic descriptions have played an increasingly important role in phonological theory with the growing body of work from the school of *Articulatory Phonology*, which suggests a framework for modeling phonological systems with continuous articulatory gestures as the basic units of speech production. Applying similar concepts to perception has been thus far a much less productive avenue in phonology, perhaps because it is much less clear what the relevant continuous entities are that need to be modeled in perception.

Dynamic accounts of phonological perception have been presented in works like Tuller et al. (1994), Case et al. (1995), Hock et al. (2003), Tuller (2004), Tuller et al. (2008), and Lancia & Winter (2013). However, they mostly deal with categorical perception of systematically varying speech stimuli without breaking down the speech input into subcomponents, as is the case in Articulatory Phonology, whereby different moving parts within the vocal tract (e.g. tongue tip and lips) are the continuous subcomponents of the speech signal.[1]

In that sense, the vast majority of dynamic perception accounts that I am aware of cover the same aspects as the attractor landscape models mentioned above (and indeed, attractor landscape models are common in dynamic perception studies): providing a unified model for the integration of continuous and discrete entities in cognition.

3.4 Making sense: Symbols and dynamics in Howard Pattee's work

Cariani (2001) and Rączaszek-Leonardi & Kelso (2008) note the writings of physicist and theoretical biologist, Howard Pattee, as a potential source of novelty in cognitive modeling (see Pattee & Raczaszek-Leonardi 2012 for a collection of Pattee's classic papers with contemporary commentary). For Pattee, the symbolic and dynamic modes of biological systems are two crucial components with

[1]Note the work in Liberman & Mattingly (1985), where perception is conceived of as continuous, albeit in a manner that is contingent on production ("the motor theory of speech perception").

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

specific roles to play: symbols are the stable forms that harness dynamic events. Symbols, according to Pattee (1987: 337), cannot exist outside of a dynamic system that they constrain.

Bear in mind that these descriptions were initially laid out to investigate biological systems in which DNA appears to be the symbolic mode that constrains the dynamics of the cell. However, as Pattee and his followers have been arguing in recent decades, this description can be extended to any *language* system. In that context, it is perhaps useful to elucidate the difference in Pattee's thought between a *language* and a *code*.

Joanna Rączaszek-Leonardi clarifies the difference between language and code in Pattee's work (Pattee & Raczaszek-Leonardi 2012: 307–310) and emphasizes the centrality of the idea that language systems are characterized by symbols that *harness* or *constrain* dynamics. The effects of *constraining*, rather than *mapping*, are, in her words, "naturally context-dependent (crucially relying on the dynamics being constrained), thus are predictable only to some degree". In contrast, coding is a relatively noise-free process in which we "map one symbolic structure onto another symbolic structure" (p. 309). "In natural language", Rączaszek-Leonardi suggests as an example, "writing is a code for spoken expressions, but it is the spoken expressions that are the level at which meaning relation should be sought". In Pattee's original analysis, this meant that the DNA bases *code* for the amino acids, while it is "the folded amino acid sequence (the protein enzyme) where the first informational constraint on dynamics occurs" (p. 310).

What makes symbols meaningful in a language system is therefore "a relation in which a symbolic structure acts to harness dynamics" (p. 309). Symbols in language systems acquire their meanings from the co-occurrence with dynamic events, i.e. via *grounding* (and later *ungrounding*; see Rączaszek-Leonardi et al. 2018). Symbols can be "coded in another set of symbols, perhaps for a better adaptation to a given transmission medium (e.g., the Morse code is better adapted to a telegraph than the alphabet) but it does not make them more, or less meaningful. A code is not a language" (Pattee & Raczaszek-Leonardi 2012: 309).

3.5 The complementarity of mind

Pattee's specific conception of language systems entails a few interesting outcomes. One of the most important ones in the context of the current work is the idea that the two modes of language – the discrete/symbolic on the one hand, and the continuous/dynamic on the other hand – require two separate complementary models. Pattee summarizes this in Pattee & Raczaszek-Leonardi (2012: 18–21):

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

> Complementary models as I define them are models of a system that may be *formally* incompatible in the sense that no one model is logically or mathematically derivable from, or reducible to the others, and all such models are necessary for a complete understanding of the system. (pp. 18-19)

Pattee is careful not to imply ontological dualism, as he means that complementarity is an "epistemic necessity", although he still finds it difficult to assume conceptual compatibility given that "conceptual categories such as 'discrete' and 'continuous' derive from different pattern recognizing regions of the brain" (p. 19). Why it is so hard to see this picture more clearly is suggested to be related to inherent limitations on what "our classical brains can model":

> The complementarity of discrete and continuous models is a fundamental aspect of the symbol-matter problem. Evolution prepared the simplest brains to distinguish discrete objects from the continuous motion of objects, thereby allowing effective sensorimotor control. Our everyday experience as well as classical physics is based on a clear and objective distinction between discrete particles and continuous motion. In modern physics, however, this clear distinction is no longer possible. Discrete particles and continuous fields, matter and energy, space and time are no longer objectively separable but depend on how we observe nature. It appears that our artificial instruments have extended our senses beyond what our classical brains can model without cognitive dissonance. It is not clear how far we can reduce this dissonance by learning new concepts. (pp. 20-21)

It is therefore pertinent to understand symbols in language systems with respect to the continuous events that they relate to. In the present work, this means that to fully understand sonority we need to address its potential functions in auditory perception and cognition, and their effects in linguistic communication. Although tightly related to top-down symbol-based generalizations, this bottom-up route is a separate process that is based on different driving forces (e.g. laws of physics rather than statistics).

3.6 Missing links: Anticipating current contributions

In an attempt to break new ground, the current project proposes two theoretical novelties that are still missing from the picture described above in Sections 3.1–3.4: (i) decomposing speech into continuous subcomponents in the acoustic signal that allow us to extend our dynamic vocabulary with perception-based entities; and (ii) suggesting a principled way to separately model continuous and

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

symbolic aspects of speech. To elaborate on (i), I present the PRiORS framework in Chapter 4. PRiORS stands for *Perceptual Regimes of Repetitive Sound*, essentially targeting a single primitive – *repetition* – at different timescales that activate two distinct *regimes* in perception. To elaborate on (ii), I present two complementary models for a single phonological principle in Chapter 6. These models are built around the distinction between *bottom-up* and *top-down* processes in cognition of speech sounds (see, e.g., Klatt 1979, Fowler 1986). The bottom-up route expects dynamic and continuous inputs, while the top-down route in perception arrives at inferences via the learned categorical and symbolic constructs of the system. The top-down route requires symbolic models that are based on the distributional history of categorizable speech sounds, while the bottom-up route requires a model that can deal with continuous entities that must obey the laws of physics and fit with the capabilities of human perception and cognition. Crucially, these two models are irreducible into one and they explicitly attempt to model two complementary aspects of cognition, considering both *bottom-up* and *top-down* inferences in perception of speech.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Part II

Novel theoretical outlooks

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

4 Perceptual regimes of repetitive sound (PRiORS)

One of the crucial components that this project suggests is a general framework for modeling the dynamic mode of auditory perception and cognition: *Perceptual Regimes of Repetitive Sound*, abbreviated as **PRiORS**. This framework is used here to account for phonological phenomena, i.e. it is used to account for some cognitive aspects of auditory perception that are manifested in linguistic systems. PRiORS reduces the rich acoustic signal to a single primitive, based solely on the notion of repetition, to account for auditory perception in terms of *temporal integration*. Crucially, a major role is played by the *rate* of repetitions, but, as PRiORS makes clear, this single quantitative modulation in rate of repetition has two qualitatively distinct effects in perception and cognition.

4.1 Time and frequency dualism

4.1.1 Time and frequency domains in mathematical representations

Very often we are interested in representing and analyzing the course of certain observed phenomena over time. This relates to physical signals, mathematical functions and any time series of data such as economic and environmental developments. These *time domain* representations are relatively straightforward as they follow the progression of a single variable. Typically, time is plotted on the x-axis, showing progression in time from left to right, while the y-axis is used to plot the magnitude of the observed variable. Time domain representation are therefore very informative with regards to the change in magnitude (or power) over time of a single variable.

Time domain representations apply well to acoustic signals, where we can observe the progression of acoustic power over time to represent and analyze sounds. In the simplest of cases, we can observe a single sine wave like the one in the oscillogram in Figure 4.1. Here, a time domain representation is very informative with respect to the signal it depicts. The power, the (fundamental)

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

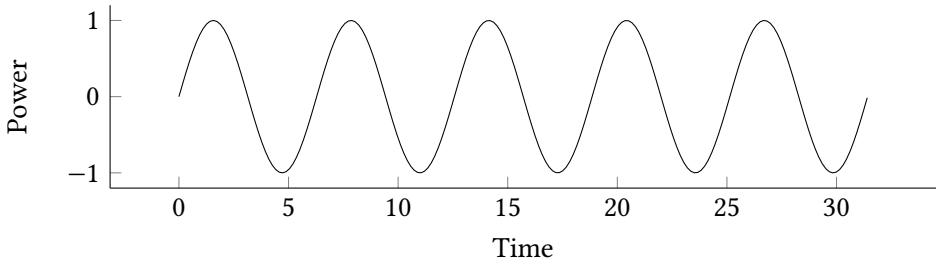

Figure 4.1: An arbitrary sine wave in a time domain *oscillogram* representation.

frequency and the duration of the signal can all be readily deduced from this representation. However, most natural sounds involve a much more complex distribution of acoustic power over different frequencies at different durations, many of them overlapping in time. A time domain representation of complex natural sounds can effectively represent the overall impact of the many subcomponents on the amount of acoustic power at different points in time. It is, however, not very well suited to being informative with respect to any of the subcomponents of complex sound. Their individual frequencies, power and durations are all bundled together when we measure acoustic power over time.

We can decompose the complex signal into its subcomponents, provided that the subcomponents can be described as periodic signals, just like the simplest non-decomposable sine wave that characterizes acoustic signals. Doing that requires a switch from the time domain to the *frequency domain*: rather than looking at the distribution of acoustic power over time, we can look at the distribution of acoustic power over different frequencies. A frequency domain representation allows us to observe the many subcomponents in terms of their frequency and power at given points in time.

Switching from time domain to frequency domain is possible with the *Fourier transform*, a mathematical transform named after the French mathematician Jean-Baptiste-Joseph Fourier, who introduced it in Fourier (1822). The Fourier transform decomposes a time series into a sum of finite series of sine or cosine functions. Note that in many contexts of acoustic analysis, the procedure is referred to as *fast Fourier transform* (FFT), which is the name given to a wide range of algorithms that perform quick calculations of the Fourier transform (see overview in Brigham 1988).

The time and frequency domains are two complementary, and, to some extent, redundant mathematical representations of the same physical reality. By representing time as a scale of different potential rates, we can shift the representation

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

of events in time into a representation of the co-activation of different rates in the frequency domain. It is perhaps useful to imagine the frequency domain as an interpretation of the time domain, whereby time is converted to one of its main effects – the time-related impression of *speed* (or *rate*). Crucially, note that both domains are available for representation of physical events, regardless of timescales.

4.1.2 Time and place theories in models of pitch perception

Models of pitch perception attempt to explain how the auditory system resolves the harmonic structure of complex signals into the sensation of pitch. Two types of theories have traditionally dominated this field. *Place* theories of pitch are based on the idea that different frequencies in the signal can excite different places along the basilar membrane. *Time* (or *temporal*) theories of pitch are based on the idea that neural firing rates exhibit sensitivity in time to periodic events, allowing them to *phase-lock* to the rate of periodicity in time.

Place theories of pitch often date back to Ohm (1843) and Helmholtz (1863), and they are in line with the notion of the frequency domain (Ohm 1843 in fact assumed that a Fourier analysis took place in the auditory system). Time theories of pitch often date back to Seebeck (1841), Wever & Bray (1930) and Schouten (1938), and they are likewise in line with time theories of pitch. De Cheveigné (2005) traced back the early roots of these two pitch theories to ancient Greece, linking writings from Pythagoras (6th century BCE) and Aristoxenos (4th century BCE) with place theories of pitch, and linking the writings of Greek mathematician Nicomachus (2nd century CE) with time theories of pitch.

Place and time theories are sometimes presented as two dueling narratives, highlighting historical differences between Seebeck (1841) and Ohm (1843), and later between Helmholtz (1863) and Schouten (1938) and many others. However, of consequence to this work is the more currently common appreciation that these two theories of pitch are, in fact, complementary and even desirably redundant to various extents (see House 1990, De Cheveigné 2005, Houtsma 1995, Oxenham 2013).

Warren & Bashford (1981) mention the findings that frequency selectivity along the basilar membrane, which is essential for place-related pitch resolution, is roughly limited to 50–16k Hz (von Békésy & Wever 1960), while phase-locking of the auditory nerve fibers' firing rate, which is essential for time-related pitch resolution, has a typical upper limit of about 5k Hz (Rose et al. 1967; see a more recent review on this topic in Verschooten et al. 2019). Warren & Bashford (1981)

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

then make the point that the range in which both place and time models overlap, around 50–5k Hz is exactly the range for optimal perception of musical pitch.

4.2 The spectral and temporal regimes of auditory perception

Auditory perception according to PRiORS is dramatically affected by two different responses to repetitions at two distinct timescales that constitute two perceptual regimes: the *temporal regime* and the *spectral regime*. The temporal regime designates the timescale at which humans can perceive successive acoustic events as isolated events in time, while exhibiting a relatively good ability to predict the upcoming event with a given steady rate, or perceive and estimate durations, in the related terminology of Fraisse (1984). These upper and lower limits of perception were described in MacDougall (1903: 321) as conditions for "the impression of rhythm", and, indeed, the timescale of the temporal regime defines the range at which musical rhythmic patterns tend to occur and auditory sensitivity to changes in rate allows us to reliably detect temporal patterns (e.g. Fraisse 1984, Farbood et al. 2013, Repp 2005). In other words, the temporal regime covers our ability to perceive the difference between *fast* and *slow*.

The spectral regime, in contrast, operates at a faster timescale, where a sequence of acoustic events repeats too fast to be perceived as isolated events (see, e.g, Miller & Taylor 1948, Broadbent & Ladefoged 1959, Efron 1973, Hirsh 1959). At these fast rates, rather than perceiving temporal intervals as occurring at non-overlapping moments in time, cognition switches to perceiving them as spectral intervals that occur at the same time, giving rise to the perception of a complex harmonic tone, whereby the harmonic partials reflect the rate of repetition in the spectral dimension (Flanagan & Guttman 1960, Stockhausen & Cardew 1959, Warren 1982). Within the spectral regime, changes in rate of repetition of periodic signals are perceived as changes in pitch. In other words, the spectral regime covers our ability to perceive the difference between *high* and *low* (pitch).[1]

In principle, the temporal regime is congruent with the notion of *time domain* and the spectral regime is congruent with the notion of *frequency domain*

[1]The notion of repetition is not limited to rate-based distinctions, but this work requires only this type of distinction to cover prosodic phenomena. To cover the acoustic qualities of segmental phenomena in speech we can analyze the lack of regular repetition at the spectral regime in at least two meaningful ways: (i) transient bursts are characteristic of speech sounds like *stops* that exhibit a non-repeating signal (or more accurately, a critically damped signal); (ii) continuous yet irregular (asynchronous) repetitions within the timescale of the spectral regime are characteristic of *aperiodic* noise that results from articulatory friction (e.g. *fricatives*).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

(see Section 4.1.1). Likewise, the two regimes correspond to the two independent anatomical and neurological processes for pitch resolution that are also based on time and frequency representations (see Section 4.1.2). However, while the time and frequency domains can independently describe the same event in mathematical terms, and while place and time theories of pitch perception may be, to a large extent, complementary and redundant, the perceptual regimes in PRiORS operate at two slightly overlapping yet mostly distinct and mutually exclusive timescales. Each perceptual regime responds to repetitions within its timescale such that a single quantitative modulation of the rate of repetition gives rise to qualitative differences in the sensation of rhythm (temporal regime), or, alternatively, the sensation of pitch (spectral regime). In other words, the temporal and spectral regimes suggest a perceptual qualia perspective, whereby the two regimes are mutually exclusive.

This last point is reminiscent of Pattee's quote from from Pattee & Raczaszek-Leonardi (2012: 21), which was given in Section 3.5, and is partially repeated here: "It appears that our artificial instruments have extended our senses beyond what our classical brains can model without cognitive dissonance." In that sense, the time and frequency domains in physics are revealed by our instruments as two overlapping representations, but in our mind, the temporal and spectral regimes are represented as two distinct and separate sensations. To be clear, auditory perception can handle the two regimes at the same time, i.e. process information with both rhythmic and periodic effects. The mutual exclusivity is in response to isolated events within their relevant timescale.

When evaluating PRiORS, it is useful to acknowledge how the auditory system is uniquely adapted to capturing reoccurrence in terms of repetition. Chowning (2001) suggests that the auditory system is far more sensitive than the visual system to differences in reoccurring structures. He reflects this ability in the visual system in terms of the capacity to detect minute differences in the spacing between otherwise identical objects (think of a typical *spot the difference* puzzle as an example), essentially equating quasi-periodicity with quasi-symmetry. Chowning (2001: 267) claims that the auditory system, unlike the visual system, "can readily detect a fraction of a percent of deviation from periodicity". This great sensitivity to repetition is linked in PRiORS to perceptual and cognitive aspects of an auditory system that is specialized in processing rhythm and pitch as the two modes of auditory temporal integration.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

4.3 Visual FFT-based simulations

It is useful to illustrate the distinction between perceptual regimes with a *Band-Limited Impulse Train* (BLIT) synthesis that produces a train of transient acoustic bursts at adjustable rates. Each burst is a single *impulse*, which is the shortest electric burst a given system can produce, with equal power across the frequency scale (a perfect impulse has acoustic power over an infinite frequency range, but the impulses in a BLIT, as the name suggests, are band-limited to human hearing ranges, between approx. 20–20k Hz).[2] The BLIT signal can be effectively visualized with standard FFT-based tools that convert signals between time domain and frequency domain representations (see Section 4.1)

Table 4.1 presents a rough sketch of the relevant timescales of the two perceptual regimes. Within each regime the effects of repetition are named differently in order to maintain a distinction that attempts to be in-line with most common uses of these terms: it is *Rhythm* when occurring within the timescale of the temporal regime vs. *Periodicity* when occurring within the timescale of the spectral regime. Table 4.1 also shows that these repetition-induced effects have boundaries. Repetitions within the temporal regime may be too slow to be perceived as rhythmic (*infra-rhythmic* perception below 30 BPM; see Fraisse 1984, Farbood et al. 2013, and see overview of tapping literature in Repp 2005). Likewise, repetitions within the spectral regime may be too fast to be perceived as periodic (*ultra-periodic* perception above 5k Hz, given that our auditory system can typically perceive frequencies up to 20k Hz, but our ability to sense discernible pitches does not typically exceed 5k Hz; see Ward 1954, Attneave & Olson 1971).

Four examples are provided in Figure 4.2, each one with three corresponding visual panels. The bottom white panel presents a 1-second long waveform (*oscillogram*) which shows the unipolar transient bursts produced by the BLIT synthesis in the time domain, going from left to right. The number of visible bursts within this 1-second interval corresponds to the rate of the BLIT in Hz. The two upper dark panels show FFT-based analyses exhibiting the dispersion of acoustic power across the audible frequency range in the frequency domain. The middle panel, often called a *spectrograph*, exhibits a 2-dimensional representation of frequency (x-axis) and power (y-axis), while the top panel, which is typically called a *spectrogram*, exhibits a 3-dimensional representation of frequency (x-axis), power

[2] One major advantage of the BLIT synthesis concerns the minimal duration of impulses that allows the simulation to reduce confounding factors regarding burst duration. Longer bursts are expected to appear as more continuous at slower rates than comparable impulses because they fill a longer portion of the intervals between onsets of recurring events (i.e. they have a longer *tail*).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table 4.1: Perceptual regimes with corresponding effects and timescales (rough sketch). *Note.* Hz = Hertz (repetitions per second); BPM = Beat Per Minute; ms = millisecond (duration of repeating intervals).

		Timescales		
Perceptual regimes	Effects	Hz	BPM	ms
Temporal	Infra-rhythmic	0–0.5	0–30	∞–2k
	Rhythm	0.5–20	30–1200	2k–50
Spectral	Periodicity	20–5k	1200–300k	50–0.2
	Ultra-periodic	5k–20k	300k–1200k	0.2–0.05

(color) and time (y-axis). The frequency x-axes of the spectrograph and the spectrogram are perfectly aligned to facilitate the interpretation of the spectrograph in the middle as a "slice", or a "still image" of the temporal representation in the spectrogram above it.

Figure 4.2(a) (top left) shows a clear rhythmic effect at 4 Hz, indicated by four bursts in the bottom oscillogram panel. A single burst appears with equal power along the (band-limited) frequency range in the spectrograph, indicated by the fairly straight horizontal green line across the middle panel. Note that the still image shown here captured a moment in time in which the power graph of the spectrograph was high. With rhythmic bursts, like the 4 Hz BLIT in Figure 4.2(a), this graph goes visibly up and down over time. Above it, a succession of 10 impulses over a short period of time (about 2.5 seconds) is visible as isolated bursts, indicated by the horizontal lines going from bottom to top in the corresponding upper spectrogram panel.

In sharp contrast, Figure 4.2(d) (bottom right) clearly shows tonal behavior at 120 Hz. There are, indeed, 120 bursts in the time-domain display of the bottom oscillogram panel, but the isolated bursts are no longer visible in the top spectrogram panel, i.e., there are no horizontal lines going from bottom to top across the upper panel (note that with the 2.5 second-long window of the spectrogram, 120 bursts per second should have resulted in 300 horizontal lines by comparison to Figure 4.2(a)). The sensation of isolated discrete bursts transitions into a sensation of continuous sound in perception at these higher rates of repetition. This can be thought of as a smearing effect that occurs above a certain threshold. This perceptual effect is neatly reflected by the two FFT-based representations in Figure 4.2(d), which display a signal with the properties of a continuous sound that has a complex harmonic structure. The middle spectrograph

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

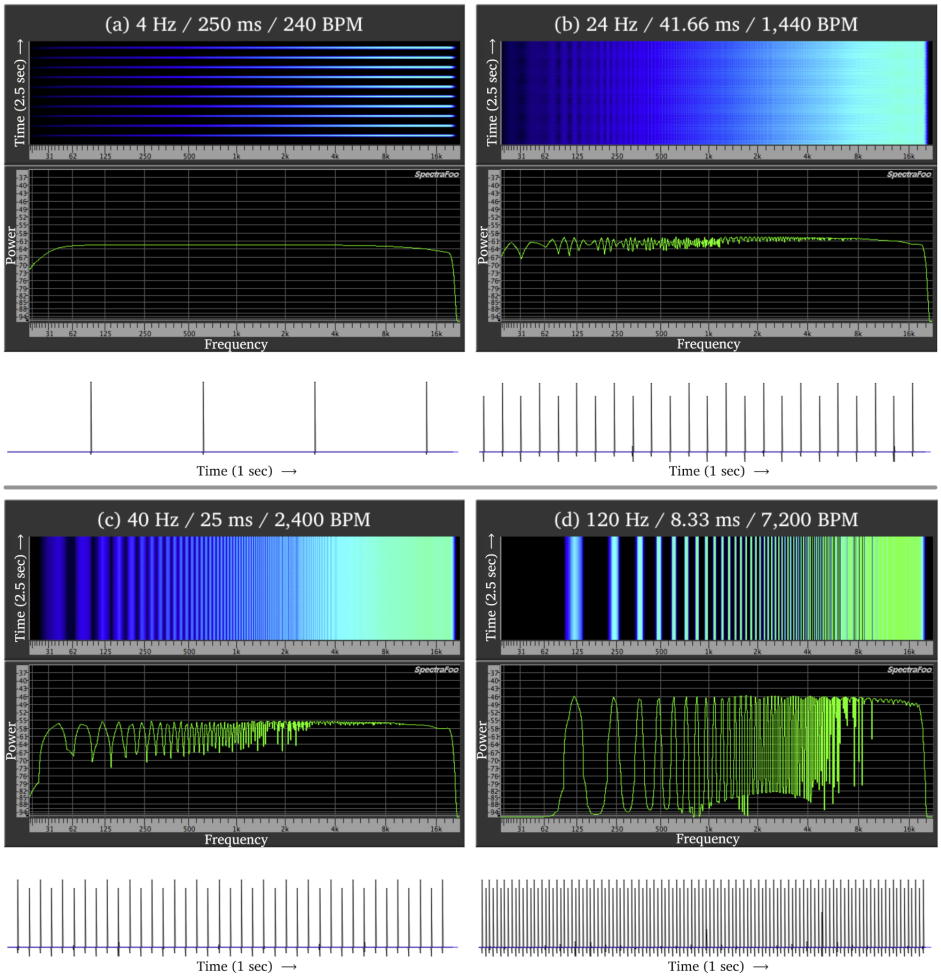

Figure 4.2: Illustration of perceptual regimes with visual analyses of acoustic impulse trains (BLIT) at different rates and different domains (see text for details).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

panel shows a series of "bumps" along the green curve, from left to right, corresponding to a series of continuous energy "poles" in the vertical representations of the upper spectrogram panel. This is a harmonic series in which the rate of repetition of the BLIT synthesis is mapped onto the fundamental frequency (F0) of the continuous sound, which is also manifested in the distance between partials in the harmonic series.[3] Specifically, the two FFT-based representations in Figure 4.2(d) show the harmonic partials in terms of continuous acoustic power at the frequencies 120 Hz, 240 Hz, 360 Hz, 480 Hz, etc. This demonstrates that at this faster timescale of the spectral regime, the sensation of repetition feeds perceptual effects of continuity and pitch, rather than of discreteness and rhythm.

The switch between regimes does not occur at once. Between the temporal and the spectral regime, we can spot a transitional range in which effects of both rhythm and periodicity are present, but neither is strong enough to completely take over. This results in a less definitive effective sensation. Figures 4.2(b–c) demonstrate this transitional range between the two distinct regimes that are illustrated by Figure 4.2(a) (for the temporal regime) and Figure 4.2(d) (for the spectral regime), as detailed above.

Figure 4.2(b) (top right) is especially well-suited for illustrating the indeterminacy of the transitional range. At a BLIT rate of 24 Hz, the impulses seem to be too fast to support a rhythmic perception of discrete bursts, and, at the same time, too slow to support the perception of a continuous harmonic (pitch-bearing) sound. The upper spectrogram panel of Figure 4.2(b) shows a combination of both horizontal lines that reflect isolated events in time, going from bottom to top, as well as vertical lines that reflect the emerging harmonic structure of a continuous complex tone (visible also as corresponding energy fluctuations in the middle spectrograph panel).

To consider the transitional phases between the two regimes, Table 4.2 presents a slightly more elaborate sketch than Table 4.1, with transitional phases at 12–50 Hz. Essentially, this emphasizes the fact that the main effects – rhythm sensation in the temporal regime and periodic sensation in the spectral regime – are optimally achieved closer to the center of each perceptual regime.

Note that the visual effects of the FFT-based representations in Figure 4.2 are calibrated to reflect human perception.[4] The shift from temporal to spectral

[3] The polarity of the impulses can also play a role with pitch perception at higher rates within the spectral regime, as demonstrated in Flanagan & Guttman (1960). For this reason, I used only unipolar impulses for which the relationship between impulse rate and frequency rate is kept stable.

[4] Here I use a commercial metering application, *SpectraFoo* by *Metric Halo* (version 4.2.3), with the default Analyzer depth setting of 4,096 points (10 Hz). The BLIT synthesis and the oscillo-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table 4.2: Rough sketch of perceptual regimes with corresponding effects and timescales (transitions included). *Note.* Hz = Hertz (repetitions per second); BPM = Beat Per Minute; ms = millisecond (duration of repeating intervals).

		Timescales		
Perceptual regimes	Effects	Hz	BPM	ms
Temporal	Infra-rhythmic	0–0.5	0–30	∞–2k
	Rhythm	0.5–12	30–720	2k–83.3
	Ultra-rhythmic	12–20	720–1200	83.3–50
Spectral	Infra-periodic	20–50	1200–3k	50–20
	Periodicity	50–5k	3k–300k	20–0.2
	Ultra-periodic	5k–20k	300k–1200k	0.2–0.05

regimes does not represent a change in any physical quality. Rather, it represents a perceptual threshold of a given system. A different system that responds to higher rates of periodicities, such as, for example, models simulating the auditory system of barn owls (see Köppl 1997), will most probably require higher frequencies to adequately represent the shift from the temporal to the spectral regime. Importantly, higher frequencies will require higher resolutions before the temporal representation becomes too fast and eventually "smears" into the spectral one.

4.4 A note about previous works: Warren and Rosen

The ideas in PRiORS are not entirely new. For one, they are not based on any new data, but on established findings in the literature from the fields of acoustics, auditory perception, neuroscience and linguistics. More specifically, previous proposals were presented in the past for frameworks of perception that are, much like PRiORS, based on delineating the unique contribution of different timescales to auditory perception. I will summarise two of these in the following.

Richard Warren, who studied temporal integration in auditory perception quite extensively, sketched a model with different perceptual effects at different timescales in Warren & Bashford (1981) and Warren (1982: 80–85). Warren & Bashford (1981) determined that 50–5k Hz is the optimal timescale for pitch perception

gram were produced with the sound design software *Plogue Bidule* (version 0.9766).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

("melodic pure pitch"), given that in this range, both place-based and time-based resolutions of pitch are available. At faster rates of 5k–16k Hz, where only place-based pitch resolution may be available, an "amelodic pure pitch" perception takes over. At slower rates, between 20–50 Hz, with only time-based pitch resolution available, the "pure pitch" sensation changes to "noisy pitch". Further down, below 20 Hz, repetitions are considered by Warren & Bashford (1981) as *Infrapitch*, as they are too slow to induce a sensation of pitch. Warren & Bashford (1981) follow Guttman & Julesz (1963) in determining 0.5 Hz as a rough lower floor for perceptual integration of acoustic events. This 0.5 Hz floor of about 2 seconds-long intervals is commonly mentioned as the lower threshold of the human ability to keep isochronous rhythm or temporally integrate events (see Fraisse 1984, Farbood et al. 2013, Repp 2005).

Approaching auditory perception of speech from a more linguistic point of view, Rosen (1992) presented a framework for describing temporal information in speech, which can also be considered as a precursor of the PRiORS framework. In his framework, Rosen (1992) divides perception into three "temporal features" at distinct timescales: *envelope* (2–50 Hz), *periodicity* (50–500 Hz) and *fine-structure* (600–1k Hz). *Envelope* covers mainly "tempo, rhythm" and "syllabicity", while *periodicity* covers mainly "stress", "intonation" and "voicing" (see Rosen 1992: 76, in which other segmental qualities are also covered).

4.5 Advantages of PRiORS

The PRiORS framework is useful for understanding various phenomena in auditory cognition and in phonological systems. It can be useful for models of speech perception that consider the contribution of auditory perception and cognition to language systems, as was detailed in Chapter 3, and especially Section 3.6. The following subsections address two major points that PRiORS can greatly help to elucidate. In Section 4.5.1 I discuss how PRiORS can dispel a lot of the mystery surrounding the phonological notion of the *syllable*, and in Section 4.5.2 I discuss PRiORS' potential for uncovering the different functions that music and speech utilize when they wield the effect of *rhythm* from the timescale of the temporal regime.

4.5.1 Universal aspects of syllabic structure

Syllables are, first and foremost, abstract units of phonological systems and they do not easily lend themselves to consistent and straightforward phonetic explanations in terms of perception and/or articulation. The PRiORS framework can

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

do a lot of heavy lifting in this regard, by providing the baseline conditions that can explain the evolutionary trajectory of syllables. According to this analysis, syllables were shaped by selection to optimally take advantage of the two perceptual regimes: carrying pitch in the spectral regime and giving rise to speech rate relations in the temporal regime. (see Räsänen et al. 2018 for a similar type of analysis). In other words, syllables have an internal segmental makeup that is optimized to carry pitch in order to exploit the distinction between low vs. high periodicity in the spectral regime, and, at the same time, they appear in sizes that allow the distance between them to give rise to speech rate effects in order to exploit the distinction between slow vs. fast rates in the temporal regime. Figure 4.3 illustrates this state of affairs with a highly generalized schematic sketch of the slower rhythmic cycle between syllables and faster periodic cycles within syllables.

PRiORS can therefore explain the universality of the typical syllable size and the universality of the preferred segmental makeup in the syllabic nucleus, which is commonly measured in linguistic terms of *sonority*. In line with PRiORS, this work claims that sonority should be understood as a measure of pitch intelligibility, acting as a defining feature of syllabic nuclei. Syllables therefore need to be long enough to allow a minimum amount of periods to be effectively perceived. For example, a very low F0 of 50 Hz, which repeats every 20 ms, requires a minimum of 3 periods (60 ms) to be perceived, while higher F0 values (which characterize most speech) repeat faster and require even shorter minimum intervals (Fyk 1987, Josephs 1967). It is of interest to note that the average duration of syllables is about 200 ms (5 Hz, 300 BPM), which is enough for adequate perception of pitch, as well as being at the center of the temporal regime, where it is optimally situated to achieve speech rate effects from the distance between syllables.

4.5.2 Speech is quasi-repetitive

The notion of repetition implies identical intervals between occurrences over time, i.e. repetition is taken to be *isochronous* unless otherwise stated. It has long been noted that pitch-inducing speech sounds are in fact not fully-isochronous, but, rather, *quasi-periodic*, as the pitch and its underlying periods are often unstable at some level. The notion of quasi-repetitiveness is mostly used to refer to an inherent *jitter* in the regularity of repeating patterns that prevent perfect isochrony. This level of jitter (or noise) in the voice is assumed to be perceptually negligible. However, on top of that there is another – much larger – source of apparent instability in repetitive structures in speech. Speech is dynamically

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

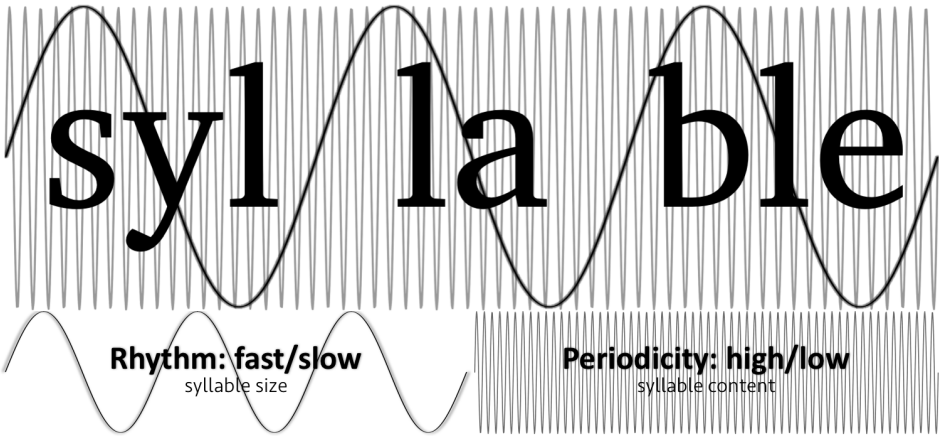

Figure 4.3: Schematic illustration of the relationship between perceptual regimes and syllabic units, using the three canonical syllables of the English word *syllable* (demonstrated with one possible underlying annotation). Segmental makeup, i.e. *sonority*, is related to the spectral regime with high-frequency (*periodic*) oscillations within syllables, while syllabic speech rate is related to the temporal regime with low-frequency (*rhythmic*) oscillations between syllables. The ratio between the low-frequency and high-frequency oscillations in this illustration is arbitrarily set to be 1:20. This is a realistic ratio for syllables such that if syllables are taken to have a typical duration of 200 ms (5 Hz), the high-frequency oscillation within it would reflect a typical F0 for adult males at 100 Hz. For simplicity, this generalized illustration shows a single rate at each timescale with isochronous repetitions (see Section 4.5.2 on the more complex picture regarding isochrony in speech).

changing all the time, in magnitudes that far exceed the levels of inherent jitter, in order to achieve perceptible goals and to effectively exploit the sensations of rhythm and pitch.

Consider for example the periods during a rising pitch contour, in which every period is shorter than the previous one. These degrees of change do not hinder the perception of coherent pitch contours, demonstrating our specialized ability to perceive dynamically-changing pitch. As long as these communicatively relevant pitch differences occur within the timescale of the spectral regime (and follow basic Gestalt principles) they invoke a reliable effect in perception. It is exactly these dynamic changes in the rate of repetition that prosody seems to exploit in speech.

A similar behavior can be observed for rhythm in speech. Speech rates do not typically appear as isochronous within the rhythm-inducing timescale of the temporal regime (see Turk & Shattuck-Hufnagel 2013 and Nolan & Jeon 2014).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

This is the temporal range which is exploited in speech for its perceptible effect on speech rate in terms of slow vs. fast. Crucially, there are no strong reasons to assume that the effect of rhythm in speech is exploited for isochrony, as it is not clear what purpose this would serve in speech. However, in order to achieve various prosodic goals such as phrasal demarcation, turn-taking management and prominence marking (among many others), asynchronous temporal relations are exploited within the scope of speech rate sensations. In other words, speech is *quasi-rhythmic* and dynamically changing at the timescale of the rhythm-inducing effects of the temporal regime in order to be effective for prosody.

Confusingly, speech makes a very different usage of the temporal regime when compared to music, which, more often than not seems to favor isochronous rhythmic patterns over meandering ones within the rhythm-inducing timescale. Musical experiences tend towards isochronous rhythms, perhaps because of the powerful ability of the perception of isochrony to create a shared clock that can be synced across separate systems and human agents, whereby different people can couple sensorimotor oscillations between one another and experience *entrainment* (see, e.g., Cummins 2009, Cummins 2015, Benichov et al. 2016, Haegens & Golumbic 2018, Kotz et al. 2018, Rouse et al. 2016, Tal et al. 2017).

In contrast to a classic case of entrainment to external clocks, languages seem to use the effects of rhythm in the temporal regime for a different set of goals that do not seem to require isochrony and should not be considered to reflect classic entrainment (see, e.g., Cummins 2012 and Meyer et al. 2020). The rhythm-inducing timescale is mostly used in speech to effectively exploit the distinction between slow and fast speech rates as useful cues in a system of speech prosody. To that end, the most (quasi-)isochronus element in speech perception should be internal rather than external (i.e. not in the speech signal itself but in the mind of the hearer). This is needed in order to allow interlocutors to effectively perceive the (largely unpredictable) external speech rates of other interlocutors.

Thus, a striking feature of the effects of repetition in language is that they make use of the two regimes by keeping repetitive elements in a constant state of flux within their effective timescales. To be communicatively useful in prosody, pitch in speech is mostly quasi-periodic and speech rate is mostly quasi-rhythmic. This is the case both in terms of the perceptually negligible jitter and the perceptually informative dynamic changes in rate.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table 4.3: Generic neural oscillations (roughly defined) across the temporal regime

Perceptual regimes	Effects	Neural Oscillations	Timescales
Temporal	Rhythm	*delta*	0.5–4 Hz
		theta	4–8 Hz
		alpha	8–12 Hz

4.6 Neural oscillations in perception and cognition

The PRiORS timescales also fit very well with the characterization of speech processing via neural oscillations of brain activity at different wave lengths (see overviews in Buzsáki 2006, Myers et al. 2019, Poeppel & Assaneo 2020). As Table 4.3 demonstrates, three distinct neural activity patterns are commonly observed within the rhythmic portion of the temporal regime, comprising a set of low frequency oscillations. Interestingly, the mid-range among the three, the *theta* frequency band at 4–8 Hz, has been often studied in conjunction with syllables, as it covers the range of durations (125–250 ms) that, indeed, characterizes the vast majority of syllables, cross-linguistically (Ding et al. 2014, 2017, Gross et al. 2013, Keitel et al. 2017, Luo et al. 2010, Poeppel & Assaneo 2020).

The link between the theta frequency and syllables is consistent with the PRiORS framework, whereby syllables and the temporal domains of cognition are assumed to have co-evolved to exploit the rhythmic effects of the temporal regime, therefore tending towards the center of this particular perceptual-cognitive sensation.

Note that the timescales of speech units do not necessarily fall into the classic division of generic wave lengths. Syllables can be quite diverse and they can be found at rates ranging from 2 Hz, with long 500 ms intervals (e.g. Chandrasekaran et al. 2009) to 20 Hz, with short 50 ms intervals (e.g. Greenberg et al. 2003). Likewise, various speech phenomena have been associated with the ranges of delta and theta waves (1–8 Hz), i.e. at intervals ranging from a little over 100 ms up to one second (e.g. Ghitza 2017, Ghitza 2013, Cummins 2012, Goswami & Leong 2013, Inbar et al. 2020, Meyer et al. 2017).

Keitel et al. (2018) focus on timescales directly extracted from statistical regularities in their speech material, rather than focusing on generic timescales like delta or theta bands. The timescales in their study reflected the rates of *phrases* (0.6–1.3 Hz), *words* (1.8–3 Hz), *syllables* (2.8–4.8 Hz), and *phonemes* (8–12.4 Hz) as

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

they appeared in their corpus of carefully read speech (by a trained, male, native British actor). Indeed, after analyzing corresponding speech tracking signals from listeners using magnetoencephalography (MEG), they found neural activity strongly correlated to these timescales. PRiORS can shed light on such results, whereby, as with the generic waves, the timescale of syllables occupies a central portion of the rhythm-inducing range of the temporal regime (2.8–4.8 Hz). Furthermore, all the different linguistic units are neatly spread across the rhythm-inducing range, defined here at 0.5–12 Hz (see Table 4.2), allowing speakers to perceive quasi-rhythmic patterns of stressed and emphasized/accented syllables at the lower end of the rhythmic perception that Keitel et al. (2018) link with "phrases" (0.6–1.3 Hz) and "words" (1.8–3 Hz), as well as perceiving quasi-rhythmic patterns of segment-size units that Keitel et al. (2018) link with "phonemes" at the highest end of rhythmic perception (8–12.4 Hz).

Note that the transitional range between regimes in PRiORS, defined here as roughly 12–50 Hz (see Table 4.2), is expected to be of little usefulness for bottom-up processing of auditory material given that at this timescale the effects of rhythm and periodicity are somewhat indeterminate. The neural oscillations that are commonly associated with this timescale are the *beta* band at about 13–30 Hz and the *gamma* band at about 30–50 Hz. Interestingly, studies such as Mai et al. (2016) and Keitel et al. (2018) find the beta and gamma rates of neural oscillation to be more closely related to top-down inferences involved in processing of syntax and semantics. This implies a functional division of labor when processing speech, whereby top-down inferences may "piggyback" on channels that are less useful for bottom-up processes.

Recent work in Tang et al. (2020) is in line with the rationale of the PRiORS framework, linking phonological outcomes with the same perceptual primitives that the PRiORS framework assumes. Tang et al. (2020) investigate the relationship between the frequency of certain genes in human populations and phoneme inventories in their respective languages. The genes they investigate are assumed to modify faithful spectral and temporal encoding in the auditory cortex. It appears that the distribution of these genes across human populations can quite reliably predict the size of stop and nasal consonant inventories that will be featured in their languages. The authors suggest that the differences in spectral and temporal precision that can explain these phonemic preferences, may be directly related to observed differences in genetic expressions.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

4.7 Shifting paradigms in linguistic theory with PRiORS

The most relevant areas of phonological theory that PRiORS can shed light on concern the notions of sonority (see Section 4.7.2), as well as the notion of rhythm in speech. Rhythm is beyond the scope of the current study, but some pointers towards the potential contribution of PRiORS are given in Section 4.7.1 below.

4.7.1 A different rhythm

The idea that the effect of rhythm in speech has the same function as rhythm in music has misled many attempts to characterize rhythm phenomena in speech. The thorniest challenge that such endeavors have to face is the search for isochrony in conversational speech that is not chanted or sung. A strict view of isochrony, which is sometimes referred to as *coordinative* or *periodic* rhythm, is very characteristic of what we typically consider to be rhythmic in musical terms – the division of time into equal parts (and the further subdivisions of those equal parts into simple fractions: 1/2, 1/3, 1/4, 1/8, 1/16, etc.).

A well-known manifestation of this search for isochrony can be found in the typological classification of *stress-timed* vs. *syllable-timed* languages (see Pike 1945, Abercrombie 1967, Dauer 1983, Lehiste 1990), whereby isochrony is supposedly maintained between syllables in syllable-timed languages and between stressed syllables in stress-timed languages. According to this idealized view, languages that do not reduce unstressed syllables (e.g. Spanish) maintain isochrony between all the syllables within a phrase, while languages like English, with secondary stress and reduction of unstressed syllables, maintain isochrony only between the stressed syllables within a phrase (such that reduced syllables do not participate in this timing scheme). This view is idealized since a straightforward isochrony of this type, in which some level of spoken language adheres to an external clock, is not to be found on the surface acoustics of speech (e.g. Arvaniti 2009, Turk & Shattuck-Hufnagel 2013).

A slightly more nuanced concept of rhythm targets the relationship between speech items, rather than the alignment of speech items to real time. It is sometimes referred to as *contrastive* or *phonological* rhythm as it addresses our tendency to perceive a strong/weak distinction between repeating items in a sequence. This conception of rhythm does not assume strict isochrony that adheres to an external clock. Instead, it uses local timing relations to reflect prominence and grouping in the speech signal (e.g. Arvaniti 2009).

Various acoustic metrics were developed in order to measure global rhythmic distinctions, with the aim of characterizing different languages. Among them are

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

measurements of the relative abundance or regularity of durations of selected units such as vowels and consonants, e.g. *%V*, *ΔV*, *ΔC* (Ramus et al. 1999), *VarcoV*, *VarcoC* (Dellwo 2006) and variants of the *Pairwise Variability Index* (PVI) (e.g. Grabe & Low 2002). These metrics manage to avoid the requirement for strict isochrony and they succeed in characterizing different languages, but this seems to be true only to some extent, with small effect sizes (the variability within languages can be very high due to various factors like speech style and methodological decisions in the measurement itself) and in manners that are inconsistent with classic rhythm typologies (see Arvaniti 2012). Lowit (2014) concluded that none of these metrics are useful in clinical settings, based on systematic comparisons between speakers with dysarthria and matched healthy participants.

Nolan & Jeon (2014) provide an overview of these problems. They suggest that language is perhaps *antirhythmic* such that isochronous patterns are not to be found in the surface acoustics, but they might be metaphorically projected in perception. This description is indicative of the fact that even the concept of *contrastive* rhythm is essentially based on comparison to an isochronous baseline, as if isochrony is an underlying goal of speech rhythm in and of itself.

PRiORS can help us make the next logical step by providing a slightly different framework for the understanding of rhythm, such that isochrony is no longer a key ingredient in its definition. Isochrony is one goal that can be achieved from rhythm effects in the temporal regime, and, indeed, this goal is exploited extensively in music. Music seems to exploit rhythm effects to achieve isochrony in order to promote entrainment, while speech seems to exploit rhythm effects in order to control the temporal dimension of prosody and effectively use the distinct sensation of slow vs. fast. Speech – unlike music – is mostly meandering in its rhythmic patterns. Expecting rhythm in speech to exhibit isochrony is akin to an expectation that every syllable would have a steady level pitch in order to count as periodic, overlooking the major role of perceived dynamic changes within each perceptual regime (see Section 4.5.2).

Rhythm in speech should therefore be understood as a moving target that can be more adequately modeled in terms of a trajectory, much like the trajectory of the F0 at the faster timescale of the spectral regime. Related ideas towards this goal can be found in the pioneering work of Pfitzinger (2001), which uses dynamic trajectories to describe local speech rate. A PRiORS-based analysis of rhythm in speech should therefore target the syllable-size fluctuations in the periodic energy curve, and model their temporal distances in terms of a dynamic trajectory in order to capture local speech rate as the main effect of rhythm in speech (for preliminary attempts, see Section 9.2.3).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

4.7.2 A new type of sonority

The human auditory system evolved to exhibit great sensitivity to (quasi-)periodic signals within the spectral regime, specializing in the perception of pitch. This is evident from the impact of pitch on our categorization of many musical sounds (e.g. Bidelman & Krishnan 2009) as well as on tone and intonation in speech (e.g. Krishnan et al. 2005). This is also evident from anatomical and neurological activity, either in terms of *place* representations in the cochlea (i.e. in spectral terms) or in terms of *timing* representations in the auditory nerve, characterized by *phase-locking* to neural firing rates (see Section 4.2).

The *vocalic* or *voiced* portions of speech can be described as a train of glottal pulses produced by vocal fold vibration, not unlike the idealized BLIT simulation in Section 4.3. This voiced component of the speech signal is the main carrier of pitch in speech and it is a striking fact about all languages that they prefer this auditory characteristic at the nucleus of their syllables (as well as the fact that all known languages exhibit a syllabic structure to begin with).

One aspect of this important dimension of linguistic sound systems is our ability to obtain good estimations of the *fundamental frequency* (F0) of complex sounds, which we take as a reliable correlate of perceived pitch height. Indeed, phonologists and phoneticians have incorporated continuous measurements of F0 as a regular part of their toolbox, and they are well aware of the fact that F0 is more robust at syllabic nuclei, where the most sonorant elements usually sustain a sufficiently long and powerful (quasi-)periodic sound (see Barnes et al. 2011, Barnes et al. 2014, Roettger & Grice 2019).

Measurements of F0 therefore cover a qualitative aspect of perceived pitch: its height in terms of the rate of repetition of the fundamental frequency. What is missing from this picture is the quantitative aspect of this auditory dimension, a description of acoustic power that targets only the pitch-inducing (vocalic/periodic) portions of the speech signal, unlike the commonplace practice to obtain the physical *intensity* of the acoustic signal as a whole. A measurement of this kind, referred to as *periodic energy*, has the promising ability to correlate with the notion of *sonority* in a way that implies causation related to *pitch intelligibility*, as it separates pitch-inducing components that favor syllabic nuclei from noise-inducing aperiodic components that favor syllabic margins.

Sonority in this sense is viewed as a measurement of the goodness of fit for syllabic nuclei, directly targeting pitch as an auditory dimension that our perceptual-cognitive systems are specialized for and that has evidently shaped the basic structure of all linguistic sound systems, given the universality of syllables with sonorous/pitch-bearing nuclei.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

The perspective that PRiORS suggests helps in redirecting our focus away from the non-discriminative nature of general acoustic intensity and other acoustic measurements that do not suggest a clear and consistent association with perception and cognition. Instead, PRiORS directs us to search for acoustic measurements that exhibit robust links to pitch-inducing phenomena in the spectral regime in order to adequately characterize sonority. As PRiORS helps elucidating, periodic sounds in the acoustic speech signal carry valuable pitch information, which, in turn, makes them privileged in terms of position within the syllable.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

5 Sonority, pitch and the Nucleus Attraction Principle (NAP)

5.1 Sonority and pitch intelligibility: A causal link

The observation that sonority summarizes an essential quality that is related to vowels and their propensity to deliver a relatively steady harmonic structure, highlighting pitch and formant information, is by no means new. Previous proposals already defined sonority as either relating to vowels in some general way, more specifically relating it to voicing or glottal fold vibration, or to the clarity/strength of formants.[1] A few previous accounts went even further, by addressing the function of this evasive vowel-centric feature, suggesting that sonority may be related to periodic energy or pitch/tone (Heselwood 1998, Ladefoged 1997, Lass 1988, Nathan 1989, Puppel 1992). What all these proposals share, explicitly or implicitly, is a recurring insight about a strong link between the preferred type of segmental material in syllabic nuclei and a set of features that conspire to optimize pitch intelligibility, a property which characterizes vowels more than consonants.

Pitch is an indispensable communicative dimension of all linguistic sound systems (Bolinger 1978, Cutler et al. 1997, House 1990, Roettger & Grice 2019), whether it is lexically determined as in linguistic *tone*, or post-lexically employed to convey intonation, i.e., the linguistic *tune* (see typological accounts of prosodic systems in Jun 2005, Jun 2015). Tones are used to distinguish lexical items while tunes are used to demarcate units, to modulate semantics (e.g. information structure and sentence modality) and to express a vast array of non-propositional meanings (e.g. discourse-pragmatic intention, emotional state, socio-indexical identity, and attitudinal stance). The importance of pitch to human communication cannot be overstated.

[1] A partial list of some prominent examples includes Sigurd (1955), Jakobson & Halle (1956), Chomsky & Halle (1968), Foley (1972), Ladefoged (1971), Allen (1973), Fujimura (1975), Donegan (1978), Ultan (1978), Price (1980), Lindblom (1983), Anderson (1986), Vennemann (1988), Levitt et al. (1991), Pierrehumbert & Talkin (1992), Fujimura & Erickson (1999), Bernhardt & Stemberger (1997), Boersma (1998), Zhang (2001), Howe & Pulleyblank (2004), Clements (2009), Sharma & Prasanna (2018).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Crucially, linguistic pitch events are known to target syllable-sized units as their "docking site", regardless of the type of pitch event, whether they are lexical tones or post-lexical tunes. These linguistic pitch events are commonly considered to associate with *Tone-Bearing Units* (see Leben 1973), that are either syllables or *moras*.[2] These associations between the text on the one hand and tone or tune on the other hand are widely assumed to be mediated by syllabic/moraic units. For example, intonation pitch contours that highlight and modulate whole words and phrases essentially target privileged syllables – *heads* (stressed syllables) and *edges* (syllables at initial and final positions of prosodic words and phrases) – to achieve their communicative goal on textual material of various sizes. This tone-bearing role of syllables and moras is the hallmark of many prominent theories regarding tone and intonation, following from Autosegmental and Autosegmental-Metrical Phonology (e.g. Liberman 1975, Goldsmith 1976, Ladd 2008, Pierrehumbert 1980).

The functionally motivated conclusion that emerges with respect to sonority is therefore that syllables require a pitch-bearing nucleus and that sonority is a scalar measure of the ability to bear pitch. In other words, sonority is, most likely, a measure of *pitch intelligibility*. This hypothesis comes with an underlying assumption that was introduced by the PRiORS theoretical framework in Chapter 4, whereby syllables are claimed to have followed an evolutionary trajectory that shaped them to optimally carry pitch in their nuclei (Section 4.5.1). Sonority, according to this description, serves as the tool that governs the requirement for intelligible pitch as a fundamental characteristic in the design of the building blocks of prosody (see Section 4.7.2).

It is important to note that this view of sonority is explicitly and exclusively based on perception, rather than articulation of speech. However, it does not exclude articulation-based description of syllables under the assumption that restrictions on syllabic structure must be derived from both the perception and the articulation of speech. A case in point is the *Articulatory Phonology* framework (see Section 3.2), with its valuable descriptions of temporal coordination and phase relations between motor gestures, which can be effectively linked to syllabic organization (see, e.g., Goldstein et al. 2007, Gafos et al. 2014, Goldstein et al. 2009, Hermes et al. 2017, Shaw et al. 2009).

[2]Moras are used to represent quantitative differences between light and heavy syllables (*weight sensitivity*, see Section 5.3.2), such that light syllables contain one mora while heavier syllables contain two (and sometimes even three) moras (see Hyman 1984, Hayes 1989, Itô 1989, McCarthy & Prince 1990, Zec 1995, Zec 2003).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

5.2 Periodic energy and sonority: Causation by transitivity

Pitch is a psychophysical phenomenon based on perception and cognition (see Plomp 1976, Plack & Oxenham 2005). We can technically obtain pitch-related measurements in terms of neurological and behavioral responses directly from perception. Such measurements are hard to accumulate in very large numbers as they require intricate lab procedures in order to collect data from each subject. Another avenue for obtaining perception-related measurements is to extract them from acoustics, i.e., not directly from the perceived sensation of a human subject but from the digitally-analyzed description of the physical sound in space. The benefits of acoustic measurements include the accessibility of recording and processing capabilities and the availability of many existing corpora, which facilitate access to large amounts and diverse types of acoustic speech data. Using acoustics to cover auditory psychophysical phenomena is not a straightforward task. It requires a consistent and reliable association between acoustics on the one hand, and perception and cognition on the other hand. This task is potentially complicated further with a complex phenomenon like pitch, which is evidently sensitive to various aspects of the rich acoustic signal as well as to our top-down expectations with regards to learned regularities of pitch behavior in the speech signal (see, e.g., Houtsma 1995, McPherson & McDermott 2018, Moore 2013, Shepard 2001: 203).

Fortunately, there are strong links between pitch and acoustic markers. This is well-known from the extensive use of acoustic F0 measurements to estimate perceived pitch height. Furthermore, pitch estimations from F0 measurements can become more reliable when dealing with specific types of audio such as speech, as in this case the bulk of pitch information comes from a single source (i.e. one speaker) within a limited range of fundamental frequencies (mostly between 75–400 Hz, rarely below 50 Hz or above 600 Hz).

To estimate perceived pitch intelligibility from acoustic signals, we need to obtain a measure of *periodic energy*, which is a measurement of the acoustic power of periodic components in the signal. It may be helpful to think of this as a measurement of general intensity that excludes the contribution of aperiodic noise and transient bursts. Measurements of periodic energy are not very different from widely-used F0 measurements that are commonly based on the ability to detect periodic components in the complex signal. Roughly speaking, rather than resolving the harmonic denominator of detected periodic components in order to estimate F0, a periodic energy meter needs to sum over their power.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

To conclude, our ability to detect periodicity in acoustic signals allows us to extract good estimates of F0 and periodic energy from speech data. We stand on firm grounds when we map these acoustic markers to perception in terms of pitch height and pitch intelligibility (respectively). Given a causal link between perceived pitch height and linguistic tone and intonation contours, it is reasonable and, indeed, commonplace, to assume by transitivity that acoustic F0 maintains a causal link to linguistic tone and intonation. Likewise, given a causal link between perceived pitch intelligibility and linguistic sonority, it should be reasonable to assume by transitivity that acoustic periodic energy maintains a causal link with the linguistically-loaded notion of sonority.

5.3 The Nucleus Attraction Principle

At the heart of all sonority-based principles lies the idea that the most sonorous segment in a sequence is contained within the nucleus of the syllable. This idea in fact postulates a link between the amount of sonority and the nucleus position of the syllable. I adopt this fundamental insight that guides all other sonority principles in the development of the Nucleus Attraction Principle. However, instead of adding further formal assumptions about non-overlapping segments with fixed sonority values and corresponding sonority slopes in symbolic time, the link between sonority and the syllabic nucleus is simply modeled as a dynamic process in real time. All the portions of the speech signal compete against each other for available nuclei in this process.

Sonority is therefore the quality that is capable of *attracting* the nucleus. The varying quantities of this quality, which temporally fluctuate along the stream of speech, determine which portions of speech are prone to succeed in attracting nuclei given their superior local sonority *mass*. The speech portions that fall between those successful attractors are syllabified in the margins of syllables, at onset and coda positions.

Crucially, NAP treats the postulated link between sonority peaks and syllabic nuclei as the result of a perceptual-cognitive process in real time, rather than describing a geometric state of affairs with symbolic discrete tools. In fact, by modelling the link between sonority and the syllabic nucleus in dynamic terms it is not necessary to add further theoretical postulates about sonority slopes or discrete segmental categories of consonants and vowels in order to determine well-formedness of syllabic structures. Syllabic ill-formedness in NAP-based models is positively correlated with the degree of nucleus competition that a given syllabified portion incurs.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

It is important to note that the informativeness of NAP-based models is not derived from identifying the winner of the nucleus competition, but from quantifying the degree of competition within different portions of speech that stand for potential syllabic parses. NAP-based models can analyze speech parts that are parsed together as a single syllabic unit in order to estimate the degree of competition they give rise to when they compete for a single nucleus. In discrete terms, NAP-based models can quantify different sequences of segments to reflect how strongly they compete for a single nucleus. Either way, the higher the degree of internal competition, the more ill-formed a syllable is predicted to result from this parse. To simplify this further with respect to the subset of instances discussed in this work (i.e. syllables with complex consonantal onset clusters), it is possible to say that the winner of the nucleus competition is always the only vowel in the structure. The determination of ill-formedness in these cases is based on quantifying the amount of competition that the winning vowel has to withstand given different consonantal clusters in the onset of the same syllable.

It should be also useful to note that we do not expect serious competition to arise from a consonant adjacent to the vowel in the same syllable, such that in a C_1C_2V syllable only C_1 is considered to be the potential competitor to V. The consonant in C_2 position has a crucial impact on the competing potential of C_1 but it is not, in and of itself, a competitor in the data presented in this study.[3] To elucidate this point, consider the case of simple CV syllables. Here, sonority levels are expected to rise from C to V continuously, with no competition for the nucleus. Nucleus competition, much like sonority slopes, has a limited impact on syllables with maximally simple onsets and/or codas, (i.e. V, CV, VC and CVC). Principles like SSP and NAP play a role chiefly when sequences of consonants are syllabified within a single syllable as complex onset or coda clusters (e.g. CCV or VCC). The phonotactics of these possible sequences are determined to a large extent by sonority principles. We interpret this aspect of cluster phonotactics such that sequences within syllables are avoided the more they increase the potential competition for the nucleus in the process of syllabifying/parsing the stream of speech.

5.3.1 Schematic NAP sketches

To understand the rationale of NAP, a series of schematic sketches are presented in Figure 5.1, accompanied by an impressionistic description. These will eventu-

[3]We narrowly expect vocoids (i.e. glides) to be able to compete for the nucleus from the vowel-adjacent C_2 position, but this case is likely circular since a glide in the nucleus position would be simply considered a (high) vowel.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

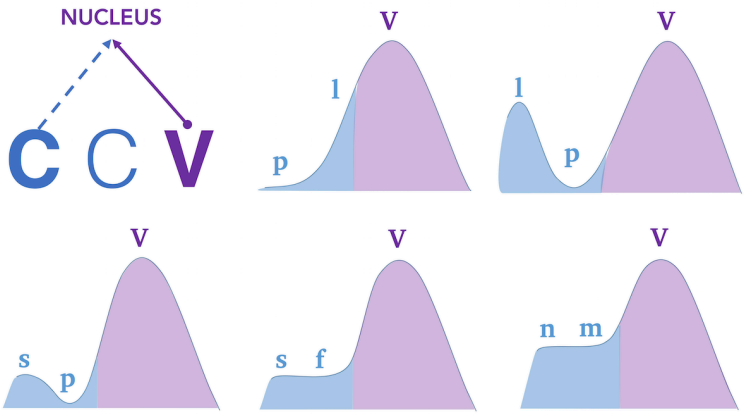

Figure 5.1: Schematic depictions of competition scenarios with symbolic CCV structures. Nucleus competition can be understood as the competition between the blue and the purple areas under the sonority curve. The two examples in the top row – *plV* and *lpV* – suggest a replication of successful traditional predictions, while the three examples in the bottom row – *spV*, *sfV* and *nmV* – suggest a divergence from inherent failures of SSP-type models (see text for more details).

ally be implemented within formal models that are described in detail in Chapter 6. The five examples with specified consonantal clusters exhibit their related sonorant energy depicted as the *area under the curve*, whereby the curve itself is an idealized depiction of schematic sonority. The purple area in each syllable in Figure 5.1 denotes the sonorant energy of the winning vowel in the nucleus position while the blue area denotes the sonorant energy of the losing portions in the onset. Consider for example the pair *plV* and *lpV*, with schematic NAP-related depictions in the top row of Figure 5.1 (and with more traditional sonority slopes in Figure 2.1). A consonantal onset cluster with a putatively well-formed rising sonority slope like *plV* should be also considered well-formed under NAP due to the very low potential of competition between the marginal minimally-sonorous onset consonant /p/ and the non-adjacent vowel that wins the competition for the nucleus. The intervening /l/ in this case only promotes a continuous rise in sonority from /p/ to V. Likewise, a consonantal onset cluster with a putatively ill-formed falling sonority slope like *lpV* should be also considered ill-formed under NAP due to the strong potential for competition between the marginal sonorous onset consonant /l/ and the non-adjacent vowel, especially given the intervening /p/ that leads to discontinuity in the sonority trajectory between /l/ and V.

Unlike the examples above, where the rationale of NAP is expected to replicate successful predictions of the SSP with cases like *plV* and *lpV*, NAP is also expected

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

to diverge from traditional sonority sequencing principles in those cases where traditional principles suffer from inherent failures, as detailed in Section 2.2.2. Consider the examples in the bottom row of Figure 5.1, which were also depicted with traditional sonority slopes in Figures 2.4 and 2.5. Under NAP, neither */s/-stop* clusters like *spV* nor voiceless obstruent plateaus like *sfV* are expected to incur a strong competition syllable-internally due to the low potential for competition between the minimally-sonorous onset consonant /s/ and the non-adjacent vowel that wins the competition (here, the intervening voiceless obstruents /p/ and /f/ retain a minimally sonorous trajectory throughout the whole onset). At the same time, a strong competition potential is predicted under NAP for nasal plateaus like *nmV* when compared to obstruent plateaus like *sfV*. This should be expected given the strong potential for competition between the marginal sonorous onset consonant /n/ and the non-adjacent winning vowel (here, the intervening nasal retains a relatively level sonorous trajectory throughout the onset).

As a rough conclusion, it is possible to suggest that by observing the potential competition between blue and purple areas in Figure 5.1, we should easily see that the two structures on the right-most side (*lpV* and *nmV*) exhibit a stronger competition potential syllable-internally in comparison to the other three structures, in a manner that is not fully predictable from their sonority slopes. For more elaborate competition-based distinctions, see Section 6.3.

5.3.2 On the roots of prosodic *attraction*

The central idea behind NAP, whereby sonority *attracts* syllabic nuclei, is, in fact, well-established in phonological theory. In various descriptions of stress systems, it is often suggested that some languages exhibit *weight sensitivity*. This is not a universal process, as stress assignment patterns vary from language to language, and not all languages even have stress to begin with. However, weight sensitivity is one of the naturally occurring stress assignment patterns that various unrelated languages exhibit (e.g. Arabic, Tibetan (Lhasa), Wolof, Finnish, Latin and many more; see Goedemans & van der Hulst 2013, and Gordon 2006: 23 for more exhaustive lists).

Weight sensitivity usually means that a language which regularly assigns the primary stress to a certain syllabic position within phonological words (e.g. initial/final syllable, etc.) may diverge from this canonical position and assign the stress to an adjacent syllable if it is *heavier* than the syllable at the canonically stressed position. This is standardly understood as *attraction* of the primary stress by the heavy syllable, where heaviness is mainly the product of a longer

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

vowel in the nucleus, and in some languages heaviness may also result from a (preferably sonorant) consonant in the coda (see, e.g., McCarthy 1979, Gordon 2006, Hayes 1980, Prince 1990). There are also analyses whereby vowel qualities that are considered more sonorous due to degree of opening (i.e. more *open*/*lower* vowels) can contribute to heaviness and attract stress (Gordon et al. 2012, Kenstowicz 1997, de Lacy 2002, Zec 1995, 2003).

Importantly, all of these notions of weight are consistent with a hierarchy of sonority. Structurally, the rime is the locus of weight phenomena, and within the rime – the nucleus is most important for weight. Segmentally, sonorants contribute more to weight than obstruents, and within sonorants, open vowels are the strongest attractors. Viewed with NAP in mind, attraction of stress in weight-sensitive systems is simply the special case of a regular procedure, whereby weight – i.e. sonority mass – attracts syllabic nuclei. In other words, given the regular process that NAP assumes, by which syllabic nuclei are attracted to sonorant energy masses, weight sensitivity is simply an extension whereby *heavy* syllabic nuclei are attracted to *heavy* sonorant energy masses.

The stressed syllable in weight-sensitive systems is maintained as highly sonorous, which makes it an optimal syllable for carrying tonal events in intonation, generally serving as the docking site for *pitch accents*. Attraction in prosody thus follows a consistent rationale: sufficiently pitch-intelligible units satisfy the requirement for a regular syllable by attracting nuclei in general, and exceptionally pitch-intelligible units may satisfy a special requirement for the stressed syllable by attracting the strongest nuclei.

A similar process also occurs post-lexically in many languages. This process is related to text-tune interaction, which can often lead to local sonority enhancements of syllables that need to carry tonal information. The most prominent cases are post-lexical prosodic enhancements through an increase in duration and/or intensity of sonorant material (alongside insertions of transitional vocoids and epenthetic vowels) serving to accommodate certain tonal events in intonation (see Roettger & Grice 2019).

To conclude, the understanding that sonority is linked to pitch via syllabic units is well established in phonology. NAP takes this understanding further than previous insights about prosodic weight and text-tune interactions in proposing a functional theory of prosody and sonority based on pitch intelligibility.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

6 NAP implementations

6.1 Complementary NAP models

NAP essentially describes a bottom-up process, illustrating the parsing of the stream of speech into syllables as the end point of a process that starts in perception. As such, NAP is designed to agree with the laws of physics and the biases of the human auditory system in order to shed light on linguistic processing. A bottom-up perspective on modelling NAP is therefore relatively straightforward as it requires a similar approach to the process NAP describes: the analysis of continuous acoustic data at the input, resulting in well-formedness predictions at the output.

A bottom-up approach for NAP models has no capacity to exploit the power of abstraction, so it essentially has no "memory". It is a mechanistic dynamic model that contains discrete symbolic entities only as the linguistic target of the task, at the end of the process determining syllabic well-formedness. This means that a bottom-up model can be only designed to analyze concrete speech tokens. Unlike traditional sonority principles and their models, a bottom-up model of NAP cannot determine the well-formedness of an abstract syllable as it is depicted in symbolic form. It will therefore give slightly different scores to different renditions of the same syllable, even by the same speaker.

A NAP-based model operating on abstracted symbolic units is used as a separate, complementary top-down model (see Chapter 3 and specifically Section 3.6). Top-down inferences are based on learned regularities and categorical abstractions that reflect linguistic experience. To that end, knowledge about consonantal inventories and the probabilities of consonantal co-occurrence and distribution with respect to position in the syllable has to be acquired and then stored in abstract symbolic forms which are available for top-down inferences. In that sense, top-down inferences in perception are based on the distributional probability of recognized symbols.

The above description of top-down inferences, which are detached from the functional aspects of the bottom-up route, echo models of the language user as a *statistical learner* (see, e.g., Christiansen & Curtin 1999, Frisch & Zawaydeh 2001, Tremblay et al. 2013) and, more specifically, they are very much in line with

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

models of *phonotactic learners* (see, e.g., Coleman & Pierrehumbert 1997, Albright 2009, Bailey & Hahn 2001, Daland et al. 2011, Hayes 2011, Hayes & Wilson 2008, Jarosz et al. 2017, Mayer & Nelson 2019, Vitevitch & Luce 2004). That said, the current project does not explore the statistical nature of top-down inferences. Instead, it operationalizes the rationale behind NAP with symbolic machinery to present what can be understood as the symbolic model of NAP and is used to estimate top-down inferences. This choice allows the presentation of a top-down model with a stronger explanatory value with regards to NAP as it uses a similar architecture to that of standard sonority principles, helping to elucidate NAP's core ideas while using a familiar vocabulary (see Section 6.2.2).

Moreover, it should be noted that since a cognitively plausible top-down architecture in this framework is based on the distributional patterns of recognizable symbols, these distributions should be "blind" to their various sources, which include a host of universal and idiosyncratic phonotactic pressures. A true top-down statistical learner is thus inherently "contaminated" by all the different sources that contribute to phonotactics in a given system, without a clear distinction between sonority and other factors. Thus, it remains an open question whether top-down inferences that target only sonority-based phonotactics can be modeled in a more direct and principled way than the one presented here with the symbolic model of NAP.

As two complementary inference routes, the top-down and bottom-up models should not be considered equal. The bottom-up route is the source of learned linguistic distinctions and it is functionally motivated by the laws of physics and the limitations of the perceptual and cognitive systems. In contrast, the top-down route is based on linguistic experience and superficial inferences that reflect the history of the symbols in the system (i.e. the distributional probabilities of recognizable recurring patterns and their extensions by analogy). In other words, top-down inferences reflect functionally motivated behaviors only indirectly, as the outcome of learning the superficial expressions of functionally-motivated (bottom-up) dynamics.

6.2 Model implementations in dynamic and symbolic terms

In order to compare the different proposals, four types of traditional sonority models are considered alongside the two NAP models. For traditional models I use the two types of sonority hierarchies that were presented in Section 2.1.1,

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

where the class of obstruents is either *collapsed* (H_{col}) into a single level or *expanded* (H_{exp}) to include distinctions between voiced and voiceless obstruents, and between stops and fricatives. Both hierarchies are applied with each of the two main variants of traditional sonority principles, the *Sonority Sequencing Principle*, SSP, and the *Minimum Sonority Distance*, MSD (see Section 2.1.2). The four traditional sonority models under discussion are therefore a combination of a sonority principle (either SSP or MSD) and a sonority hierarchy (either H_{col} or H_{exp}). Accordingly, they are referred to as SSP_{col} , SSP_{exp}, MSD_{col}, and MSD_{exp}.

The two NAP models use periodic energy as the correlate of sonority, and periodic energy is applied either continuously through acoustics (bottom-up model), or in a discrete manner using symbols (top-down model). These two NAP models are referred to as NAP_{td} for the top-down model and NAP_{bu} for the bottom-up one.

To demonstrate the different sonority models, this study focuses on complex onset clusters of the general form CCV, where C denotes consonants in onset position and V denotes a vowel in nucleus position. Traditional sonority models inspect the sonority slope of the onset cluster to determine well-formedness of CCV syllables, while NAP-based models apply the notion of *competition* to determine well-formedness.

In the following sections, I will elaborate on the methods for obtaining well-formedness scores, starting with the ordinal scores obtained from the four traditional sonority models (Section 6.2.1), and the symbolic NAP model NAP_{td} (Section 6.2.2). The implementation of the continuous model NAP_{bu} follows in Section 6.2.4. Finally, this chapter concludes with a short overview of key advantages of NAP over traditional sonority principles (Section 6.3).

6.2.1 Traditional sonority models

Implementation of traditional sonority principles like the SSP is based on a calculation of the sonority slope over a given sequence of segments. Speech segments in these frameworks have fixed index values on the sonority hierarchy, based on their class membership, as in the H_{col} and H_{exp} hierarchies (see Table 2.1). These sonority index values are usually expressed in terms of integers since they reflect an ordinal scale, and, for this reason, the mathematical operations that these models employ should be restricted to basic arithmetic functions of addition and subtraction. Sonority slopes can be therefore obtained straightforwardly by a subtraction between the corresponding sonority indices of two adjacent consonants. In onset clusters with two consonants (CCV) this can be simply achieved by the formula $C_2 - C_1$, which yields positive results for rising sonority slopes, negative

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

results for falling sonority slopes, or a zero for plateaus. This calculation is applied to the two SSP models, SSP_{col} and SSP_{exp} (see examples in Table 6.2).

The exact same formula is also used to obtain scores for the Minimum Sonority Distance models, MSD_{col} and MSD_{exp}, which elaborate on the well-formedness of onset rises. MSD models differ from the SSP in the interpretation of positive values (that reflect rising sonority slopes). While under the SSP all positive scores map to a single score (i.e. all rises are well-formed to the same extent), under the MSD higher positive scores are preferred over lower positive scores to reflect the preference for a larger sonority distance (or a steeper slope) in a rising onset configuration (see examples in Table 6.2).

6.2.2 The top-down symbolic NAP model

The symbolic version of NAP, which is used to derive predictions for the top-down NAP (NAP_{td}), shares a similar architecture with common SSP-based models. Crucially, it also reflects the novelties of the current proposal, both in terms of the sonority hierarchy it assumes, and in terms of the design of the sonority principle. NAP_{td} uses a sonority hierarchy that is based on the periodic energy potential of different phoneme classes as the basis of distinct categorical patterning (see Section 6.2.2.1). Furthermore, NAP_{td} models syllabic well-formedness with the notion of nucleus competition, rather than the formal notion of sonority slopes as in traditional SSP-type models (see Section 6.2.2.2).

6.2.2.1 The sonority hierarchy in NAP_{td}

The symbolic sonority hierarchy in NAP uses the basic ratio between periodic and aperiodic energy in the speech signal to divide all speech sounds into three distinct groups, reflecting the coarse, yet reliable differences in potential periodic energy mass of different abstract speech sound categories. To achieve that, we rely on the following set of general characteristics: (i) the main source of periodic energy in speech stems from vocal fold vibrations when voicing occurs; (ii) aperiodic energy in speech is mostly the result of the turbulent airflow resulting from articulatory friction (i.e. fricatives) and from articulatory closure in oral stops, which often results in transient bursts when released (see Rosen 1992).

Thus, the ratio between periodic and aperiodic components in speech sounds readily yields the following three distinct groups: (i) voiceless obstruents that consist of mostly aperiodic energy and are the least sonorous type of speech sounds; (ii) sonorant consonants and vowels that consist of mostly periodic energy and are the most sonorous type of speech sounds; as well as (iii) voiced

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

obstruents that consist of both periodic and aperiodic energy and belong in the middle of this ternary scale (see 6.1).

$$\text{Voiceless obstruents} < \text{Voiced obstruents} < \text{Sonorants} \tag{6.1}$$

A further distinction in NAP's sonority hierarchy is based on the general presence or absence of articulatory contact. A free and open vocal tract contributes to a potentially stronger and longer vocalic signal that can qualitatively enhance the potential periodic energy mass. This distinction effectively separates the sonorants into *sonorant vocoids* (glides and vowels) and *sonorant contoids* (nasals and liquids).[1] See Table 6.1 for the full sonority hierarchy in the symbolic model of NAP.

Table 6.1: The symbolic sonority hierarchy in NAP_{td}. *Note.* Index values reflect the ordinal ranking of categories in the sonority hierarchy. The distinctions between categories in the symbolic NAP hierarchy are based on the characteristic ratio between periodic and aperiodic energy, and on articulatory contact, both taken to reflect the potential of the periodic energy mass, i.e., the potential for nucleus attraction.

Sonority index	Segmental classes	Periodic: Aperiodic	Articulatory contact
4	Sonorant vocoids (*glides*, *vowels*)	1:0	–
3	Sonorant contoids (*nasals*, *liquids*)	1:0	+
2	Voiced obstruents (*stops*, *fricatives*)	1:1	+
1	Voiceless obstruents (*stops*, *fricatives*)	0:1	+

The symbolic sonority hierarchy in NAP reconciles perceptual and articulatory approaches to sonority by modelling their mutual contribution to enhancing pitch intelligibility (or periodic energy mass, in acoustic terms). This hierarchy is similar to a few proposals for sonority hierarchies that combined levels of voicing/periodicity with degree of vocal tract opening (e.g. Lass 1988, Miller 2012, and Sharma & Prasanna 2018). Such hierarchies may also be seen as compatible with source-filter models of speech (Fant 1970), where the *source* controls voicing and the *filter* controls opening (e.g. Puppel 1992).

[1]Note that some rhotics, which are traditionally considered liquids, may in fact belong with the vocoid consonants (e.g. most of the English rhotics, especially in coda position). However, we can ignore this issue here since as rhotics are not included in the data investigated in this paper.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

The complete 4-place sonority hierarchy of NAP_{td} in Table 6.1 also reflects a basic typology of nucleus types, which supports the use of this scale as a qualitative measure for nucleus attraction potentials. Sonorant vocoids like glides and vowels can attract the nucleus in all languages we know (a glide is considered a vowel when syllabified in the nucleus position), while sonorant contoids like nasals or liquids can be syllabic (i.e. attract the nucleus) only in a subset of languages, of which a smaller subset may allow obstruents to attract nuclei (but see Easterday 2019 for some divergent patterns with syllabic obstruents relative to syllabic liquids).

6.2.2.2 NAP_{td} implementation

When assessing C_1C_2V syllables under the NAP framework, we essentially aim to measure the competition potential between C_1 and V given C_2. In and of itself, C_2 is not considered a competitor due to its proximity to the vowel, as discussed in Section (5.3). The issue of competition may be therefore expressed by the following questions: (i) what is the potential periodic energy mass of C_1 (i.e. how sonorous is C_1, or what is the intercept of the cluster that determines the starting point of the slope); (ii) how much of the energy in C_1 is potentially lost, gained or maintained in C_2, before peaking at the vowel (i.e. what is the sonority slope). Assessing this relationship between C_1 and V given C_2 can be achieved by the combination of two subtraction formulas: (i) a calculation of the difference between C_1 and the non-adjacent vowel, to reflect the potential strength of C_1 in terms of the intercept relative to the nucleus; (ii) a calculation of the slope between adjacent C_1 and C_2, as in SSP-based models, to reflect the trajectories of fluctuating energy towards the peak. This can be summarized with the formula in (6.2).[2]

$$(V - C_1) + (C_2 - C_1) \tag{6.2}$$

6.2.3 Ordinal sonority scores

Table 6.2 (page 68) demonstrates and compares the scores of the five ordinal models (2*X*SSP, 2*X*MSD and NAP_{td}) with different CCV cluster types. It shows that the main difference between the two sonority hierarchies, H_{exp} and H_{col}, concerns fricative-stop clusters like the */s/-stop* cluster *spV*, which are considered as either an onset fall (with the H_{exp} hierarchy) or an onset plateau (with the H_{col} hierarchy). When the MSD is applied, the two sonority hierarchies also show differences in ranking within onset rises, given their different treatment

[2] A somewhat similar calculation can be found in Fullwood's (2014) *Sonority Angle*.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

of obstruents. In models that use the H_{exp} hierarchy there are four levels of obstruents (voiced and voiceless stops and fricatives) which are collapsed into one level in models that use the H_{col} hierarchy. This results in five distinct sonority rise scores in the MSD_{exp} model, but only two in the MSD_{col} model (where some of the trends also differ, e.g. *smV* vs. *vlV* in the two MSD-based models).

Unlike traditional models, the predictions of NAP_{td} are not grouped into levels that reflect the rough angle of the sonority slope in terms of falls, rises and plateaus. The raw score of the NAP_{td} formula is taken as reflective of the nucleus competition potential such that higher scores denote weaker competition and are thus better-formed. The top-down NAP model allows scores within a range that goes from -3 for the most ill-formed syllable up to 6 for the most well-formed, although a more relevant range to consider, given that glides are excluded from this set, is between -1 and 5. These scores are not immediately comparable to the traditional model scores, but some interesting departures from the traditional models can be observed in Table 6.2. For example, NAP_{td} considers the onset rise in the sonorous cluster *mlV* to be equally as ill-formed as the inverse fall, *lmV*. Both of these clusters pattern with nasal plateaus (e.g. *nmV*), where they all receive the same relatively low value of 1. At the same time, voiceless clusters pattern in with well-formed combinations (scoring 3) although they may include sonority plateaus (e.g. *sfV*) or sonority falls (e.g. *spV*) in traditional model terms.

6.2.4 The bottom-up dynamic NAP model

There are various ways to calculate an estimation of the nucleus competition potential within syllables based on the periodic energy in the acoustic signal. The method presented here has the advantage of not relying on segmental landmarks that are categorical abstractions of the type that is not assumed to be available in the bottom-up route (see Sections 3.6 and 6.1). See also Chapter 3 and especially Section 3.1 for more on the problems related to the assumption of discrete segments in continuous signals of speech.

The periodic energy data that was extracted from acoustic recordings of speech is viewed in terms of a mass, i.e., the area under the periodic energy curve, integrating duration and power as the two linked dimensions of quantity in sound (see Turk & Sawusch 1996 on interactions between duration and intensity in linguistic perception contexts). Summing is essentially different from averaging, as well as from peak extraction, in how much strength is assigned to the dimension of duration in the abstract measurement of quantity (duration is absent from peak extraction, it is normalized in averages, and it is strongly influencing the sum).

Table 6.2: Ordinal sonority scores. *Note.* Well-formedness scores with ordinal models. The table demonstrates the predictions we obtain using the two traditional sonority hierarchies, H_{col} and H_{exp}, with each of the two traditional sonority principles, SSP and MSD. Numbers in brackets next to "Rise" reflect MSD's ranking of onset rises by distance – higher values indicate better-formed rises. The scores derived from NAP_{td} on the right column are taken to directly reflect the nucleus competition potential, where higher scores are better-formed.

	Traditional sonority principles				Symbolic NAP
Onset	H_{exp} hierarchy		H_{col} hierarchy		$\textbf{NAP}_{\textbf{td}}$
clusters	$C_2 - C_1$	***SSP(MSD)$_{exp}$***	$C_2 - C_1$	***SSP(MSD)$_{col}$***	$(V - C_1) + (C_2 - C_1)$
plV	$6 - 1 = 5$	**Rise (5)**	$3 - 1 = 2$	**Rise (2)**	$(4 - 1) + (3 - 1) = \mathbf{5}$
flV	$6 - 2 = 4$	**Rise (4)**	$3 - 1 = 2$	**Rise (2)**	$(4 - 1) + (3 - 1) = \mathbf{5}$
smV	$5 - 2 = 3$	**Rise (3)**	$2 - 1 = 1$	**Rise (1)**	$(4 - 1) + (3 - 1) = \mathbf{5}$
vlV	$6 - 4 = 2$	**Rise (2)**	$3 - 1 = 2$	**Rise (2)**	$(4 - 2) + (3 - 2) = \mathbf{3}$
mlV	$6 - 5 = 1$	**Rise (1)**	$3 - 2 = 1$	**Rise (1)**	$(4 - 3) + (3 - 3) = \mathbf{1}$
sfV	$2 - 2 = 0$	**Plateau**	$1 - 1 = 0$	**Plateau**	$(4 - 1) + (1 - 1) = \mathbf{3}$
zvV	$3 - 3 = 0$	**Plateau**	$1 - 1 = 0$	**Plateau**	$(4 - 2) + (2 - 2) = \mathbf{2}$
nmV	$5 - 5 = 0$	**Plateau**	$2 - 2 = 0$	**Plateau**	$(4 - 3) + (3 - 3) = \mathbf{1}$
spV	$1 - 2 = -1$	**Fall**	$1 - 1 = 0$	**Plateau**	$(4 - 1) + (1 - 1) = \mathbf{3}$
lmV	$5 - 6 = -1$	**Fall**	$2 - 3 = -1$	**Fall**	$(4 - 3) + (3 - 3) = \mathbf{1}$
mzV	$4 - 5 = -1$	**Fall**	$1 - 2 = -1$	**Fall**	$(4 - 3) + (2 - 3) = \mathbf{0}$
lvV	$4 - 6 = -2$	**Fall**	$2 - 4 = -2$	**Fall**	$(4 - 3) + (2 - 3) = \mathbf{0}$
msV	$2 - 5 = -3$	**Fall**	$1 - 2 = -1$	**Fall**	$(4 - 3) + (1 - 3) = \mathbf{-1}$
npV	$1 - 5 = -4$	**Fall**	$1 - 2 = -1$	**Fall**	$(4 - 3) + (1 - 3) = \mathbf{-1}$
lpV	$1 - 6 = -5$	**Fall**	$1 - 3 = -2$	**Fall**	$(4 - 3) + (1 - 3) = \mathbf{-1}$

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Importantly, only summing strategies are capable of uncovering the quantitative difference between two sounds that have similar amplitude envelopes yet differ in duration.

The contribution of duration to sonority was convincingly illustrated in the seminal work of Price (1980). Price showed that disyllabic English words like *polite* /pəlʌɪt/ were perceived when the duration of the sonorant /l/ in the superficially related monosyllabic word *plight* /plʌɪt/ was manipulated. Thus, an increase in the duration of the sonorant essentially leads to the perception of another syllable. More supporting evidence on the interaction between duration and syllabic parsing can be found in Dupoux et al. (1999), who showed differences in perception between Japanese and French speakers, and in Berent et al. (2007) as well as Wilson et al. (2014), who analyzed patterns of misperception of Russian onset clusters by English speakers.

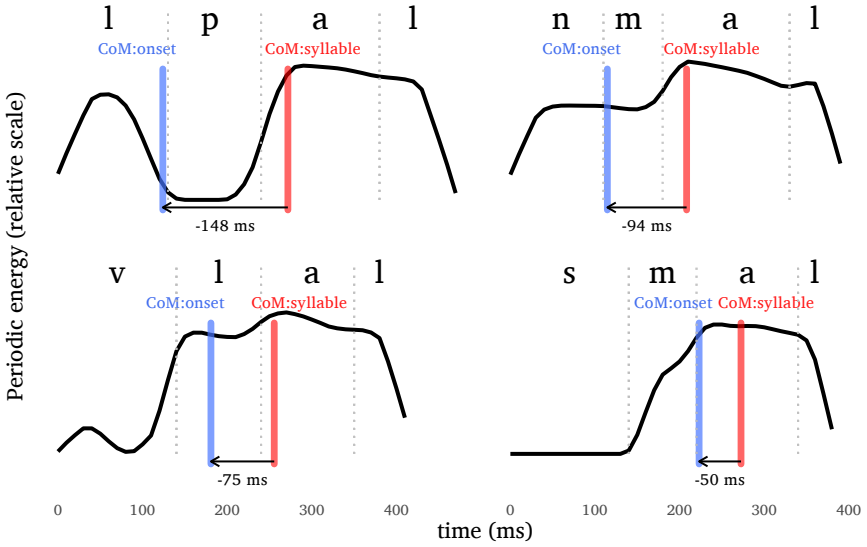

Figure 6.1: Smoothed periodic energy curve (black) of the four syllables from the experimental stimuli – *lpal*, *nmal*, *vlal*, and *smal*. The red vertical line denotes the center of periodic mass of the entire syllable ($CoM_{syllable}$), the blue vertical line denotes the center of periodic mass of the left portion (CoM_{onset}). Grey dotted vertical lines and annotated text denote segmental intervals by manual segmentation (for exposition purposes only). The distance between the two CoM landmarks is indicative of the energy displacement away from the syllabic center, reflecting the nucleus competition potential within the syllable (see details on this measurement in Section 7.2.3)

It is therefore useful to locate the *center of mass* within regions of interest as a

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

measurement that is sensitive to the two axes of periodic energy mass – duration (x-axis) and power (y-axis). The center of mass can be viewed as the point in time in which the area under the curve is split into two equal parts. The location of the center of mass in time (x-axis) is attracted to the peak of the curve (on the y-axis), where it is expected to be found given a perfectly symmetrical shape. However, the center of mass most often diverges from the peak of rise-fall curves so as to reflect asymmetries in the overall distribution of mass. Identification of the center of mass of the periodic energy curve (henceforth CoM) follows a methodology that was introduced with the *Tonal Center of Gravity* (Barnes et al. 2012), in calculating a weighted average time point that uses a continuous time series as the weighting term. The equation in (6.3) is used to locate the average point in time (t), weighted by continuous periodic energy (per) at discrete time points:

$$\mathrm{CoM} = \frac{\sum_i \mathrm{per}_i \, t_i}{\sum_i \mathrm{per}_i} \tag{6.3}$$

The location of the center of periodic energy mass of the entire syllable (henceforth $\mathrm{CoM}_{\mathrm{syllable}}$) guides us to the point in time where the periodic mass of all the competing forces within that syllable are split into two equal parts. Once we obtain this reference point we can repeat this process within the resulting left-side portion, i.e., from the beginning of the syllable up to $\mathrm{CoM}_{\mathrm{syllable}}$, to focus on the onset position (henceforth $\mathrm{CoM}_{\mathrm{onset}}$). We therefore measure the center of mass twice – first for the entire syllable (resulting in $\mathrm{CoM}_{\mathrm{syllable}}$) and then for the left portion of the first measurement (resulting in $\mathrm{CoM}_{\mathrm{onset}}$). The distance between $\mathrm{CoM}_{\mathrm{syllable}}$ and $\mathrm{CoM}_{\mathrm{onset}}$ is indicative of the amount of displacement of energy away from the center of the syllable, which in turn reflects the degree of nucleus competition (see Figure 6.1).

The center of mass is capable of capturing both components of a two-dimensional mass by considering the non-linear shape of the periodic energy curve. The leftward displacement of $\mathrm{CoM}_{\mathrm{onset}}$ relative to $\mathrm{CoM}_{\mathrm{syllable}}$ is affected by the distance, the amplitude, and the amount of discontinuity between the periodic energy at the onset and the center of mass of the entire syllable. Any increase in the above results in a larger distance between the two centers of mass, as Figure 6.1 demonstrates.

6.3 NAP advantages

Before turning to the experimental evidence in Chapter 7, the potential advantages of NAP over traditional models can already be demonstrated with four ex-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

amples that illustrate major differences in the expected model predictions. Consider the clusters in the syllables *spV* (an */s/-stop* cluster), *sfV* (a voiceless fricative plateau), *nmV* (a nasal plateau) and *npV* (a sonority fall from sonorant to voiceless obstruent). In traditional sonority slope terms, all of these clusters are either highly ill-formed (with sonority falls) or border line ill-formed (with sonority plateaus). Predictions may slightly differ with different sonority hierarchies, such that these examples can represent three sonority plateaus and one fall with the H_{col} hierarchy ($npV < spV = sfV = nmv$), or two plateaus and two falls with the H_{exp} hierarchy ($npv = spV < sfV = nmV$).

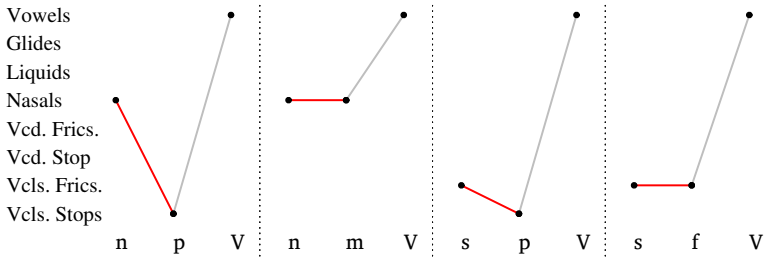

Figure 6.2: Schematic depiction of the sonority slopes of four different onset clusters. The solid red line which denotes the sonority slope of the onset clusters is a plateau in the case of *nmV* and *sfV* and it is falling in the case of *npV* and *spV*. Note that these determinations are based solely on the angle of the red line, regardless of its overall height.

Figure 6.2 schematizes these four examples with traditional sonority slopes, using red lines to denote the portion of the trajectory that represents the relevant slope of the consonantal clusters. These red slopes are level in the onset plateaus *sfV* and *nmV* and falling in the onsets *npV* and *spV* (note again that with the H_{col} hierarchy, *spV* can be considered a plateau; see Figure 2.4). The representation of sonority slopes in Figure 6.2 highlights the irrelevance of the overall height of the slope in traditional sonority formalizations – only the general trend of the slope matters for the characterization of well-formedness.

In contrast to the traditional approach, NAP is explicitly concerned with energetic quantities that compete for the nucleus (as described in Section 5.3). The scores of NAP_{td} reflect the estimated degree of competition such that lower values imply more competition (= worse-formed). In Table 6.2, both *sfV* and *spV* receive the relatively high value 3 (i.e. relatively well-formed). *nmV* receives a lower score of 1 (i.e. relatively ill-formed), and *npV* is almost at the bottom of the NAP_{td} scale with -1 (i.e. clearly ill-formed).

Unlike the symbol-based ordinal scores of NAP_{td}, the continuous signal-based NAP_{bu} makes no a priori predictions via symbols. A few concrete audio stimuli

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

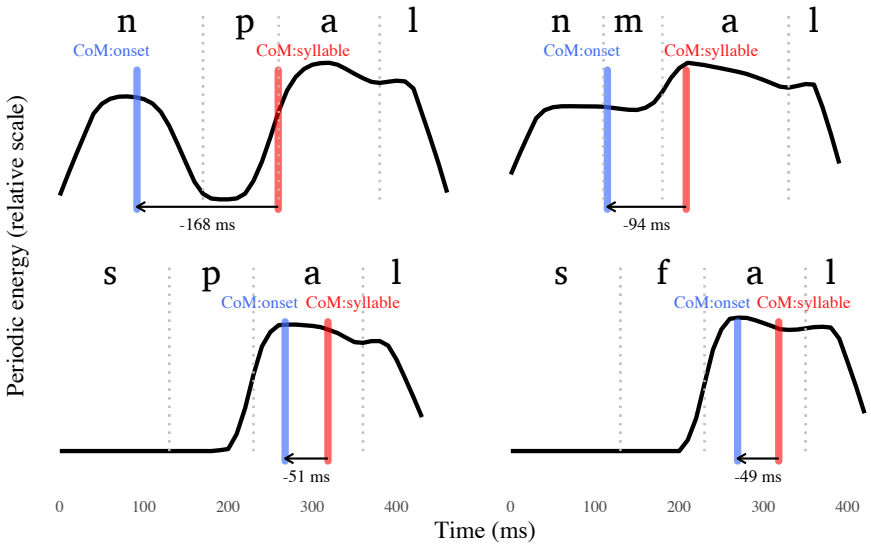

Figure 6.3: Smoothed periodic energy curve (black) of the same four syllables as in Figure 6.2, taken from the experimental stimuli (see Chapter 7). Plot details are described in Figure 6.1.

that were measured in the context of the experiment (see Chapter 7) can nevertheless be shown here to reflect the exact same trend as in the symbolic NAP model. Figure 6.3 shows the periodic energy curve of the four examples with annotated landmarks. As before, the vertical red line denotes the center of periodic mass of the entire syllable ($CoM_{syllable}$), while the vertical blue line denotes the center of periodic mass of the left half of the syllabic mass (CoM_{onset}). A greater distance between the two lines implies more competition (= worse-formed). Here, the distance between $CoM_{syllable}$ and CoM_{onset} is around 50 ms for the two voiceless clusters (***sf****al* and ***sp****al*), it is close to 100 ms for the nasal plateau ***nm****al*, and above 150 ms for the nasal-initial falling sonority slope in ***np****al*.

In NAP terms, the two voiceless clusters *spV* and *sfV* have only minimal, if any, sonorant energy (effectively zero periodic mass) that would make their onset a serious competitor for the nucleus, regardless of the slope. Therefore, even if *spV* exhibits a sonority fall it should not pattern with *npV* in terms of ill-formedness. Likewise, if we consider *spV* as a plateau, neither this nor *sfV* should pattern with *nmV* just because they are all considered plateaus. The two nasal-initial clusters, *nmV* and *npV*, should in fact be considered as much worse-formed than the the two /s/-initial voiceless clusters given their distribution within and across languages. Previous works by Greenberg (1978), Lindblom (1983), Lombardi (1995,

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

1991), Kreitman (2008, 2010) have basically confirmed (although with some considerable differences) that voiceless initial consonant clusters are less *marked* (more common) than voiced clusters, and both types of clusters are less marked than (the less common) voiced-voiceless initial clusters. Such a hierarchy of well-formedness is neatly captured by the rationale and results of NAP, while traditional sonority models regularly make predictions that contradict it to at least some extent.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Part III

Evidence in support of the Nucleus Attraction Principle

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

7 Experimental study

In order to assess NAP-based predictions in situations where both bottom-up and top-down inferences contribute to speech processing, an experimental procedure was designed to collect behavioral responses using a perception task. In what follows I present three experiments: Experiment 1 is a short exploratory pilot study with 12 German-speaking subjects; Experiment 2 is a confirmatory study with 51 German-speaking subjects; and Experiment 3 is a confirmatory study with 33 Hebrew-speaking subjects. This chapter starts by describing the rationale of the experimental design (Section 7.1) before presenting the linguistic and acoustic materials used in the experiments (Section 7.2) and the perception task procedures (Section 7.3). The predictions of the different models are then summarized in Section 7.4, followed by descriptions of the experimental design (Section 7.5), participants (Section 7.6) and our data analysis strategies (Section 7.7). The results and related discussions follow in Section 7.8.

Important notes:

- The design of the model implementations and the ensuing experiments were co-authored with Bruno Nicenboim (University of Potsdam and Tilburg University), who also contributed greatly to the statistical analyses of the results.
- Major parts of this chapter were also published in Albert & Nicenboim (2022).
- The original code and all the materials and data can be found online in an *Open Science Framework* repository at https://osf.io/y477r/.
- The experiments complied with the June 1964 Declaration of Helsinki (carried out by the World Medical Association and entitled "Ethical Principles for Medical Research Involving Human Subjects"), as last revised in accordance with German Research Foundation (DFG) guidelines for experiments with unimpaired adult populations. The ethics approval was obtained by the Principal Investigator (Prof. Dr. Martine Grice). Informed consent from the participants was obtained before each experimental session.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

7.1 Rationale

The goal of the experimental procedure is to tap into the cognitive cost of syllabification processes. To that end, we devised a forced-choice task that allows us to systematically compare response times of forced categorical decisions. Response times are linked with cognitive cost, which, in the context of this task, is understood as the result of nucleus competition. The working assumption is that more competition within a structure makes it cognitively harder for this structure to be parsed as a single syllable, which is reflected in slower processing altogether.

This design uses nonce words to test specific consonantal combinations in structures that either feature two vowels and no consonantal sequences (typically considered to be disyllabic forms) or one vowel with a word-initial sequence (more likely to be considered as monosyllabic forms). This experimental design is reminiscent of many experiments on sonority effects that Iris Berent and her colleagues have published, starting with the seminal Berent et al. (2007).[1] The premise of many of the tasks that Berent et al. test in the context of traditional sonority principles has a slightly different rationale than the one used for NAP, although with very similar predictions: For Berent et al., an ill-formed sonority onset fall, as in the monosyllable *lbV*, is more likely to be confused with disyllabic *lə.bV* when compared with well-formed monosyllable *blV* and its disyllabic counterpart, *bə.lV* (the schwa /ə/ in these examples denotes a generic epenthetic weak vowel). This misperception and confusion between alternatives is expected to be systematically greater with worse-formed sonority clusters, which leads to a drop in categorical accuracy (e.g. "correct" identification of syllable number, or correct detection of similarity in a same/different task) accompanied by a scalar increase in response time (I return to Berent's work in the general discussion in Section 10.2).

Comparable experimental assumptions regarding misperception of consonantal clusters can be found in related works on perception of non-native clusters such as Dupoux et al. (1999) and Davidson & Shaw (2012), including also tasks that utilized the production of such clusters (e.g. Davidson 2010 and Wilson et al. 2014).

To test the different predictions of the six sonority models (SSP_{col}, SSP_{exp}, MSD_{col}, MSD_{exp}, NAP_{td} and NAP_{bu}), we designed a perception task that prompts

[1]Examples of further publications by Berent et al. with various experimental settings that test sonority effects in perception with behavioral data include: Berent et al. (2008, 2010, 2011), Berent, Lennertz & Balaban (2012), Berent et al. (2013), Tamási & Berent (2014), Zhao & Berent (2015), Lennertz & Berent (2015). The following examples also include neurological data: Berent et al. (2014), Gómez et al. (2014), Berent et al. (2015).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

meta-linguistic syllable count judgement with 29 experimental target items. Participants were presented with a collection of speech items that were systematically produced with one or two vowels for each combination of consonants in our set. Only the single-vowel productions were considered as targets, and an accurate response to our targets is always the monosyllabic option (note that the term "accuracy" is used here to describe participants' responses with respect to predictions). By focusing on the response time of "correct" responses to the target words we essentially measure the time it took participants to decide that a given single-vowel stimulus is monosyllabic. We can therefore interpret the reaction times of monosyllabic responses to single-vowel targets as reflective of the processing cost of assigning one nucleus to a given target stimulus with one vowel.

We assume with NAP-based models that this processing cost is tightly related to the nucleus competition between different portions of a syllable, such that response times will reflect the degree of nucleus competition within syllables (more competition = slower responses = worse-formed sequence). Traditional sonority models interpret the processing cost as related to well-formedness in terms of sonority slopes, such that worse-formed clusters are more likely to be misperceived and take longer to process (e.g. Berent et al. 2007, Berent, Lennertz & Balaban 2012, Berent et al. 2008, 2009, Lennertz 2010, Maïonchi-Pino et al. 2015, Sung 2016, Young & Wilson 2017.

The SSP derives a ternary ordinal hierarchy of complex onset well-formedness scores: onset rise > onset plateau > onset fall. This essentially predicts that response times will pattern into three groups, in line with the sonority slope of the onset clusters. MSD models derive a slightly more elaborate ordinal hierarchy, where onset rises with a small sonority distance pattern below onset rises with a larger sonority distance. The latter are predicted to evoke the fastest responses in MSD models.

Note that since the bottom-up predictions of NAP are derived via measurements of acoustic signals of particular productions rather than from fixed symbolic predictions, the assumption that all things other than the controlled variable are equal in the experimental stimuli should hold also for a large degree of variation that occurs in natural speech. Thus, if a certain segment in one item is slightly longer, shorter, louder or softer than in other comparable tokens, bottom-up NAP is designed to directly account for this variation, while the other symbol-based ordinal models essentially assume that such variation is mostly negligible. This allows us to opt for a slightly more ecologically valid experimental paradigm, by using natural speech recordings that were designed and selected to

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

sound as similar as possible, rather than using synthesized speech, which would have allowed a higher degree of similarity between tokens.

7.2 Materials

The experimental design is focused on onset consonantal clusters with two members. These CC combinations are composed from a set of consonants with one of two major *place of articulation* types: either *coronal* or *labial*. This allows us to avoid articulatory effects that may arise from *homorganic* sequences (i.e. adjacent consonants that share the same place of articulation, and may coalesce to some extent as a result) while exploiting both directions of each combination – coronal-labial (back-to-front) and labial-coronal (front-to-back). From an articulatory point of view, there is also an advantage in the fact that the two places of articulation use different main articulators – the tongue tip reaches the palate in coronals, while the lower lip reaches the upper lip or teeth in labials. This relative articulatory independence helps to reduce co-articulation effects of adjacent gestures in consonantal clusters.

The consonantal classes in this study include *stops*, *fricatives*, *nasals*, and *liquids* to reflect the main *manner of articulation* classes in traditional sonority hierarchies (excluding *glides*). The list of considerations and criteria that were used in constructing the experimental stimulus set is presented in Section 7.2.1.

Table 7.1: Experimental stimulus set: CC types. *Note.* VS = voiceless stops; F– = voiceless fricatives; F+ = voiced fricatives; N = nasals; L = liquids; cor = coronal; lab = labial; * = voicing disagreement between obstruents; ** = no labial liquid; *** = dorsal stop /k/ (see list in Section 7.2.1).

	C_1							
	F–		F+		N		L	
C_2	cor-lab	lab-cor	cor-lab	lab-cor	cor-lab	lab-cor	cor-lab	lab-cor
VS	sp, ʃp	ft	*	*	np	mt	lp	lk***
F–	sf, ʃf	fs	*	*	nf	ms	lf	**
F+	*	*	zv	vz	nv	mz	lv	**
N	sm, ʃm	fn	zm	vn	nm	mn	lm	**
L	**	fl	**	vl	**	ml	**	**

Table 7.1 presents the 29 CC types in the experimental set, reflecting 16 different combinations of *manner* classes (16 unique cells in Table 7.1, irrespective of

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

differences in place of articulation). Of the 16 cluster types, 7–8 are considered onset falls, 3–4 are considered onset plateaus (11 total), and 5 are considered onset rises.[2]

Of the 29 different clusters, only three clusters regularly occur in German words (/ʃp, ʃm, fl/), while six clusters are attested to some degree in German loanwords (/sp, sf, sm, vl, zv, ml/; see Van de Vijver & Baer-Henney 2012), and one cluster (/ʃf/) may be considered as similar to German licit clusters with a voiced obstruent following a voiceless one (i.e., /ʃv/ and /cv/). Thus, the experimental set contains 19 clusters that are unattested in German words. These unattested clusters appear in 13 of the 16 unique cluster types. The other three are rising sonority clusters with a liquid in C_2, /fl, vl, ml/, that are attested in German complex onsets to some degree, yet only marginally so in the case of /vl/ and /ml/.

More clusters out of the 29 different cluster types in Table 7.1 occur regularly in Modern Hebrew (see Asherov & Bat-El 2019). These include all of the eight sibilant-initial clusters, /sp, ʃp, sf ʃf, sm, ʃm, zm, zv/, and two liquid-second clusters, /fl, vl/. The voiceless cluster /ft/ and the /m/-initial clusters /ml, mn/ are marginally attested in Modern Hebrew Asherov & Bat-El (2019: 75, 86). Thus, the experimental set contains 16 cluster types that are unattested in Hebrew words. These unattested CC types appear in 12 of the 16 unique combinations in Table 7.1, excluding the three rising sonority clusters with a liquid in C_2 (e.g. /fl, vl, ml/), and the *fricative-stop* clusters (including /ft/), although note that /ft/ and /ml/ are only marginally attested in Hebrew complex onsets.

The different CC sequences were embedded within a /CCal/ word-like frame, with a recurring *-al* rime. These /CCal/ tokens were produced with a single vowel, intended to yield monosyllabic items that resemble typical content words (i.e. prosodically heavier than a single light syllable; see, e.g., Demuth 1996). Two disyllabic counterparts were prepared for each CC type – one with an epenthetic vowel, /CəCal/, and another with a prothetic vowel, /əCCal/ (a more accurate annotation should be /(ʔ)əCCal/, given that the presence of an initial glottal stop was not controlled for). Note that the schwa in the stimulus set recorded by speaker AA was produced as a weak (unstressed) /e/ vowel from the 5-vowel inventory of Modern Hebrew, while in the stimulus set recorded by speaker HN it was produced as a typical German schwa. The entire word set eventually includes 29 single-vowel target types and 58 associated bi-vocalic filler types, adding up to 87 different word-like stimuli.

[2]Depending on whether fricatives are considered higher or similar in sonority to stops, clusters of the type *fricative-stop* may be considered as either an onset fall or an onset plateau.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

7.2.1 Segmental considerations

The following list summarizes concerns that were taken into consideration when constructing the stimulus set (see full set in Table 7.1):

- Glides were excluded from the experimental set due to their complex status, which is dependent on both structure and theory. A glide (sometimes referred to as a *semi-vowel*) is considered a vowel when it is in the nucleus position. A glide immediately adjacent to a nuclear vowel may be analyzed as a vowel in the nucleus, or as a consonant in the onset or coda positions, depending on language and analysis (namely, this depends on whether the language is considered to feature *diphthongs* or not, in itself not always a simple determination). Furthermore, clusters with glides in C_1 are predicted to be ill-formed in all the models we consider, while clusters with glides in C_2 are predicted to be well-formed in all of them. We therefore also do not expect glides to be very informative in the context of this study.

- For the class of liquids, only the lateral /l/ is used, disregarding the subclass of *rhotics* that are phonetically very varied and highly inconsistent between different languages in terms of phonetic detail (see, e.g., Lindau 1985, Ladefoged & Maddieson 1996, Wiese 2001). In that context, it is important to note that the set of stimuli used in this study was created with the intention of being used on speakers of many different languages in which the relevant segments can map to native segments to a comparable degree. There are therefore no liquid plateaus in the experimental set.

- The alveolar /s/ is used for the class of voiceless sibilants (voiceless coronal fricatives). In C_1 positions, the post-alveolar /ʃ/ is also used to control for potential language-specific effects that may appear due to specific restrictions on /s/. For example, in German /ʃC/ onset clusters can be licit, while /sC/ onset clusters occur only marginally in loanwords.

- Stops are used only in C_2 position, and only voiceless stops are used in order to keep the size of the stimulus set reasonably small. Stops in C_1 position are avoided since it is also the phrase-initial position of the stimuli, which is practically devoid of acoustic cues for the closure phase of the stop. Within the stream of speech, the movement of articulators towards the target of a stop's closure phase leaves auditory traces from the preceding segment and into the closure of the stop, containing important

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

information about the identity of the stop (e.g. Barry 1984). In that sense, a stop in C_1 position at the beginning of a phrase contains only a transient release burst. Furthermore, note that all the stop-initial clusters (with the exclusion of *stop-stop* plateaus) are generally well-formed according to all sonority models tested here, such that their added value in this comparison would have been smaller than their cost (in terms of the size of the stimulus set).

- The set includes one instance of the dorsal consonant /k/ instead of the coronal /t/ as an alternative to a labial-coronal cluster, which would have required a labial liquid. Instead, the *coronal-dorsal* cluster /lk/ is used to retain the same direction of a labial-coronal cluster – both are front-to-back in terms of places of articulation. There are no other dorsals in the set (i.e. no fricative, nasal or liquid dorsal), as these tend to be relatively more marked and less consistent between languages.

- Lastly, sequences of obstruents that differ in voicing are avoided due to the cross-linguistic tendency of obstruent clusters to agree in voicing (see, e.g., Cho 1990), although note that German allows /ʃv/ and /cv/ clusters while banning /ʃf/.

7.2.2 Audio recordings

Audio stimuli for the experiment were recorded by a phonetically trained native Hebrew speaker, AA (the author), and a phonetically trained native German speaker, HN, in a sound-attenuated booth at the phonetics laboratory of the University of Cologne. Speech was recorded via a head-mounted headset condenser microphone (AKG C420), capturing mono digital audio files at a resolution of 44.1 kHz sample-rate and 24 bit depth with a *Metric Halo MIO 2882* audio interface. Selected audio takes were treated in the original high resolution for DC offset correction and compression of ultra low frequencies under 52 Hz (to compensate for some room reverberation effects). Audio was then downgraded from 24 to 16 bit depth with *Goodhertz Good Dither* dithering to be used in the perception task running on *OpenSesame* 3.1.9 (Mathôt et al. 2012). The audio that was submitted to analyses by the *APP Detector* (see Section 7.2.3) was also downgraded in sample-rate to 16k Hz. Finally, all audio takes, at all resolutions, were normalized to the same RMS target of −20 dBFS (*dB Full Scale*).

To record the stimuli, the combined 87 word-like stimuli (29 targets and 58 fillers) were embedded within carrier sentences in non-final position and produced with default declarative intonation, in order to maintain consistent prosody.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Carrier sentences were also designed to minimize potential effects of resyllabification as well as co-articulation by controlling the segmental makeup immediately preceding and following target words (see examples in (1–2)).

(1) **/ze maʁ.gíʃ *CCal* ka.ʁé.ga/** (Hebrew: 'it feels (like) *CCal* at the moment')

(2) **er muss *CCal* kaufen** (German: 'he must buy *CCal*')

The original sentence elicitation lists are available at the OSF repository in mixed Hebrew, German and phonemic transcripts in the form of the PowerPoint presentations that were used in this self-paced task (see link in the opening notes of Chapter 7).

7.2.3 Obtaining periodic energy data

Continuous measurements of periodic energy from acoustic signals were extracted for the experiments using the *Aperiodicity, Periodicity and Pitch Detector* (APP Detector), a computer code that was introduced in Deshmukh & Espy-Wilson (2003) and developed in subsequent publications (Deshmukh et al. 2005, Vishnubhotla 2007). The APP Detector has the ability to measure the spectral distribution of periodic energy from digital audio files with a 16k Hz sample-rate, effectively measuring periodic energy up to 8k Hz (more than sufficient for speech, see Section 5.2). The periodic energy data was exported from the APP Detector's Matlab analysis tables into R (R Core Team 2018) for further data manipulation, visualization, modelling, and statistical analysis.

To obtain the periodic energy curve, it is necessary to first sum over the different frequencies that the APP Detector measures at each time point (every 10 ms) to create a time series of *periodic power*. Next, a smoothed curve is fitted to the periodic power time series with *Tukey's (Running Median) Smoothing* ("3RS3R"), to eliminate small-scale fluctuations in the periodic power curve. Finally, the periodic power time series is log-transformed to yield *periodic energy*, see Equation (7.3).

Within the log-transform function we can plug a value that reflects the threshold of effective voicing periodicity to set a meaningful zero for the periodic energy floor – the *periodic floor* in Equation (7.3). This is similar to the standard *dB SPL* measurement (SPL stands for *Sound Pressure Level*), which plugs a generic value that represents the threshold of human hearing in terms of sound pressure into the denominator of the log-transform function. In this way, SPL suggests a shared reference for different dB measurements that use the zero value to denote the low end of human hearing. In the case of the current periodic energy

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

measurement the threshold of the floor is not a universal determination but a calibration that allows us to take the audio quality and the inner-workings of the APP Detector into account. The effective periodicity threshold for the log-transform of the periodic energy time series was determined by extracting the maximal periodic power value obtained for voiceless portions in the given set. To be sure that there was no marginal voicing in these samples, only voiceless C_1 consonants that precede another voiceless consonant in C_2 were measured. In this way, the value 0 in our periodic energy curve is optimally calibrated to reflect the low end of pitch-related periodic components in the signal.

$$\text{periodic energy} = 10 \log_{10}\left(\frac{\text{periodic power}}{\text{periodic floor}}\right) \tag{7.3}$$

Note that the periodic energy curve is smoothed further with Local Polynomial Regression Fitting (*loess*) in the figures shown in this chapter. This is only used for additional visual clarity. All the acoustic analyses are based on the periodic energy curve before this final aesthetic smoothing (and after the other processes mentioned above). The codes of all the above processes are available at the OSF repository (see link in the opening notes of Chapter 7).

7.3 Perception task procedures

Recall that the experiments were designed as a forced-choice 2-alternative perception task, where accuracy and response time information were collected. To normalize response times, the countdown in each trial started in the middle of the transition from C_2 to /a/, illustrated with the location of the dash in *(ə)C_1(ə)**C_2–al***. This zero time point was determined individually for each one of the 87 stimuli, capitalizing on the fact that all stimuli share the rime *al*, which is fully predictable in the context of the experiment, in contrast to the unpredictability of preceding material (the predictability of the *al* rime was assumed to become evident already in the training phase, before any data were collected for analysis). Manual segmentations conducted by the author were used to determine this point for each target. Eventually, response times shorter than 100 ms (i.e. 100 ms after the zero point between C_2 and /a/) were considered as too fast to be valid and were therefore excluded. This threshold led to only one observation being excluded from Experiment 2.

Participants were seated in a quiet room in front of a laptop computer (a MacBook Air 13-inch, Early 2014) running the experiment on OpenSesame 3.1.9 (Mathôt et al. 2012), where they listened to the stimuli through a set of closed

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

headphones (Sennheiser HD 201), fed directly from the laptop's internal audio interface. After verifying that participants shared a standard understanding of the notion of the syllable with a few examples of words in their language (German or Hebrew) with one and two syllables (e.g. German *See*, *Spaß*, *Quark*, *Angst* vs. *Schu-le*, *Kin-der*, *Bre-zel*, *Pflau-men*), they were instructed to listen to nonce words in an "unknown" foreign language. Nonce words were used and a foreign language was mentioned in order to increase reliance on bottom-up processing in the task as much as possible. The meta-linguistic task may otherwise strongly have favoured top-down inferences. To that end, it was important to use recordings of a speaker with a foreign native language compared to the participants' L1. The German-speaking listeners in Experiments 1–2 heard speech recording of a native Hebrew speaker and the Hebrew-speaking listeners in Experiment 3 heard speech recording of a native German speaker.

Participants were instructed to respond quickly and accurately whether they heard one or two syllables by using their left and right index fingers to choose 1 or 2 at the location of the "F" (for 1) and "J" (for 2) keys on a QWERTY keyboard layout (relevant keys were covered with salient red-on-white "1" and "2" stickers).

A training session of ten trials preceded the experimental blocks, allowing the participants to familiarize themselves with the task, and allowing the experimenter to adjust listening volume and monitor potential problems and misunderstandings regarding the task.

7.4 Summary of predictions

The full set of predictions for the 29 experimental targets is presented for all the symbol-based ordinal models (SSP_{col}, SSP_{exp}, MSD_{col}, MSD_{exp} and NAP_{td}) in Table 7.2, and for the signal-based continuous model (NAP_{bu}) in Figures 7.1–7.2. Note that the scores of NAP_{bu} are presented on a continuous ratio scale, with specific predictions for each token and consequential intervals between scores. The scores in NAP_{bu} are not a generalization (nor are they based on averages). Rather, they were extracted from the specific set of recordings, and they are expected to vary to some extent when measuring different tokens. NAP_{bu} scores are presented for the two sets of stimuli used in the experiments: a set spoken by a native Hebrew speaker (Figure 7.1) and a set spoken by a native German speaker (Figure 7.2).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

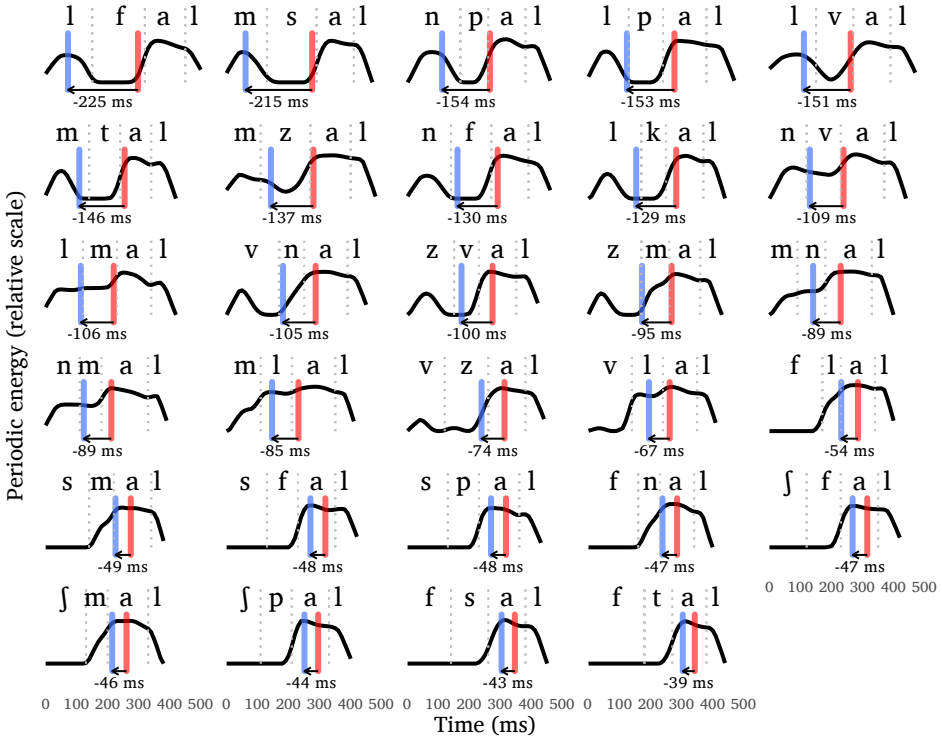

Figure 7.1: AA set (Hebrew speaker). Well-formedness scores in the continuous NAP_{bu} model shown in terms of the distance between the center of mass of the entire syllable, $CoM_{syllable}$ (red vertical lines), and the center of mass of the left portion, CoM_{onset} (blue vertical lines). See Section 6.2.4 for details. Periodic energy is represented by the black curve. Grey dotted vertical lines and annotated text denote segmental intervals by manual segmentation (for exposition purposes only). Items are ordered by score (from worse- to better-formed), going from left-to-right and from top-to-bottom.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table 7.2: Well-formedness scores for the 29 experimental items using the five ordinal models that are based on symbolic phonemes: $SSP_{col/exp}$, $MSD_{col/exp}$, and NAP_{td}. Positive values indicate a rise (rs), negative values a fall (fl), and 0 a plateau (plt). *Note.* Higher values predict better-formed onset clusters in an ordinal scale (i.e. magnitude of differences between values cannot be inferred from these models).

Onset cluster types	SSP_{col}	SSP_{exp}	MSD_{col}	MSD_{exp}	NAP_{td}
fl	1 (rs)	1 (rs)	2 (rs)	4 (rs)	5
sm, ʃm, fn	1 (rs)	1 (rs)	1 (rs)	3 (rs)	5
vl	1 (rs)	1 (rs)	2 (rs)	2 (rs)	3
zm, vn	1 (rs)	1 (rs)	1 (rs)	1 (rs)	3
ml	1 (rs)	1 (rs)	1 (rs)	1 (rs)	1
sf, ʃf, fs	0 (plt)	0 (plt)	0 (plt)	0 (plt)	3
zv, vz	0 (plt)	0 (plt)	0 (plt)	0 (plt)	2
nm, mn	0 (plt)	0 (plt)	0 (plt)	0 (plt)	1
sp, ʃp, ft	0 (plt)	−1 (fl)	0 (plt)	−1 (fl)	3
lm	−1 (fl)	−1 (fl)	−1 (fl)	−1 (fl)	1
mz, nv, lv	−1 (fl)	−1 (fl)	−1 (fl)	−1 (fl)	0
ms, nf, np, mt, lf, lp, lk	−1 (fl)	−1 (fl)	−1 (fl)	−1 (fl)	−1

7.5 Designs

The details in the following analyses address three separate experiments: *Experiment 1*, an exploratory pilot experiment with 12 German-speaking subjects listening to stimulus set AA (Hebrew speaker); *Experiment 2*, a confirmatory experiment with 51 German-speaking subjects listening to stimulus set AA; and *Experiment 3*, a confirmatory experiment with 33 Hebrew-speaking subjects listening to stimulus set HN (German speaker).

Given the various novelties in this proposal, the methodologies for data collection, data extraction, and model implementation were first tested on a small body of real data that we collected before finalizing our methodologies (namely, the model implementations in Chapter 6 and the various procedural details in Section 7.2). We used this exploratory study to test our methodologies and to explore the possibilities for properly estimating nucleus competition in each of the NAP models.

We also used the exploratory pilot study to verify that the number of participants is large enough with respect to the size of expected effects. With 12 participants, we could already observe clear effects (see Section 7.8). To be confident

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

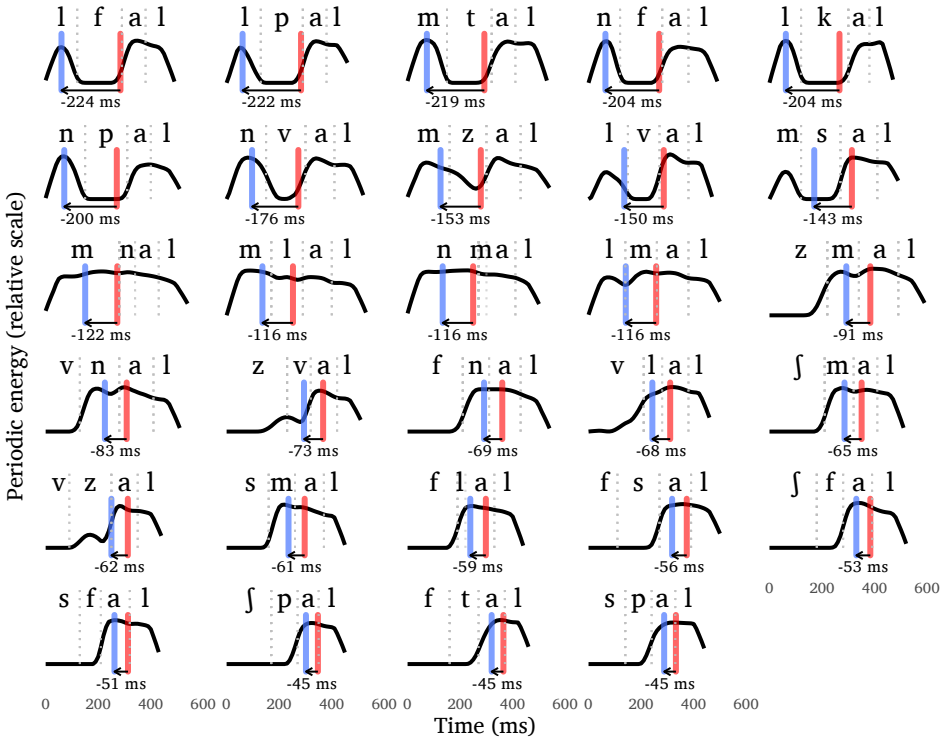

Figure 7.2: HN set (German speaker). See previous figure (Figure 7.1) for plot details.

that we have enough power to compare the models, we aimed at 50 participants in the confirmatory studies (note that this goal was partially reached in Experiment 3 due to the COVID-19 pandemic).

The exploratory pilot study was conducted in two versions, each with half of the fillers and all of the targets in one block, yielding a total of 58 data points per subject (29 fillers + 29 targets, no repetitions). The two different versions were evenly split between participants (each version was presented to six participants).

Experiments 2 and 3 are the main confirmatory studies conducted after finalizing our hypotheses and methodologies with the data from Experiment 1. The difference between Experiments 2 and 3 concerns the native language of the subjects, and, as a consequence, the stimulus set in use. Experiment 2 tested German-speaking subjects on stimulus set AA, featuring a Hebrew speaker, while Experiment 3 tested Hebrew-speaking subjects on stimulus set HN, featuring a German

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

speaker (see explanation in Section 7.1). Each experimental block in Experiments 2-3 consisted of two repetitions of the target words ($2 \times 29 = 58$) and one trial of each filler word (1×58). The experiment consisted of two blocks with randomized trials, generating altogether four repetitions of the target words ($4 \times 29 = 116$) and two repetitions of the filler words ($2 \times 58 = 116$), yielding a total of 232 data points per subject.

7.6 Participants

7.6.1 Experiment 1

The exploratory pilot study consisted of 12 subjects (two males and ten females), all native German-speaking students from the Technische Hochschule Köln, who volunteered to participate in the study. The experiment was administered in a quiet room at one the institute's buildings in Cologne. The mean age of participants in the pilot study was 25 (21–30 range).

7.6.2 Experiment 2

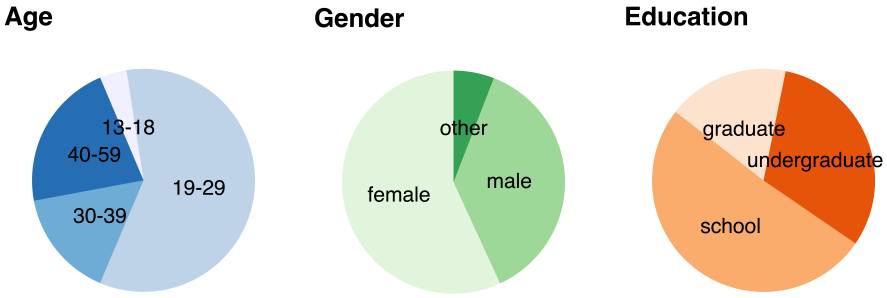

Figure 7.3: Participants in Experiment 2 ($n = 51$). Education categories refer to academic achievements ("school" = academic degree not yet acquired).

Fifty-one native German speakers (who did not participate in the exploratory pilot study) participated in Experiment 2, of which 48 were monolingual (the 3 bilingual speakers had Polish, Low German, and Hebrew as their heritage language). 49 participants were right-handed. See more details on age, gender and education of participants in Figure 7.3.

Of the 51 participants, 34 were students at the University of Cologne who took part in the experiment at the sound-attenuated booth of the phonetics laboratory.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

The other 17 participants took part in the experiment at three different locations – all small quiet rooms within private apartments. All subjects were paid five Euros for their participation.

We excluded the responses from one participant who failed in our participant inclusion criterion requiring accuracy of at least 75% with bi-vocalic fillers. The bi-vocalic fillers of the forms /CəCal/ and /əCCal/ link correct responses to the disyllabic choice (2), and we expect relatively few monosyllabic choices (1) in response to stimuli with two separate vowels. Indeed, the overall average accuracy of all 51 participants, when responding to bi-vocalic filler stimuli, was 96%. The excluded participant achieved a much lower accuracy score for bi-vocalic fillers, almost approaching chance-level with 65%.

7.6.3 Experiment 3

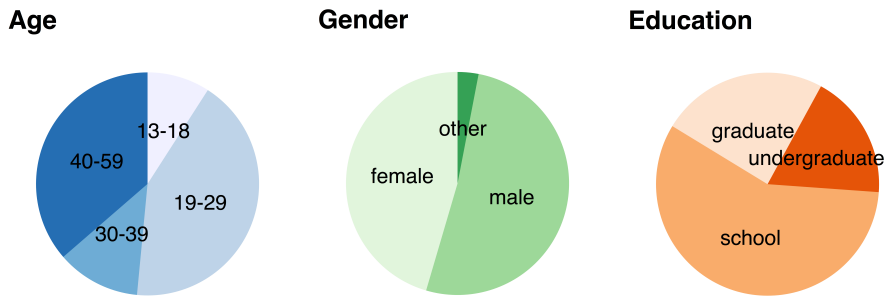

Figure 7.4: Participants in Experiment 3 (n = 33). Education categories refer to academic achievements ("school" = academic degree not yet acquired).

Thirty-three native Hebrew speakers participated in Experiment 3, of which 28 were monolingual (the five bilinguals were also native speakers of English, Russian and Spanish). 28 participants were right-handed. See more details on age, gender and education of participants in Figure 7.4.

The data collection in Experiment 3 was more diverse, and, perhaps therefore also more "noisy" than in Experiment 2. The first round of data collection took place in 2019 with student volunteers from Tel Aviv University and The Hebrew University of Jerusalem. The second round of data collection took place during the early phases of the global COVID-19 pandemic, which resulted in fewer overall participants and the use of different ad-hoc and suboptimal locations to administer the experiment.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

7.7 Data analysis

We used a Bayesian data analysis approach implemented in the probabilistic programming language *Stan* (Stan Development Team 2018b) using the model wrapper package *brms* (Bürkner 2017, Bürkner 2018) in *R* (R Core Team 2018).[3] An important motivation for using the Bayesian approach is that it facilitates fitting fully hierarchical models with the so-called "maximal random effect structure", which provide the most conservative estimates of uncertainty (Schielzeth & Forstmeier 2009). In all our models, we used regularizing priors (detailed below). These priors are minimally informative and have the objective of yielding stable inferences (Chung et al. 2013, Gelman et al. 2008, 2017). Nicenboim & Vasishth (2016) and Vasishth et al. (2018) discuss the Bayesian approach in detail in the context of psycholinguistics and phonetics. We fitted the models with four chains and 4000 iterations each, of which 1000 iterations were the warm-up phase. In order to assess convergence, we verified that there were no divergent transitions, that all the $\hat{R}$ (the between- to within-chain variances) were close to one, that the number of effective sample size was at least 10% of the number of post-warmup samples, and visually inspected the chains.

For the statistical models, we took into account that the traditional sonority models and the top-down version of NAP (i.e. SSP_{col} , SSP_{exp}, MSD_{col}, MSD_{exp}, and NAP_{td}) are ordinal models, while the bottom-up version of NAP (NAP_{bu}) is a continuous model. The ordinal models predict that certain groups of onset clusters will be better or worse-formed than other group depending on an ordinal score, but they do not assume that the score will be equidistant with respect to its effect on the response variable, log-transformed response times. For this reason, the discrete scores of these models are assumed to have a monotonic effect on the log-response time in our task, that is, having a monotonically increasing or decreasing relationship with the log-response time, while the distance between groups is estimated from the data (Bürkner & Charpentier 2018).

[3]The complete list of *R* packages and versions that we used is: R (Version 3.6.3; R Core Team 2018) and the R-packages *brms* (Version 2.16.3; Bürkner 2017, Bürkner 2018), *Cairo* (Version 1.5.12; Urbanek & Horner 2020), *dplyr* (Version 0.8.5; Wickham, François, et al. 2020), *ggplot2* (Version 3.3.0; Wickham, Chang, et al. 2020), *ggrepel* (Version 0.8.2; Slowikowski 2019), *hexbin* (Version 1.28.1; Carr et al. 2018), *loo* (Version 2.4.1; Yao et al. 2017), *purrr* (Version 0.3.4; Henry & Wickham 2020), *R.matlab* (Version 3.6.2; Bengtsson 2018), *Rcpp* (Version 1.0.4.6; Eddelbuettel & François 2011, Eddelbuettel & Balamuta 2017), *readr* (Version 1.3.1; Wickham et al. 2018), *rstan* (Version 2.19.3; Stan Development Team 2018a), *StanHeaders* (Version 2.21.0.1; Stan Development Team 2018c), *stringr* (Version 1.4.0; Wickham 2019), and *tidyr* (Version 1.0.2; Wickham & Henry 2018).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

In contrast, NAP_{bu} provides scores on a *ratio scale*, in which the distance between scores is also taken to be informative (as opposed to the *ordinal* scales of the other models), which is modeled with a continuous predictor which is assumed to have a linear relationship with the log-response times. Finally, as a baseline, we fitted a "null" model which assumes no relationship between the stimuli and the response times.

All the models included a random intercept and slope by subjects (except for the null model that included only a random intercept) and the following weakly regularizing priors: $\text{Normal}(6, 2)$ for the intercept, $\text{Normal}(0, 1)$ for the slope, $\text{Normal}_{+}(0, 1)$ for the variance components, and $lkj(2)$ for the correlation between by-participant adjustments. The ordinal models also have a Dirichlet prior for the simplex vector that represents the distance between the categories set to one for each of its parameters.

We evaluated the models in three different ways: (i) estimation, (ii) descriptive adequacy, and (iii) model comparison.

Estimation: We report mean estimates and 95% quantile-based Bayesian credible intervals. A 95% Bayesian credible interval is interpreted such that it contains the true value with 95% probability given the data and the model (see, for example, Jaynes & Kempthorne 1976, Morey et al. 2016).

Descriptive adequacy: We used posterior predictive checks to examine the descriptive adequacy or "fit" of the models (Shiffrin et al. 2008). The observed data should look plausible under the posterior predictive distribution of the models. The posterior predictive distribution of each model is composed of simulated datasets generated based on the posterior distributions of its parameters. Given the posterior of the parameters of the model, the posterior predictive distribution shows how similar data may look. Achieving descriptive adequacy means that the current data could have been predicted with the model. It is important to notice that a good fit, that is, passing a test of descriptive adequacy, is not strong evidence in favor of a model. In contrast, a major failure in descriptive adequacy can be interpreted as strong evidence against a model (Shiffrin et al. 2008). Thus, we use posterior predictive checks to assess whether the model behavior is reasonable and in which situations it is not (see Gelman et al. 2013 for further discussion).

Model comparison: For model comparison, we examine the out-of-sample predictive accuracy of the different models using k-fold ($k = 15$) cross-validation

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

stratified by subjects.[4] Cross-validation evaluates the different models with respect to their predictive accuracy, that is, how well the models generalize to new data.

7.8 Results

7.8.1 Estimations

For all the models, the well-formedness score shows a clear effect on response times, with lower scores yielding longer log-transformed response times (see Table 7.3).

Notice that the posterior of the effect of well-formedness, $\hat{\beta}$, is not comparable across models. For the ordinal models, it represents the average increase (or decrease) in the dependent variable associated with two neighboring factor levels, or in other words, $\hat{\beta}$ multiplied by the number of categories minus one represents the increase in log-scale between the first and the last category. This means that it is highly affected by the number of categories. For the continuous bottom-up model, NAP_{bu}, β represents, the increase in log-scale for one unit in the well-formedness scale. To give some concrete examples from set AA, there are 24 units (since their NAP scores are -129 and -153, respectively) between /lpal/ and /lkal/; and there are 81 units between /lkal/ and /spal/ (-48 and -129, respectively). However, for all the models, $\hat{\beta}$ is negative, indicating that well-formedness is associated with faster responses. See Appendix A for the complete output of the models.

The results shown here reflect the final state of the models in the exploratory stage, which is the same as the state of the models in the confirmatory stage. Importantly, the results of the confirmatory studies, Experiments 2–3, which are statistically much more robust, remain consistent with those of Experiment 1, which had a relatively small number of observations. As such, Experiment 1 was not designed to distinguish between the models and it will not be considered in the further presentation of results.

7.8.2 Descriptive adequacy

The model fits of the different models are shown in Figures 7.5–7.11. The plots in these figures present the dispersion of the average response time results, depicted

[4] Pareto smoothed importance sampling approximation to leave-one-out cross-validation (implemented in the package `loo`, `vehtariPracticalBayesianModel2017`, Vehtari et al. 2015) failed to yield stable estimates.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table 7.3: Estimations.

	$\hat{\beta}$	95% CrI	
Experiment 1			
SSP_{col}	−0.18	[−0.27,	−0.087]
SSP_{exp}	−0.1	[−0.19,	−0.011]
MSD_{col}	−0.13	[−0.2,	−0.056]
MSD_{exp}	−0.052	[−0.11,	0.0016]
NAP_{td}	−0.079	[−0.13,	−0.03]
NAP_{bu}	−0.003	[−0.0054,	−0.00068]
Experiment 2			
SSP_{col}	−0.14	[−0.17,	−0.11]
SSP_{exp}	−0.066	[−0.084,	−0.048]
MSD_{col}	−0.099	[−0.12,	−0.078]
MSD_{exp}	−0.039	[−0.049,	−0.03]
NAP_{td}	−0.071	[−0.085,	−0.058]
NAP_{bu}	−0.0027	[−0.0032,	−0.0021]
Experiment 3			
SSP_{col}	−0.066	[−0.096,	−0.036]
SSP_{exp}	−0.043	[−0.065,	−0.02]
MSD_{col}	−0.045	[−0.066,	−0.024]
MSD_{exp}	−0.02	[−0.032,	−0.009]
NAP_{td}	−0.032	[−0.048,	−0.017]
NAP_{bu}	−0.00097	[−0.0015,	−0.00049]

as red points for related CC clusters, vis-à-vis each models' predictions in the form of distributions, depicted with blue violins. The order of the stimuli, from left to right, follows from the models' scores such that predictions for better-formed clusters appear further to the right. Recall that scores in the NAP_{bu} model yield slightly different predictions for each stimulus set (AA vs. HN).

7.8.2.1 Null models

The null models are shown in Figure 7.5 as baselines in the respective experiments (the order of stimuli along the x-axis follows the NAP_{bu} scores, but in a forced ordinal scale, with equidistant intervals). The slight differences in predictions for different clusters are due to individual differences in the accuracy. Recall

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

that we subset the response times conditional on the monosyllabic response (1) to the forced-choice task. This means that when participants give more monosyllabic answers for a specific cluster, their adjusted intercept will have a greater influence on the predictions of the model for that cluster. In addition, clusters with fewer monosyllabic responses show more variability in their predictions (e.g. /lf/ vs. /fl/ in the AA set, on the left side of Figure 7.5).

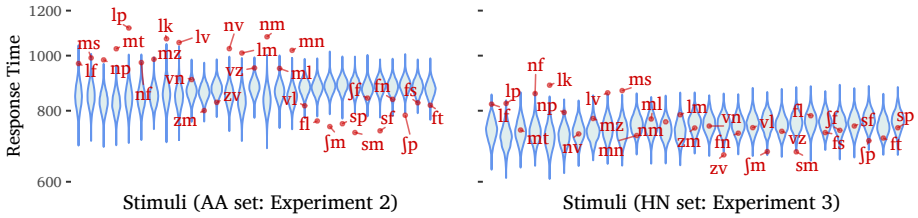

Figure 7.5: Null model fit. Observed mean log-transformed response times are depicted with red points, distribution of simulated means based on the null model are depicted with blue violins.

7.8.2.2 SSP and MSD models

We consider a good fit in the case of the ordinal models to be roughly characterized by the following three criteria: (i) the data are contained within the predictions, i.e., the red points appear within the respective violins; (ii) the data are consistent within each predicted level, i.e., the vertical dispersion of red points pattern together around the same area within each level (preferably in the middle of the distribution); and (iii) the model predictors are not redundant, i.e., the violins of the different model levels show little overlap between them.

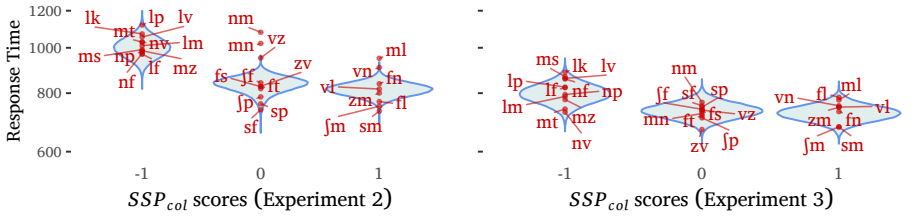

Figure 7.6: SSP_{col} model fit. Stimuli ordered from left to right according to their score in the model in ascending well-formedness (other details are the same as above).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

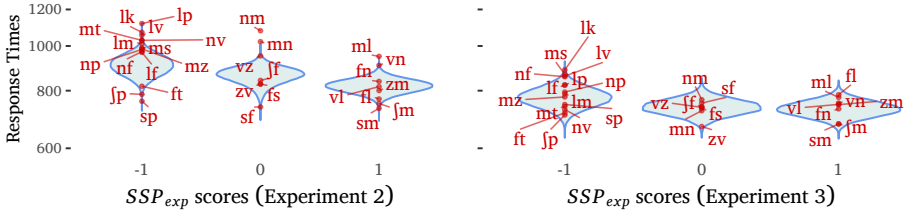

Figure 7.7: SSP_{exp} model fit (plot details are the same as above).

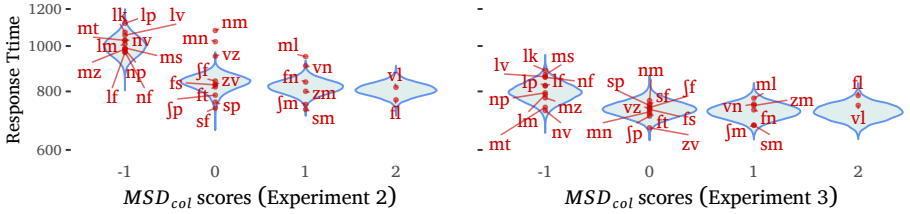

Figure 7.8: MSD_{col} model fit (plot details are the same as above).

A quick glance at the four plots for Experiment 2, in the left panels of Figures 7.6–7.9, reveals a common failure of all the traditional sonority models to contain the nasal plateaus (/mn/ and /nm/) within their predicted distribution alongside all the other plateaus (0 model score in all figures). Furthermore, the data within the 0 plateau levels appears to be broadly dispersed for the German-speaking subjects in Experiment 2 (plots on the left side) but quite well centered for the Hebrew-speaking subjects in Experiment 3 (right plots).

A comparison of the left-most violin in Figures 7.6–7.9 highlights some differences between the two sonority hierarchies H_{col} (*SSP/MSD*$_{col}$) and H_{exp} (SSP/MSD_{exp}). The left-most violins reflect the onset fall levels of the SSP and MSD models. For the Hebrew-speaking subjects in Experiment 3 (right panels), there was no clear

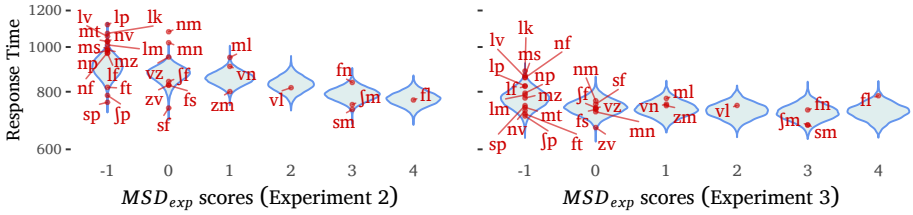

Figure 7.9: MSD_{exp} model fit (plot details are the same as above).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

difference between the two sonority hierarchies and a similar, broad distribution appears in all fits of sonority falls. In contrast, the reponse times of German-speaking subjects in Experiment 2 exhibit a bimodal distribution in the falling onsets of sonority models that use the H_{exp} hierarchy (SSP/MSD_{exp}), whereby *fricative-stop* clusters /ʃp, sp, ft/ are considered to be highly ill-formed onset falls.

This suggests that the H_{col} hierarchy (where all obstruents are grouped into one class on the sonority hierarchy such that *fricative-stop* clusters are considered plateaus) is better than the H_{exp} hierarchy in treating *fricative-stop* clusters. This can be deduced from the better model fits for onset sonority falls and plateaus when the H_{col} hierarchy is applied (*SSP/MSD*$_{col}$ vs. SSP/MSD_{exp}). However, the difference between the two sonority hierarchies also plays a role in the grouping of onset rises when the MSD-based models are taken into account.

The violins in the right panel of each plot, reflecting well-formed onset rises with positive model scores, present three types of grouping across the four models. The two SSP models (*SSP*$_{col/exp}$) make identical predictions with respect to onset rises, lumping all rises into one category (1 in Figures 7.6–7.7). This, again, results in a broader distribution for the German-speaking subjects in Experiment 2 (left panels) compared to the Hebrew-speaking subjects in Experiment 3 (right panels).

The MSD models present multiple levels of well-formedness for onset rises. MSD_{col} exhibits two levels of rises (1–2 in Figure 7.8) while MSD_{exp} exhibits four levels of rises (1–4 in Figure 7.9). This elaboration seems to be beneficial in fitting the scores of the German-speaking subjects to the 4 rise levels of MSD_{exp}, but less so for MSD_{col}. Furthermore, the additional levels of the MSD are redundant, or even worse, for the fits of the scores of the Hebrew-speaking subjects in Experiment 3 (right plots).

To conclude, an observation of the model fits of the four traditional sonority models in the two confirmatory studies reveals a mixed picture. The H_{col} hierarchy (in models *SSP/MSD*$_{col}$) appears to result in a better fit with onset falls and plateaus, especially for the German-speaking subjects. The competing H_{exp} hierarchy appears to be advantageous when fitting the scores of rising onset slopes, but mostly with MSD_{exp} and only for the German-speaking subjects in Experiment 2, where sonority falls exhibit an undesirable bimodal distribution.

7.8.2.3 NAP models

Although NAP_{td} is an ordinal model like all the traditional sonority models, it follows a different rationale (see Section 6.2.2), whereby the scores of the model estimate nucleus competition to reflect well-formedness.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

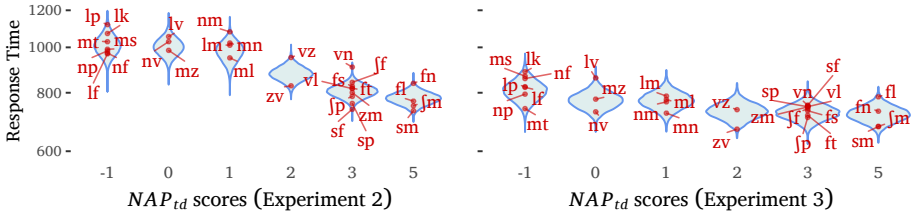

Figure 7.10: NAP_{td} model fit (plot details are the same as above).

Figure 7.10 shows that NAP_{td} succeeds in containing all the data (points) within the respective predictions (blue violins) in both experiments, making NAP_{td} the only model to achieve such coverage. NAP_{td} appears to exhibit some redundancy, as suggested by the relatively large degrees of overlap between some of the predictive distributions of the model. This is apparent from the overlap between violins in the left side (worse-formed) of the model fit with Experiment 2 (left panel), and between violins in the right side (better-formed) of the model fit with Experiment 3 (right panel) in Figure 7.10.

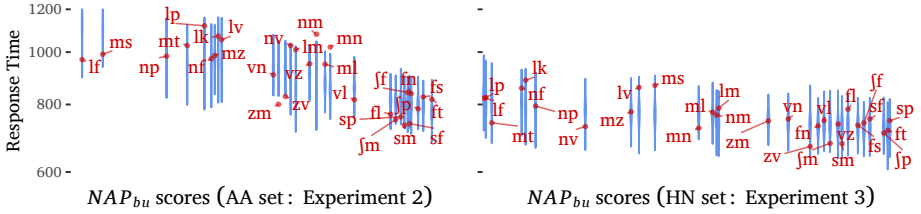

Figure 7.11: NAP_{bu} model fit (plot details are the same as above).

NAP_{bu} is different from all the other models in that it presents scores that are specific to each token in a continuous ratio scale, rather than an ordinal scale (i.e. the distances between scores in the model are also predicted). Importantly, the expected correlation between response time and ill-formedness appears to hold for the model fits of NAP_{bu} in Figure 7.11.

Our criteria for goodness of fit based on the plot analyses (see Subsection 7.8.2.2) are not all valid when evaluating NAP_{bu} since we have no classes and no vertical dispersion of data (points) within levels, and since the horizontal overlap of predictions (violins) between levels requires a different interpretation. However, the criterion for inclusion of data points within the violins of the models' predictions naturally also holds for the NAP_{bu} fit, which fails to include the data for the nasal plateaus /nm/ and /mn/ within the respective predictive distribution

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

in Experiment 2 (a failure that is shared by all the traditional models in Experiment 2; see Section 7.8.2.2). Furthermore, in Experiment 2 NAP_{bu} also fails to include the /z/-initial clusters – /zm/ and /zv/ – within their respective predictive distribution.

The failures in the fit of the NAP_{bu} model with German-speaking subjects in Experiment 2 can be split into two types: (i) nasal-initial clusters – *nval*, *nmal*, and *mnal* – which received results on a par with the slowest responses in the data, reflecting an overestimation of well-formedness by the model, and; (ii) syllables beginning with a voiced sibilant – *zval* and *zmal* – which received results that pattern with faster responses, reflecting an underestimation of well-formedness by the model.

These results may be taken to suggest language-specific top-down effects of German. In German, sibilants are regularly unvoiced/devoiced at edges of clusters, while nasals, on the other hand, can be syllabic. In that sense, German-speaking listeners may be more prone to considering marginal sibilance as a voiceless nucleus repeller and nasality as a potential nucleus attractor. Compare this with Hebrew (the native language of the subjects in Experiment 3), in which nasals cannot be syllabic and voiced sibilants are common in marginal cluster edges.

7.8.3 Model comparison

While the model fits give us an insight into the behavior of each model with respect to the data, they are not well-suited to a comparison of different models against a consistent criterion. To do this, we ran out-of-sample predictions using cross-validation, thereby testing the ability of each model to predict unseen items.

7.8.3.1 Experiment 2

A bird's eye view of all the six model fits in Experiment 2 is available in Figure 7.12. The results of the model comparison from Experiment 2 are available in Table 7.4 and reveal a clear advantage of NAP_{td} over all other models. The main metric in the table is the *elpd* score, which stands for *expected log-predictive density* (higher score indicating better predictive accuracy). The raw values are transformed to more informative values that measure the distance from the best score in terms of *Difference in elpd*. The size of this difference can be compared to the size of a standard error of difference, *Difference SE*.

The difference of NAP_{td} from the next three models – SSP_{col}, MSD_{col} and NAP_{bu} – is about 6 standard errors (considering that the difference is around 90

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

elpd and the corresponding standard error is around 15), reflecting a very robust lead for NAP_{td}. The small differences between the next three models (SSP_{col}, MSD_{col} and NAP_{bu}) make them all indistinguishable in the second place. The two traditional models that are based on the H_{exp} hierarchy – SSP/MSD_{exp} – are similar to each other in last place and only marginally better than the null model.

The right-most column in Table 7.4, *weight*, shows model averaging via stacking of predictive distributions. Stacking maximizes the potential elpd score by pulling the predictions of all the different models together. The values under the *weight* column represent the relative contribution of each model to this combined optimal model. NAP_{td} alone contributes the lion's share with 65% and NAP_{bu} comes second with 14%. This is notable as both NAP models are essentially based on the same principle, lending support to the idea that the two models are essentially complementary. The other traditional models contribute 8% (SSP_{col}) and 3% (MSD_{col}) to this picture, less than the 9% that the null model manages to contribute.

Table 7.4: All models comparison: Experiment 2. *Note.* The table is ordered by the expected log-predictive density (elpd) score of the models, with a higher score indicating better predictive accuracy. The highest scored model is used as a baseline for the difference in elpd and the difference standard error (SE). The column weight represents the weights of the individual models that maximize the total elpd score of all the models.

model	elpd	Difference in elpd	Difference SE	weight
NAP_{td}	−28595	0.00	0.00	0.65
SSP_{col}	−28685	−89.56	14.30	0.08
NAP_{bu}	−28686	−90.91	15.45	0.14
MSD_{col}	−28689	−93.85	14.07	0.03
MSD_{exp}	−28796	−200.32	20.14	≈ 0
SSP_{exp}	−28806	−211.12	20.20	≈ 0
Null	−28850	−255.00	23.69	0.09

7.8.3.2 Experiment 3

A bird's eye view of all the six model fits in Experiment 3 is available in Figure 7.13. The results of the model comparison from Experiment 3 (see Table 7.5) reveal a borderline advantage of NAP_{td} over other models. The difference in elpd scores from the next two models – MSD_{col} and NAP_{bu} – is only about 2 standard errors

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

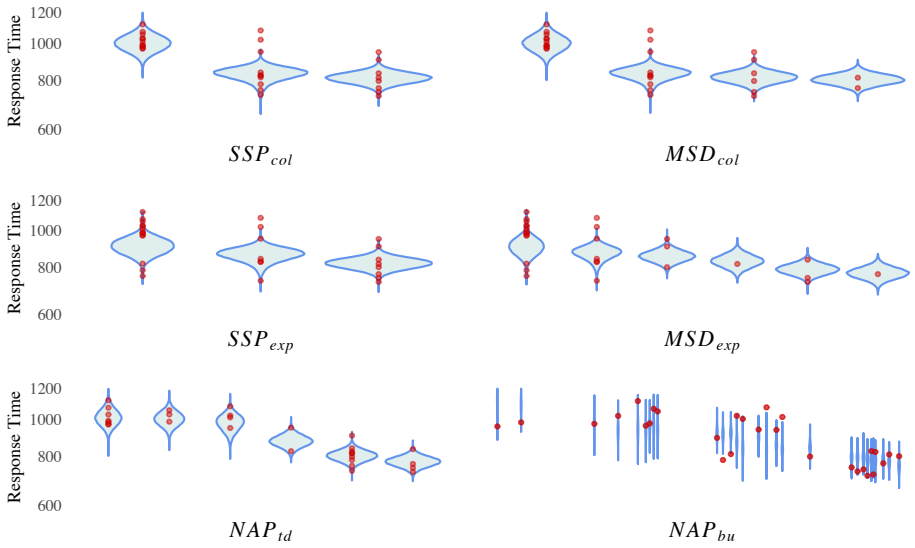

Figure 7.12: Experiment 2: all sonority model fits (unspecified cluster types, see detailed versions above). Observed mean log-transformed response times are depicted with red points; distribution of simulated means based on the model are depicted with blue violins. Stimuli are ordered from left to right according to their score in a given model in ascending well-formedness.

(considering the difference at around 10 elpd and the corresponding standard error at around 5).

SSP_{col} is more clearly distinguishable from NAP_{td}, with a difference that is almost 3 standard errors (about 20:7). MSD_{col} and NAP_{bu} are barely distinguishable from SSP_{col} and NAP_{td}. The two traditional models that are based on the H_{exp} hierarchy – SSP/MSD_{exp} – are, again, very clearly the worst in the comparison.

The *weight* values of Experiment 3 in Table 7.5 show that, again, NAP_{td} alone provides the biggest relative contribution to a combined optimal model, with 61%. MSD_{col} covers almost the entire remaining space with 37%, leaving NAP_{bu} and all the other traditional models with zero additional contribution.

7.8.4 Summary of results

The results of the confirmatory studies, Experiments 2-3, can be summarized as follows: (i) all of the sonority models we tested are capable of explaining the response time data for different consonant clusters to a reasonable extent; (ii) the symbolic top-down NAP model, NAP_{td}, outperforms all the the other models; (iii)

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table 7.5: All models comparison: Experiment 3. *Note.* (details are the same as above).

model	elpd	Difference in elpd	Difference SE	weight
NAP_{td}	−22981	0.00	0.00	0.61
NAP_{bu}	−22990	−9.58	4.82	≈ 0
MSD_{col}	−22991	−10.01	6.89	0.37
SSP_{col}	−23000	−19.77	6.99	≈ 0
SSP_{exp}	−23027	−46.22	10.25	≈ 0
MSD_{exp}	−23031	−50.44	10.10	≈ 0
Null	−23053	−72.62	12.52	0.01

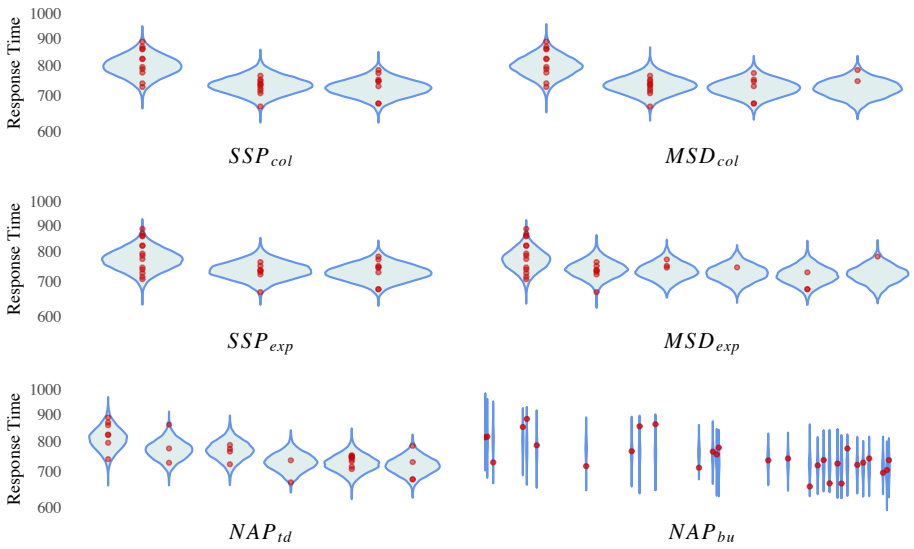

Figure 7.13: Experiment 3: all sonority model fits (plot details are the same as above).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

some interesting differences between the H_{col} and H_{exp} sonority hierarchies were observed and the advantages of the minimal H_{col} sonority hierarchy proved to be more effective.

Experiment 3 exhibits most of the general trends found in Experiment 2, albeit in a less compelling way. The Hebrew speakers in Experiment 3 tended to respond relatively fast to ill-formed structures. One path of explanation for these discrepancies can be found in the differences between the ambient languages. We expect language-specific differences to account for some of the differences between the experiments, as was mentioned in Section 7.5. Specifically, the difference between nasals in Hebrew and German as well as the difference between voiced sibilants in Hebrew and German was suggested as explanatory in Section 7.8.2.3.

On top of that, we suspect that differences between the experiments were also due to the various sources of noise that were introduced in the process. These include the smaller group of participants and the diverse physical locations in which Experiment 3 was administered (see Section 7.6.3). The results may be taken to support this with a larger standard deviation for the by-subject adjustments to the intercept for the models of Experiment 3 in comparison with Experiment 2 (e.g. $\hat{\sigma}_{\alpha} = 0.32\ [0.25, 0.41]$ in Experiment 3 vs. $\hat{\sigma}_{\alpha} = 0.21\ [0.17, 0.26]$ in Experiment 2, when comparing the null models, see the full models in Appendix A).

The success of our NAP models relative to the traditional models in predicting the data can be mainly attributed to the following traits of NAP: (i) all the voiceless-initial onset clusters, including onset falls and plateaus (e.g. /sp/ and /sf/), are relatively well-formed in NAP, correctly predicting the patterning together of such data with faster response times (at the low-right parts of the plots); (ii) onset rises (like /ml/), nasal plateaus (/nm/ and /mn/), and onset falls (like /lm/) pattern together as similar and relatively ill-formed in NAP, correctly predicting the data, as sonorant-initial plateaus and rises do not tend to pattern with (better-formed) obstruent-initial plateaus and rises.

A superficial formal generalization that can illustrate these results in symbolic terms may be that the sonority *intercept* of onset clusters appears to be (at least) as impactful as the sonority *slope* in determining syllabic well-formedness (i.e. the starting level of the onset cluster is at least as predictive of well-formedness as the angle of the cluster's slope).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

8 Corpus study

A useful aspect of the symbolic interpretation of NAP is that it can be employed for diachronic descriptions of historic sound change, where processes tend to be phonologized over time in ways that lend themselves to symbolic descriptions such as deletion, insertion and category change of individual segments. Thus, NAP-based predictions can be tested against the prevailing SSP-based predictions in cases of diachronic sound change where syllabic well-formedness is assumed to play a role.

Traditional sonority-based principles have often been invoked with relation to Modern Hebrew (MH) phonotactics, as they have been for many other languages that were studied with the toolbox of mainstream phonological research in recent decades (for examples from Modern Hebrew see Adam 2002, Asherov & Bat-El 2019, Bat-El 1994, 1996, 2002, 2012b, Bolozky 1978, 2006, 2009, Cohen 2009, Faust 2014, 2015, Kreitman 2008, Laks et al. 2016, Schwarzwald 2005). One prominent feature of MH is that complex onsets of consonant clusters are often formed morpheme-initially in the plural inflection of many nouns, where sonority seems to play a crucial role in determining which sequences of consonants would be considered well-formed enough to allow complex onset clusters to occur.

The data for this corpus study are derived from the *Living Lexicon of Hebrew Nouns* (LLHN; Bolozky & Becker 2006). The LLHN is a tabulated collection of 12,043 Hebrew nouns based on a normative MH dictionary, the *Even-Shoshan Dictionary* (Even-Shoshan 2003), with phonemic transcriptions in IPA of colloquial singular and plural forms, provided by the LLHN authors as a highly generalized depiction of MH around the turn of the century.

This study targets the *Segholate* class, which comprises a very large group of Hebrew nouns, with 1,016 entries in the LLHN (close to 10% of the entire list). Segholates feature many frequently used words like *ké.lev* (‘dog’), *pé.ʁaχ* (‘flower’) and *jé.led* (‘kid’). Consonant clusters appear morpheme-initially in the plural inflections of Segholates if the two initial consonants can be syllabified together in a well-formed complex onset. To illustrate this with the three examples above, consider the potential sequences /kl/ and /pʁ/ from ***k***é.***l***ev and ***p***é.***ʁ***aχ (respectively), that make a well-formed complex onset (rising sonority), in contrast to the potential sequence /jl/ from ***j***é.***l***ed, that makes an ill-formed complex onset

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

(falling sonority). As a result, the plural forms ***kl****a.v-ím* ('dog-PL') and ***pʁ****a.χ-ím* ('flower-PL') allow a complex onset cluster, while the plural form ***je.l****a.d-ím* ('kid-PL') does not (****jl****a.d-ím*).

In the remainder of this chapter, I provide the relevant background on Modern Hebrew (Sections 8.2–8.4) and outline the preparation of the study corpus (Sections 8.5–8.7), before presenting a descriptive analysis of the data (Sections 8.8–8.10) and concluding with a short discussion in Section 8.11. I start in the next section (8.1) with a description of the limitations of the corpus, to help clarify the scope of this study.

Important notes:

- Major parts of this chapter were also published in Albert (2022).
- The corpus study is fully replicable from the LLHN public data file and the R code made available in an *Open Science Framework* repository at the following link: https://osf.io/wuf3j/.

8.1 Limitations of the corpus study

It is important to clarify the limitations of this corpus study at the outset. This is not a survey of MH phonotactics, nor can it be regarded as such. The point of this study is to observe sonority-related phonotactics in phonologized Modern Hebrew forms, based on systematic divergence from the Biblical Hebrew norm. The class of Segholate nouns presents an opportunity to limit the scope of this question and make it more manageable in terms of the size of the data set. Segholates are both unique and abundant at the same time. Their uniqueness makes it easier to cover an exhaustive list of confounding factors to screen out forms that are not informative with respect to the question at hand. Their abundance assures us that even after we reduce the size of the Segholate set due to exclusions, we will still remain with a rich enough set of tokens that contains many varied examples of the systematic alternation of interest for the study of sonority-related phonotactics.

Segholates are frequent nouns that are distinctively of older Hebrew origin. The Segholate class is not a productive host for new nouns (see Bolozky 2020). As such, Segholates may reflect some facts about the phonology of Biblical Hebrew rather than MH. For example, only sibilants are possible fricatives in C_1 of Segholates and the bilabial stops /p, b/ never occur in the C_2 position of Segholates. These generalizations are a legacy of the old spirantization rule of Biblical Hebrew, which is mostly maintained as a morphological alternation in MH (see

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Albert 2019). It should not be taken to mean that there are no /f/-initial and /χ/-initial nouns in MH, or that bilabial stops are illicit in C_2 positions in MH.

This corpus is, therefore, very useful for the following type of observation: given that MH tends to delete the reduced vowel of Biblical Hebrew (the *mobile schwa*; see Section 8.3), we can learn about the phonotactics of MH by systematically tracking which types of consonants around this position allow a cluster formation or, otherwise, block it with an epenthetic vowel.

8.2 Historic sound change and the Hebrew languages

One very interesting diachronic process in the context of sonority is cluster formation due to loss of vocalic elements, such as the loss of *yers* in the Slavic language family (e.g. Rubach & Booij 1990, Gouskova & Becker 2013, Scheer 2007). The loss of vocalic elements creates new phonological environments where often two consonants that where initially in different syllabic positions and/or different syllables, end up as members of a tautosyllabic cluster, forming complex onsets or complex codas. Phonotactic principles are expected to restrict certain clusters such that some segmental sequences will end up following a different path of historic sound change in order to prevent illicit clusters from occurring. This is often achieved by inserting the language's default vocalic element – its epenthetic vowel – between the two consonants.

Historic loss of vocalic elements can therefore serve as a window into language-specific criteria for syllabic well-formedness in terms of licit and illicit consonantal combinations. In what follows, NAP-based and SSP-based predictions are tested against data from Modern Hebrew (MH), which – given the characteristics detailed below – serves as a hotbed for the emergence of phonotactic universals (see Adam 2002, Albert 2014, Bat-El 2005).

This situation in MH is unique. On the one hand, MH is based on centuries-old classical Hebrew varieties that were preserved via writing systems and niche roles that spoken Hebrew traditions kept filling (chiefly in religious contexts). On the other hand, as a natural language with a community of native speakers, MH is a brand new language from the late nineteenth century, with only few generations of native speakers. Thus, unlike more typical historic trajectories, MH cannot be described as the result of direct evolution from classical Hebrew varieties (see Blanc 1957, Fellman 1973, Morag 1959).

The reliance of MH on textual sources (see Myhill 2004) contributed greatly to the perseverance of old Hebrew morphology, but was less determinant in preserving the phonology of old Hebrew. The resolution of MH phonology by the

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

new Hebrew speakers, especially given the rich morphological structure of Hebrew grammar, provides us with a rare opportunity to observe accelerated and well-documented phonological patterns that resemble historical sound-change, which – under more typical conditions – would have taken many generations to establish.

Note that there are different periods of old Hebrew that contributed to MH: Biblical Hebrew (spoken around 1200–300 BCE), Mishnaic or Rabbinic Hebrew (from around 300 BCE to 600 CE) and Medieval Hebrew (mostly written around 600–1300 CE). An important role during Medieval Hebrew was played by what is known as Tiberian Hebrew or Masoretic Hebrew (7th to 10th century CE) which was crucial in developing the intricate writing system that is still in use in MH to a large extent. Since this study does not deal with historical Hebrew varieties and the differences between them, in what follows I simply refer to all the old varieties of Hebrew under the cover term *Biblical Hebrew*, which is abbreviated as BH.

8.3 Consonantal clusters in Modern Hebrew

The phonology of MH can be roughly described as a combination of the native phonologies of the new MH speakers (varieties of Yiddish, as well as a myriad of Slavic, Arabic, Germanic, Romance, and other languages), and their various traditions for mapping Hebrew graphemes to sounds in religious reading contexts, where Hebrew often remained in use.

One striking feature of MH phonology that sets it apart from BH is its much broader tolerance towards consonantal clusters. In BH, tautosyllabic consonantal clusters were limited to final coda positions as a result of morpho-phonological processes, most often when the suffix *-t* was attached to a consonant-final base of a verb in the feminine inflection (e.g. *ka.táv-t* 'write.PST-2SG.F'). Morpheme-initial consonants in BH were regularly separated by a vowel to avoid complex (tautosyllabic) onset clusters.[1] In contrast to the restrictive phonology of BH with respect to complex onset clusters, MH speakers seem to prefer clusters in many unstressed morpheme-initial positions, as various studies have already noted before (e.g. Rosén 1956, Albert et al. 2013, Asherov & Bat-El 2019, Bat-El 2008, Bolozky 1978, 2006, Cohen-Gross 2015, Laufer 1991, Schwarzwald 2005).

[1]Consonantal sequences in middle positions of BH words occur frequently, yet they are mostly considered as belonging to two different syllables (i.e. *heterosyllabic*), not forming a tautosyllabic complex cluster.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

BH featured a reduced (short) vowel in unstressed positions, which the Tiberian scholars marked with a unique diacritic termed *schwa*, which inspired the naming of the phonetic schwa, although they are quite different (see Laufer 2019 for a short overview of the two terms). The schwa in the Tiberian writing system has two main interpretations: it is either a short vowel (*mobile schwa*) or no vowel (*silent schwa*). The silent schwa is restricted to coda positions, to indicate that the consonantal grapheme has no following vowel. The mobile schwa occurs in onsets, indicating a reduced vowel after the consonantal grapheme. In MH there are no phonologically reduced vowels in unstressed positions (not considering post-lexical prosody) such that the reduced vowel of BH – the mobile schwa – tends to be deleted in MH.

Hebrew words typically combine affixation with vocalic changes in the base morpheme when inflected. This also includes the movement of stress towards the suffix in order to keep the strong syllable at the final edge of the prosodic word. This regular stress shift towards the final edge has implications for the beginning of the prosodic structure as well, considering that unstressed initial syllables are more prone to reduction processes. Furthermore, when an inflectional suffix is added to a base morpheme, the deletion of a vowel from the base can offset the overall increase in size of prosodic word due to affixation. These are perhaps the main contributors to the relative abundance of morpheme-initial complex clusters in MH (Asherov & Bat-El 2019, Bat-El 2008).

Regardless of the sources of MH phonotactic patterns, MH speakers produce many complex onset clusters (C_1C_2V) across the lexicon. Moreover, the variety of possible onset clusters in MH is relatively large, allowing more combinations than most Germanic and Romance languages exhibit, including practically any combination of two obstruents in a complex onset cluster (i.e. *stop-stop*, *stop-fricative*, *fricative-fricative* and *fricative-stop*).

Importantly, the tendency towards cluster formation can be blocked with MH's epenthetic vowel /e/ to avoid certain ill-formed CC combinations in complex onsets, thus serving as a window into the phonology of MH, with a specific view to its phonotactic landscape. The literature on the subject of clusters in MH points at the crucial role that sonority seems to play in blocking cluster formation. For example, it has often been noticed that the sonorant consonants of the system (/m, n, l, ʁ, j/)[2] do not form a cluster with a following consonant whenever they

[2]The historic labiovelar glide /w/, which was native to BH phonology, has merged with the voiced labiodental fricative /v/ in MH. That said, the glide /w/ has a marginal phonemic status in MH as a distinctive consonant in many common loanwords from both English (e.g. ***wáj.faj*** 'WiFi') and Arabic (e.g. ***wá.la*** 'indeed').

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

are morpheme-initial, i.e. in C1 position (e.g. Rosén 1956, Asherov & Bat-El 2019, Bolozky 2006, Schwarzwald 2005).

(1) ***ka.χól*** → ***kχu.l-ím*** ('blue-PL.M')

(2) ***ja.ʁók*** → ***je.ʁu.k-ím*** ('green-PL.M')

Examples (1–2) demonstrate this with two color adjectives that share the same vocalic template – C_1a.C_2óC_3 – while differing in their consonantal makeup. The base morpheme of inflected adjectives in the C_1a.C_2óC_3 pattern deletes its first vowel /a/ and changes the quality of its second vowel, which is no longer in the stressed syllable, by raising from /o/ to /u/. As a result, C_1**a**.C_2**ó**C_3 becomes C_1C_2**u**.C_3-ím. This is apparent in (1) but note that in (2) the epenthetic vowel /e/ is inserted between C_1 and C_2 in the plural inflection, yielding trisyllabic C_1**e**.C_2u.C_3-ím. This is done in order to avoid an otherwise ill-formed onset cluster that would be headed by a highly sonorant glide /j/ (****jʁu.kím***).

8.4 Segholates in Modern Hebrew

Segholates form a special class of Hebrew nouns due to their unique stress pattern (see Bat-El 2012a). In citation form, when the base morpheme is devoid of affixes (the singular masculine forms by default), Segholates exhibit a penultimate stress unlike typical nouns of Hebrew origin, which standardly exhibit a final stress (*iambic* pattern). At the same time, Segholates do behave like typical Hebrew nouns in that they exhibit the standard final stress pattern with inflected forms, whereby the stress shifts from the base morpheme to the suffix. This divergence from the norm in the bare citation form of Segholates is related to historic processes within BH (see Yeverechyahu & Bat-El 2020) and is maintained by the lexical stress system of MH, which tolerates varying stress assignments, including some apparent tendencies towards penultimate stress (*trochaic* pattern), despite the strong iambic preference of BH (Bat-El 2005).

When the default plural suffixes *-im* or *-ot* are added to typical Segholates, the stress shifts to the end, and the first vowel of the base morpheme deletes. As a result, the first two consonants of Segholates tend to form a complex onset cluster morpheme-initially when plural suffixes are added. However, if the resulting C_1C_2 sequence constitutes an ill-formed complex onset cluster, the formation of a cluster is blocked by the epenthetic vowel of MH, /e/. See Tables 8.1–8.2 for various examples of these two main routes in plural inflections of disyllabic Segholates, resulting in either onset clusters (Table 8.1) or epenthesis (Table 8.2) morpheme-initially.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table 8.1: Cluster formation in MH Segholates. *Note.* Here and elsewhere, the "Complex onset" column refers to the data in the LLHN such that "✓" indicates a complex onset cluster in plural inflections and "✗" indicates that a vowel appears between C_1 and C_2 in the plural inflection.

Singular	Plural	Gloss	C_1C_2	Complex onset
pé.ʁaχ	***pʁa.χ-ím***	'flower'	pʁ	✓
dé.let	***dla.t-ót***	'door'	dl	✓
pá.χad	***pχa.d-ím***	'fear'	pχ	✓
kó.tel	***kta.l-ím***	'wall'	kt	✓
ʃé.ka	***ʃka.-ím***	'socket'	ʃk	✓
sé.feʁ	***sfa.ʁ-ím***	'book'	sf	✓
ʃé.meʃ	***ʃma.ʃ-ót***	'sun'	ʃm	✓
vé.ʁed	***vʁa.d-ím***	'rose'	vʁ	✓

Table 8.2: Vowel epenthesis in MH Segholates

Singular	Plural	Gloss	C_1C_2	Complex onset
ʁé.geʃ	***ʁe.ga.ʃ-ót***	'feeling'	ʁg	✗
ʁó.tev	***ʁe.ta.v-ím***	'sauce'	ʁt	✗
lé.χem	***le.χa.m-ím***	'bread'	lχ	✗
má.χat	***me.χa.t-ím***	'needle'	mχ	✗
né.meʃ	***ne.ma.ʃ-ím***	'freckle'	nm	✗
mé.laχ	***me.la.χ-ím***	'salt'	ml	✗

The vast majority of Segholates are disyllabic. The most common vocalic pattern in Segholates is the C**é**.C**e**(C) pattern with two /e/ vowels in the citation form (e.g. ***dé.let*** in Table 13). Other vocalic patterns in citation form in Tables 8.1–8.2 include C**á**.C**a**(C) (e.g., ***má.χat***, ***pá.χad***), C**ó**.C**e**(C) (e.g., ***ʁó.tev***, ***kó.tel***) and C**é**.C**a**(C) (e.g. ***pé.ʁaχ***, ***mé.laχ***, ***ʃé.ka***).

8.5 Epenthesis verification

The epenthetic status of the vowel that appears between C_1 and C_2 in inflected Segholates can be independently verified via systematic resyllabification processes in MH. For example, when preceded by a proclitic such as the definite arti-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

cle *(h)a-*, the epenthetic vowel can disappear if C_1 resyllabifies as the coda of *(h)a-* leaving C_2 in a *simple onset* position: ha-C_1.C_2V… This scenario allows consonantal sequences to surface without an intervening vowel as they no longer constitute a tautosyllabic complex onset (see Bolozky 2006: 227). This procedure yields forms like those given in Table 8.3, demonstrating heterosyllabic sequences for all the same cases that exhibit an epenthetic vowel to block a tautosyllabic onset cluster in Table 8.2.

Importantly, if a non-epenthetic vowel appears between C_1 and C_2 of Segholates, as detailed in the following section, it will not delete in any of these environments, including environments that do not require a vowel to break complex tautosyllablic clusters, as demonstrated for the forms in Table 8.4.

Table 8.3: Number inflection in Segholates with epenthtic vowels. *Note.* The epenthetic vowels between C_1 and C_2 are not mandatory in the Det+Plural forms, where they are not required to break a complex onset cluster, and they are free to delete as shown here.

Singular	Plural	Det+Plural	Gloss	C_1C_2	CO[a]	EV[b]
ʁé.geʃ	*ʁe.ga.ʃ-ót*	*(h)a-ʁ.ga.ʃ-ót*	‘feeling’	ʁg	✗	✓
ʁó.tev	*ʁe.ta.v-ím*	*(h)a-ʁ.ta.v-ím*	‘sauce’	ʁt	✗	✓
lé.χem	*le.χa.m-ím*	*(h)a-l.χa.m-ím*	‘bread’	lχ	✗	✓
má.χat	*me.χa.t-ím*	*(h)a-m.χa.t-ím*	‘needle’	mχ	✗	✓
né.meʃ	*ne.ma.ʃ-ím*	*(h)a-n.ma.ʃ-ím*	‘freckle’	nm	✗	✓
mé.laχ	*me.la.χ-ím*	*(h)a-m.la.χ-ím*	‘salt’	ml	✗	✓

[a] Complex onset
[b] Epenthetic vowel

It is of interest to note that consonantal sequences which cannot appear as word-initial tautosyllabic clusters (e.g. /lχ/ in illicit ****lχ****a.m-ím*) can, at the same time, appear as sequences with no intervening vowel if they are heterosyllabic (e.g. *(h)a****l.χ****a.m-ím*). This is an independent validation that the ill-formedness of the structures in the corpus is not simply due to adjacency, but involves restrictions on adjacency in the context of syllabic structure. Hence, this verification process also serves as an independent validation that syllabic well-formedness, and more specifically sonority, are justifiably invoked in this case.

Crucially, the Segholate forms that reveal sensitivity to sonority-related phonotactics must be those that either allow a complex onset cluster in the plural inflection, thus deleting the first vowel that surfaces in the singular form (as in

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table 8.4: Number inflection in Segholates with non-epenthtic vowels. *Note.* The non-epenthetic vowels between C_1 and C_2 are mandatory. They are expected to surface regardless of syllabic structure.

Singular	Plural	Det+Plural	Gloss	C_1C_2	CO[a]	EV[b]
ʁó.maχ	*ʁo.ma.χ-ím*	*(h)a-.ʁo.ma.χ-ím*	'lance'	ʁm	✗	✗
nó.feʃ	*no.fa.ʃ-ím*	*(h)a-.no.fa.ʃ-ím*	'vacation'	nf	✗	✗
nó.saχ	*no.sa.χ-ím*	*(h)a-.no.sa.χ-ím*	'wording'	ns	✗	✗
χé.ʁev	*χa.ʁa.v-ót*	*(h)a-.χa.ʁa.v-ót*	'sword'	χʁ	✗	✗
kó.va	*ko.va.(ʔ)-ím*	*(h)a-.ko.va.(ʔ)-ím*	'hat'	kv	✗	✗
ʃó.ʁeʃ	*ʃo.ʁa.ʃ-ím*	*(h)a-.ʃo.ʁa.ʃ-ím*	'root'	ʃʁ	✗	✗

[a]Complex onset
[b]Epenthetic vowel

Table 8.1), or, alternatively, require a vowel that can be shown to be an epenthetic vowel (as shown in Table 8.3). In any of these cases, no mandatory vowel is expected between C_1 and C_2 when the Segholate noun is preceded by a (C)V proclitic.

8.6 Confounding factors

The forms that fail in the general epenthetic vowel test (see Section 8.5) were ultimately excluded from the corpus study since their behavior across the number inflection is not expected to be reflective of sonority-based phonotactics. Apart from a few idiosyncratic forms which are covered in Section 8.6.5, the vast majority of these exclusions stem from structural and segmental factors, not related to sonority, which I consider as *confounding factors*. The following Sections 8.6.1–8.6.5 cover the various confounding factors that lead to exclusion from the study corpus.

8.6.1 Final rime merge

As described above, the condition for cluster formation in Segholates is related to a morpheme-initial adjustment (vowel deletion) that offsets the additional vowel of suffixes when inflected to plural forms. While this pattern is the most prominent in Segholates, it is not the only one. A large subset of Segholates (336 nouns, about a third of all LLHN Segholates) makes the adjustment morpheme-finally,

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

mostly replacing the final VC rime of the singular form with the VC suffix *-im* or *-ot*.

This final rime merge happens almost exclusively with Segholates that end in /Vt/ (*et* or *at*), that is, either with a feminine suffix such as *-et* in *gvé.ʁet* ('lady'; lit. 'man-SG.F') or with a templatic particle such as C_1a.C_2é.C_3**et** in *da.lé.k**et*** ('inflammation'). The plural suffix in these cases is almost always *-ot* such that it either replaces the singular-feminine suffix *-Vt* with the plural-feminine suffix *-ot*, or, alternatively, it merges with the templatic final *-Vt* of the base morpheme rather than being concatenated to it. These two processes are superficially identical in that the change from singular to plural requires only the replacement of the final vowel while retaining the following coda /t/, and without altering the morpheme-initial structure (see examples in Table 8.5).

Note that due to the fact that this final *-Vt* particle is appended to the default tri-consonantal root, these Segholates tend to stand out because they are mostly either trisyllabic or include a complex onset cluster in their singular citation form to accommodate this extra material (see examples in Table 8.5).

Table 8.5: Fixed morpheme-initial forms: changes in final rather than initial vowel. *Note.* Parentheses in the "Complex onset" condition are due to the lack of morpheme-initial change between the singular and plural inflections.

Singular	Plural	Gloss	C_1C_2	Complex onset
***k**a.**s**é.**f**et*	***k**a.**s**a.**f**-**ót***	'safe'	ks	(✗)
***ʁ**a.**k**é.**v**et*	***ʁ**a.**k**a.**v**-**ót***	'train'	ʁk	(✗)
***d**a.**l**é.**k**et*	***d**a.**l**a.**k**-**ót***	'inflamation'	dl	(✗)
***kt**ó.**v**et*	***kt**o.**v**-**ót***	'address'	kt	(✓)
***gv**é.**ʁ**et*	***gv**a.**ʁ**-**ót***	'lady'	gv	(✓)
***kn**é.**s**et*	***kn**a.**s**-**ót***	'assembly'	kn	(✓)

One Segholate exception in the LLHN uses the plural suffix *-im* to replace the final *et* portion of the base morpheme: *ʃi.bó.l**et*** → *ʃi.bo.l-**ím*** 'stalk (of grain)-PL'. Two other Segholate exceptions in the LLHN delete the final vowel of base morphemes that end with *en* and take *-im* as their plural suffix, yielding: *ci.pó.ʁ**en*** → *ci.poʁ.n-**ím*** 'clove-PL' and *mik.tó.ʁ**en*** → *mik.toʁ.n-**ím*** 'jacket-PL'. These, and the more typical patterns where the final -Vt portion of the base morpheme is replaced by the plural suffix *-ot*, are excluded from the study as they do not exhibit morpheme-initial epenthesis or cluster formation when the plural suffix is added.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

8.6.2 Non-typical plurals

Segholate nouns that lack any plural inflection were excluded from the study. The LLHN lists 49 of 1,016 Segholates (about 5%) without plurals. These include mass nouns like *ʃá.χat* 'hay' and *té.va* 'nature'. Moreover, there are 9 Segholates in the LLHN with an irregular plural inflection, derived from the old dual inflection of Hebrew. These trigger adjustments of the base morpheme that differ from the regular pattern. MH retained this restricted version of the dual morphology of BH, a number inflection that is common in Semitic languages, alongside the more general singular and plural inflections. The dual suffix is used in MH with a limited set of nouns of Hebrew origin, with varying semantics (either general plural, exactly two, or even a mass noun interpretation).

Table 8.6: Segholates with the historical dual suffix

Singular	Plural	Gloss	C_1C_2	Complex onset
té.lef	***tl****a.f-á.(j)im*	'hoof'	tl	✓
gé.ʁev	***gaʁ****.b-á.(j)im*	'sock'	gʁ	✗
bé.ʁeχ	***biʁ****.k-á.(j)im*	'knee'	bʁ	✗
ké.ʁen	***kaʁ****.n-á.(j)im*	'horn'	kʁ	✗

The *-á(j)im* dual suffix features two vowels with inherent penultimate stress, unlike the regular *-ot* or *-im* of the plural suffixes, that appear within the (typically stressed) word-final syllable. Importantly, the effect of the dual suffix on morpheme-initial cluster formation cannot be related to sonority. Four of the nine Segholates with dual suffixes in the LLHN exhibit a potentially well-formed obstruent-sonorant cluster, yet only one of those four exhibits a cluster in the plural inflection with the dual suffix since these forms evidently allow the deletion of the second vowel from the base morpheme, thus exhibiting C_2C_3 clusters rather than C_1C_2, regardless of sonority (see Table 8.6).

8.6.3 Gutturals

A major confounding factor to consider with respect to the expected phonotactics of Segholates is related to the segmental identity of the first two consonants, C_1 and C_2. Specifically, consider cases in which C_1 features one of the four historic gutturals of BH – /ʔ, h, ʕ, ħ/ – or if C_2 features a member of the glottal(ized) subset of the historic gutturals: /ʔ, h, ʕ/. The exception to this broad generalization concerns the historical voiceless pharyngeal fricative /ħ/, which typically

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

surfaces in MH as the dorsal fricative /χ/ that can participate in MH Segholate clusters when it is in the C_2 position (see Tables 8.7–8.8).

Table 8.7: Cluster avoidance with historic gutturals in C_1. *Note.* Consonants within parentheses are optional; starred consonants denote historical sounds; ">" marks change.

Singular	Plural	Gloss	C_1C_2	Complex onset
*(**ʔ**)é.ʁec*	*(**ʔ**)a.ʁa.c-ót*	'land'	(ʔ)ʁ	✗
*(**h**)é.vel*	*(**h**)a.va.l-ím*	'nonsense'	(h)v	✗
*(*ʕ>**ʔ**)é.ʁev*	*(*ʕ>**ʔ**)a.ʁa.v-ím*	'evening'	(ʔ)ʁ	✗
*(***ħ**>)**χ**é.ʁev*	*(***ħ**>)**χ**a.ʁa.v-ót*	'sword'	χʁ	✗

The cause of this peculiar behavior is related to the fact that the historic gutturals of BH have undergone major phonological changes in MH, where they are still denoted by unique graphemes in the writing system (see Bolozky 1978, Faust 2019, Gafter 2019, Schwarzwald 2005). The historic glottal stop /ʔ/ and the voiced pharyngeal fricative /ʕ/ both tend to have no consonantal interpretation in MH, mostly alternating between no consonant and a glottal stop on phonetic rather than phonological grounds. Likewise, the glottal fricative /h/ alternates between a glottal fricative or stop, or no consonant. Therefore, these glottal(ized) gutturals do not canonically participate in consonantal clusters in MH, as they simply do not even have a stable consonantal interpretation.

Table 8.8: Cluster avoidance with glottal(ized) historic gutturals in C_2. *Note.* The details are the same as Table 8.7 above.

Singular	Plural	Gloss	C_1C_2	Complex onset
*tó.(**ʔ**)aʁ*	*te.(**ʔ**)a.ʁ-ím*	'title'	t(ʔ)	✗
*sá.(**h**)aʁ*	*se.(**h**)a.ʁ-ím*	'crescent'	s(h)	✗
*ʃá.(*ʕ>**ʔ**)aʁ*	*ʃe.(*ʕ>**ʔ**)a.ʁ-ím*	'gate'	ʃ(ʔ)	✗
*ʃá.(***ħ**>)**χ**af*	***ʃ(*ħ>)χ**a.f-ím*	'seagull'	ʃχ	✓

The fate of the voiceless pharyngeal fricative /ħ/ is different as it merged with the uvular-velar fricative /χ/ of MH, which also corresponds to the historic spirantized counterpart of /k/ (Adam 2002, Albert 2019, Barkai 1975, Bolozky 1978, 2013). Importantly, /ħ/ is the only historic guttural in this set that is consistently

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

mapped to a consonant in MH. Furthermore, unlike the glottal stop and the fricative, which are restricted to simple onsets in MH, /χ/ can be also found in complex onsets and codas (e.g. *sχa.vá* 'rag', *ma.táχ-t* 'stretch.PST-2SG.F'), although rarely at the margins of clusters. This general behavior of /χ/ is apparent also in MH Segholates. When /χ/ is in C_1 position of a Segholate it behaves like the other historic gutturals, essentially avoiding /χC/ complex onset clusters with /χ/ at their margin (see Table 8.7). However, when /χ/ is in C_2 position it behaves much like a typical obstruent in MH, potentially forming clusters with other obstruents in C_1 (see Table 8.8).

To conclude, Segholates featuring one of the four historic gutturals – /ʔ, h, ʕ, ħ/ – in C_1, or one of the three historic gutturals – /ʔ, h, ʕ/ – in C_2, were excluded from this study. Out of the 1,016 Segholate entries in the LLHN, 125 feature historic gutturals in C_1 and 68 Segholates feature historic /ʔ, h, ʕ/ in C_2. One word, *(ʔ)ó.(h)el* 'tent', exhibits historic gutturals in both C_1 and C_2 positions, bringing the total of guttural exclusions to 192 out of 1,016 Segholates (about 19%) in the LLHN.

8.6.4 Glides

Segholates with the glide /j/ in C_2 position should be also excluded from the study corpus, as they inconsistently vary between allowing and avoiding a morpheme-initial cluster in plurals, regardless of sonority. For Instance, consider the following two examples with a voiceless stop in C_1: (i) *ˈka.(j)ic* → *kejˈc-im* 'summer-PL'; (ii) *ˈta.(j)iʃ* → ***tj***a.ˈʃ-im 'billy goat-PL'. A cluster is formed in (ii) with /tj/ but not in (i) with comparable /kj/. The disyllabic structure is maintained in both scenarios thanks to the glide's ability to occupy the coda of the first syllable in plural inflections as in (i). Segholates with /j/ in C_2 were thus completely excluded from the study. The LLHN lists 19 such Segholates with a glide in C_2 (3 of which also have a guttural in C_1).

8.6.5 Other exclusions

8.6.5.1 Loanwords

Segholates are defined for the purpose of this study as in the LLHN, that is, as nouns with penultimate stress in their bare (citation) form and with a stress shift towards the suffix under inflection. Only words of Hebrew origin demonstrate this type of behavior, as loanwords do not shift the stress to the final syllable with plural inflections. For example, the word ***mé.teʁ***, which is the adapted form of the loanword 'meter', fits with the most common Segholate pattern, Cé.CeC, yet as

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

a loanword it retains the position of stress on the initial syllable when inflected to plural (i.e. ***mét.ʁ-im*** 'meter-PL'), therefore not giving way to the deletion of the initial vowel and deleting the second vowel instead (although note that the sequence /mt/ is nevertheless not expected to form a cluster).[3] Thus, even when superficial similarities to Hebrew Segholates are striking, loanwords follow a different path in MH morpho-phonology (see Bat-El 1994, Cohen 2009). Loanwords are not considered as Segholates in the LLHN, such that no further exclusion was needed. Furthermore, I am not aware of another example of a Segholate-like loanword beyond *meter*, as detailed above.

8.6.5.2 OCP

Another confounding factor is related to dissimilatory processes in articulation rather than sonority, often linked to the notion of the *Obligatory Contour Principle* (OCP) in the phonological literature going back to Leben (1973) and McCarthy (1979). According to this, clusters are avoided if both consonants are identical. However, since voicing differences between otherwise identical obstruents do not appear to have a relevant effect on coordination of articulatory gestures, any cluster in which the two consonants share the same place and manner of articulation is avoided, essentially also targeting sequences of two near-identical stops or fricatives as they may still differ in voicing.

Of all the Segholates in the LLHN, only the noun *té.deʁ* 'frequency', exhibits two consonants that share the same manner and place of articulation, /t/ and /d/. Here, the epenthetic vowel in the plural inflection, *te.da.ʁ-ím* 'frequency-PL', should be attributed to OCP rather than to sonority. This single case was excluded from the study corpus. It is in fact not surprising that only one case was found to exhibit this problem, as Hebrew, along with other Semitic languages, tends to keep the first two consonants of lexical *roots* phonetically distinct (Yeverechyahu 2019).

8.6.5.3 Idiosyncrasies

After considering structural and segmental generalizations that can affect the phonotactics of inflected Segholates irrespective of sonority, there are still 35

[3]Two further notes regarding 'meter-PL' in MH: (i) the choice between the two potential syllabifications – *mé.tʁim* vs. *mét.ʁim* – is inconsequential for this study and it will not be pursued here; (ii) the Hebraized version of the plural *met.ʁím*, where the stress does move to the final position as it does with nouns of Hebrew origin, may be also attested in hypercorrect speech. The latter could be due to the fact that this loanword has an exceptional Hebrew-like form and is widely used (moreover, it is very often used with number inflections), thus increasing the probability that speakers will not treat it like other loanwords.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Segholates that feature a non-epenthetic vowel in the plural inflection, without an apparent independent explanation for this behavior, other than, perhaps, lexicalized exceptions (although note that 33 of the 35 Segholates feature /o/ as the first vowel of the base morpheme).

Table 8.9: Segholates with and without non-epenthetic vowels in the plural inflection. *Note.* Forms that either allow a complex onset cluster, or, alternatively, introduce an epenthtic vowel in the plural inflection are considered as valid forms in this study. The two invalid forms in this example (*nó.feʃ* and *ʃó.ʁeʃ*) feature a non-epenthetic vowel.

Singular	Plural	Det+Plural	Gloss	C_1C_2	CO[a]	EV[b]
né.feʃ	***ne.fa.ʃ-ót***	*(h)a-**n.fa.ʃ-ót***	'soul'	nf	✗	✓
nó.feʃ	***no.fa.ʃ-ím***	*(h)a-.**no.fa.ʃ-ím***	'vacation'	nf	✗	✗
ʃé.ʁec	***ʃʁa.c-ím***	*(h)a-**ʃ.ʁa.c-ím***	'vermin'	ʃʁ	✓	(-)
ʃó.ʁeʃ	***ʃo.ʁa.ʃ-ím***	*(h)a-.**ʃo.ʁa.ʃ-ím***	'root'	ʃʁ	✗	✗

[a]Complex onset
[b]Epenthetic vowel

For example, consider the items in Table 8.9. Compare the expected *obstruent–sonorant* cluster /ʃʁ/ in ***ʃé.ʁec*** → ***ʃʁa.c-ím*** ('vermin-PL') with the non-epenthetic vowel in the exact same consonantal sequence type when it appears in the noun ***ʃó.ʁeʃ*** → ***ʃo.ʁa.ʃ-ím*** ('root-PL'). Likewise, compare the opposite *sonorant–obstruent* sequence /nf/ with a typical epenthetic vowel in ***né.feʃ*** → ***ne.fa.ʃ-ót*** ('soul-PL') and with the non-epenthetic vowel in the same cluster type in ***nó.feʃ*** → ***no.fa.ʃ-ím*** ('vacation-PL'). Note also how the forms with non-epenthetic vowel in the plural inflection are accordingly not expected to change their initial vowel across inflections in Table 8.9, even when resyllabification of an initial consonantal sequence is possible following the proclitic *(h)a-*, which results in related examples like in *(h)a-**ʃo.ʁa.ʃ-ím*** and *(h)a-**no.fa.ʃ-ím***, not **(h)a-**ʃ.ʁa.ʃ-ím*** or **(h)a-**n.fa.ʃ-ím*** (compare with *(h)a-**ʃ.ʁa.c-ím*** and *(h)a-**n.fa.ʃ-ót*** in cases where the plural exhibits a cluster or an epenthetic vowel).[4]

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table 8.10: C_1C_2 types and tokens in the Segholate study corpus. *Note.* S– = voiceless stops; S+ = voiced stops; A– = voiceless affricates; F– = voiceless fricatives; F+ = voiced fricatives; N = nasals; L = liquids, G = glides; Frics. = Fricatives; Affrics. = Affricates. Numbers in superscript represent the number of word tokens per C_1C_2 type. Colored cells mark sequences of obstruents that differ in voicing (see Section 8.8.3). See Appendix B for the full list of word tokens.

	C_2						
	Voiceless			Voiced		Sonorants	
C_1	Stops	Affrics.	Frics.	Stops	Frics.	Nasals	Liquids
S–	kt^{9} pt^{5} tk^{4}	kc^{4} pc^{1}	kf^{3} ks^{4} kʃ^{6} kχ^{1} ps^{4} pʃ^{2} pχ^{2} tf^{5} tχ^{3} tʃ^{1}	kd^{1} pg^{3}	kv^{8} tv^{2}	km^{5} kn^{1} tm^{2} tn^{2}	kl^{7} kʁ^{11} pl^{6} pʁ^{5} tl^{1} tʁ^{4}
A–	–	–	cf^{4}	cd^{1}	cv^{3}	cm^{4}	cl^{1} cʁ^{1}
F–	sk^{1} st^{1} ʃk^{7} ʃt^{4}	ʃc^{1}	sf^{3} sχ^{7} ʃf^{3} ʃs^{1} ʃχ^{6}	sd^{2} sg^{4} ʃd^{1} ʃg^{1}	sv^{2} ʃv^{3}	sm^{2} ʃm^{4} ʃn^{2}	sl^{2} sʁ^{3} ʃl^{5} ʃʁ^{1}
S+	bk^{1} bt^{1} dk^{1}	bc^{1}	bs^{1} bχ^{1} df^{4} dʃ^{2} dχ^{4} gf^{2} gʃ^{2} gχ^{1}	bd^{1} bg^{1} dg^{2} gd^{3}	dv^{2} gv^{4} gz^{4}	dm^{2} gm^{1}	bʁ^{2} dl^{4} dʁ^{2} gl^{2} gʁ^{2}
F+	vt^{1}	–	vs^{1} vʃ^{1} zf^{2} zχ^{2}	–	zv^{2}	zm^{1}	vʁ^{1} zl^{1} zʁ^{5}
N	mt^{3} nk^{3} nt^{6}	mc^{1} nc^{2}	ms^{2} mʃ^{2} mχ^{5} nf^{4} ns^{1} nʃ^{5} nχ^{5}	mg^{1} nd^{1} ng^{4}	mz^{2} nv^{3} nz^{3}	mn^{1} nm^{2}	ml^{4} mʁ^{3}
L	lk^{2} lt^{1} ʁk^{3} ʁt^{5}	ʁc^{2}	lf^{1} ls^{1} lʃ^{1} lχ^{1} ʁf^{2} ʁs^{3} ʁʃ^{3} ʁχ^{6}	ʁg^{3}	lv^{2} ʁv^{3}	ʁm^{3}	–
G	jk^{1} jt^{1}	jc^{1}	jf^{1} jʃ^{1} jχ^{1}	jd^{1} jg^{1}	jz^{1}	–	jl^{1} jʁ^{1}

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

8.7 Final corpus of Modern Hebrew Segholates

The preparation of the Segholate dataset, using the different criteria detailed in Sections 8.5–8.6, resulted in 381 different singular–plural pairs of Segholates that represent 381 C_1C_2 tokens in the study (see Table 8.10 for a summary of this distribution, and see Appendix B for the full list of words). Note that in this context, C_1C_2 refers to the two initial consonants of alternating Segholates, without making reference to potential clusters and epenthetic vowels, which will be examined in the following Descriptive Analysis (Section 8.8). These tokens consist of 144 unique C_1C_2 types at the level of segmental description (i.e. 144 unique C_1C_2 combinations in Table 8.10), and 50 unique C_1C_2 types at the level of segmental-class description (i.e. 50 unique non-empty cells in Table 8.10).

The 144 types and 381 tokens in Table 8.10 provide a large and diverse set of C_1C_2 combinations that behave in one of two possible ways in the plural inflection: either they form a complex onset cluster, or they introduce an epenthetic vowel. Crucially, the working hypothesis for the set in Table 8.10 is that the choice between a cluster or epenthesis in the plural inflection is directly related to syllabic well-formedness in terms of sonority-based restrictions, serving as a window into the top-down, sonority-based phonotactics of MH. Thus, there should be a cut-off point of well-formedness in the predictions of symbolic sonority models that is linked to the tendency to either form a complex onset cluster or to break the sequence with an epenthetic vowel.

8.8 Descriptive analysis

The MH data exhibits a binary distinction between two phonotactic alternatives, either permitting or avoiding complex onset clusters in the initial position of Segholate plural inflections. For the current analysis, this binary distinction is mapped onto the N-ary scores of the different sonority models (Sections 8.8.1–8.8.2). This mapping makes it possible to provide a descriptive observation and analysis of the fit between each of the four sonority models and the MH data (Sections 8.9–8.10). Before the analysis itself, I will also present an explanation of the treatment of voicing assimilation processes in Section 8.8.3. A discussion in Section 8.11 concludes this chapter.

[4]The determination of syllabic boundaries in the case of *obstruent-sonorant* sequences in the middle of a prosodic word (like *(h)a-ʃ.ʁa.c-ím* in the examples discussed) is potentially arguable as it could also be *(h)a-ʃʁa.c-ím*. However, these discrepancies have no implications for the current observation.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

8.8.1 Sonority hierarchies with Modern Hebrew considered

Table 2.1 in Section 2.1 (also repeated below in Table 8.11) demonstrated two sonority hierarchies that represent two ends of the spectrum of potential (and common) treatments of the obstruent class in sonority hierarchies: the H_{col} hierarchy, which collapses all obstruents into a single class; and the H_{exp} hierarchy, which maximally expands obstruents by employing both *voicing* distinctions and *manner of articulation* distinctions between the two major classes, *stops* and *fricatives*.

Since the following analysis concerns MH, a few specific additions are in place. First, the class of *affricates* was added to the subtypes of obstruents in order to account for the MH affricate – the voiceless alveolar /c/ (also regularly annotated as t͡s in standard IPA). The voiced counterpart, /d͡z/, is also taken into account in the following study (see Section 8.8.3).

Second, another hierarchy is suggested in anticipation of the most suitable obstruent configuration for MH: the H_{MH} hierarchy, which partially expands obstruents by employing only *voicing* distinctions. This is in line with the fact that MH tolerates various *obstruent-obstruent* complex clusters such that differences between stops and fricatives do not appear to play a role (see Section 8.3). At the same time, MH is known to exhibit voicing assimilation between obstruents (see Section 8.8.3) such that this distinction does appear to be playing a role, as the descriptions and analyses in Sections 8.9–8.10 reveal.

Table 8.11 presents all the above-mentioned additions to the sonority hierarchies that were considered thus far. It demonstrates the three different hierarchies, H_{col}, H_{exp} and H_{MH}, and it shows them all with the addition of the affricates class between stops and fricatives, although note that this is consequential only in the case of the H_{exp} hierarchy.

Note also that the symbolic NAP-based model, NAP_{td}, remains unchanged from when it was introduced in Section 6.2.2. This is the case because the addition of the affricate class plays no role in the sonority hierarchy of the symbolic NAP_{td} model, which only considers voicing to be distinctive between obstruents. Furthermore, the NAP_{td} model assumes one basic, fixed and universal sonority hierarchy. Contrary to the typical approach in SSP-based models, NAP-based models are not compatible with the notion of language-specific sonority hierarchies. Instead, NAP is committed to a universal view of sonority which is explicitly based on pitch intelligibility in perception and periodic energy in the acoustic signal. NAP_{td} links this quality with different symbolic discrete speech sounds in terms of their potential to deliver periodic energy, yielding a relatively coarse separation of all speech sounds into 4 groups (see Section 6.2.2).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table 8.11: Traditional sonority hierarchies (with MH-related information). *Note.* Index values reflect the ordinal ranking of categories in different sonority hierarchies. The voiced affricate in parentheses is the voiced allophone of the voiceless /c/ in MH (see text for details).

Sonority index values			Segmental classes	Phonemic examples (MH)
H_{col}	H_{MH}	H_{exp}		
5	6	10	Vowels	/u, i, o, e, a/
4	5	9	Glides	/w, j/
3	4	8	Liquids	/l, ʁ/
2	3	7	Nasals	/m, n/
1	**2**	**6**	Voiced Fricatives	/v, z/
1	**2**	**5**	Voiced Affricates	(d͡z)
1	**2**	**4**	Voiced Stops	/b, d, g/
1	**1**	**3**	Voiceless Fricatives	/f, s, χ/
1	**1**	**2**	Voiceless Affricates	/c/
1	**1**	**1**	Voiceless Stops	/p, t, k/

8.8.2 Mapping sonority scores to Modern Hebrew data

Traditional SSP-based models focus on the sonority slope of the consonantal sequence, which they rate with a ternary ordinal scale capturing the distinction between sonority slope types: *falls*, *rises* and *plateaus*. As discussed in Asherov & Bat-El (2019), the location of the SSP cut-off point for well-formedness in MH should be found between onset falls on the one hand, and plateaus and rises on the other hand, since *stop-stop* and *fricative-fricative* clusters are known to be licit in MH. Therefore, onset sonority plateaus and rises should be well-formed in MH (thus allowing complex onset clusters), while onset sonority falls should be ill-formed (thus promoting vowel epenthesis to break illicit clusters).

In contrast to the SSP-based models that provide scores on a ternary ordinal scale, the scores in NAP_{td} give a numerical estimation of competition potential based on symbolic representations. The formula used here to derive NAP_{td} scores (see Sections 6.2.2–6.2.3 and Table 6.2) assigns higher numerical values to better-formed combinations, assuming that they represent a weaker competition potential for the nucleus. The search for a cut-off point with NAP_{td} is therefore the search for a number that reliably separates well-formed cluster formations (higher scores) from ill-formed epenthesis cases (lower scores). For the case of MH and the NAP_{td} scale, the cut-off point was found between 1 and 2 such that

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

NAP_{td} scores equal to 1 and below (down to −3) are ill-formed, while NAP_{td} scores that are equal to 2 and above (up to 5) are well-formed, as we shall see in Section 8.9.

Table 8.12: Sonority cut-off points for well-formed complex onsets in MH

$SSP_{col/exp/MH}$	NAP_{td}	Complex onset
plateau/rise	2 – 5	✓
fall	(-3) – 1	✗

Table 8.12 summarizes this expected mapping scheme with color codes that will remain effective throughout this chapter: green for well-formed sequences and red for ill-formed ones. Recall that well-formed cases are those found in the LLHN-based study corpus that have a complex cluster in plural Segholates. The ill-formed cases are the rest of the plural Segholates found in the LLHN-based study corpus, which appear with an epenthetic vowel. Section 8.9 observes this mapping from a "bird's eye view" via bar plots, before going into a more in-depth analysis of the successes and failures of the models with respect to the segmental content of the sequences (Section 8.10).

8.8.3 A note about voicing assimilation processes

Table 8.10 above consisted of colored cells with obstruent sequences that differ in voicing. Purple cells feature voiced-initial sequences that are followed by a voiceless obstruent, while blue cells feature voiceless-initial sequences that are followed by a voiced obstruent. MH is typically considered to exhibit regressive voicing assimilation between adjacent obstruents (see Barkai 1972). The picture is in fact more complex and less dichotomous, not only in terms of the likelihood, but also the degree and even directionality of voicing assimilation in MH (see Bolozky 2006, Kreitman 2010, Mizrachi 2019).

Be that as it may, all of the sequences with obstruents that differ in voicing have the potential to agree in voicing as a result of voicing assimilation processes. Table 8.13 summarizes the possible effects of voicing assimilation processes on the well-formedness predictions of the different sonority models. The top row, *No V.A.*, demonstrates the well-formedness status of those sequences when no voicing assimilation takes place. The two bottom rows, *Reg. V.A.* and *Prog. V.A.*,

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

demonstrate the results of regressive and progressive voicing assimilation (respectively). As can be seen, the two possible directions yield identical results vis-à-vis the well-formedness predictions of the different models. In other words, what matters in this context is only whether the sequences of obstruents do or do not agree in voicing.

The color codes of the fonts in Table 8.13 are in line with the coloring of cells in Table 8.10: the purple sequences are canonically *voiced-voiceless* and the blue sequences are canonically *voiceless-voiced*. Note that the predictions of the SSP_{col} model (right column in Table 8.13) are the same for all the sequences in all conditions. This is the case because the H_{col} hierarchy considers all obstruents as a single sonority class, regardless of voicing and manner distinctions between stops and fricatives. As a result, all of these sequences are evaluated as sonority plateaus (which are well-formed in this context; see Section 8.8.2).

The picture is different in all the other sonority models since they are based on sonority hierarchies that use voicing distinctions. When no voicing assimilation takes place (top row in Table 8.13), the blue *voiceless-voiced* clusters are considered as well-formed in all these models. The SSP_{exp} and SSP_{MH} models evaluate them as onset rises and NAP_{td} scores for these clusters are larger than 1 (which is the NAP_{td} threshold of well-formedness in this context; see Section 8.8.2). At the same time, the purple *voiced-voiceless* clusters are considered ill-formed since they are onset falls in SSP_{exp} and SSP_{MH} terms, and the NAP_{td} score of these clusters is not larger than 1.

Table 8.13 shows the potential implications of voicing assimilation on the well-formedness predictions of the different models. The crucial effect is that in all the models that make voicing distinctions, the purple *voiced-voiceless* clusters have the potential to change from ill-formed to well-formed. This is a complete description of events for both the SSP_{MH} and the NAP_{td} models. The picture is slightly more complex for the SSP_{exp} model, which makes a further manner-based distinction between obstruents, namely separating fricatives from stops. This means that in the SSP_{exp} model, all *fricative-stop* sequences are evaluated as ill-formed onset falls in sequences that agree in voicing. As a result, one of the canonically ill-formed purple sequences remains ill-formed after voicing assimilation and five canonically well-formed sequences become ill-formed (left column and two bottom rows in Table 8.13).

The vast majority of consequences summarized in Table 8.13 are unchanged or improved in terms of well-formedness, when considering the switch from no voicing assimilation (top row) to one of the two voicing assimilation patterns (regressive or progressive). Only a small subset of cases exhibits the few worse-formed scenarios when voicing assimilation occurs. Considering that voicing as-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table 8.13: Voicing assimilation scenarios. *Note.* V.A. = Voicing Assimilation; Reg. = Regressive; Prog. = Progressive. Sequences in parentheses are not very likely (/χ/ does not tend to alternate in voicing). The symbols "✗" and "✓" reflect the binary model predictions for well-formedness: in SSP models, plateaus and rises are well-formed while falls are ill-formed; in the NAP_{td} model, scores larger than 1 are considered well-formed (see Section 8.8.2). The color codes are consistent with Table 8.10. Bold purple font in the regressive voicing assimiliation scenario in the middle (Reg. V.A.) is used to highlight the C_1 devoicing process that was eventually taken into account. See text in Section 8.8.3 for more details.

	SSP_{exp}		SSP_{MH} / NAP_{td}		SSP_{col}
	✗	✓	✗	✓	✓
No V.A.	bk bt dk bc bs bχ df dʃ dχ gf gʃ gχ vt vs vʃ zf zχ	kd pg kv tv cd cv sd sg ʃd ʃg sv ʃv	bk bt dk bc bs bχ df dʃ dχ gf gʃ gχ vt vs vʃ zf zχ	kd pg kv tv cd cv sd sg ʃd ʃg sv ʃv	kd pg kv tv cd cv sd sg ʃd ʃg sv ʃv bk bt dk bc bs bχ df dʃ dχ gf gʃ gχ vt vs vʃ zf zχ
Reg. V.A.	d͡zd zd zg ʒd ʒg **ft**	gd bg gv dv d͡zv zv ʒv **pk pt tk pc** **ps pχ tf tʃ** **tχ kf kʃ kχ** **fs fʃ sf sχ**		gd bg gv dv d͡zd d͡zv zd zg ʒd ʒg zv ʒv **pk pt tk pc** **ps pχ tf tʃ** **tχ kf kʃ kχ** **ft fs fʃ sf sχ**	gd bg gv dv d͡zd d͡zv zd zg ʒd ʒg zv ʒv **pk pt tk pc** **ps pχ tf tʃ** **tχ kf kʃ kχ** **ft fs fʃ sf sχ**
Prog. V.A.	ct st sk ʃt ʃk vd	kt pk kf tf cf sf ʃf bg bd dg bd͡z bz (bʁ) dv dʒ (dʁ) gv gʒ (gʁ) vz vʒ zv (zʁ)		kt pk kf tf ct cf st sk ʃt ʃk sf ʃf bg bd dg bd͡z bz (bʁ) dv dʒ (dʁ) gv gʒ (gʁ) vd vz vʒ zv (zʁ)	kt pk kf tf ct cf st sk ʃt ʃk sf ʃf bg bd dg bd͡z bz (bʁ) dv dʒ (dʁ) gv gʒ (gʁ) vd vz vʒ zv (zʁ)

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

similation is an optional process in MH, the pressure to assimilate in voicing should be weaker if the result is a substantially worse-formed syllable. A reasonable simplification of these facts is to consider the potential of C_1 to devoice, as would be most typically expected in MH. Bolozky (2006: 232) already noted this systematic alternation between MH word-initial *voiced-voiceless* sequences. According to his account, the voiced obstruent in C_1 can retain its voicing with a following epenthetic vowel (resembling a more archaic and prescriptively correct pronunciation), but in most speech contexts when the sequence appears as a cluster, there is a devoicing of C_1.

The outcome of this consideration can be seen in Table 8.13. Devoicing potentials of C_1 are marked in bold for the purple sequences in the middle row of the regressive assimilation scenario. This means that the corpus does not consider the theoretical possibility of the five *fricative-stop* sequences in blue, at the left column of the SSP_{exp} model, to change from well-formed to ill-formed (essentially not disadvantaging the SSP_{exp} model for this less likely and rather negligible possibility).

The propensity to devoice C_1 in *voiced–voiceless* sequences of obstruents is therefore taken into account in the following analysis. In order to consider this variation, the study corpus used here elaborates on the data sourced from the LLHN by providing two forms for plural inflections with a *voiced–voiceless* sequence of obstruents: plurals with an epenthetic vowel, whereby C_1 remains voiced; and plurals with a complex onset cluster, whereby C_1 is devoiced. For example, the singular *zá.χal* 'larva' is expected to yield two possible plural forms: *ze.χa.l-ím* or *sχa.l-ím* 'larva-PL'. In this way, both potential options can be accounted for, regardless of any independent determination about the frequency and likelihood of certain devoiced forms.[5]

8.9 Model fits

The following plots in Figures 8.1–8.4 show the 144 different CC types distributed along the scores of each sonority model. The x-axis in each plot displays the different scores of the given model while the y-axis reflects the amount of CC types that received this score. The color codes reflect the status of C_1 and C_2 in plural Segholates in the LLHN, with the addition of the Variable category in

[5]Note that in the LLHN, most Segholates with *voiced–voiceless* C_1C_2 sequences of obstruents appear with a complex onset cluster in their plural inflections. The only exceptions that appear with an epenthetic vowel in their plural inflections are the sequence types *bc*, *vs*, *vʃ* and *vt*.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

orange, to account for the CC types that can potentially devoice their C_1, and thus change from ill-formed to well-formed complex onset clusters (see Section 8.8.3).

For simplicity and clarity, the plots in Figures 8.1–8.4 show only the distribution of different CC *types* in the corpus. For completeness, the same distribution is presented using the 381 different CC *tokens* (i.e. different lexical items) in Appendix C. Importantly, the differences between the two descriptions of the data are negligible.

The distribution of the data in the following plots (Figures 8.1–8.4) shows that all of the sonority models have a relatively good fit with the bulk of the data. This is true even for the worst fitting models that employ the two extreme sonority hierarchies, i.e. H_{col}, which collapses all obstruents into one class, resulting in the SSP_{col} model (see Figure 8.1), and H_{exp}, which expands the class of obstruents to include all distinctions based on manner of articulation (*stop* < *affricate* < *fricative*) and on voicing (*voiceless* < *voiced*), resulting in the SSP_{exp} model (see Figure 8.2).

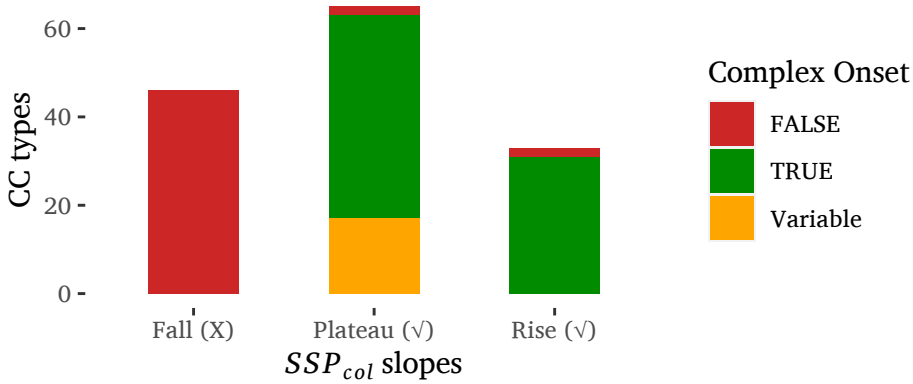

Figure 8.1: Fit of CC types between the SSP_{col} model (x-axis) and the corpus data (color).

Figure 8.1 shows that the SSP_{col} model manages to allocate all of the well-formed onset clusters in the data (green color) to either sonority plateaus or sonority rises, as expected. It exhibits a marginal failure with the allocation of the ill-formed onset clusters (red color) to sonority falls, evident from the red portions at the top of the Plateau and Rise bars. These are, in fact, the sequences of the types /ml, mʁ/ (sonority rises) and /mn, nm/ (sonority plateaus), that all the traditional sonority models fail to predict here. Furthermore, SSP_{col} is incapable of reflecting the potential variation due to voicing assimilation processes

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

(orange color) since any obstruent sequence in SSP_{col} has to be considered a plateau, which, in the context of MH, means that a well-formed onset cluster is expected, irrespective of voicing assimilation.

The fit of SSP_{exp} in Figure 8.2 has the same marginal problem that SSP_{col} exhibits with respect to allocation of the ill-formed onset clusters (red color) to sonority falls, given the red portions at the top of the Plateau and Rise bars. SSP_{exp} introduces a new marginal problem given that some well-formed onset clusters are now allocated to the ill-formed sonority fall category (green portion at the bottom of the Fall bar). These are, in fact, the *fricative-stop* sequences which are all */s/-stop* clusters here. On the other hand, SSP_{exp} succeeds where the SSP_{col} failed in accounting for variation due to voicing assimilation. The potentially varying items (in orange) are allocated to F↔P or F↔R (excluding one case at the very bottom of the Fall bar), indicating their ability to change between an ill-formed sonority fall and a well-formed sonority plateau or rise, respectively.

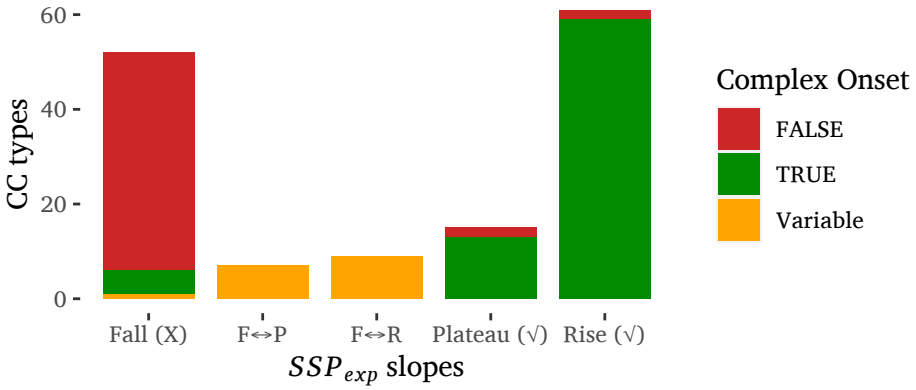

Figure 8.2: Fit of CC types between the SSP_{exp} model (x-axis) and the corpus data (color). *F↔P* can vary between *Fall* and *Plateau* and *F↔R* can vary between *Fall* and *Rise* (both ✗↔✓) due to voicing assimilation.

Not surprisingly, the combination of the H_{col} and H_{exp} hierarchies into a hierarchy that is more specifically tailored to account for distinctions relevant to MH speakers manages to yield the best SSP model in this study: SSP_{MH}. As evident from the plot in Figure 8.3, the only problem that persists in SSP_{MH} is the incomplete allocation of ill-formed clusters in red to the falling onset category, resulting in some of them being allocated to the supposedly well-formed categories of sonority plateaus and falls (these are the nasal-initial plateaus and

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

rises, as mentioned above). At the same time, SSP_{MH} manages to retain the success of SSP_{col} (Figure 8.1) in allocating all the well-formed clusters (in green) to either sonority plateaus or rises. Furthermore, SSP_{MH} manages to retain the success of SSP_{exp} (Figure 8.2) in accounting for variation due to voicing assimilation, where devoiced CC clusters can change score from ill-formed falls to well-formed plateaus.

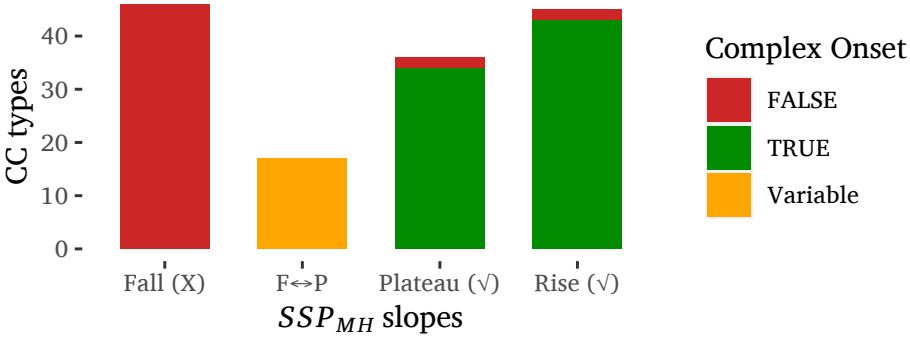

Figure 8.3: Fit of CC types between the SSP_{MH} model (x-axis) and the corpus data (color). *F↔P* can vary between *Fall* and *Plateau* (✗↔✓) due to voicing assimilation.

The most successful fit among the four models in this comparison is found with the NAP_{td} model in Figure 8.4. As expected, all the well-formed clusters in the data (green) are allocated to well-formedness scores of 2 and above, while all the ill-formed clusters in the data (red) are allocated to scores of 1 and below. Likewise, the potentially varying clusters (orange) switch from an ill-formed value of 1 with the voiced-initial clusters to a well-formed value of 3 with the devoiced versions.

8.10 Model analyses

In what follows, the general model fits reported above are examined for the segmental content that underlies their successes and failures, starting with cases that successfully predict the data and are shared by all models (Section 8.10.1), and moving on to the subsets of cases in which SSP models are incongruent with the data (Section 8.10.2).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

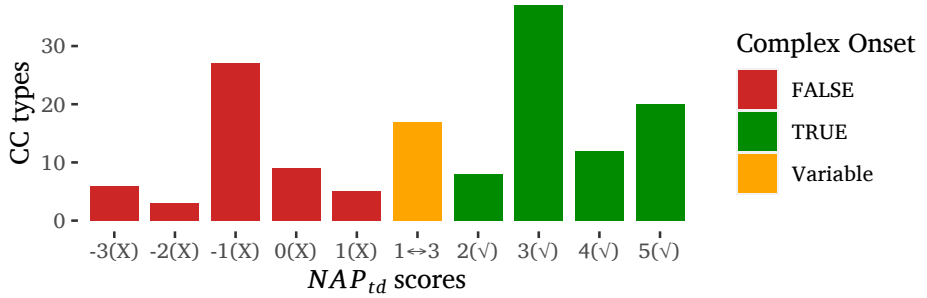

Figure 8.4: Fit of CC types between the NAP_{td} model (x-axis) and the corpus data (color). *1↔3* can vary between scores *1* and *3* (✗↔✓) due to voicing assimilation.

8.10.1 Congruent predictions

Table 8.14 exhibits all the cases that are fully congruent between the data and all the different sonority models. These represent about 82% of the types (118/144) and about 86% of the tokens (329/381) in the dataset. Evidently, for the vast majority of the items in the corpus, all the sonority models are capable of explaining the data and provide scores that are congruent with the data.

The items in Subsets (I–IV) in Table 8.14 are all well-formed given that they are either analyzed as having sonority plateaus or rises in SSP models, or obtain a well-formedness score of 2 and above in the NAP_{td} model. The differences between those sets concern specific assignments and scores, but are redundant in the binary distinction of ill-formed vs. well-formed onsets.

8.10.2 Incongruent predictions

Table 8.15 focuses on the relatively fewer cases of incongruence between scores obtained from the different SSP models and the corpus data (only the scores of the NAP_{td} model were fully congruent with the corpus, see Section 8.9). These incongruent cases account for 18% of the types (26/144) and 14% of the tokens (52/381) in the dataset.

The items in Subset (I) in Table 8.15 are all types of licit */s/-stop* clusters in MH that are produced as complex onset clusters in the plural inflection. Subset (I) is correctly predicted to be well-formed by the SSP_{col} model, in which all the different combinations of obstruent clusters are considered as plateaus. Likewise, the SSP_{MH} model successfully predicts the well-formedness of Subset (I) since it

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table 8.14: Congruence between all sonority models and complex onsets in the MH data

	SSP_{col}	SSP_{exp}	SSP_{MH}	NAP_{td}	Complex onset
I.	**rise**	**rise**	**rise**	**3 – 5**	✓
	bʁ, cl, cm, cʁ, dl, dm, dʁ, gl, gm, gʁ, kl, km, kn, kʁ, pl, pʁ pʁ, sl, sm, sʁ, ʃl, ʃm, ʃn, ʃʁ, tl, tm, tn, tʁ, vʁ, zl, zm, zʁ				
II.	**plateau**	**rise**	**rise**	**2 – 4**	✓
	cd, cv, kd, kv, pg, sd, sg, sv, ʃd, ʃg, ʃv, tv				
III.	**plateau**	**rise**	**plateau**	**2 – 4**	✓
	cf, dv, gv, gz, kc, kf, ks, kʃ, kχ, pc, ps, pʃ, pχ, tf, tʃ, tχ				
IV.	**plateau**	**plateau**	**plateau**	**2 – 3**	✓
	bd, bg, dg, gd, kt, pt, sf, sχ, ʃf, ʃs, ʃχ, tk, zv				
V.	**fall**	**fall**	**fall**	**(-3) – 1**	X
	jc, jd, jf, jg, jk, jl, jʁ, jʃ, jt, jz, jχ, lf, lk, ls, lʃ, lt, lv, lχ, mc mg, ms, mʃ, mt, mz, mχ, nc, nd, nf, ng, nk, ns, nʃ, nt, nv nz, nχ, ʁc, ʁf, ʁg, ʁk, ʁm, ʁs, ʁʃ, ʁt, ʁv, ʁχ				

Note. See Appendix B for the full list of word tokens.

does not make a distinction between stops and fricatives (only voicing is distinctive between obstruents in SSP_{MH}). The SSP_{exp} model fails with subset (I) as it predicts that the clusters will be ill-formed due to the onset sonority fall they incur when stops and fricatives pattern separately on the corresponding sonority scale.

The case of Subsets (IIa–c) in Table 8.15 is of particular interest and requires some elaboration. The sequences in these sets exhibit a *voiced–voiceless* pattern of obstruents that is prone to devoice C_1 due to typical voicing assimilation processes (associated with a switch between ill-formed and well-formed complex onset clusters; see Section 8.8.3). The SSP_{col} model cannot capture this variation since all obstruents belong to the same level in this model. All the other models that do indeed make a distinction between voiceless and voiced (obstruents) succeed in capturing this variation to a large extent. However, because the SSP_{exp}

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table 8.15: Incongruence between SSP models and complex onsets in the MH data

	SSP_{col}	SSP_{exp}	SSP_{MH}	NAP_{td}	Complex onset
I.	plateau	fall	plateau	3	✓
	sk, st, ʃc, ʃk, ʃt				
IIa.	plateau	fall↔rise	fall↔plateau	1 ↔ 3	X ↔ √
	bc, bs, bχ, df, dʃ, dχ, gf, gʃ, gχ				
IIb.	plateau	fall↔plateau	fall↔plateau	1 ↔ 3	X ↔ √
	bk, bt, dk, vs, vʃ, zf, zχ				
IIc.	plateau	fall	fall↔plateau	1 ↔ 3	X ↔ √
	vt				
III.	plateau	plateau	plateau	1	X
	mn, nm				
IV.	rise	rise	rise	1	X
	ml, mʁ				

Note. See Appendix B for the full list of word tokens.

model also makes the distinction between stops and fricatives, it fails to capture the potentially devoiced sequence *vtV*↔*ftV* in Subset (IIc). The SSP_{MH} model, in slight contrast, provides a more uniform picture of a switch between *fall* and *plateau* for all the items in Subsets (IIa–c), much like the successful prediction of the NAP_{td} model, where the values for all items in Subset (IIa–c) uniformly switch from 1 to 3, below and above the threshold of well-formedness, respectively. Essentially, SSP_{MH}, NAP_{td} and, to a certain extent also SSP_{exp}, reflect the expected variation whereby the clusters in Subsets (IIa–c) are predicted to block cluster formation if no voicing assimilation takes place, yet allow complex onsets if devoicing occurs.

Lastly, as reflected in the persistent red portions at the top of the *Plateau* and *Rise* bars in Figures 8.1–8.3, all three SSP-based models tested here fail in predicting the ill-formedness of the sonorant plateaus in Subset (III) and the sonorant-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

initial rises in Subset (IV). Both cases are blocked from surfacing as complex clusters in MH, even though all SSP models consider them as well-formed.[6] In contrast, NAP_{td} assigns a low score of 1 to these clusters (putting them in the range of ill-formed onset clusters), thus correctly predicting that they will be systematically avoided through insertion of an epenthetic vowel.

8.11 Discussion

Formal sonority models can do a lot of heavy lifting with a very simple principle which reduces sonority-based phonotactics to the angle of the sonority slope, but this simplicity comes at a price. It gives the notion of slopes either too much or too little power. Thus, sonority slopes at the lower ends of the sonority hierarchy, such as the notorious */s/-stop* clusters, receive too much power in traditional sonority models, which mostly judge them to be ill-formed, despite of their relative abundance (Goad 2016, Morelli 2003, Steriade 1999). Likewise, sonority slopes at the higher ends of the hierarchy, such as the highly uncommon sonorant rises and plateaus, are considered to be well-formed although they are quite rare (Greenberg 1978).

This corpus study showed that the strictly symbolic model NAP_{td} is the appropriate model for dealing with annotated corpus data of highly abstracted prototypical phonemic transcriptions. The failures of the SSP models compared to NAP_{td} in analyzing the MH data are marginal in quantity, but they are not randomly distributed. These failures were also evident in the experimental study in Chapter 7 and they exhibit the same distinct problems that were highlighted as being an inherent part of traditional sonority sequencing principles in Section 2.2. These problems can be demonstrated with failures to predict the ill-formedness of sonorant-initial onset clusters, which do not present a falling sonority slope (e.g. /nm, ml/), as well as difficulties to predict the well-formedness of some obstruent-initial onset clusters that do not present a rising sonority slope (e.g. */s/-stop* clusters). Note that although SSP_{exp} and SSP_{MH} are capable of considering */s/-stop* clusters as well-formed, they still score them with the border-line well-formedness of plateaus, which may not be the best reflection of the relatively robust well-formed status of */s/-stop* clusters. The NAP_{td} model consistently fares better in accounting for these cases, while at the same time replicating the success of traditional models for the vast majority of cases in which the SSP already provides useful predictions.

[6] Note that even if the NAP_{td} hierarchy would have been used with the SSP, the epenthetic vowel in the cases of /ml, mʁ, mn, nm/ would not have been successfully predicted as all of these sequence types would have been regarded as well-formed onset plateaus.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Part IV

Further contributions of periodic energy to the study of prosody

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

9 Prosodic analysis with periodic energy (ProPer)

The conceptualization of sonority with causal links to pitch perception has direct implications on models that cover prosodic phenomena. If acoustic periodic energy is strongly associated with the notion of sonority, then the major fluctuations along the periodic energy curve should reflect an underlying syllabic structure in the speech signal. This is very similar to a relatively common practice (mentioned in Section 2.2.3), in which the amplitude envelope of the acoustic signal is used to automatically detect syllables (see, e.g., Pfitzinger et al. 1996, Galves et al. 2002, Nakajima et al. 2017, Patha et al. 2016, Port et al. 1996, Räsänen et al. 2018, Tilsen & Arvaniti 2013, Wang & Narayanan 2007). This is typically done by filtering some frequency bands that discriminate in favor of the low-mid range, where most periodic energy in speech is typically found. The periodic energy curve is thus similar to an amplitude modulation curve that is specialized for the detection of syllabic nuclei in acoustic signals.

Having a continuous measure of the duration and the power of the acoustic correlate of sonority is akin to having a continuous representation of an important *syllabic essence*. This is valuable for modeling various aspects of prosodic structure that go far beyond automatic syllable detection, and include acoustic manifestations of *speech rate* and *prominence* aspects of speech. In other words, if we take periodic energy to be the acoustic correlate of sonority, we can deduce from it where syllables are located by observing the location of the major fluctuations on the periodic energy curve. Moreover, we can compute the temporal distance between different syllables in order to measure speech rate, and we can furthermore deduce how prosodically strong each syllable is with respect to other syllables in the same utterance to estimate effects of prominence (e.g. lexical *stress* and post-lexical *accents*).

The advantages mentioned thus far concern only the periodic energy time series. Measuring periodic energy in correspondence with the F0 of the speech signal can unlock a host of other advantages for prosodic analysis. Recall that the values of F0 measurement denote the rate of the fundamental frequency, essentially capturing the *quality* of the pitch sensation in terms of *high* vs. *low* frequen-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

cies. Periodic energy provides the *quantity* component of the same sensation that F0 describes from a qualitative perspective. The two measurements are therefore fully compatible, and their interaction is meaningful in any model that attempts to characterize perceived pitch. Thus, regardless of any link to the linguistic notion of sonority, the interaction between F0 and periodic energy should lead to more comprehensive representations of pitch in speech and beyond.

To test these goals and operationalize them, a set of tools for prosodic analysis based on periodic energy was developed using Praat (Boersma & Weenink 2019) and R (R Core Team 2018) code, that are combined together in a coherent workflow which we call *ProPer*, standing for ***Pro**sodic analysis with **Per**iodic energy* (see notes at the end of this subsection on collaborators and the availability of ProPer).[1] The ProPer tools essentially reduce the acoustic signal into two parallel interacting time series of periodic energy and F0 in order to describe various phenomena in speech prosody by visualization and quantification procedures.

The following presentation should be regarded as a showcase for an independent project that is still under development. It is relevant in the context of this book since it is the direct result of the main claim behind this work – that the quantitative dimension of pitch perception is the basis of sonority, and, as such, has the potential to account for many aspects of prosody that have been thus far hard to model. In the remainder of this chapter I present the various capabilities and advantages that ProPer currently has to offer, without providing a great amount of technical detail (as all technical details can be seen and inspected in the public repository mentioned in the notes below, and many of them may likely change over time). I start by describing how periodic energy data is obtained in ProPer (Section 9.1), and continue by showing how these data can be exploited on their own (Section 9.2), as well as in interaction with F0 data to further enhance our inventory of prosodic analysis tools (Section 9.3).

Important notes:

- The *ProPer* toolbox has been developed in collaboration with Francesco Cangemi and has benefitted from active contributions by T. Mark Ellison and Martine Grice (all from the University of Cologne).
- The ProPer workflow is an open-source project, freely available via an *Open Science Framework* repository at: https://osf.io/28ea5/.

[1]The complete list of *R* packages and versions currently used in ProPer is: R (Version 3.6.3; R Core Team 2018) and the R-packages *Cairo* (Version 1.5.12; Urbanek & Horner 2020), *dplyr* (Version 0.8.5; Wickham, François, et al. 2020), *ggplot2* (Version 3.3.0; Wickham 2016), *purrr* (Version 0.3.4; Henry & Wickham 2020), *rPraat* (Version 1.3.1; Boril 2020), *seewave* (Version 2.1.6; Sueur et al. 2020), *stringr* (Version 1.4.0; Wickham 2019), *tuneR* (Version 1.3.3; Ligges et al. 2018), and *zoo* (Version 1.8.7; Zeileis et al. 2020).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

9.1 Obtaining periodic energy data in ProPer

The current method used to obtain periodic energy data is not inherent to the ProPer workflow, and is expected to change whenever improved methods will become available. It is already different from the method used in the experimental study of NAP (Chapter 7), where the *APP Detector* (Deshmukh et al. 2005, see Section 7.2.3) was used. The current ProPer workflow uses Praat's signal processing abilities to extract the raw data and obtain the periodic energy curve, as detailed below.[2]

We use Praat's autocorrelation analysis to detect periodicity. With autocorrelation, the signal is compared to itself at given time points. Periodic signals are generally more similar to themselves than aperiodic signals such that the level of similarity in the autocorrelation function serve as a very good indication of periodicity in the signal. There are various ways to extract this data from Praat, either directly from a *Harmonicity* object that computes the *harmonics-to-noise ratio* (HNR), or, as we have chosen to do here, from the *strength* value associated with each *pitch candidate* in Praat's *Pitch* object (on a scale of 0–1). We choose the highest strength from up to 15 pitch candidates between 40–1k Hz at each time point (every 1 ms) to determine the *similarity index* (or *periodic fraction*). The full details of this implementation are available in the public release of the ProPer workflow (see notes at the end of the introduction of this chapter).

The similarity index is not indicative of acoustic power and it may give the same values to signals with very different underlying acoustic power. The similarity index values, always ranging from 0 to 1, need to be multiplied by the general acoustic power of the signal in order to express the power of the periodic component, as shown in (9.2). Before doing so, we need to run the inverse function using the formula in (9.1), in order to recover the acoustic power from the intensity measurements of Praat (that are log-transformed to present values in dB SPL).

$$\text{acoustic power} = 4 \times 10^{-10} \times 10^{\frac{\text{intensity}}{10}} \tag{9.1}$$

$$\text{periodic power} = \text{acoustic power} \times \text{similarity index} \tag{9.2}$$

Demonstrations of these data can be viewed in the plot in Figure 9.2, with an example taken from Albert et al. (2019), which is also available at the public ProPer repository (the following example is named "joe_7" in the examples of the ProPer repository). The audio recording is a rendition of the expression *can I*

[2] I thank Paul Boersma for his kind help in solving some of the issues related to Praat via personal communication. Any possible misunderstanding in the interpretation is my own.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

ask you a question?, spontaneously uttered by a morning show host (specifically, Joe Scarborough on MSNBC's *Morning Joe*), and made available to the public at the *TV News Archive*.[3] Figure 9.1 displays a waveform representation of the acoustic signal, and Figure 9.2 shows the *similarity index* in the red dotted line, the *acoustic power* in the dashed blue line, and the resulting *periodic power* in the solid purple curve. Note that scales are normalized to fit the entire plot.

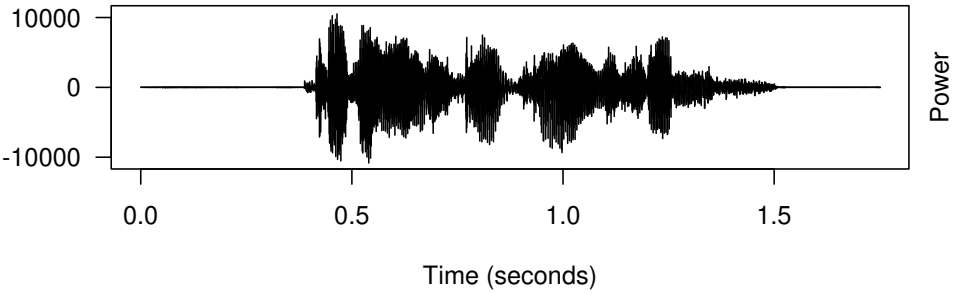

Figure 9.1: Waveform representation (a time-domain oscillogram) of the audio example used in the following Figures 9.2–9.11.

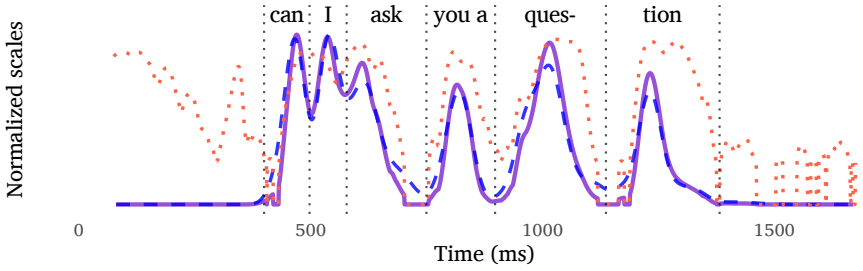

Figure 9.2: Examples of the *similarity index* (red dotted line), *acoustic power* (dashed blue line), and resulting *periodic power* (solid purple curve). Dotted vertical lines and annotations were manually added by the author for exposition purposes. See text for more details.

Note how the purple periodic power curve in Figure 9.2 appears to overlap with the dashed blue curve of the general acoustic power in vocalic portions (the high peaks), but not in the voiceless obstruent portions, where the blue curve shows some energy but the purple curve reaches the floor due to the aperiodicity of the signal (e.g. /s/ in *ask* and *question*).

To obtain the *periodic energy* curve, we log-transform the periodic power values in a similar way as with the *APP Detector*, as explained in Section 7.2.3 and the

[3] https://archive.org/details/tv

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

function in (7.3), repeated here in (9.3). Given the varying conditions of different audio recordings, we need to estimate the threshold of effective pitch sensation on the periodic power scale to set the *periodic floor* variable in the denominator of the log transform in Equation (9.3) and thus set the zero value of the resulting periodic energy curve. The estimation of the periodic floor can be achieved, for example, by sampling voiceless portions in the same dataset to track how high they reach on the periodic power scale.

$$\text{periodic energy} = 10 \log_{10}\left(\frac{\text{periodic power}}{\text{periodic floor}}\right) \tag{9.3}$$

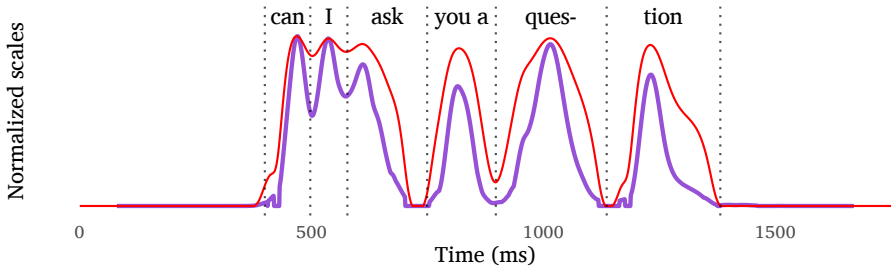

Figure 9.3: Examples of the *periodic power* (solid purple curve) and the log-transformed *periodic energy*, smoothed with a 20 Hz low-pass filter. Other details are the same as for the previous plot. See text for more details.

This can be viewed in the plot in Figure 9.3, which continues with the same audio example, and with the same periodic power curve in purple as in Figure 9.2. The red curve that is added is the *periodic energy* curve after the log-transform function and a 20 Hz low-pass filter that smooths the final periodic energy curve. Again, note that the scales are normalized to fit the plot.

The log-transform is not only useful for setting the floor, it is also a widely used approach to dealing with perception of quantities at various domains and dimensions. In acoustics, it is common to log-transform both the frequency and power scales under the general assumption that differences at the high ends of these scales have a smaller effect than differences of the same absolute size at the lower ends of the scale (e.g. a 100 Hz difference is perceptually salient between 200 and 300 Hz, but it is negligible when occurring between 15 and 15.1kHz). Indeed, it is easy to see how the differences at the higher ends of the purple periodic power curve are diminished in the red periodic energy curve, and, likewise, differences at the lower ends of the purple periodic power curve are enhanced in the red periodic energy curve.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

The periodic energy curve in ProPer is smoothed with low-pass filters at 4 different frequencies from 5 Hz (capturing syllable-size fluctuations of 200 ms-long intervals) to 20 Hz (capturing segment-size fluctuations of down to 50 ms-long intervals). Two values in between, at 8 and 12 Hz, are also automatically extracted to cover intervals of 125 and 83 ms (respectively).

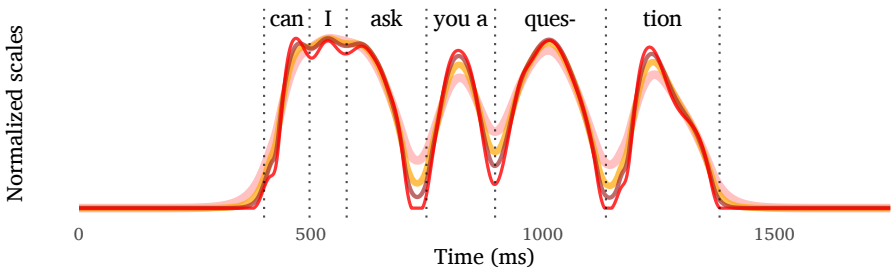

Figure 9.4: Examples of the 4 levels of *periodic energy* smoothings, with low-pass filters at 20 Hz (red), 12 Hz (brown), 8 Hz (orange) and 5 Hz (pink). Other details are the same as for the previous plots. See text for more details.

The plot in Figure 9.4 demonstrates the four levels of smoothing applied in ProPer to the periodic energy curve. This is the same audio example, here with the same red curve of periodic energy with a 20 Hz low-pass filter as in the previous figure (9.3). The added curves show gradually more smoothed behavior by small drops in the frequency of the low-pass filter, from 20 Hz in red, to 12 Hz in brown, to 8 Hz in orange, and down to 5 Hz in the thicker pink colored curve. The least smoothed version (the red curve with 20 Hz low-pass filter) is considered the default and will be used in the remainder of this demonstration.

9.2 Prosodic measurements based on periodic energy

Periodic energy already makes several important measurements available. First of these is the ability to detect the major fluctuations in the curve reflecting different syllables (Section 9.2.1). With syllabic intervals in place, it is possible to measure the strength of each syllable in terms of the periodic energy *mass*, i.e. the integral of duration and power, which is the area under the periodic energy curve (Section 9.2.2). We can then locate the *center of mass* of each syllable – a crucial landmark for many of the following computations, such as the *speech rate* trajectory (Section 9.2.3), which only requires the periodic energy curve.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

9.2.1 Boundary detection

The automatic boundary detector in ProPer is based on the fluctuations of the periodic energy curve. We use the 2nd derivative of the periodic energy curve to locate turning points. Positive local peaks in the 2nd derivative are indicative of relevant turning points in the periodic energy curve, from sharp drops to more subtle *shoulders*. The 2nd derivative undergoes dynamic smoothing in this process. It starts with a very strong smooth of 1 Hz low-pass filtering and repeats the search with incremental steps allowing higher frequencies to control the low-pass filter – effectively reducing the level of smoothing – until the expected amount of boundaries is successfully detected (or until the smoothing reaches 40 Hz low-pass filtering).

As implied above, this algorithm expects a certain number of syllables for each token. If the data is separately annotated (e.g. using Praat's TextGrid to demarcate syllables), it is possible to use this information to derive expectations for syllables. Otherwise, an automatic expectation can be produced given an adjustable average syllable size. The algorithm can take advantage of separately annotated syllabic intervals in another useful way: if syllabic boundaries were segmented by a separate process and fed to ProPer, the automatic boundary detector can avoid the detection of boundaries when they are too far from a pre-segmented boundary, and it can add a boundary if – at the end of the automatic detection process – there are pre-segmented boundaries that have no automatic boundary in their vicinity. It is also possible to completely force the given segmentation on the automatic detector, but that will result in many suboptimal boundaries that are slightly off the periodic energy minima.

The boundary detection algorithm in ProPer thus allows the whole spectrum of behaviors, from fully automatic (signal-based) detection, all the way to fully pre-segmented boundaries, as well as options that incorporate the two. These combined processes use pre-segmented boundaries to inform the signal-based automatic detector, offering an optimal boundary detection in the current system: choosing the desirable periodic energy minima only where a boundary is required, while not missing any crucial boundary that the periodic energy curve cannot detect on its own.

Figure 9.5 demonstrates this with the the same red periodic energy curve as in Figures 9.3–9.4. At the bottom of the plot, two derivative curves fluctuate above and below zero. The green curve is the raw 2nd derivative (i.e. the acceleration curve of the periodic energy trajectory). The purple curve is the dynamically smoothed copy of the 2nd derivative. Automatic boundaries appear in thick red vertical lines and they are located at the positive high peak maxima along the pur-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

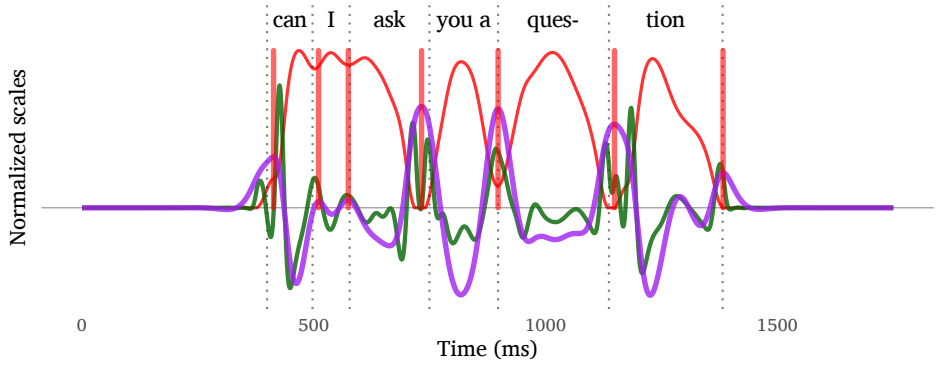

Figure 9.5: A demonstration of the ProPer boundary detector with the same audio example as above, including the same manual segmentation boundaries in dotted vertical black lines, and the same periodic energy curve in red. Vertical red lines denote the boundaries of the ProPer boundary detection algorithm. The curves fluctuating above and below zero are derivatives of the red periodic energy curve: the green curve shows the raw 2nd derivative and the purple curve shows the dynamically smoothed 2nd derivative that is used in the boundary detection algorithm.

ple curve. The dynamically smoothed purple curve is initially highly smoothed (1 Hz low-pass filter) and it stops the process of "unsmoothing" (gradually raising the cut-off frequency of the low-pass filter) as soon as it reached a sufficient number of valid positive peaks on the purple curve. The expected number of boundaries in this example is seven, and it is derived from the number of the manually annotated boundaries provided via a Praat TextGrid (black dotted vertical lines in the plot). Note that since the manual segmentation into syllabic intervals was available for the automatic boundary detection algorithm, it "knew" not to place a boundary in the middle of the last syllable (*-tion*), although the shoulder of the final nasal on the red periodic energy curve was pronounced enough to be detected by the purple 2nd derivative curve as a boundary (a positive peak on the purple curve).

Crucially, ProPer does not require an input of discrete segmental or syllabic intervals in order to work, but, as explained above, it can make use of such standard segmentation information when available. The actual preferred strategy in this respect should be always tied to a specific task. For example, a different preference should be made if it is important to avoid discrete assumption in the model, or, if it is more important to target a specific syllable in a corpus of elicited speech. Another consideration in this respect is related to statistical power. With a relatively small dataset, small deviations can have a big impact on the results,

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

so a separate syllabic segmentation may be a good way to reduce inconsistencies that could result from problematic boundary placement. However, if a relatively big amount of data is considered, small deviations due to suboptimal boundary detection should be more easily identified as noise, and the ability to process big data without a separate segmentation process can become a crucial advantage.

9.2.2 Mass

Once interval boundaries are finalized, it is possible to characterize different aspects of the signal based on the syllable-sized intervals, the most immediate of which is the estimation of prosodic strength. The area under the periodic energy curve between two boundaries is termed *mass* in ProPer (Albert et al. 2022). It is the integral of duration and power, that are often measured as two separate cues to prominence. Typically, acoustic intensity is measured for its contribution to prominence rather than periodic energy. The switch to periodic energy instead of the more general intensity is supported by the the current proposal that periodic energy is related to sonority (Section 5.1), and the relatively established link between sonority and syllable weight (Section 5.3.2), sharing, among others, the idea that voiceless obstruents hardly contribute to syllable weight and prosodic prominence.[4]

Typical usages of intensity in order to measure cues to prominence or sonority tend to employ them within regions of interest in one of the following two ways: (i) extracting peak values from the intensity curve (either minima or maxima, e.g. Parker 2008); or (ii) calculating an average value over the intensity curve (probably the more common strategy of the two). These kinds of measurements either ignore the interaction of duration and power (i), or normalize over the contribution of duration (ii). The periodic energy mass employs a different strategy of summing – rather than averaging or peak tracking – which accounts for duration and power together in a single variable that attempts to capture the overall prosodic strength. This move towards summing was discussed in Section 6.2.4 citing seminal works that provided evidence for the interaction of duration with sonority (Price 1980) and with the perception of loudness in linguistic contexts (Turk & Sawusch 1996).

It is noteworthy to add that duration and power are two abstract aspects of acoustic quantity. They are abstract in the sense that we never experience one

[4]Note that emphasis in service of prosodic prominence (where lexical *stress* and post-lexical *accents* play a role) is different from a selective emphasis that is intended to improve clarity of communication by reducing potential ambiguity. In the latter case, any portion of the speech signal – including voiceless portions – may be the target of emphasis, depending on various contextual variables that have little to do with prosodic prominence.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

without the other. In perception, acoustic quantity is always expressed by the combination of duration and power. The mass measurement in ProPer aims to capture that, while retaining the ability to disintegrate mass into the two sub-components: interval duration and mean periodic energy (which remain, indeed, interesting to observe as well).

Roessig et al. (2022) found ProPer's mass measurement to be the second best predictor of the occurrence of pitch accent (second only to "F0 mean"). Mass was tested in that part of the study alongside 14 other competing acoustic and articulatory measurements. Those included also what might be considered as the two sub-components of mass, "RMS amplitude" and "vowel duration".

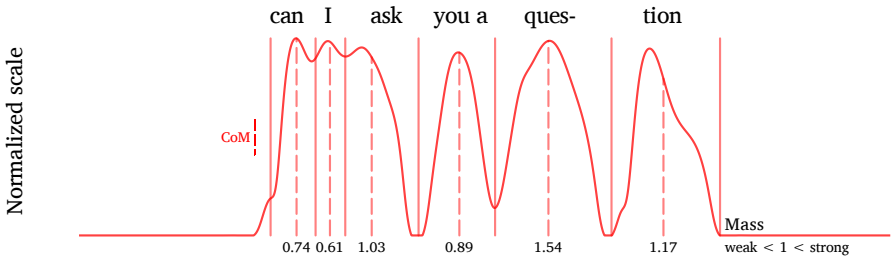

Figure 9.6: A demonstration of *Mass* and *Center of Mass* (CoM) with the same speech example as above. Mass values (relative scale) are presented in numbers below each syllabic interval. The dashed vertical red lines show the position of the CoM within intervals. Other details are the same as for the previous plots. See text for more details.

Note that the raw mass values are, in and of themselves, not very informative. The absolute values are contingent on various degrees of freedom in the adjustment of the periodic energy curve (see Section 9.1), and on the resolution of the dataset (e.g. a data point every 1 ms should yield mass values that are about ten times higher than a data point every 10 ms). In order to calculate the mass values in an informative way, it is useful to calculate relative mass values, representing the prosodic strength of syllables relative to other syllables in the utterance. To achieve this, the area under the periodic energy curve of the entire utterance is calculated and then divided by the number of syllabic intervals in the utterance. The resulting value is the utterance's average mass for a single syllable, which can then be compared against each syllable by calculating the mass of each observed syllable relative to the average value (i.e. observed mass divided by average mass). The resulting values are centered around 1, which is exactly average, such that weak syllables exhibit mass values lower than 1 and strong

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

syllables exhibit mass values higher than 1. Figure 9.6 presents the mass values of each syllabic interval at the bottom of the plot.

The dashed vertical red lines in the middle of syllabic intervals in Figure 9.6 denote the *Center of Mass* (CoM) at each interval. This is a weighted average calculation which finds the average time point weighted by the corresponding periodic energy curve, within each syllabic interval. The center of mass splits the area under the periodic energy curve into two equal parts (within an interval). CoM was introduced in the implementation of the NAP_{bu} model (Section 6.2.4) and equation (6.3), repeated here in (9.4). Note that per = periodic energy and t = time. As we shall see below, the Center of Mass is an essential landmark for many ProPer tools.

$$\text{CoM} = \frac{\sum_i \text{per}_i \, t_i}{\sum_i \text{per}_i} \tag{9.4}$$

9.2.3 Speech rate

In line with the PRiORS framework, presented in Chapter 4 (and specifically Section 4.7.1), the ProPer toolbox views rhythm in speech much like F0 on a slower timescale, that is, as a moving target that exploits dynamic change for communicative effect. Rhythm in speech according to this understanding should be adequately modeled as a trajectory, reminiscent of the local speech rate curves in Pfitzinger (2001).

In keeping with the PRiORS understanding that rhythm trajectories are mechanically related to F0, the speech rate measurements in ProPer are based on temporal distances between anchors rather than on the duration of the intervals. This difference should yield similar results in the majority of cases, but differences are also to be expected (and are yet to be explored).

The speech rate trajectory (see thick green curve in Figure 9.7) is calculated from the temporal distance between successive CoMs, which serve as robust anchors in this context. The continuous curve is based on a smoothed interpolation over these CoM-distance values. The speech rate curve goes up to designate faster rates (shorter distance from the previous CoM) and down for slower rates (larger distance from the previous CoM). The full implementation is available at the public ProPer repository. Note that the speech rate curve starts at the first CoM in Figure 9.7, even though it has no previous CoM to calculate distance from. To overcome this problem, the first syllable is measured for its duration relative to the duration of the longest interval in the same utterance. The status of this initial value should be therefore considered as speculative and experimental at this stage.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

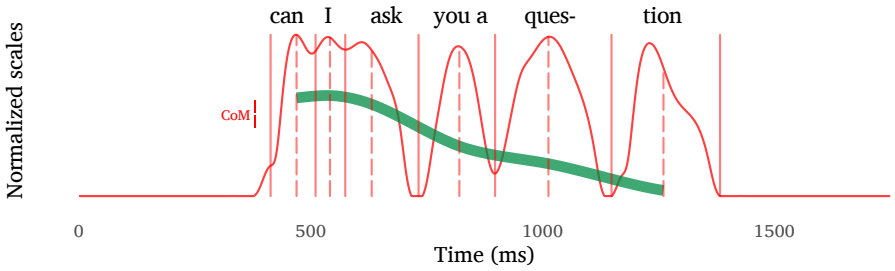

Figure 9.7: A demonstration of the speech rate curve in green, based on the distance between successive CoMs (up = faster; down = slower). Other details are the same as for the previous plots. See text for more details.

9.3 Interactions between F0 and periodic energy

The ProPer tools considered thus far were based solely on the periodic energy curve. As was already mentioned in the opening of this chapter, there are further advantages for the study of prosody that can be unlocked when considering the interaction of periodic energy with the corresponding F0 of the speech signal. These advantages include improvements of the visual representation of F0 data with *periograms* (Section 9.3.1), as well as novel methods to characterize the F0 trajectory between syllables with *ΔF0* (Section 9.3.2) and within syllables with *synchrony* (Section 9.3.3).

9.3.1 Periograms

The first type of interaction between periodic energy and F0 is designed to enrich visual representations of pitch by adding a 3rd informative dimension to the standard visual representations of F0. We call these representations *periograms* (Albert et al. 2018) to echo the 3 dimensions of the spectrogram representation which shows time and frequency on the x/y axes, while representing power in terms of color differences. Most standard visual representations of pitch show a 2-dimensional plot of the F0 trajectory, whereby the x-axis represents time and the y-axis represents frequency. The F0 trajectory in itself is binary – it is either present or absent (*on* or *off*). Figure 9.8 shows the running example with a standard F0 representation.

Periograms enrich the standard representation by adding the power dimension in terms of intuitive changes in visual appearance of the F0 curve. In periograms, the width and darkness of the F0 curve change to reflect the underlying periodic

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

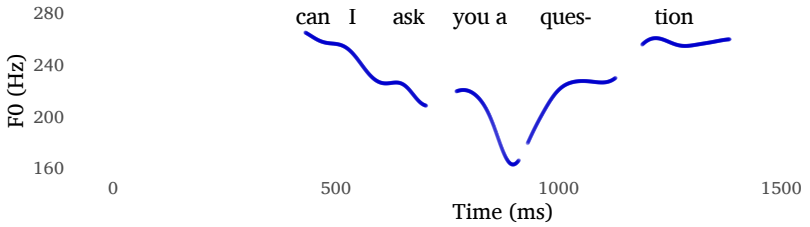

Figure 9.8: Standard "binary" F0 representation. F0 in blue is either present or absent across the 2-dimensional plane, with time on the x-axis and frequency on the y-axis. Other details are the same as for the previous plots. See text for more details.

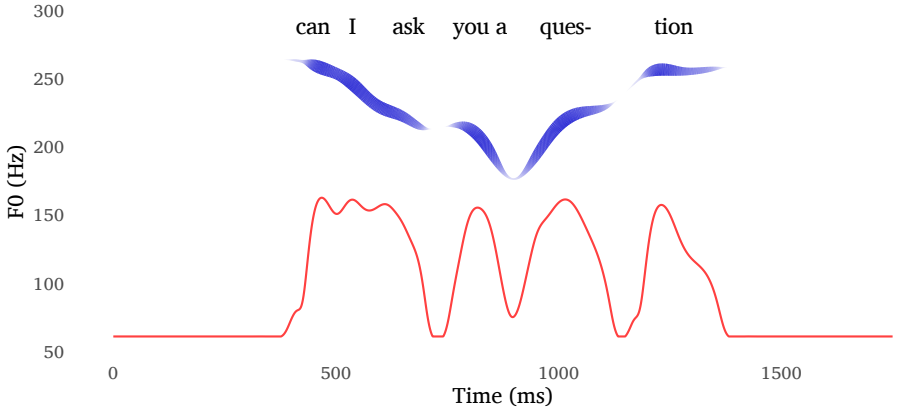

Figure 9.9: A periogram representation. Note how the red periodic energy curve in the lower half of the plot modulates the appearance of the F0 curve in blue in the upper half of the plot. Note also that the frequency values on the y-axis correspond only to F0 at the upper half, not to periodic energy at the lower half. Other details are the same as for the previous plots. See text for more details.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

energy. The F0 curve changes gradually from thin and transparent on the weak end (when the corresponding periodic energy curve is low), to wide and dark on the strong end (when the corresponding periodic energy curve is high). Figure 9.9 shows a periogram representation of the running example.

9.3.2 ΔF0

The following ProPer tools, *ΔF0* and *synchrony*, are designed to characterize F0 shape within and across syllables using metrics that build on the interaction between F0 and periodic energy. The first one is rather straightforward: *ΔF0* (*Delta F0*) extracts the F0 values at the centers of mass and measures the difference in frequency between successive syllables. The ΔF0 values therefore reflect the change in F0 between syllables by computing the difference from the previous syllable. The ΔF0 values are computed in absolute terms (Hz), but they are also transformed to a speaker-specific relative scale where we divide the raw ΔF0 values by the speaker's F0 range, considering all tokens from that speaker. The relative measurement is presented in percentages (see demonstration of ΔF0 in Figure 9.10).

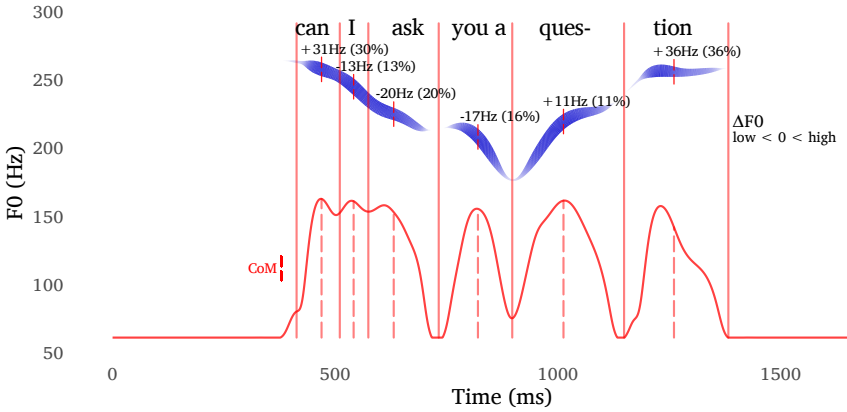

Figure 9.10: A demonstration of ΔF0 data reflecting change in F0 between syllables. The location of the CoMs is indicated by a short red dashed line on the F0 trajectory, to show where F0 values were extracted. ΔF0 values are superimposed above the F0 curve (note the negative and positive signs, which extend also to the values in percentages). Other details are the same as for the previous plots. See text for more details.

Note that this measurement shares methodological aspects with the measurement of speech rate (see Section 9.2.3), as both speech rate and ΔF0 focus on

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

differences between successive CoMs, regarding either their temporal distance (for speech rate) or their spectral distance in F0 (for ΔF0). Relatedly, the utterance-initial syllable cannot provide ΔF0 data that is based on the difference from the previous syllable. Instead, the ΔF0 of the first syllable computes the difference in F0 from the speaker's median F0 value, considering all tokens from that speaker. In this way, the ΔF0 of the first syllable can quite reliably show when speakers start an utterance with low or, more commonly, high pitch.

9.3.3 Synchrony

While ΔF0 is a good indication of long-distance outcomes in terms of pitch change, it is not able to characterize the shape of the F0 trajectory locally, within syllables. For this we designed the complementary measurement termed *synchrony* (Cangemi et al. 2019), which is capable of characterizing the trend of F0 within syllables (rising/falling/level pitch) by taking the non-linear shape of the curves into account.

To achieve this goal, another landmark needs to be extracted from the F0 curve within each syllabic interval. This is very similar to the CoM measurement, being an average point in time, weighted by corresponding curves of acoustic data. The methodology takes inspiration from the *Tonal Center of Gravity* (Barnes et al. 2012), for which an average time point, weighted by F0, is computed to replace the more typical landmark of the F0 peak in standard intonation research.[5]

It is important to note that periodic energy and F0 curves are essentially very different. The periodic energy curve represents a quantity that goes all the way down to zero, while the F0 curve represents a quality with values typically between 50–600 Hz. The interpretation of the periodic energy contribution to our CoM procedure is therefore straightforward, as can be seen in the CoM function in (9.4). However, since F0 does not represent a quantity it needs to be used with caution as it is not immediately clear what it means to sum and average over qualities rather than quantities. For that reason, it makes good sense to call this measurement the *Center of Gravity* (CoG), in keeping with Barnes et al. (2012), and retaining a useful distinction between *mass*, which relates to quantity, and *gravity*, which relates to the shape of the F0 slope. Importantly, to reliably reflect

[5]The F0 peak is commonly used in measurements of *tonal alignment* (e.g. Arvaniti et al. 2006), which calculate temporal distance from a selected F0 peak to an anchor in the segmented speech stream (usually the stressed syllable). The F0 peak is likewise used in standard *scaling* measurements, which calculate the spectral distance in F0 between a selected F0 peak and a previous low turning point on the F0 curve (or any other anchor in the annotation of segmented speech).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

the general slope of the non-linear F0 curve, the CoG measurement requires a few adjustments.

The function for CoG is given in (9.5). As before, t is time and *per* stands for periodic energy. There are two adjustments in the CoG function: (i) we multiply F0 by the corresponding periodic energy (using a normalized 0–1 scale) to account for the strength of F0 at each observed point in time; and (ii) instead of directly using F0 we subtract the constant *F*0floor, which corrects for the problematic distance between the floor of the F0 curve and the never-attained zero value. The first adjustment makes sure that the magnitude of an F0 inflection and its influence on the outcome can be diminished when the signal is weak, based on the underlying periodic energy. For the second adjustment we need to define the *F*0floor variable as detailed below.

$$\text{CoG} = \frac{\sum_i (F0 - F0\text{floor})_i \, \text{per}_i \, t_i}{\sum_i (F0 - F0\text{floor})_i \, \text{per}_i} \tag{9.5}$$

Once we have extracted the two landmarks of CoM and CoG within each syllabic interval we can compute synchrony simply by measuring the temporal distance between these two centers (see examples in Figure 9.11). More rightward displacement of CoG relative to CoM reflects a more rising F0 trend. Likewise, more leftward displacement of CoG relative to CoM reflects a more falling F0 trend. At values around zero the two centers are in synchrony, meaning that the F0 contour is either level, or includes a symmetric rise-fall or fall-rise F0 movement syllable-internally (the additional measure of ΔF0 is needed for complete interpretations of zero synchrony values).

Note that the raw synchrony values are given in absolute terms of milliseconds (ms) and are therefore affected by the overall duration of the interval. To eliminate this effect, relative synchrony values in percentage are given, by dividing the raw synchrony value with the duration of the interval. This allows a more reliable and consistent representation of the angles of the F0 slope.

Without setting any floor for the F0 curve in the CoG function, the values of CoG would show very little sensitivity to the F0 slope. For example, a noticeably rising F0 slope of 50 Hz from 400 to 450 Hz will be computed as having a fixed "quantity" of 400 Hz and a much smaller change of 50 Hz on top of that. Correcting the floor here would mean to designate the minimum F0 as the relevant zero of this interval (400 Hz in this example), so that the change in 50 Hz will become noticeable in the CoG function (50 Hz out of 50 rather than 450 Hz). In fact, it may easily become too noticeable since any change in a trajectory that has its minimal F0 value set to zero can greatly affect the result of the CoG function. Even a negligible rise of 5 Hz would be exaggerated in scale if we simply

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

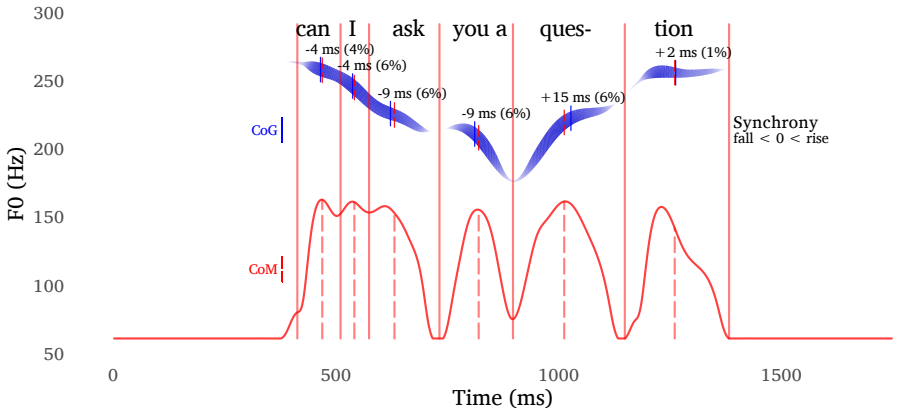

Figure 9.11: A demonstration of synchrony data reflecting change in F0 within syllables. The location of the CoMs is indicated by dashed red vertical lines (under the periodic energy curve and on top of the F0 trajectory). The location of the CoGs is indicated by short blue vertical lines on top of the F0 trajectory. The distance between the two centers yields the synchrony values that are superimposed above the F0 curve (note the negative and positive signs, which extend also to the values in percentages). Other details are the same as the previous plots. See text for more details.

choose the minimum F0 value as our zero for each interval. To solve this problem, the *F0*floor variable in the CoG function computes a certain fixed size to take the floor slightly below the local minimum F0 in each interval. This fixed size is set at 10% of a speaker's F0 range, relative to all tokens produced by that same speaker.

Roessig et al. (2022) found the synchrony measurement to be the third best predictor of different types of pitch accents (closely following two classic measurements, "peak alignment" and "tonal onglide", that are both based on annotated segmental landmarks and F0 turning points). Synchrony was tested in that part of their study alongside 18 other competing acoustic and articulatory measurements.

9.4 ProPer prospects

A brief overview of the ProPer toolbox was shown here to present the benefits of incorporating periodic energy into prosodic research. ProPer is a work in progress but a number of studies have already used the ProPer toolbox in various ways, which can help to evaluate the methodology: Albert et al. (2018),

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Albert (2018), Albert et al. (2019, 2022), Cangemi et al. (2019), Ventura et al. (2019), Sbranna et al. (2021a,b), Lialiou et al. (2021), Savino et al. (2021), Jeon & Nichols (2022) and Roessig et al. (2022). The presentation of ProPer in this chapter is an important opportunity to present the ProPer tools in a context that fully illustrates the rationale behind them, as well as the rationale behind this work: from the theoretical PRiORS framework presented in Chapter 4, which suggests new ways to conceptualize perception models in speech, to the proposals that redefine sonority as a measure of pitch intelligibility in perception, with periodic energy as its acoustic correlate (Chapter 5), all the way to the relevant contribution that periodic energy can make for the study of various prosodic phenomena using the ProPer toolbox (Chapter 9).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Part V

Conclusion

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

10 General discussion

The results of the experimental study (Chapter 7) and the corpus study (Chapter 8), alongside the promising outlook of the ProPer toolbox (Chapter 9), provide strong support for the synergy of proposals laid out in this work. These include the general PRiORS framework for models of auditory perception in linguistic contexts (Chapter 4), the specific treatment of sonority with direct links to perception of pitch and the modeling of syllabic well-formedness with the *Nucleus Attraction Principle* (Chapter 5), as well as the dual-route modeling strategy that considers both top-down and bottom-up inferences with complementary models that can successfully integrate symbolic and dynamic aspects of speech (Chapter 6).

A few interesting issues deserve elaboration given the above. In Section 10.1 I discuss the phonotactic division of labor with respect to sonority, which is defined here in very explicit terms, resulting in a narrower approach to what sonority should and should not account for. The discussion uses the case of */s/-stop* clusters to illustrate this division of labor, making it of special interest as it provides an explanation for the preference of */s/-stop* clusters over other obstruent clusters that has heretofore been lacking. Section 10.2 is devoted to the classic *nature vs. nurture* debate as applied to sonority. Here I explicate the contribution of this work to answering the question of what underlies the universality of sonority-based restrictions. Finally, in Section 10.3 I discuss the complementarity of symbolic/discrete and dynamic/continuous modes in cognitive modeling, suggesting that the top-down–bottom-up distinction exhibits a better fit with the discrete–continuous dichotomy than the classic phonetics–phonology dichotomy. I finish the book with a brief description of directions for future work in Section 10.4.

10.1 Phonotactic division of labor

As already mentioned in Chapter 2 (and especially Section 2.2), sonority has been widely used to explain practically any type of phonotactic phenomenon, since there is nothing in the standard theory that commits the formal concept of sonority to any specific effect in the perception or articulation of speech. The position

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

taken in this work is very different, drawing explicit links between sonority and the auditory perception of pitch. As a result, sonority in this work is a more specific and more narrowly defined concept. This is important since it is very unlikely that one force underlies all the different phonotactic phenomena. However, given that there is no consensus on its phonetic basis, sonority has become too often the lightning rod for unrelated phonotactic phenomena. A more well-defined notion of sonority therefore allows us to achieve a better understanding of the phonotactic division of labor between different forces that play a role in the processing of speech. */s/-stop* clusters make a good case in point.

10.1.1 Towards a holistic account of /s/-stop clusters

NAP's account of the well-formedness of */s/-stop* clusters does not suffice to explain this phonotactic phenomenon since there is nothing in NAP specific to sibilants or stops that would justify assigning a special status to the particular obstruent combination of a sibilant followed by a stop. In fact, any voiceless element is practically invisible to NAP as it is only sensitive to portions of the speech signal that contain sufficient periodic energy. Indeed, the predictions of NAP, which were corroborated by experimental results (in Section 7.8), expect non-sibilant counterparts of /s/, like /f/ in the cluster *ftV*, to pattern with *spV* and *ʃpV*. Furthermore, NAP_{bu} successfully predicted that all the voiceless-initial clusters in the experiment – including the */s/-stop* clusters – generally pattern together as well-formed, as far as sonority-based restrictions are concerned. This may suffice to explain why */s/-stop* clusters are tolerated, but not why they are so often preferred over other obstruent combinations. The complete phonotactic story of */s/-stop* clusters thus requires an integrative explanation, in which sonority only plays a limited role.

First, there are various reasons to assume that *fricative-stop* clusters are better-formed than *stop-stop* clusters. This generalization is traditionally captured in abstract formal phonological constraints like the *Obligatory Contour Principle* (OCP; going back to Leben 1973, and Goldsmith 1976), which acts as a general dissimilatory requirement banning two successive units of the same type. The OCP in this case may be the reflection of an articulatory disadvantage of the *stop-stop* configuration since it should be harder to coordinate two successive closure and release gestures within the span of a complex onset due to reasons of aerodynamics.

Note that this also leads to a disadvantage of *stop-stop* from a perceptual point of view, since the first stop in a *stop-stop* configuration is released into the closure phase of the following stop (see Surprenant & Goldstein 1998). The release of a

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

stop burst into a silent closure phase of another stop, rather than the periodic signal of a vowel, means that many of the acoustic cues to the identity of the first stop consonant are severely attenuated (see Fujimura et al. 1978).

This explanation is essentially based on the concept of perceptual *cue robustness* (Wright 2004), which is less relevant to syllabic organization, but rather based on adjacency between speech sounds and their chances of being recovered given transitions between them. As Ohala and Kawasaki-Fukumori (Ohala & Kawasaki-Fukumori 1997: 361) concluded, "the degree of salience of modulations created by segmental transitions", rather than sonority and syllabicity, is the determinant factor of many phonotactic constraints.

Wright's (2004) *cue robustness* is also critical for the remaining explanation regarding the phonotactic advantage of */s/-stop* clusters over comparable non-sibilant *fricative-stop* clusters, e.g., *spV* vs. *ftV*. Here, the notion of cue robustness serves to explain why sibilants, with their salient and distinctive high frequency aperiodic energy, stand out more than other fricatives, thus allowing more effective recoverability from relatively weak marginal positions (i.e. distant from the vocalic nucleus).

The three phonotactic perspectives are complementary, and although they do not represent an exhaustive list of phonotactic pressures, we need at least these three – *sonority*, *articulatory dissimilation*, and *cue robustness* – in order to properly appreciate the phonotactic phenomenon of */s/-stop* clusters. According to this more holistic account, */s/-stop* clusters are relatively well-formed in terms of sonority because they are not competing for the nucleus, they are well-formed in terms of articulatory coordination complexity due to the two dissimilar successive gestures and, finally, they are robust in terms of their acoustic cues: stops in C_2 can be released into a vowel to optimize the effect of the burst in the release phase, while sibilants retain strong cues to their identity thanks to their unique spectral profile.

10.1.2 Revisiting extrasyllabicity

Recall the common extrasyllabic accounts of sibilants in */s/-stop* clusters, discussed in Section 2.2.2, in which marginal sibilants are given a unique status with respect to syllabification to explain why they are not predicted by traditional sonority accounts. NAP-based accounts present an advantage because they do not need to carve out exceptions in order to theoretically remove sibilants from syllables that are not predicted by the model. Under NAP, those sibilants can remain in the structure as members of a well-formed syllable.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

On the other hand, NAP-based accounts are actually compatible with the kinematic findings in Hermes et al. (2013), which were taken to support extrasyllabic accounts (having found unique articulatory coordination patterns for sibilants in cluster-initial position in Italian). In NAP-based accounts, sonority has prosodic roles to play in carrying the pitch and the overall prosodic strength at the nucleus of the syllable. Marginal voiceless elements can, therefore, be timed with different considerations in NAP-based accounts. For example, it may be beneficial to prolong duration of marginal voiceless elements to increase their recoverability without the risk of increased nucleus competition. This would, indeed, result in some unique timing patterns of marginal sibilants in complex onsets while still fitting comfortably with the rationale of NAP.

10.2 Universality of sonority

A consistent interest within theoretical phonology concerns the universality of sonority-based principles. An impressive volume of publications devoted to this question can be found in the works of Iris Berent and her colleagues, starting with Berent et al. (2007), and followed by many subsequent studies (e.g. Berent et al. 2008, 2015, Berent, Lennertz & Balaban 2012, Berent, Lennertz & Rosselli 2012, Berent et al. 2011, 2014, 2013, Berent 2017, Gómez et al. 2014, Lennertz & Berent 2015, Zhao & Berent 2015). Berent et al. collected mostly behavioral data from perception tasks, where subjects of various different language backgrounds were found to adhere to the SSP, even when presented with combinations that are not attested in their language. The patterns under Berent's consistent scrutiny are usually limited to a set of initial clusters with an onset rise (e.g. *blif*), an onset plateau (e.g. *bdif*) and an onset fall (e.g. *lbif*). Since */s/-clusters* and sonorant plateaus are absent from these studies, Berent's experimental results with SSP-based models are largely compatible with NAP, as the hierarchy *blif* (3) > *bdif* (2) > *lbif* (0) is maintained in NAP_{td} (model scores in brackets).[1]

Berent and her colleagues interpret these findings as supporting the innateness hypothesis, assuming that all humans share a universal linguistic knowledge, which is genetically encoded (the *Universal Grammar* in generative traditions). The universality of sonority principles thus implies innate knowledge of ordinal sonority hierarchies that map onto a discrete representation of the speech signal, with mechanisms that compute the sonority slopes within syllables to determine well-formedness.

[1] NAP_{bu} cannot make such determinations based on symbolic representations, but it should be expected to generally follow the same trends in the vast majority of cases.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

The interpretation of Berent's findings has been a matter of interest in the literature. Some responses, like Daland et al. (2011) and Hayes (2011), have argued that the universal phonotactic behaviors that Berent et al. present can be shown to result from speakers' ability to generalize categories and distributions from the attested lexicon, and use analogy and probabilities to predict unattested forms. Such models can successfully apply statistical learning methods based on the lexicon, without a requirement for prior formal knowledge of sonority (e.g., Jurafsky & Martin 2009, Albright 2009, Bailey & Hahn 2001, Coleman & Pierrehumbert 1997, Futrell et al. 2017, Hayes 2011, Hayes & Wilson 2008, Jarosz et al. 2017, Mayer & Nelson 2019, Vitevitch & Luce 2004, and Mirea & Bicknell 2019).

While it is relatively clear that statistical learners reflect top-down inferences, it is perhaps less obvious that connectionist models, such as Goldsmith (1992), Laks (1995), Smolensky et al. (2014) and Tupper & Fry (2012), also seem to be quite compatible with what is considered here as top-down phonology. Connectionist models can be historically related to an opposition to the classic symbol-based models (see Section 3.1). However, the inputs and outputs of these models are expressed in discrete symbols and they are designed to capture generalizations in terms of the weights of connections in the system, which may serve as a good mechanistic description of top-down operations.

In contrast to traditional sonority principles, NAP was designed to be compatible with general cognitive processes and auditory perception, such that no unique assumptions are required for postulating an innate formal knowledge of sonority. Sonority-based patterns in NAP arise from the general cognitive process that underlies the parsing of the speech stream into syllables with a pitch-bearing nucleus (i.e. nucleus competition). This requirement for pitch-bearing units may be explained in evolutionary timescales as the inevitable result of the important role of pitch in speech communication (Bolinger 1978, Cutler et al. 1997, House 1990, Roettger & Grice 2019) and the observation that tune-text integration occurs with syllable-sized units (e.g. Goldsmith 1976, Ladd 2008, Liberman 1975, Pierrehumbert 1980).

The PRiORS framework in Chapter 4, and especially its take on universal aspects of syllabic structure (Section 4.5.1), can contribute greatly to explanations regarding the universality of syllables, both in terms of their typical duration, which is governed by the temporal regime of perception, and their internal segmental makeup in terms of sonority and pitch perception, which are governed by the timescale of the spectral regime.

The NAP approach appears capable of synthesizing the different views on the origins of universal sonority. The bottom-up model of NAP can explain the universality of sonority as the natural development of communication systems that

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

exploit pitch perception as they shape language systems. The top-down model of NAP is, at the same time, very much in line with statistical phonotactic learners, in which the regularities of language can be deduced from the symbolic abstractions that reflect the speakers' knowledge in stable forms. Top-down inferences reflect the history of the distribution of recognized symbols as they appear in the lexicon of the ambient language. They only indirectly express the functional aspects that we see in the bottom-up route since they reflect the surface manifestations of the functionally-motivated (bottom-up) dynamics.

To conclude, bottom-up NAP combines the innateness claims for formal sonority universals with a more general explanation that is based on the workings of the perceptual and cognitive systems and the evolution of languages as pitch-bearing communication systems. At the same time, top-down NAP is in line with the rationale of statistical learners and the mechanics of connectionist models. These explanations require symbolic interpretation of the signal that abstract from variable dynamic events into stable forms (e.g. consonants, vowels, phonological features) in order to learn and generalize over their distributions.

10.3 Reshuffling dichotomies in linguistic models

This work rejects the classic dichotomy between phonetics and phonology, whereby continuous phenomena are considered phonetic, while phonology is exclusively modeled with discrete terms (see Section 3.1). As was already mentioned in Section 3.2, the integration of dynamic aspects into phonological models has already shown that a phonetics–phonology dichotomy does not fit well with a classic continuous–discrete dichotomy, as we have good reasons to incorporate continuous entities into phonology alongside discrete units, and we have good models to simulate this integration (e.g. Articulatory Phonology and attractor landscape models).

The present work suggests further avenues to integrate dynamics and continuity in perception-based models of phonology, alongside discrete symbolic entities. In this work, the effects of processes that respond to signal-based continuous stimuli are modeled as bottom-up processes, while the effects of processes that are initiated by symbol-based discrete units are separately modeled as top-down processes. As such, this work suggests that the continuous–discrete dichotomy in phonology should be linked to the bottom-up–top-down dichotomy, and that both of these distinct types of processes need to coexist in language systems. It is therefore important to highlight the difference between them.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Bottom-up routes in perception are based on continuous stimuli and they are functional in the sense that they adhere to the laws of physics and to the limitations of the perceptual and cognitive systems of the agents. Bottom-up processes that seem to systematically characterize language processing may be taken to imply an evolutionary benefit for reliable communication.

In contrast, top-down inferences in perception are based on the history of symbolic representations that speakers learn from experience. This learning ability has its own universal functional limitations (e.g. memory-related capacities), but the learned links between the dynamic and symbolic modes can be largely arbitrary, as they rely on the superficial history of co-occurrence, systematically presented by a given language system (see Section 3.4). These symbols and their probabilistic distributions may be constantly updated, in a Bayesian fashion, reflecting knowledge about the distribution of categorically analyzable units of speech, and contributing to what is typically considered to be *phonological knowledge*.

The notion of sonority and its contribution to linguistic sound systems was modeled in this book with the assumption that the two different routes – bottom-up and top-down – are both active when speech inferences take place. The bottom-up model uses continuous data (periodic energy), dynamic principles (attraction and competition), and functional motivation (syllables carry pitch information) to model sonority. The top-down model is based on generalizations over the discrete segmental units in the system and their distribution given bottom-up sonority restrictions (note that the top-down NAP model is not a true statistical learner for reasons that are explained in Section 6.1).

The results of the perception experiments may be taken to support the importance of both routes, given the evident relative success of both the NAP_{td} and the NAP_{bu} model. More specifically, the results of the model comparison in Experiment 2 (see Section 7.8.3) suggest further support for the complementarity of the top-down and bottom-up models. Table 7.4 showed the combined contribution of all the different sonority models to a maximized score, which reflects the combined ability of the models to predict unseen forms (see details in Section 7.7). Model comparison in Experiment 2 showed that the combined contribution of the two NAP models exhibits the highest degree of complementarity among all models (65% for NAP_{td} and 14% for NAP_{bu}), even though they represent the same principle. This is a desirable result for the present framework, which advocates the need for two complementary models to better account for phonological phenomena.

The division of language perception into signal-based models that adhere to the laws of physics and auditory perception, and symbol-based models that ad-

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

here to probability-based inferences in cognitive systems, can have profound implications. For one, it should allow us to extend traditional models of phonology to be readily compatible with models in other related scientific fields, providing more opportunities to share vocabularies and models across disciplines.

10.4 Directions for future work

The novelties that are proposed here for models of sonority, and more generally, for models of phonology and auditory perception, will need to amass more supporting evidence from multiple sources in order to be more widely adopted and consequently developed further. I hope to have laid the foundations for such potential long-term contributions with this book.

There are many different threads this work leaves open. Among them are improved characterizations of the competition procedure in NAP. The method presented here for NAP_{bu}, using the *Center of Mass* calculations (see Section 6.2.4), is not a model of the cognitive process itself, but, rather, an estimation of its result. A more robust and cognitively-plausible measurement would surely improve our bottom-up model of competition for the syllabic nucleus based on NAP.

Furthermore, there is a potentially vast uncharted ground yet to explore by combining the PRiORS theoretical backbone (see Chapter 4) with the methodology of the ProPer toolbox (see Chapter 9). Most immediately, this relates to the study of prosody, where the continuous information of F0 and periodic energy, and their interactions, can be effectively exploited to model the major prosodic effects, namely *intonation*, *prominence* and *speech rate*.

Hopefully, the findings and approaches presented in this work will be able to deliver more valuable insights into old and new problems in phonology and linguistic theory.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Appendix A: Complete output of the Bayesian models

The complete results of the exploratory (Experiment 1) and confirmatory models (Experiment 2–3) are presented in Tables A.1–A.21.

Notice that the parameters are not entirely comparable across models: (i) The intercept, α, represents the mean log-RT of the first category for the ordinal models, but it is the grand mean for the continuous model, NAP_{bu}. (ii) The size of the effect of well-formedness, β, represents the distance between two adjacent categories had they been equidistant for the ordinal models but it is the increase in log-scale for one unit in the well-formedness scale for the continuous model, NAP_{bu}. (iii) The parameter vector, ζ, (present only in ordinal models) represents the normalized distances between consecutive predictor categories, so that the distance between the first and last category is 1. (iv) For all models, the variance components are comparable: σ represents the scale of the log-normal likelihood (or standard deviation of the distribution on the log scale), σ_α and σ_β represent the by-participant adjustment to the intercept and slope respectively, and $\rho_{\alpha,\beta}$ represents the correlation between by-participant intercept and slope.

For each parameter, `Bulk ESS` and `Tail ESS` are effective sample size measures, and `Rhat` is the potential scale reduction factor on split chains (at convergence, `Rhat`$= 1$, and `ESS` $> 10\%$ of post-warmup samples $= 1200$).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table A.1: Results from the exploratory model examining the results of the NAP_{bu} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.51	6.26	6.76	1.00	3422	3521
$\hat{\beta}$	0.00	−0.01	0.00	1.00	4599	4614
$\hat{\sigma}$	0.37	0.33	0.40	1.00	9928	5390
$\hat{\sigma}_{\alpha}$	0.23	0.04	0.58	1.00	2120	1734
$\hat{\sigma}_{\beta}$	0.00	0.00	0.01	1.00	2615	2548
$\hat{\rho}_{\alpha,\beta}$	−0.09	−0.83	0.72	1.00	5437	4908

Table A.2: Results from the exploratory model examining the results of the *Null* model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.76	6.50	7.04	1.00	1699	1920
$\hat{\sigma}$	0.39	0.35	0.43	1.00	4180	4097
$\hat{\sigma}_{\alpha}$	0.27	0.12	0.63	1.00	1408	1524

Table A.3: Results from the exploratory model examining the results of the SSP_{exp} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.88	6.58	7.16	1.00	2469	3101
$\hat{\beta}$	−0.10	−0.19	−0.01	1.00	3931	3852
$\hat{\zeta}_1$	0.67	0.21	0.98	1.00	6903	3137
$\hat{\zeta}_2$	0.33	0.02	0.79	1.00	6903	3137
$\hat{\sigma}$	0.38	0.34	0.41	1.00	8599	5422
$\hat{\sigma}_{\alpha}$	0.31	0.14	0.67	1.00	2464	3230
$\hat{\sigma}_{\beta}$	0.06	0.00	0.19	1.00	2426	3329
$\hat{\rho}_{\alpha,\beta}$	−0.26	−0.90	0.66	1.00	5457	4726

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table A.4: Results from the exploratory model examining the results of the SSP_{col} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	7.02	6.70	7.35	1.00	2313	2875
$\hat{\beta}$	−0.18	−0.27	−0.09	1.00	4371	3662
$\hat{\zeta}_1$	0.84	0.60	0.99	1.00	4927	2647
$\hat{\zeta}_2$	0.16	0.01	0.40	1.00	4927	2647
$\hat{\sigma}$	0.36	0.33	0.40	1.00	8316	5110
$\hat{\sigma}_\alpha$	0.33	0.15	0.73	1.00	2677	3583
$\hat{\sigma}_\beta$	0.06	0.00	0.20	1.00	2359	2925
$\hat{\rho}_{\alpha,\beta}$	−0.30	−0.91	0.61	1.00	5286	5596

Table A.5: Results from the exploratory model examining the results of the MSD_{exp} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.86	6.57	7.15	1.00	2507	3179
$\hat{\beta}$	−0.05	−0.11	0.00	1.00	3741	3191
$\hat{\zeta}_1$	0.33	0.04	0.64	1.00	6564	3212
$\hat{\zeta}_2$	0.13	0.01	0.41	1.00	7609	4539
$\hat{\zeta}_3$	0.18	0.01	0.52	1.00	8598	5011
$\hat{\zeta}_4$	0.18	0.01	0.52	1.00	8300	5200
$\hat{\zeta}_5$	0.17	0.01	0.47	1.00	8367	4663
$\hat{\sigma}$	0.37	0.34	0.41	1.00	8786	5483
$\hat{\sigma}_\alpha$	0.32	0.15	0.69	1.00	2779	4136
$\hat{\sigma}_\beta$	0.05	0.00	0.13	1.00	2161	2411
$\hat{\rho}_{\alpha,\beta}$	−0.35	−0.92	0.54	1.00	6063	5421

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table A.6: Results from the exploratory model examining the results of the MSD_{col} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	7.01	6.69	7.31	1.00	2215	3274
$\hat{\beta}$	−0.13	−0.20	−0.06	1.00	3873	3367
$\hat{\zeta}_1$	0.73	0.48	0.94	1.00	7306	5031
$\hat{\zeta}_2$	0.13	0.01	0.34	1.00	7999	5527
$\hat{\zeta}_3$	0.14	0.01	0.37	1.00	6914	4640
$\hat{\sigma}$	0.36	0.33	0.40	1.00	9150	5817
$\hat{\sigma}_\alpha$	0.33	0.15	0.71	1.00	2431	3409
$\hat{\sigma}_\beta$	0.05	0.00	0.15	1.00	2492	3790
$\hat{\rho}_{\alpha,\beta}$	−0.30	−0.91	0.62	1.00	5388	4968

Table A.7: Results from the exploratory model examining the results of the NAP_{td} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	7.02	6.69	7.34	1.00	2337	3522
$\hat{\beta}$	−0.08	−0.13	−0.03	1.00	3475	3721
$\hat{\zeta}_1$	0.12	0.00	0.34	1.00	6743	3888
$\hat{\zeta}_2$	0.13	0.00	0.37	1.00	6403	3807
$\hat{\zeta}_3$	0.39	0.06	0.72	1.00	6840	4077
$\hat{\zeta}_4$	0.20	0.01	0.52	1.00	6718	5006
$\hat{\zeta}_5$	0.16	0.01	0.37	1.00	7022	3686
$\hat{\sigma}$	0.36	0.32	0.39	1.00	9611	6104
$\hat{\sigma}_\alpha$	0.35	0.16	0.76	1.00	2598	3969
$\hat{\sigma}_\beta$	0.04	0.00	0.12	1.00	1993	2749
$\hat{\rho}_{\alpha,\beta}$	−0.40	−0.93	0.48	1.00	4377	5097

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table A.8: Results from Experiment 2 model examining the results of the NAP_{bu} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.48	6.42	6.54	1.00	4241	6483
$\hat{\beta}$	0.00	0.00	0.00	1.00	8424	9130
$\hat{\sigma}$	0.37	0.36	0.38	1.00	25283	8277
$\hat{\sigma}_{\alpha}$	0.20	0.16	0.26	1.00	5968	7794
$\hat{\sigma}_{\beta}$	0.00	0.00	0.00	1.00	4577	5943
$\hat{\rho}_{\alpha,\beta}$	0.12	−0.25	0.46	1.00	4703	7104

Table A.9: Results from Experiment 2 model examining the results of the *Null* model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.69	6.63	6.76	1.00	753	1402
$\hat{\sigma}$	0.39	0.38	0.39	1.00	12050	9344
$\hat{\sigma}_{\alpha}$	0.21	0.17	0.26	1.00	941	1718

Table A.10: Results from Experiment 2 model examining the results of the SSP_{exp} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.76	6.69	6.82	1.00	1405	2886
$\hat{\beta}$	−0.07	−0.08	−0.05	1.00	7932	7602
$\hat{\zeta}_1$	0.55	0.34	0.76	1.00	5261	6751
$\hat{\zeta}_2$	0.45	0.24	0.66	1.00	5261	6751
$\hat{\sigma}$	0.38	0.37	0.39	1.00	13103	8310
$\hat{\sigma}_{\alpha}$	0.22	0.18	0.28	1.00	2113	4788
$\hat{\sigma}_{\beta}$	0.03	0.00	0.06	1.00	1467	2084
$\hat{\rho}_{\alpha,\beta}$	−0.30	−0.75	0.30	1.00	10180	5395

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table A.11: Results from Experiment 2 model examining the results of the SSP_{col} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.91	6.83	7.00	1.00	2823	4693
$\hat{\beta}$	−0.14	−0.17	−0.11	1.00	4828	7415
$\hat{\zeta}_1$	0.91	0.83	0.98	1.00	9893	5241
$\hat{\zeta}_2$	0.09	0.02	0.17	1.00	9893	5241
$\hat{\sigma}$	0.37	0.36	0.38	1.00	20139	8355
$\hat{\sigma}_\alpha$	0.27	0.22	0.34	1.00	3389	5328
$\hat{\sigma}_\beta$	0.08	0.06	0.11	1.00	3885	5590
$\hat{\rho}_{\alpha,\beta}$	−0.60	−0.79	−0.33	1.00	6511	7972

Table A.12: Results from Experiment 2 model examining the results of the MSD_{exp} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.76	6.69	6.83	1.00	1207	2483
$\hat{\beta}$	−0.04	−0.05	−0.03	1.00	9059	8906
$\hat{\zeta}_1$	0.32	0.18	0.46	1.00	11717	8196
$\hat{\zeta}_2$	0.07	0.00	0.20	1.00	10623	5950
$\hat{\zeta}_3$	0.25	0.03	0.51	1.00	11737	6303
$\hat{\zeta}_4$	0.27	0.03	0.52	1.00	12246	6832
$\hat{\zeta}_5$	0.10	0.00	0.27	1.00	13303	7245
$\hat{\sigma}$	0.38	0.37	0.39	1.00	17200	8875
$\hat{\sigma}_\alpha$	0.22	0.18	0.28	1.00	2036	4080
$\hat{\sigma}_\beta$	0.01	0.00	0.03	1.00	2493	3373
$\hat{\rho}_{\alpha,\beta}$	−0.45	−0.85	0.11	1.00	10053	7010

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table A.13: Results from Experiment 2 model examining the results of the MSD_{col} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.91	6.83	7.00	1.00	3182	4917
$\hat{\beta}$	−0.10	−0.12	−0.08	1.00	5746	8322
$\hat{\zeta}_1$	0.86	0.77	0.95	1.00	12442	6840
$\hat{\zeta}_2$	0.07	0.01	0.15	1.00	13136	6788
$\hat{\zeta}_3$	0.07	0.00	0.17	1.00	13613	7719
$\hat{\sigma}$	0.37	0.36	0.38	1.00	20521	8436
$\hat{\sigma}_\alpha$	0.27	0.22	0.34	1.00	4490	6821
$\hat{\sigma}_\beta$	0.06	0.04	0.08	1.00	4531	7459
$\hat{\rho}_{\alpha,\beta}$	−0.59	−0.79	−0.34	1.00	7168	8915

Table A.14: Results from Experiment 2 model examining the results of the NAP_{td} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.94	6.85	7.02	1.00	2816	4806
$\hat{\beta}$	−0.07	−0.09	−0.06	1.00	4432	6643
$\hat{\zeta}_1$	0.10	0.01	0.22	1.00	10863	6583
$\hat{\zeta}_2$	0.07	0.00	0.18	1.00	11702	7035
$\hat{\zeta}_3$	0.48	0.34	0.62	1.00	16602	9264
$\hat{\zeta}_4$	0.24	0.13	0.36	1.00	17005	8633
$\hat{\zeta}_5$	0.11	0.03	0.18	1.00	13157	5839
$\hat{\sigma}$	0.36	0.35	0.37	1.00	19667	8740
$\hat{\sigma}_\alpha$	0.28	0.23	0.35	1.00	3831	5793
$\hat{\sigma}_\beta$	0.04	0.03	0.05	1.00	4381	6643
$\hat{\rho}_{\alpha,\beta}$	−0.64	−0.81	−0.40	1.00	6906	8639

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table A.15: Results from Experiment 3 model examining the results of the NAP_{bu} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.43	6.32	6.55	1.00	1320	2587
$\hat{\beta}$	0.00	0.00	0.00	1.00	4426	5286
$\hat{\sigma}$	0.38	0.37	0.39	1.00	16733	5993
$\hat{\sigma}_{\alpha}$	0.32	0.24	0.43	1.00	2186	3603
$\hat{\sigma}_{\beta}$	0.00	0.00	0.00	1.00	3905	5283
$\hat{\rho}_{\alpha,\beta}$	0.15	−0.26	0.52	1.00	3634	4740

Table A.16: Results from Experiment 3 model examining the results of the *Null* model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.53	6.42	6.64	1.01	339	647
$\hat{\sigma}$	0.39	0.38	0.40	1.00	3590	4226
$\hat{\sigma}_{\alpha}$	0.32	0.25	0.41	1.00	576	983

Table A.17: Results from Experiment 3 model examining the results of the SSP_{exp} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.57	6.46	6.69	1.00	879	1658
$\hat{\beta}$	−0.04	−0.07	−0.02	1.00	5233	5514
$\hat{\zeta}_1$	0.84	0.61	0.99	1.00	7170	4354
$\hat{\zeta}_2$	0.16	0.01	0.39	1.00	7170	4354
$\hat{\sigma}$	0.38	0.37	0.39	1.00	15130	5447
$\hat{\sigma}_{\alpha}$	0.33	0.26	0.43	1.00	1600	2901
$\hat{\sigma}_{\beta}$	0.05	0.03	0.08	1.00	3175	4883
$\hat{\rho}_{\alpha,\beta}$	−0.21	−0.60	0.24	1.00	6966	6286

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table A.18: Results from Experiment 3 model examining the results of the SSP_{col} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.62	6.49	6.74	1.00	1042	1848
$\hat{\beta}$	−0.07	−0.10	−0.04	1.00	2548	3659
$\hat{\zeta}_1$	0.91	0.77	1.00	1.00	6966	3706
$\hat{\zeta}_2$	0.09	0.00	0.23	1.00	6966	3706
$\hat{\sigma}$	0.38	0.37	0.39	1.00	13379	6069
$\hat{\sigma}_\alpha$	0.36	0.28	0.46	1.00	1599	3025
$\hat{\sigma}_\beta$	0.07	0.05	0.10	1.00	3090	4429
$\hat{\rho}_{\alpha,\beta}$	−0.43	−0.72	−0.06	1.00	3567	5039

Table A.19: Results from Experiment 3 model examining the results of the MSD_{exp} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.57	6.45	6.69	1.00	1101	1762
$\hat{\beta}$	−0.02	−0.03	−0.01	1.00	5495	6179
$\hat{\zeta}_1$	0.63	0.41	0.83	1.00	8412	5254
$\hat{\zeta}_2$	0.08	0.00	0.26	1.00	8409	3902
$\hat{\zeta}_3$	0.12	0.00	0.33	1.00	8357	5302
$\hat{\zeta}_4$	0.11	0.00	0.31	1.00	8465	5234
$\hat{\zeta}_5$	0.06	0.00	0.20	1.00	10235	5349
$\hat{\sigma}$	0.38	0.37	0.39	1.00	12449	5474
$\hat{\sigma}_\alpha$	0.33	0.26	0.43	1.00	1778	3012
$\hat{\sigma}_\beta$	0.02	0.01	0.04	1.00	3317	4125
$\hat{\rho}_{\alpha,\beta}$	−0.23	−0.61	0.22	1.00	6490	5979

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Table A.20: Results from Experiment 3 model examining the results of the MSD_{col} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.62	6.49	6.74	1.00	682	1299
$\hat{\beta}$	−0.04	−0.07	−0.02	1.00	2356	4065
$\hat{\zeta}_1$	0.87	0.72	0.97	1.00	9025	4916
$\hat{\zeta}_2$	0.08	0.00	0.21	1.00	7847	4930
$\hat{\zeta}_3$	0.05	0.00	0.16	1.00	10180	4781
$\hat{\sigma}$	0.38	0.37	0.39	1.00	14667	5153
$\hat{\sigma}_\alpha$	0.36	0.28	0.46	1.00	1625	3090
$\hat{\sigma}_\beta$	0.05	0.03	0.07	1.00	2849	4976
$\hat{\rho}_{\alpha,\beta}$	−0.43	−0.72	−0.05	1.00	3649	5390

Table A.21: Results from Experiment 3 model examining the results of the NAP_{td} model. See text for the interpretation of the parameters and column names.

Parameter	Estimate	l-95% CI	u-95% CI	Rhat	Bulk ESS	Tail ESS
$\hat{\alpha}$	6.64	6.51	6.77	1.00	928	1670
$\hat{\beta}$	−0.03	−0.05	−0.02	1.00	2191	3841
$\hat{\zeta}_1$	0.36	0.17	0.53	1.00	7212	4742
$\hat{\zeta}_2$	0.10	0.00	0.29	1.00	6872	4497
$\hat{\zeta}_3$	0.40	0.20	0.59	1.00	9288	6920
$\hat{\zeta}_4$	0.08	0.00	0.24	1.00	9474	6006
$\hat{\zeta}_5$	0.06	0.00	0.17	1.00	9796	5460
$\hat{\sigma}$	0.38	0.37	0.39	1.00	12640	5343
$\hat{\sigma}_\alpha$	0.36	0.28	0.47	1.00	1821	2456
$\hat{\sigma}_\beta$	0.04	0.03	0.05	1.00	2516	4376
$\hat{\rho}_{\alpha,\beta}$	−0.43	−0.71	−0.08	1.00	3396	4574

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Appendix B: Alphabetized list of MH Segholate nouns in the corpus study

(1) bé.ca
(2) bé.dek
(3) bé.ged
(4) bé.ka
(5) bé.ʁez
(6) bé.ten
(7) bó.ʁeg
(8) bó.sem
(9) bó.χan
(10) cé.dek
(11) cé.fa
(12) cé.fek
(13) cé.fi
(14) cé.lem
(15) cé.maχ
(16) cé.med
(17) cé.meʁ
(18) cé.va
(19) cé.veʁ
(20) cé.vet
(21) có.fen
(22) có.met
(23) có.ʁeχ
(24) dá.χaf
(25) dé.fek
(26) dé.gel
(27) dé.gem
(28) dé.kel
(29) dé.lef
(30) dé.lek
(31) dé.let
(32) dé.ma
(33) dé.ʁeg
(34) dé.ʁeχ
(35) dé.ʃe
(36) dé.ʃen
(37) dé.vek
(38) dé.veʁ
(39) dé.χi
(40) dó.fek
(41) dó.fen
(42) dó.fi
(43) dó.lev
(44) dó.men
(45) dó.χak
(46) dó.χan
(47) gá.χal
(48) gé.dem
(49) gé.fen
(50) gé.led
(51) gé.ʁed
(52) gé.ʃem
(53) gé.ʃeʁ
(54) gé.va
(55) gé.veʁ
(56) gé.ves
(57) gé.za
(58) gé.zel
(59) gé.zem
(60) gé.zeʁ
(61) gó.del
(62) gó.deʃ
(63) gó.feʁ
(64) gó.lem
(65) gó.me
(66) gó.ʁen
(67) gó.vah
(68) já.χas
(69) jé.ceʁ
(70) jé.da
(71) jé.ga
(72) jé.kev
(73) jé.led
(74) jé.ʁaχ
(75) jé.ʃa
(76) jé.teʁ
(77) jé.za
(78) jó.fi
(79) ká.χaʃ
(80) ké.caχ
(81) ké.cef
(82) ké.ceʁ
(83) ké.cev
(84) ké.dem

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

(85) ké.fel
(86) ké.fel
(87) ké.la
(88) ké.laχ
(89) ké.laχ
(90) ké.le
(91) ké.les
(92) ké.let
(93) ké.lev
(94) ké.maχ
(95) ké.meʃ
(96) ké.met
(97) ké.nes
(98) ké.ʁa
(99) ké.ʁa
(100) ké.ʁaχ
(101) ké.ʁem
(102) ké.ʁen
(103) ké.ʁes
(104) ké.ʁes
(105) ké.ʁes
(106) ké.ʁeʃ
(107) ké.ʁet
(108) ké.ʁeχ
(109) ké.sef
(110) ké.sem
(111) ké.set
(112) ké.set
(113) ké.ʃel
(114) ké.ʃeʁ
(115) ké.ʃet
(116) ké.ʃev
(117) ké.ta
(118) ké.tel
(119) ké.tem
(120) ké.teʁ
(121) ké.va
(122) ké.vel
(123) ké.veʁ
(124) ké.ves
(125) ké.ves
(126) ké.veʃ
(127) kó.feʁ
(128) kó.mec
(129) kó.meʁ
(130) kó.ʃeʁ
(131) kó.ʃi
(132) kó.tel
(133) kó.tel
(134) kó.ten
(135) kó.teʁ
(136) kó.tev
(137) kó.vec
(138) kó.ved
(139) lá.χac
(140) lá.χan
(141) lá.χaʃ
(142) lé.fet
(143) lé.kaχ
(144) lé.ket
(145) lé.set
(146) lé.ʃem
(147) lé.tet
(148) lé.ved
(149) lé.vet
(150) lé.χem
(151) má.χac
(152) má.χak
(153) má.χat
(154) mé.caχ
(155) mé.ged
(156) mé.laχ
(157) mé.lel
(158) mé.let
(159) mé.leχ
(160) mé.na
(161) mé.ʁec
(162) mé.ʁed
(163) mé.ʁek
(164) mé.seg
(165) mé.seʁ
(166) mé.ʃek
(167) mé.ʃeχ
(168) mé.taχ
(169) mé.teg
(170) mé.tek
(171) mé.zaχ
(172) mé.zeg
(173) mé.χeʁ
(174) mé.χes
(175) ná.χac
(176) ná.χal
(177) né.caχ
(178) né.ceʁ
(179) né.deʁ
(180) né.faχ
(181) né.fec
(182) né.fel
(183) né.feʃ
(184) né.ga
(185) né.gef
(186) né.gen
(187) né.gev
(188) né.ka
(189) né.keʁ
(190) né.kev
(191) né.mek
(192) né.meʃ
(193) né.seχ
(194) né.ʃef
(195) né.ʃek
(196) né.ʃel
(197) né.ʃeʁ
(198) né.ʃeχ
(199) né.ta
(200) né.taχ
(201) né.tek
(202) né.tel
(203) né.tez
(204) né.teχ

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

(205) né.veg
(206) né.vel
(207) né.vet
(208) né.zek
(209) né.zem
(210) né.zeʁ
(211) né.χed
(212) né.χel
(213) né.χes
(214) pá.χad
(215) pá.χaz
(216) pé.ca
(217) pé.ga
(218) pé.geʁ
(219) pé.geʃ
(220) pé.laχ
(221) pé.le
(222) pé.leg
(223) pé.les
(224) pé.let
(225) pé.leχ
(226) pé.ʁaχ
(227) pé.ʁe
(228) pé.ʁec
(229) pé.ʁek
(230) pé.ʁeʃ
(231) pé.sa
(232) pé.saχ
(233) pé.sek
(234) pé.sel
(235) pé.ʃa
(236) pé.ʃeʁ
(237) pé.taχ
(238) pé.tek
(239) pé.tel
(240) pé.tem
(241) pé.ti
(242) ʁá.χam
(243) ʁá.χaʃ
(244) ʁé.caχ
(245) ʁé.cef
(246) ʁé.feʃ
(247) ʁé.fet
(248) ʁé.ga
(249) ʁé.geʃ
(250) ʁé.gev
(251) ʁé.ka
(252) ʁé.kaχ
(253) ʁé.kev
(254) ʁé.mec
(255) ʁé.mes
(256) ʁé.mez
(257) ʁé.sek
(258) ʁé.sen
(259) ʁé.ses
(260) ʁé.ʃef
(261) ʁé.ʃet
(262) ʁé.ta
(263) ʁé.tek
(264) ʁé.tet
(265) ʁé.tet
(266) ʁé.va
(267) ʁé.vaχ
(268) ʁé.χem
(269) ʁé.χes
(270) ʁé.χeʃ
(271) ʁé.χev
(272) ʁó.ʃem
(273) ʁó.tev
(274) ʁó.ved
(275) sá.χaf
(276) sá.χaʁ
(277) sá.χav
(278) sé.dek
(279) sé.deʁ
(280) sé.faχ
(281) sé.fel
(282) sé.feʁ
(283) sé.gel
(284) sé.gen
(285) sé.geʁ
(286) sé.gev
(287) sé.keʁ
(288) sé.la
(289) sé.lek
(290) sé.mel
(291) sé.meχ
(292) sé.ʁaχ
(293) sé.ʁen
(294) sé.ʁet
(295) sé.teʁ
(296) sé.vel
(297) sé.veʁ
(298) sé.χel
(299) sé.χel
(300) sé.χem
(301) sé.χeʁ
(302) ʃá.χac
(303) ʃá.χaf
(304) ʃá.χak
(305) ʃá.χal
(306) ʃé.cef
(307) ʃé.deʁ
(308) ʃé.fa
(309) ʃé.fel
(310) ʃé.feχ
(311) ʃé.geʁ
(312) ʃé.ka
(313) ʃé.kec
(314) ʃé.kef
(315) ʃé.kel
(316) ʃé.keʁ
(317) ʃé.ket
(318) ʃé.laχ
(319) ʃé.led
(320) ʃé.lef
(321) ʃé.leg
(322) ʃé.let
(323) ʃé.ma
(324) ʃé.mec

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

(325) ʃé.meʁ
(326) ʃé.meʃ
(327) ʃé.nec
(328) ʃé.net
(329) ʃé.ʁec
(330) ʃé.sa
(331) ʃé.taχ
(332) ʃé.tef
(333) ʃé.tel
(334) ʃé.ten
(335) ʃé.vaχ
(336) ʃé.veʁ
(337) ʃé.vet
(338) ʃé.χem
(339) ʃé.χev
(340) ʃó.ket
(341) tá.χaʃ
(342) tá.χav
(343) té.faχ
(344) té.feʁ
(345) té.fes
(346) té.ka
(347) té.ken
(348) té.keʁ
(349) té.kes
(350) té.lem
(351) té.ma
(352) té.meχ
(353) té.na
(354) té.ne
(355) té.ʁed
(356) té.ʁef
(357) té.ʁeʃ
(358) té.ʃeʁ
(359) té.vaχ
(360) té.ven
(361) tó.fes
(362) tó.fet
(363) tó.ʁen
(364) tó.χen
(365) vé.ʁed
(366) vé.set
(367) vé.ʃet
(368) vé.tek
(369) zá.χal
(370) zé.fek
(371) zé.fet
(372) zé.lef
(373) zé.meʁ
(374) zé.ʁa
(375) zé.ʁaχ
(376) zé.ʁed
(377) zé.ʁem
(378) zé.ʁet
(379) zé.vaχ
(380) zé.vel
(381) ze.χeʁ

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Appendix C: Model fits of the corpus data using CC tokens

Figures C.1–C.4 are equivalent to Figures 8.1–8.4. The latter show the distribution of CC types and the former (here) show the distribution of different CC tokens (i.e. different lexical items). The differences between the two descriptions of the data are negligible.

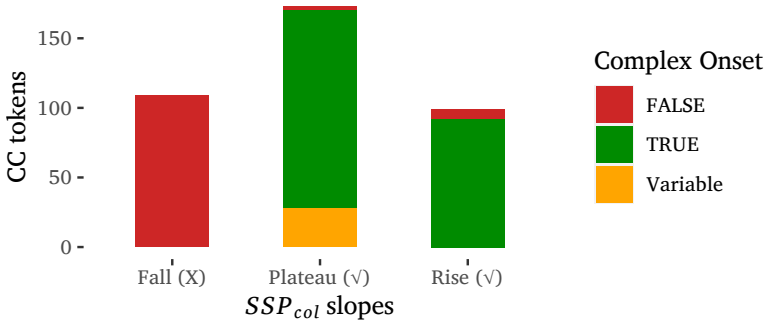

Figure C.1: Fit of CC tokens between the SSP_{col} model (x-axis) and the corpus data (color).

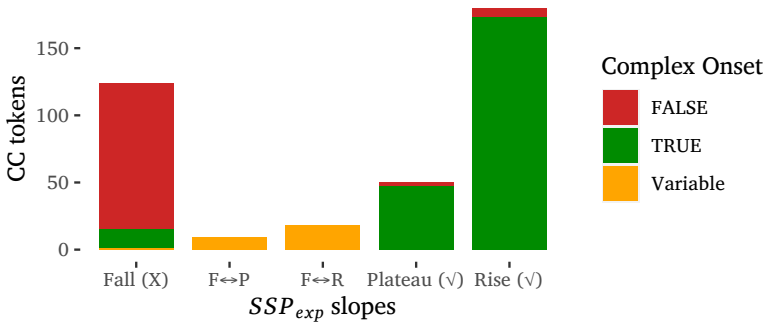

Figure C.2: Fit of CC tokens between the SSP_{exp} model (x-axis) and the corpus data (color). *F↔P* can vary between *Fall* & *Plateau* and *F↔R* can vary between *Fall* & *Rise* (both *X↔√*) due to voicing assimilation

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

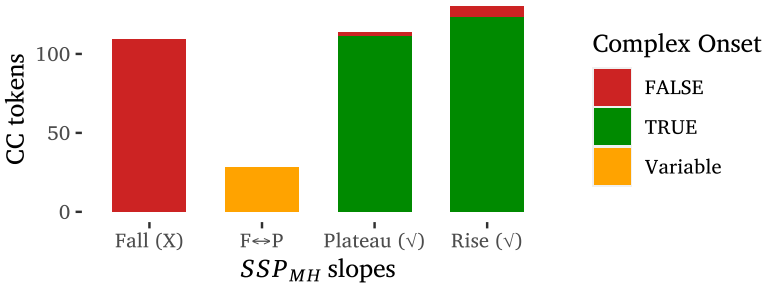

Figure C.3: Fit of CC tokens between the SSP_{MH} model (x-axis) and the corpus data (color). *F↔P* can vary between *Fall* & *Plateau* (*X↔√*) due to voicing assimilation

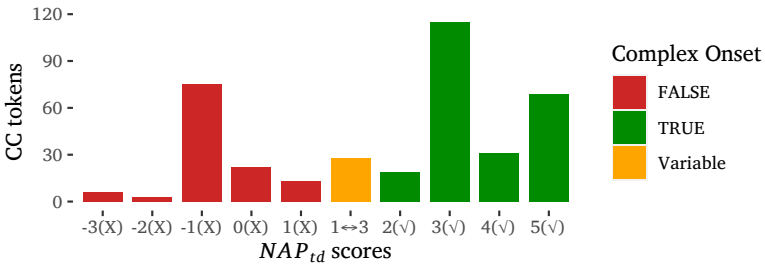

Figure C.4: Fit of CC tokens between the NAP_{td} model (x-axis) and the corpus data (color). *1↔3* can vary between scores *1* & *3* (*X↔√*) due to voicing assimilation

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

References

Abercrombie, David. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Adam, Galit. 2002. *From variable to optimal grammar: Evidence from language acquisition*. Tel-Aviv University. (Doctoral dissertation).

Albert, Aviad. 2014. *Phonotactic universals in Modern Hebrew: Evidence for prosodic alignment of stops*. Tel-Aviv University. (MA thesis).

Albert, Aviad. 2018. A tonal conspiracy: A perceptually-motivated acoustic model of prosodic prominence. In *The 2nd International Conference "Prominence in Language" (ICPL). Cologne, Germany [oral presentation]*.

Albert, Aviad. 2019. The state of stop–fricative alternation in Modern Hebrew. *Brill's Journal of Afroasiatic Languages and Linguistics* 11(1). 135–161. DOI: 10.1163/18776930-01101010.

Albert, Aviad. 2022. Sonority in Modern Hebrew. *Radical: A Journal of Phonology* 4. 531–594.

Albert, Aviad, Francesco Cangemi & Martine Grice. 2018. Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration. In *Proc. 9th International Conference on Speech Prosody*, 804–808.

Albert, Aviad, Francesco Cangemi & Martine Grice. 2019. Can you draw me a question? In *The 2nd Prosody Visualisation Challenge, ICPhS. Melbourne, Australia [poster presentation]*. DOI: 10.13140/RG.2.2.15700.14729.

Albert, Aviad, Maria Lialiou, Simona Sbranna & Francesco Cangemi. 2022. Improved acoustic characterization of prosodic prominence using periodic energy mass. In *The third International Conference Prominence in Language (ICPL), Cologne, Germany [poster presentation]*.

Albert, Aviad, Brian MacWhinney, Bracha Nir & Shuly Wintner. 2013. The Hebrew CHILDES corpus: Transcription and morphological analysis. *Language resources and evaluation* 47(4). 973–1005.

Albert, Aviad & Bruno Nicenboim. 2022. Modeling sonority in terms of pitch intelligibility with the Nucleus Attraction Principle. *Cognitive Science* 46(7). e13161. DOI: 10.1111/cogs.13161.

Albright, Adam. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26(01). 9. DOI: 10.1017/s0952675709001705.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Allen, William Sidney. 1973. *Accent and rhythm: Prosodic features of Latin and Greek: A study in theory and reconstruction*. Cambridge: Cambridge University Press.

Anderson, John. 1986. Suprasegmental dependencies. In Jacques Durand (ed.), *Dependency and non-linear phonology*, 55–133. London: Croom Helm.

Arvaniti, Amalia. 2009. Rhythm, timing and the timing of rhythm. *Phonetica* 66(1-2). 46–63. DOI: 10.1159/000208930.

Arvaniti, Amalia. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40(3). 351–373. DOI: 10.1016/j.wocn.2012.02.003.

Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen. 2006. Tonal association and tonal alignment: Evidence from Greek polar questions and contrastive statements. *Language and Speech* 49(4). 421–450.

Asherov, Daniel & Outi Bat-El. 2019. Syllable structure and complex onsets in Modern Hebrew. *Brill's Journal of Afroasiatic Languages and Linguistics* 11(1). 69–95. DOI: 10.1163/18776930-01101007.

Attneave, Fred & Richard K. Olson. 1971. Pitch as a medium: A new approach to psychophysical scaling. *The American journal of psychology* 84(2). 147–166.

Bailey, Todd M. & Ulrike Hahn. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44(4). 568–591.

Bao, Hua & Issa MS Panahi. 2010. Psychoacoustic active noise control with ITU-R 468 noise weighting and its sound quality analysis. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 4323–4326.

Barkai, Malachi. 1972. *Problems in the phonology of Israeli Hebrew*. Ann Arbor, Mich.: University Microfilms International.

Barkai, Malachi. 1975. On phonological representations, rules, and opacity. *Lingua* 37(4). 363–376.

Barnes, Jonathan, Alejna Brugos, Nanette Veilleux & Stefanie Shattuck-Hufnagel. 2011. Voiceless intervals and perceptual completion in F0 contours: Evidence from scaling perception in American English. In *Proc. 16th ICPhS, Hong Kong, China*, 108–111.

Barnes, Jonathan, Alejna Brugos, Nanette Veilleux & Stefanie Shattuck-Hufnagel. 2014. Segmental influences on the perception of pitch accent scaling in English. In *Proceedings of 7th Speech Prosody Conference*, 1125–1129.

Barnes, Jonathan, Nanette Veilleux, Alejna Brugos & Stefanie Shattuck-Hufnagel. 2012. Tonal center of gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology* 3(2). 337–383. DOI: 10.1515/lp-2012-0017.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Baroni, Antonio. 2014. On the importance of being noticed: The role of acoustic salience in phonotactics (and casual speech). *Language Sciences* 46. 18–36. DOI: 10.1016/j.langsci.2014.06.004.

Barry, William J. 1984. Place-of-articulation information in the closure voicing of plosives. *The Journal of the Acoustical Society of America* 76(4). 1245–1247. DOI: 10.1121/1.391375.

Bat-El, Outi. 1994. Stem modification and cluster transfer in Modern Hebrew. *Natural Language & Linguistic Theory* 12(4). 571–596.

Bat-El, Outi. 1996. Selecting the best of the worst: The grammar of Hebrew blends. *Phonology* 13(03). 283–328.

Bat-El, Outi. 2002. True truncation in colloquial Hebrew imperatives. *Language* 78(4). 651–683.

Bat-El, Outi. 2005. The emergence of the trochaic foot in Hebrew hypocoristics. *Phonology* 22(02). 115. DOI: 10.1017/s0952675705000515.

Bat-El, Outi. 2008. Morphologically conditioned V–Ø alternation in Hebrew. *Current issues in generative Hebrew linguistics* 134. 27–59.

Bat-El, Outi. 2012a. Prosodic alternations in Modern Hebrew segolates. In Malka Muchnik & Tsvi Sadan (eds.), *Studies in Modern Hebrew and Jewish Languages Presented to Ora (Rodrigue) Schwarzwald*, 116–129. Jerusalem: Carmel.

Bat-El, Outi. 2012b. The sonority dispersion principle in the acquisition of Hebrew word final codas. In Stephen G. Parker (ed.), *The sonority controversy*, 319–344. Berlin; Boston: De Gruyter Mouton.

Bates, Elizabeth A. & Jeffrey L. Elman. 2002. Connectionism and the study of change. In Mark H. Johnson, Yuko Munakata & Rick O. Gilmore (eds.), *Brain development and cognition: A reader*, 2nd. Oxford: Blackwell.

Bengtsson, Henrik. 2018. *R. Matlab: Read and Write MAT Files and Call MATLAB from Within R*. R package version 3.6.2. https://CRAN.R-project.org/package=R.matlab.

Benichov, Jonathan I., Eitan Globerson & Ofer Tchernichovski. 2016. Finding the beat: From socially coordinated vocalizations in songbirds to rhythmic entrainment in humans. *Frontiers in Human Neuroscience* 10(255). DOI: 10.3389/fnhum.2016.00255.

Berent, Iris. 2017. On the origins of phonology. *Current Directions in Psychological Science* 26(2). 132–139.

Berent, Iris, Evan Balaban, Tracy Lennertz & Vered Vaknin-Nusbaum. 2010. Phonological universals constrain the processing of nonspeech stimuli. *Journal of Experimental Psychology: General* 139(3). 418–435. DOI: 10 . 1037 / a0020094.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Berent, Iris, Anna-Katharine -. K. Brem, Xu Zhao, Erica Seligson, Hong Pan, Jane Epstein, Emily Stern, Albert M. Galaburda & Alvaro Pascual-Leone. 2015. Role of the motor system in language knowledge. *Proceedings of the National Academy of Sciences* 112(7). 1983–1988. DOI: 10.1073/pnas.1416851112.

Berent, Iris, Tracy Lennertz & Evan Balaban. 2012. Language universals and misidentification: A two way street. *Language and speech* 55(Pt 3). 311–330. DOI: 10.1177/0023830911417804.

Berent, Iris, Tracy Lennertz, Jongho Jun, Miguel A. Moreno & Paul Smolensky. 2008. Language universals in human brains. *Proceedings of the National Academy of Sciences* 105(14). 5321–5325.

Berent, Iris, Tracy Lennertz & Monica Rosselli. 2012. Universal linguistic pressures and their solutions: Evidence from Spanish. *The Mental Lexicon* 7(3). 275–305. DOI: 10.1075/ml.7.3.02ber.

Berent, Iris, Tracy Lennertz & Paul Smolensky. 2011. Syllable markedness and misperception: It's a two-way street. In Eric Raimy & Charles E. Cairns (eds.), *Handbook of the syllable*. Leiden; Boston: Brill.

Berent, Iris, Tracy Lennertz, Paul Smolensky & Vered Vaknin. 2009. Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology* 26(1). 75–108. DOI: 10.1017/S0952675709001729.

Berent, Iris, Hong Pan, Xu Zhao, Jane Epstein, Monica L. Bennett, Vibhas Deshpande, Ravi Teja Seethamraju & Emily Stern. 2014. Language universals engage Broca's area. *PLoS ONE* 9(4). 1–10. DOI: 10.1371/journal.pone.0095155.

Berent, Iris, Donca Steriade, Tracy Lennertz & Vered Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104(3). 591–630.

Berent, Iris, Vered Vaknin-Nusbaum, Evan Balaban & Albert M. Galaburda. 2013. Phonological generalizations in dyslexia: The phonological grammar may not be impaired. *Cogn Neuropsychol* 30(5). 285–310. DOI: 10.1080/02643294.2013.863182.

Bernhardt, Barbara Handford & Joseph P. Stemberger. 1997. *Handbook of phonological development: From the perspective of constraint-based nonlinear phonology*. New York: Academic Press.

Bidelman, Gavin M. & Ananthanarayan Krishnan. 2009. Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *The Journal of Neuroscience* 29(42). 13165–13171.

Blanc, Haim. 1957. Hebrew in Israel: Trends and problems. *Middle East Journal* 11(4). 397–409. https://www.jstor.org/stable/4322951.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Blevins, Juliette. 1995. The syllable in phonological theory. In John A. Goldsmith (ed.), *The handbook of phonological theory*, 206–244. Cambridge, MA & Oxford: Blackwell.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics/I-FOTT.

Boersma, Paul & David Weenink. 2019. *Praat: Doing phonetics by computer*. http://www.praat.org/.

Bolinger, Dwight L. J. H. 1978. Intonation across languages. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language*, vol. 2: Phonology, 471–524. Stanford, Cal.: Stanford University Press.

Bolozky, Shmuel. 1978. Some aspects of Modern Hebrew phonology. In Ruth Aronson Berman (ed.), *Modern Hebrew structure*, 11–67. University Publishing Projects Tel Aviv.

Bolozky, Shmuel. 2006. A note on initial consonant clusters in Israeli Hebrew. *Hebrew studies* 47(1). 227–235. DOI: 10.1353/hbr.2006.0017.

Bolozky, Shmuel. 2009. Colloquial Hebrew imperatives revisited. *Language Sciences* 31(2). 136–143.

Bolozky, Shmuel. 2013. bgdkpt consonants: Modern Hebrew. In Geoffrey Khan (ed.), *Encyclopedia of Hebrew language and linguistics*, 262–268. Leiden: Brill. DOI: 10.1163/2212-4241_ehll_EHLL_COM_00000763.

Bolozky, Shmuel. 2020. *Dictionary of Hebrew nouns: 14,000 Hebrew nouns and adjectives, classified into 998 patterns, with grammatical information*. Jerusalem: Rubin Mass Publishers.

Bolozky, Shmuel & Michael Becker. 2006. *Living lexicon of Hebrew nouns*. https://becker.phonologist.org/LLHN/.

Boril, Tomas. 2020. *rPraat: Interface to Praat*. R package version 1.3.1. https://CRAN.R-project.org/package=rPraat.

Brigham, E. Oran. 1988. *The fast Fourier transform and its applications*. Engelwood Cliffs: Prentice-Hall.

Broadbent, Donald E. & Peter Ladefoged. 1959. Auditory perception of temporal order. *The Journal of the Acoustical Society of America* 31(11). 1539. DOI: 10.1121/1.1907662.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory Phonology: An overview. *Phonetica* 49(3-4). 155–180.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). 1–28. DOI: 10.18637/jss.v080.i01.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* 10(1). 395–411.

Bürkner, Paul-Christian & Emmanuel Charpentier. 2018. *Modeling monotonic effects of ordinal predictors in Bayesian regression models*. preprint. PsyArXiv. DOI: 10.31234/osf.io/9qkhj. https://osf.io/9qkhj (8 October, 2019).

Buzsáki, György. 2006. *Rhythms of the brain*. New York: Oxford University Press.

Cairns, Charles E. & Mark H. Feinstein. 1982. Markedness and the theory of syllable structure. *Linguistic Inquiry* 13(2). 193–225.

Cangemi, Francesco, Aviad Albert & Martine Grice. 2019. Modelling intonation: Beyond segments and tonal targets. In *Proceedings of the international congress of phonetic sciences, Melbourne, Australia (ICPhS 2019)*, 572–576.

Cariani, Peter. 2001. Symbols and dynamics in the brain. *Biosystems* 60(1-3). 59–83. DOI: 10.1016/S0303-2647(01)00108-3.

Carr, Dan, ported by Nicholas Lewin-Koh, Martin Maechler & contains copies of lattice functions written by Deepayan Sarkar. 2018. *hexbin: Hexagonal Binning Routines*. R package version 1.27.2. https://CRAN.R-project.org/package=hexbin.

Case, Pamela, Betty Tuller, Mingzhou Ding & J. A. Scott Kelso. 1995. Evaluation of a dynamical model of speech perception. *Attention, Perception, & Psychophysics* 57(7). 977–988.

Chandrasekaran, Chandramouli, Andrea Trubanova, Sébastien Stillittano, Alice Caplier & Asif A. Ghazanfar. 2009. The natural statistics of audiovisual speech. *PLoS Computational Biology* 5(7). e1000436. DOI: 10.1371/journal.pcbi.1000436.

Cho, Young-mee Yu. 1990. A typology of voicing assimilation. In *Proceedings of the 9th West Coast Conference on Formal Linguistics*, vol. 9, 141–156.

Chomsky, Noam. 2021. Simplicity and the form of grammars. *Journal of Language Modelling* 9(1). 5–15. DOI: 10.15398/jlm.v9i1.257.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Chowning, John M. 2001. Perceptual fusion and auditory perspective. In Perry R. Cook (ed.), *Music, cognition, and computerized sound: An introduction to psychoacoustics*, 261–275. Cambridge, Mass.: MIT Press.

Christiansen, Morten H. & Suzanne L. Curtin. 1999. The power of statistical learning: No need for algebraic rules. In *Proceedings of the 21st annual conference of the cognitive science society*, vol. 114, 119.

Chung, Y, Andrew Gelman, S Rabe-Hesketh, J Liu & V Dorie. 2013. Weakly informative prior for point estimation of covariance matrices in hierarchical models. *Manuscript submitted for publication* 40(2). 136–157. DOI: 10.3102/1076998615570945.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Clements, George N. 1990. The role of the sonority cycle in core syllabification. *Papers in laboratory phonology* 1. 283–333.

Clements, George N. 1992. The sonority cycle and syllable organization. In Wolfgang U. Dressler (ed.), *Phonologica 1988: Proceedings of the 6th international phonology meeting*, 63–76. Cambridge: Cambridge University Press.

Clements, George N. 2009. Does sonority have a phonetic basis? In Eric Raimy & Charles E. Cairns (eds.), *Contemporary views on architecture and representations in phonology* (Current Studies in Linguistics), 165–175. Cambridge, Mass.; London, England: A Bradford Book The MIT Press.

Cohen, Evan-Gary -. G. 2009. *The role of similarity in phonology: Evidence from loanword adaptation in Hebrew*. Tel Aviv University. (Doctoral dissertation).

Cohen-Gross, Dalia. 2015. The syllable structure in the Modern Hebrew noun and adjective system. *Hebrew Studies* 56. 175–190. DOI: 10.1353/hbr.2015.0000.

Coleman, John & Janet Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. In *Eprint arxiv: Cmp-lg/9707017*.

Cummins, Fred. 2009. Rhythm as an affordance for the entrainment of movement. *Phonetica* 66(1-2). 15–28. DOI: 10.1159/000208928.

Cummins, Fred. 2012. Oscillators and syllables: A cautionary note. *Frontiers in psychology* 3. 364. DOI: 10.3389/fpsyg.2012.00364.

Cummins, Fred. 2015. Rhythm and speech. In Melissa A. Redford (ed.), *The handbook of speech production*. Malden, Ma: Wiley Blackwell.

Cummins, Robert. 1985. *The nature of psychological explanation*. Cambridge, Mass.: MIT Press.

Cutler, Anne, Delphine Dahan & Wilma Van Donselaar. 1997. Prosody in the comprehension of spoken language: A literature review. *Language and speech* 40(2). 141–201.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis & Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* 28(02). 197–234. DOI: 10.1017/s0952675711000145.

Dauer, Rebecca M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of phonetics* 11(1). 51–62.

Davidson, Lisa. 2010. Phonetic bases of similarities in cross-language production: Evidence from English and Catalan. *Journal of Phonetics* 38(2). 272–288.

Davidson, Lisa & Jason A. Shaw. 2012. Sources of illusion in consonant cluster perception. *Journal of Phonetics* 40(2). 234–248. DOI: 10.1016/j.wocn.2011.11.005.

De Cheveigné, Alain. 2005. Pitch perception models. In Christopher J. Plack, Andrew J. Oxenham & Richard R. Fay (eds.), *Pitch: Neural coding and perception*, 169–233. New York: Springer.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

de Brosses, Charles. 1765. *Traité de la formation méchanique des langues, et des principes physiques de l'étymologie*. Paris: Chez Saillant, Vincent, Desaint.

de Lacy, Paul Valiant. 2002. *The formal expression of markedness*. University of Massachusetts Amherst. (Doctoral dissertation).

Dellwo, Volker. 2006. Rhythm and speech rate: A variation coefficient for deltaC. In Paweł Karnowski & Imre Szigeti (eds.), *Language and language-processing: Proceedings of the 38th Linguistics Colloquium, Piliscsaba*. Frankfurt/Main: Peter Lang.

Demuth, Katherine. 1996. The prosodic structure of early words. In Katherine Demuth & James L. Morgan (eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 171–184. Mahwah, N.J.: Lawrence Erlbaum Associates.

Deshmukh, Om & Carol Espy-Wilson. 2003. Detection of periodicity and aperiodicity in speech signal based on temporal information. In *The 15th International Congress of Phonetic Sciences, Barcelona, Spain*.

Deshmukh, Om, Carol Y. Espy-Wilson, Ariel Salomon & Jawahar Singh. 2005. Use of temporal information: Detection of periodicity, aperiodicity, and pitch in speech. *IEEE Transactions on Speech and Audio Processing* 13(5). 776–786.

Ding, Nai, Monita Chatterjee & Jonathan Z. Simon. 2014. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroimage* 88. 41–46. DOI: 10.1016/j.neuroimage.2013.10.054.

Ding, Nai, Aniruddh D. Patel, Lin Chen, Henry Butler, Cheng Luo & David Poeppel. 2017. Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews* 81. 181–187. DOI: 10.1016/j.neubiorev.2017.02.011.

Donegan, Patricia J. 1978. *On the natural phonology of vowels*. Ohio State University. (Doctoral dissertation).

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25(6). 1568–1578. DOI: 10.1037/0096-1523.25.6.1568.

Easterday, Shelece. 2019. *Highly complex syllable structure: A typological and diachronic study* (Studies in Laboratory Phonology). Berlin: Language Science Press. http://langsci-press.org/catalog/book/249.

Eddelbuettel, Dirk & James Joseph Balamuta. 2017. Extending extitR with extitC++: A Brief Introduction to extitRcpp. *PeerJ Preprints* 5. e3188v1. DOI: 10.7287/peerj.preprints.3188v1.

Eddelbuettel, Dirk & Romain François. 2011. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software* 40(8). 1–18. DOI: 10.18637/jss.v040.i08. http://www.jstatsoft.org/v40/i08/.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Efron, Robert. 1973. Conservation of temporal information by perceptual systems. *Perception & Psychophysics* 14(3). 518–530. DOI: 10.3758/BF03211193.

Even-Shoshan, Avraham. 2003. *Milon Even-Shoshan (Even-Shoshan dictionary)*. Jerusalem: ha-Milon he-Hadash b.'a.m. (The New Dictionary inc.) [Hebrew].

Fant, Gunnar. 1970. *Acoustic theory of speech production: With calculations based on X-ray studies of Russian articulations*. The Hague: Mouton.

Fant, Gunnar, Anita Kruckenberg & Johan Liljencrants. 2000. The source-filter frame of prominence. *Phonetica* 57(2-4). 113–127.

Farbood, Morwaread Mary, Gary Marcus & David Poeppel. 2013. Temporal dynamics and the identification of musical key. *Journal of Experimental Psychology: Human Perception and Performance* 39(4). 911. DOI: 10.1037/a0031087.

Faust, Noam. 2014. Where it's [at]: A phonological effect of phasal boundaries in the construct state of Modern Hebrew. *Lingua* 150. 315–331. DOI: 10.1016/j.lingua.2014.08.001.

Faust, Noam. 2015. A novel, combined approach to Semitic word-formation. *Journal of Semitic studies* 60(2). 287–316. DOI: 10.1093/jss/fgv001.

Faust, Noam. 2019. Gutturals in general Israeli Hebrew. *Brill's Journal of Afroasiatic Languages and Linguistics* 11(1). 162–181.

Fellman, Jack. 1973. Concerning the "revival" of the Hebrew language. *Anthropological Linguistics* 15(5). 250–257. https://www.jstor.org/stable/30029347.

Flanagan, James L. & Newman Guttman. 1960. On the pitch of periodic pulses. *The Journal of the Acoustical Society of America* 32(10). 1308–1319. DOI: 10.1121/1.1907900.

Fletcher, Harvey & Wilden A. Munson. 1933. Loudness, its definition, measurement and calculation. *Bell System Technical Journal* 12(4). 377–430.

Fodor, Jerry A. 1983. *The modularity of mind*. Cambridge, Massachusetts: The MIT Press.

Fodor, Jerry A. & Zenon W. Pylyshyn. 1988. Connectionism and cognitive architecture: A critical analysis. *Cognition* 28(1-2). 3–71.

Foley, James. 1972. Rule precursors and phonological change by meta-rule. In Robert P. Stockwell & Ronald K. S. Macaulay (eds.), *Linguistic change and generative theory*, 96–100. Bloomington: Indiana University Press.

Fourier, Jean-Baptiste-Joseph. 1822. *Théorie analytique de la chaleur*. Paris: F. Didot.

Fowler, Carol A. 1980. Coarticulation and theories of extrinsic timing. *Journal of Phonetics* 8(1). 113–133. DOI: 10.1016/S0095-4470(19)31446-9.

Fowler, Carol A. 1986. An event approach to the study of speech perception from a direct–realist perspective. *Journal of phonetics* 14(1). 3–28. DOI: 10.1016/S0095-4470(19)30607-2.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Fraisse, Paul. 1984. Perception and estimation of time. *Annual review of psychology* 35(1). 1–37.

Frisch, Stefan A. & Bushra Adnan Zawaydeh. 2001. The psychological reality of OCP-Place in Arabic. *Language* 77(1). 91–106.

Fudge, Erik C. 1969. Syllables. *Journal of linguistics* 5(02). 253–286.

Fujimura, Osamu. 1975. Syllable as a unit of speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 23(1). 82–87.

Fujimura, Osamu & Donna Erickson. 1999. Acoustic phonetics. In William J. Hardcastle & John Laver (eds.), *The handbook of phonetic sciences*, 65–115. Oxford, UK; Malden, Mass.: Blackwell Publishers.

Fujimura, Osamu, Marian J. Macchi & Lynn A. Streeter. 1978. Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech* 21(4). 337–346. DOI: 10.1177/002383097802100408.

Fullwood, Michelle Alison. 2014. The perceptual dimensions of sonority-driven epenthesis. In *Proceedings of the annual meetings on phonology*, vol. 1.

Futrell, Richard, Adam Albright, Peter Graff & Timothy J. O'Donnell. 2017. A generative model of phonotactics. *Transactions of the Association for Computational Linguistics* 5. 73–86.

Fyk, Janina. 1987. Duration of tones required for satisfactory precision of pitch matching. *Bulletin of the Council for Research in Music Education* 91. 38–44.

Gafos, Adamantios I. 2006. Dynamics in grammar: Comment on ladd and ernestus & baayen. *Laboratory phonology* 8(4). 51–79.

Gafos, Adamantios I. & Stefan Benus. 2006. Dynamics of phonological cognition. *Cognitive science* 30(5). 905–943.

Gafos, Adamantios I., Simon Charlow, Jason Shaw & Philip Hoole. 2014. Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics* 44(1). 152–166. DOI: 10.1016/j.wocn.2013.11.007.

Gafter, Roey J. 2019. Modern Hebrew sociophonetics. *Brill's Journal of Afroasiatic Languages and Linguistics* 11(1). 226–242.

Galves, Antonio, Jesus Garcia, Denise Duarte & Charlotte Galves. 2002. Sonority as a basis for rhythmic class discrimination. In *Speech Prosody 2002, International Conference*.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, Donald B. Rubin, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari & Donald B. Rubin. 2013. *Bayesian Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC. DOI: 10.1201/b16018. https://www.taylorfrancis.com/books/9780429113079 (27 August, 2019).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Gelman, Andrew, Aleks Jakulin, Maria Grazia Pittau & Yu-Sung Su. 2008. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* 2(4). 1360–1383. DOI: 10.1214/08-AOAS191.

Gelman, Andrew, D. Simpson & M. Betancourt. 2017. The prior can often only Be understood in the context of the likelihood. *Entropy* 19. 555. DOI: 10.3390/e19100555.

Ghitza, Oded. 2013. The theta-syllable: A unit of speech information defined by cortical function. *Frontiers in psychology* 4. 138.

Ghitza, Oded. 2017. Acoustic-driven delta rhythms as prosodic markers. *Language, Cognition and Neuroscience* 32(5). 545–561. DOI: 10.1080/23273798.2016.1232419.

Goad, Heather. 2011. The representation of sc clusters. In M. van Oostendorp, C. Ewen, E. Hume & K. Rice (eds.), *The Blackwell companion to phonology*, vol. 2, 898–923. Oxford: Wiley-Blackwell.

Goad, Heather. 2016. Sonority and the unusual behaviour of /s/. In Martin J. Ball & Nicole Müller (eds.), *Challenging sonority: Cross-linguistic evidence*, 21–44. Sheffield; Bristol: Equinox Publishing Ltd.

Goedemans, Rob & Harry van der Hulst. 2013. Weight-sensitive stress. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/15.

Goldsmith, John. 1992. Local modeling in phonology. In Steven Davis (ed.), *Connectionism: Theory and practice*, 229–246. Oxford; New York: Oxford University Press.

Goldsmith, John A. 1976. *Autosegmental phonology*. Massachusetts Institute of Technology. (Doctoral dissertation).

Goldstein, Louis, Ioana Chitoran & Elisabeth Selkirk. 2007. Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. In *Proceedings of the XVIth International Congress of Phonetic Sciences*, 241–244.

Goldstein, Louis, Hosung Nam, Elliot Saltzman & Ioana Chitoran. 2009. Coupled oscillator planning model of speech timing and syllable structure. In C. Fant, M. Gunnar, Hiroya Fujisaki & Jiaxuan Shen (eds.), *Frontiers in phonetics and speech science*, 239–250. Beijing: The Commercial Press.

Gómez, David Maximiliano, Iris Berent, Silvia Benavides-Varela, Ricardo A. H. Bion, Luigi Cattarossi, Marina Nespor & Jacques Mehler. 2014. Language universals at birth. *Proceedings of the National Academy of Sciences* 111(16). 5837–5841. DOI: 10.1073/pnas.1318261111.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Gordon, Matthew, Edita Ghushchyan, Bradley McDonnell, Daisy Rosenblum & Patricia A. Shaw. 2012. Sonority and central vowels: A cross-linguistic phonetic study. In Stephen G. Parker (ed.), *The sonority controversy*, 219–254. Berlin; Boston: De Gruyter Mouton.

Gordon, Matthew Kelly. 2006. *Syllable weight: Phonetics, phonology, typology* (Studies in Linguistics). New York: Routledge.

Goswami, Usha & Victoria Leong. 2013. Speech rhythm and temporal structure: Converging perspectives. *Laboratory Phonology* 4(1). 67–92.

Gouskova, Maria & Michael Becker. 2013. Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language & Linguistic Theory* 31(3). 735–765.

Grabe, Esther & Ee Ling Low. 2002. Durational variability in speech and the rhythm class hypothesis. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory phonology*, vol. 7, 515–546. DOI: 10.1515/9783110197105.

Greenberg, Joseph H. 1978. Some generalizations concerning initial and final consonant sequences. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language*, vol. 2: Phonology, 243–279. Stanford, Cal.: Stanford University Press.

Greenberg, Steven, Hannah Carvey, Leah Hitchcock & Shuangyu Chang. 2003. Temporal properties of spontaneous speech: A syllable-centric perspective. *Journal of Phonetics* 31(3-4). 465–485.

Gross, Joachim, Nienke Hoogenboom, Gregor Thut, Philippe Schyns, Stefano Panzeri, Pascal Belin & Simon Garrod. 2013. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology* 11(12). e1001752. DOI: 10.1371/journal.pbio.1001752.

Guttman, Newman & Bela Julesz. 1963. Lower limits of auditory periodicity analysis. *The Journal of the Acoustical Society of America* 35(4). 610. DOI: 10.1121/1.1918551.

Haegens, Saskia & Elana Zion Golumbic. 2018. Rhythmic facilitation of sensory processing: A critical review. *Neuroscience & Biobehavioral Reviews* 86. 150–165. DOI: 10.1016/j.neubiorev.2017.12.002.

Haken, Hermann. 1990. Synergetics as a tool for the conceptualization and mathematization of cognition and behaviour: How far can we go? In Hermann Haken & Michael Stadler (eds.), *Synergetics of Cognition: Proceedings of the International Symposium at Schloß Elmau, Bavaria, June 4–8, 1989*. Berlin, Heidelberg: Springer.

Haken, Hermann, J. A. Scott Kelso & Heinz Bunz. 1985. A theoretical model of phase transitions in human hand movements. *Biological Cybernetics* 51(5). 347–356.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Halle, Morris. 1954. The strategy of phonemics. *Word* 10(2-3). 197–209.
Hankamer, Jorge & Judith Aissen. 1974. The sonority hierarchy. In Anthony Bruck, Robert A. Fox & Michael W. L. La Galy (eds.), *Papers from the parasession on natural phonology*, 131–145. Chicago: Chicago Linguistic Society.
Harnad, Stevan. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena* 42(1-3). 335–346. DOI: 10.1016/0167-2789(90)90087-6.
Harris, James W. 1983. *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, Mass.: MIT Press.
Harris, John. 2007. Representation. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*, 119–137. Cambridge: Cambridge University Press.
Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic inquiry* 20(2). 253–306.
Hayes, Bruce. 2011. Interpreting sonority-projection experiments: The role of phonotactic modeling. In *Proceedings of the 17th international congress of phonetic sciences*, 835–838.
Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry* 39(3). 379–440.
Hayes, Bruce Philip. 1980. *A metrical theory of stress rules*. Massachusetts Institute of Technology. (Doctoral dissertation).
Heffner, Roe-Merrill. 1969. *General phonetics*. Madison, Wi: University of Wisconsin Press.
Hellman, Rhona P. 1972. Asymmetry of masking between noise and tone. *Perception & Psychophysics* 11(3). 241–246.
Helmholtz, Hermann von. 1863. *Die Lehre von den Tonempfindungen als physiologische Grundlage fur die Theorie der Musik*. Braunschweig: F. Vieweg und Sohn.
Henke, Eric K., Ellen M. Kaisse & Richard Wright. 2012. Is the sonority sequencing principle an epiphenomenon? In Stephen G. Parker (ed.), *The sonority controversy*, 65–100. Berlin; Boston: De Gruyter Mouton.
Henry, Lionel & Hadley Wickham. 2020. *purrr: Functional Programming Tools*. R package version 0.3.4. https://CRAN.R-project.org/package=purrr.
Hermes, Anne, Doris Mücke & Bastian Auris. 2017. The variability of syllable patterns in Tashlhiyt Berber and Polish. *Journal of Phonetics* 64. 127–144. DOI: 10.1016/j.wocn.2017.05.004.
Hermes, Anne, Doris Mücke & Martine Grice. 2013. Gestural coordination of Italian word-initial clusters: The case of 'impure s'. *Phonology* 30(01). 1–25. DOI: 10.1017/s095267571300002x.
Heselwood, Barry. 1998. An unusual kind of sonority and its implications for phonetic theory. *Working papers in linguistics and phonetics* 6. 68–80.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Hirsh, Ira J. 1959. Auditory perception of temporal order. *The Journal of the Acoustical Society of America* 31(6). 759–767. DOI: 10.1121/1.1907782.

Hock, Howard S., Gregor Schöner & Martin Giese. 2003. The dynamical foundations of motion pattern formation: Stability, selective adaptation, and perceptual continuity. *Perception & Psychophysics* 65(3). 429–457. DOI: 10.3758/BF03194574.

Hooper, Joan B. 1976. *An introduction to natural generative phonology*. New York: Academic Press.

House, David. 1990. *Tonal perception in speech*. Lund: Lund university press.

Houtsma, Adrianus J. M. 1995. Pitch perception. In Brian C. J. Moore (ed.), *Hearing*, 2nd edn. (Handbook of Perception and Cognition), 267–295. San Diego: Academic Press.

Howe, Darin & Douglas George Pulleyblank. 2004. Harmonic scales as faithfulness. *The Canadian Journal of Linguistics/La revue canadienne de linguistique* 49(1). 1–49.

Hyman, Larry M. 1984. *A theory of phonological weight*. Dordrecht: Foris Publications.

Inbar, Maya, Eitan Grossman & Ayelet N. Landau. 2020. Sequences of Intonation Units form a ~1 Hz rhythm. *Scientific reports* 10(1). 1–9. DOI: 10.1038/s41598-020-72739-4.

Itô, Junko. 1989. A prosodic theory of epenthesis. *Natural Language & Linguistic Theory* 7(2). 217–259.

ITU-R. 2015. *Algorithms to measure audio programme loudness and true-peak audio level: ITU-R BS.1770-4*. Tech. rep. Radiocommunication Sector of International Telecommunication Union.

Jakobson, Roman & Morris Halle. 1956. *Fundamentals of language*. The Hague: Mouton.

Jany, Carmen, Matthew Gordon, Carlos M. Nash & Nobutaka Takara. 2007. How universal is the sonority hierarchy?: A cross-linguistic acoustic study. In *16th international congress of phonetic sciences, saarbrücken*, 1401–1404.

Jarosz, Gaja, Shira Calamaro & Jason Zentz. 2017. Input frequency and the acquisition of syllable structure in Polish. *Language acquisition* 24(4). 361–399.

Jaynes, E. T. & Oscar Kempthorne. 1976. Confidence intervals vs. Bayesian intervals. In William Leonard Harper & Clifford Alan Hooker (eds.), *Foundations of probability theory, statistical inference, and statistical theories of science*, vol. 6b (The University of Western Ontario Series in Philosophy of Science), 175–257. Dordrecht: Springer Netherlands. DOI: 10.1007/978-94-010-1436-6_6.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Jeon, Hae-Sung & Stephen Nichols. 2022. Investigating prosodic variation in British English varieties using ProPer. In *Proccedings of Interspeech 2022, Incheon, Korea.*

Jespersen, Otto. 1899. *Fonetik: En systematisk Fremstilling af Læren om Sproglyd.* København: Det Schøbergse Forlag.

Joanisse, Marc Francis. 2000. *Connectionist phonology.* University of Southern California. (Doctoral dissertation).

Josephs, Jesse J. 1967. *The physics of musical sound.* Princeton, N.J.: Published for the Commission on College Physics [by] D. Van Nostrand Co.

Jun, Sun-Ah (ed.). 2005. *Prosodic typology: The phonology of intonation and phrasing.* Oxford: Oxford University Press.

Jun, Sun-Ah (ed.). 2015. *Prosodic typology II: The phonology of intonation and phrasing.* Oxford: Oxford University Press.

Jurafsky, Daniel S. & James H. Martin. 2009. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition.* 2nd. Upper Saddle River, N.J.; London: Pearson Education, Inc.

Kaye, Jonathan. 1992. *Do you believe in magic? The story of s+c sequences.* Tech. rep. SOAS working papers in Linguistics & Phonetics.

Keitel, Anne, Joachim Gross & Christoph Kayser. 2018. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biology* 16(3). e2004473.

Keitel, Anne, Robin AA Ince, Joachim Gross & Christoph Kayser. 2017. Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *Neuroimage* 147. 32–42. DOI: 10.1016/j.neuroimage.2016.11.062.

Kelso, J. A. Scott. 1997. *Dynamic patterns: The self-organization of brain and behavior.* Cambridge, MA: MIT Press.

Kenstowicz, Michael. 1997. Quality-sensitive stress. *Rivista di linguistica* 9. 157–188.

Kenstowicz, Michael J. 1994. *Phonology in generative grammar.* Cambridge, MA: Blackwell.

Kiparsky, Paul. 1979. Metrical structure assignment is cyclic. *Linguistic inquiry* 10(3). 421–441.

Klatt, Dennis H. 1979. Speech perception: A model of acoustic–phonetic analysis and lexical access. *Journal of phonetics* 7(3). 279–312. DOI: 10.1016/S0095-4470(19)31059-9.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Köppl, Christine. 1997. Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, Tyto alba. *Journal of Neuroscience* 17(9). 3312–3321. DOI: 10.1523/JNEUROSCI.17-09-03312.1997.

Kotz, Sonja A., Andrea Ravignani & William T. Fitch. 2018. The evolution of rhythm processing. *Trends in cognitive sciences* 22(10). 896–910. DOI: 10.1016/j.tics.2018.08.002.

Kreitman, Rina. 2008. *The phonetics and phonology of onset clusters: The case of Modern Hebrew*. Cornell University. (Doctoral dissertation).

Kreitman, Rina. 2010. Mixed voicing word-initial onset clusters. *Laboratory Phonology 10* 4(4). 169.

Krishnan, Ananthanarayan, Yisheng Xu, Jackson Gandour & Peter Cariani. 2005. Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research* 25(1). 161–168.

Ladd, Bob. 2011. Phonetics in phonology. In John A. Goldsmith, Jason Riggle & Alan C. L. Yu (eds.), *The handbook of phonological theory*, 2nd edn., 348–373. Malden, MA: Wiley-Blackwell.

Ladd, D. Robert. 2008. *Intonational phonology*. Cambridge: Cambridge University Press.

Ladefoged, Peter. 1971. *Preliminaries to linguistic phonetics*. Chicago; London: University of Chicago Press.

Ladefoged, Peter. 1975. *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Ladefoged, Peter. 1997. Linguistic phonetic descriptions. In William J. Hardcastle & John Laver (eds.), *The handbook of phonetic sciences*, 589–618. Oxford; Cambridge, Mass.: Blackwell.

Ladefoged, Peter & Ian Maddieson. 1996. Rhotics. In *The sounds of the world's languages*. Oxford: Blackwell.

Laks, Bernard. 1995. A connectionist account of French syllabification. *Lingua* 95(1). 51–76.

Laks, Lior, Evan-Gary Cohen & Stav Azulay-Amar. 2016. Paradigm uniformity and the locus of derivation: The case of vowel epenthesis in Hebrew verbs. *Lingua* 170. 1–22. DOI: 10.1016/j.lingua.2015.10.004.

Lancia, Leonardo & Bodo Winter. 2013. The interaction between competition, learning, and habituation dynamics in speech perception. *Laboratory Phonology* 4(1). 221–257.

Lass, Roger. 1988. *Phonology: An introduction to basic concepts* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Laufer, Asher. 1991. Phoneme combinations: Phonotactics. In Moshe Goshen-Gottstein, Shelomo Morag & Simha Kogut (eds.), *Shai le-Hayim Rabin: Asupat mehkere lashon li-khevodo bi-melot lo shiv'im ve-hamesh*, 179–193. Jerusalem: Akademon.

Laufer, Asher. 2019. The origin of the IPA schwa. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, 1908–1911.

Leben, William Ronald. 1973. *Suprasegmental phonology*. Massachusetts Institute of Technology. (Doctoral dissertation).

Lehiste, Ilse. 1990. Phonetic investigation of metrical structure orally produced poetry. *Journal of Phonetics* 18(2). 123–133. DOI: 10.1016/S0095-4470(19)30397-3.

Lennertz, Tracy & Iris Berent. 2015. On the sonority levels of fricatives and stops. *The Mental Lexicon* 10(1). 88–132. DOI: 10.1075/ml.10.1.04len.

Lennertz, Tracy Jordan. 2010. *People's knowledge of phonological universals: Evidence from fricatives and stops*. Northeastern University. (Doctoral dissertation).

Levitt, Andrea, Alice F. Healy & David W. Fendrich. 1991. Syllable-internal structure and the sonority hierarchy: Differential evidence from lexical decision, naming, and reading. *Journal of Psycholinguistic Research* 20(4). 337–363.

Lialiou, Maria, Aviad Albert, Alexandra Vella & Martine Grice. 2021. Periodic energy mass on head and edge tones in Maltese wh-constructions. In *The 1st International Conference on Tone and Intonation (TAI). Sonderborg, Denmark [poster presentation]*.

Liberman, Alvin M. & Ignatius G. Mattingly. 1985. The motor theory of speech perception revised. *Cognition* 21(1). 1–36. DOI: 10.1016/0010-0277(85)90021-6.

Liberman, Mark Yoffe. 1975. *The intonational system of English*. Massachusetts Institute of Technology. (Doctoral dissertation).

Ligges, Uwe, Sebastian Krey, Olaf Mersmann & Sarah Schnackenberg. 2018. *tuneR: Analysis of Music and Speech*. https://CRAN.R-project.org/package=tuneR.

Lindau, Mona. 1985. The story of /r/. In Victoria Fromkin (ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 157–168. Orlando: Academic Press.

Lindblom, Björn. 1983. Economy of speech gestures. In Peter F. Macneilage (ed.), *The production of speech*, 217–245. New York; Heidelberg: Springer-Verlag.

Lombardi, Linda. 1991. *Laryngeal features and laryngeal neutralization*. University of Massachusetts Amherst. (Doctoral dissertation).

Lombardi, Linda. 1995. Laryngeal neutralization and syllable wellformedness. *Natural Language & Linguistic Theory* 13(1). 39–74.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Lowenstamm, Jean. 1981. On the maximal cluster approach to syllable structure. *Linguistic Inquiry* 12(4). 575–604.

Lowit, Anja. 2014. Quantification of rhythm problems in disordered speech: A re-evaluation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1658). 20130404. DOI: 10.1098/rstb.2013.0404.

Lund, Thomas & Esben Skovenborg. 2014. Loudness vs. speech normalization in film and drama for broadcast. In *Annual Technical Conference & Exhibition, SMPTE 2014*, 1–14.

Luo, Huan, Zuxiang Liu & David Poeppel. 2010. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology* 8(8). e1000445.

MacDougall, R. 1903. The structure of simple rhythm forms. *The Psychological Review: Monograph Supplements* 4(1). 309–412.

Mai, Guangting, James W. Minett & William S-Y Wang. 2016. Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *Neuroimage* 133. 516–528. DOI: 10.1016/j.neuroimage.2016.02.064.

Maïonchi-Pino, Norbert, Yasuyuki Taki, Annie Magnan, Satoru Yokoyama, Jean Écalle, Kei Takahashi, Hiroshi Hashizume & Ryuta Kawashima. 2015. Sonority-related markedness drives the misperception of unattested onset clusters in French listeners. *L'Année psychologique* 115(2). 197–222.

Mathôt, Sebastiaan, Daniel Schreij & Jan Theeuwes. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods* 44(2). 314–324.

Mayer, Connor & Max Nelson. 2019. Phonotactic learning with neural language models. In *Proceedings of the Society for Computation in Linguistics*.

McCarthy, John J. & Alan S. Prince. 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language & Linguistic Theory* 8(2). 209–283.

McCarthy, John Joseph. 1979. *Formal problems in Semitic phonology and morphology*. Massachusetts Institute of Technology. (Doctoral dissertation).

McPherson, Malinda J. & Josh H. McDermott. 2018. Diversity in pitch perception revealed by task dependence. *Nature Human Behaviour* 2(1). 52–66. DOI: 10.1038/s41562-017-0261-8.

Menzerath, Paul & Armando de Lacerda. 1933. *Koartikulation, Steuerung und Lautabgrenzung eine experimentelle Untersuchung*. Berlin; Bonn: F. Dümmlers.

Meyer, Lars, Molly J. Henry, Phoebe Gaston, Noura Schmuck & Angela D. Friederici. 2017. Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cerebral Cortex* 27(9). 4293–4302. DOI: 10.1093/cercor/bhw228.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Meyer, Lars, Yue Sun & Andrea E. Martin. 2020. Synchronous, but not entrained: Exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience* 35(9). 1089–1099. DOI: 10 . 1080 / 23273798.2019.1693050.

Miller, Brett L. K. 2012. Sonority and the larynx. In Stephen G. Parker (ed.), *The sonority controversy*, 257–288. Berlin; Boston: De Gruyter Mouton.

Miller, George A. & Walter G. Taylor. 1948. The perception of repeated bursts of noise. *The Journal of the Acoustical Society of America* 20(2). 171–182.

Mirea, Nicole & Klinton Bicknell. 2019. Using LSTMs to assess the obligatoriness of phonological distinctive features for phonotactic learning. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, 1595–1605.

Mizrachi, Avi. 2019. A note on Modern Hebrew voicing assimilation. *Brill's Journal of Afroasiatic Languages and Linguistics* 11(1). 49–56.

Mohanan, Karuvannur P. 1986. *The theory of lexical phonology*. Dordrecht u.a.: Reidel.

Moore, Brian C. J. 2013. *An introduction to the psychology of hearing*. Leiden, Boston: Brill.

Morag, Shelmo. 1959. Planned and unplanned development in Modern Hebrew. *Lingua* 8. 247–263. DOI: 10.1016/0024-3841(59)90025-7.

Morelli, Frida. 2003. The relative harmony of /s+stop/ onsets: Obstruent clusters and the sonority sequencing. In Caroline Féry & Ruben van de Vijver (eds.), *The syllable in optimality theory*, 356–371. Cambridge: Cambridge University Press.

Morey, Richard D, Rink Hoekstra, Jeffrey N Rouder, Michael D Lee & Eric-Jan Wagenmakers. 2016. The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review* 23. 103–123. DOI: 10.3758/s13423-015-0947-8.

Myers, Brett R., Miriam D. Lense & Reyna L. Gordon. 2019. Pushing the envelope: Developments in neural entrainment to speech and the biological underpinnings of prosody perception. *Brain sciences* 9(3). 70.

Myhill, John. 2004. A parameterized view of the concept of 'correctness'. *Multilingua* 23(4). 389–416.

Nakajima, Yoshitaka, Kazuo Ueda, Shota Fujimaru, Hirotoshi Motomura & Yuki Ohsaka. 2017. English phonology and an acoustic language universal. *Scientific Reports* 7. 46049. DOI: 10.1038/srep46049.

Nam, Hosung, Louis Goldstein & Elliot Saltzman. 2009. Self-organization of syllable structure: A coupled oscillator model. In François Pellegrino, Egidio Marsico & Ioana Chitoran und Christophe Coupé (eds.), *Approaches to phonological complexity*, 297–328. Berlin: Mouton de Gruyter.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Nathan, Geoffrey S. 1989. Preliminaries to a theory of phonological substance: The substance of sonority. In Roberta L. Corrigan, Fred R. Eckman & Michael Noonan (eds.), *Linguistic categorization*, 55–67. Amsterdam; Philadelphia: John Benjamins.

Nicenboim, Bruno & Shravan Vasishth. 2016. Statistical methods for linguistic research: Foundational Ideas - Part II. *Language and Linguistics Compass* 10(11). 591–613. DOI: 10.1111/lnc3.12207.

Nolan, Francis & Hae-Sung Jeon. 2014. Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1658). 20130396.

Ohala, John J. 1992. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In (Papers from the Parasession on the Syllable), 319–338. Chicago: Chicago Linguistic Society.

Ohala, John J. & Haruko Kawasaki. 1984. Prosodic phonology and phonetics. *Phonology Yearbook* 1. 113–127.

Ohala, John J. & Haruko Kawasaki-Fukumori. 1997. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In Stig Eliasson & Ernst Håkon Jahr (eds.), *Language and its Ecology: Essays in memory of Einar Haugen*, 343–365. Berlin; New York: Walter de Gruyter.

Ohm, Georg Simon. 1843. Ueber die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. *Annalen der Physik* 135(8). 513–565. DOI: 10.1002/andp.18431350802.

Olender, Adam. 2013. Acoustic evidence for word-initial /s/+stop sequences as onset clusters: "Perceptual bond" as a cross-linguistic predictor of prothesis. In *44th Poznań Linguistic Meeting*.

Olsen, Kirk N., Catherine J. Stevens & Julien Tardieu. 2010. Loudness change in response to dynamic acoustic intensity. *Journal of Experimental Psychology: Human Perception and Performance* 36(6). 1631.

Oxenham, Andrew J. 2013. Revisiting place and temporal theories of pitch. *Acoustical science and technology* 34(6). 388–396. DOI: 10.1250/ast.34.388.

Parker, Stephen G. 2002. *Quantifying the sonority hierarchy*. University of Massachusetts, Amherst. (Doctoral dissertation).

Parker, Stephen G. 2008. Sound level protrusions as physical correlates of sonority. *Journal of phonetics* 36(1). 55–90. DOI: 10.1016/j.wocn.2007.09.003.

Parker, Stephen G. (ed.). 2012. *The sonority controversy*. Berlin; Boston: De Gruyter Mouton.

Parker, Stephen G. 2017. Sounding out sonority. *Language and Linguistics Compass* 11(9). e12248. DOI: 10.1111/lnc3.12248.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Parker, Stephen G. 2018. *A bibliography of resources on sonority*. https://web.archive.org/web/20220124004246/https://diu.edu/wp-content/uploads/steve_parker/bibliography-sonority-alphabetical.pdf.

Patha, Sreedhar, Yegnanarayana Bayya & Suryakanth V. Gangashetty. 2016. Syllable nucleus and boundary detection in noisy conditions. In *Proceedings of Speech Prosody*, vol. 8, 2016–74.

Pattee, Howard H. 1987. Instabilitites and information in biological self-organization. In F. Eugene Yates (ed.), *Self-organizing systems: The emergence of order*, 325–338. New York; London: Plenum Press.

Pattee, Howard H. & Joanna Raczaszek-Leonardi. 2012. *Laws, language and life: Howard Pattee's classic papers on the physics of symbols with contemporary commentary* (Biosemiotics). Dordrecht: Springer.

Pfitzinger, Hartmut R. 2001. *Phonetische Analyse der Sprechgeschwindigkeit*. Universität München. (Doctoral dissertation).

Pfitzinger, Hartmut R., Susanne Burger & Sebastian Heid. 1996. Syllable detection in read and spontaneous speech. In *Proceedings of the Fourth International Conference on Spoken Language (ICSLP) Philadephia*, vol. 2, 1261–1264.

Pierrehumbert, Janet. 1980. *The phonetics and phonology of English intonation*. Massachusetts Institute of Technology. (Doctoral dissertation).

Pierrehumbert, Janet & David Talkin. 1992. Lenition of /h/ and glottal stop. In Gerard J. Docherty & Robert Ladd (eds.), *Papers in laboratory phonology ii: Gesture, segment, prosody*, 90–117. Cambridge, UK: Cambridge University Press.

Pike, Kenneth L. 1943. *Phonetics: A critical analysis of phonetic theory and a technic for practical description of sounds*. Ann Arbor: University of Michigan Press.

Pike, Kenneth L. 1945. *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.

Pinker, Steven & Alan Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28(1-2). 73–193. DOI: 10.1016/0010-0277(88)90032-7.

Plack, Christopher J. & Andrew J. Oxenham. 2005. The psychophysics of pitch. In Christopher J. Plack, Andrew J. Oxenham, Richard R. Fay & Arthur N. Popper (eds.), *Pitch: Neural coding and perception*, 7–55. New York: Springer.

Plack, Christopher J. L. & Robert P. Carlyon. 1995. Loudness perception and intensity coding. In Brian C. J. Moore (ed.), *Hearing*, 2nd (Handbook of Perception and Cognition), 123–160. San Diego: Academic Press.

Plomp, Reinier. 1976. *Aspects of tone sensation: A psychophysical study*. London; New York: Academic Press.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Poeppel, David & M. Florencia Assaneo. 2020. Speech rhythms and their neural foundations. *Nature Reviews Neuroscience* 21(6). 322–334. DOI: 10.1038/s41583-020-0304-4.

Port, Robert, Fred Cummins & Michael Gasser. 1996. A dynamic approach to rhythm in language: Toward a temporal phonology. In *Proceedings of the Chicago linguistic society*.

Port, Robert F. & Timothy Van Gelder. 1995. *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, Mass.: MIT press.

Price, Patti Jo. 1980. Sonority and syllabicity: Acoustic correlates of perception. *Phonetica* 37(5-6). 327–343.

Prince, Alan. 1990. Quantitative consequences of rhythmic organization. *Papers from the annual regional meeting, Chicago Linguistic Society* 26(2). 355–398.

Prince, Alan & Paul Smolensky. 2004. *Optimality theory: Constraint interaction in generative grammar*. Oxford: Blackwell Publishing.

Puppel, Stanislaw. 1992. The sonority hierarchy in a source-filter dependency framework. In Jacek Fisiak & Stanislaw Puppel (eds.), *Phonological investigations*, 467–483. Amsterdam; Philadelphia: John Benjamins.

Pylyshyn, Zenon W. 1985. *Computation and cognition : Toward a foundation for cognitive science*. 2nd. Cambridge, Mass; London: MIT Press.

R Core Team. 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Rączaszek, Joanna, Betty Tuller, Lewis P. Shapiro, Pamela Case & Scott Kelso. 1999. Categorization of ambiguous sentences as a function of a changing prosodic parameter: A dynamical approach. *Journal of Psycholinguistic Research* 28(4). 367–393. DOI: 10.1023/A:1023289031747.

Rączaszek-Leonardi, Joanna & J. A. Scott Kelso. 2008. Reconciling symbolic and dynamic aspects of language: Toward a dynamic psycholinguistics. *New Ideas in Psychology* 26(2). 193–207. DOI: 10.1016/j.newideapsych.2007.07.003.

Rączaszek-Leonardi, Joanna, Iris Nomikou, Katharina J. Rohlfing & Terrence W. Deacon. 2018. Language development from an ecological perspective: Ecologically valid ways to abstract symbols. *Ecological Psychology* 30(1). 39–73.

Ramus, Franck, Marina Nespor & Jacques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3). 265–292. DOI: 10.1016/S0010-0277(00)00101-3.

Räsänen, Okko, Gabriel Doyle & Michael C. Frank. 2018. Pre-linguistic segmentation of speech into syllable-like units. *Cognition* 171. 130–150.

Repp, Bruno H. 2005. Sensorimotor synchronization: A review of the tapping literature. *Psychonomic bulletin & review* 12(6). 969–992.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Rialland, Annie. 1994. The phonology and phonetics of extrasyllabicity in French. In Patricia A. Keating (ed.), *Phonological structure and phonetic form*, 136–159. Cambridge: Cambridge University Press.

Roessig, Simon & Doris Mücke. 2019. Modeling dimensions of prosodic prominence. *Frontiers in Communication* 4. 44.

Roessig, Simon, Doris Mücke & Martine Grice. 2019. The dynamics of intonation: Categorical and continuous variation in an attractor-based model. *PLoS ONE* 14(5). e0216859. DOI: 10.1371/journal.pone.0216859.

Roessig, Simon, Bodo Winter & Doris Mücke. 2022. Tracing the phonetic space of prosodic focus marking. *Frontiers in Artificial Intelligence* 5. DOI: 10.3389/frai.2022.842546.

Roettger, Timo B. & Martine Grice. 2019. The tune drives the text: Competing information channels of speech shape phonological systems. *Language Dynamics and Change* 9(2). 265–298. DOI: 10.1163/22105832-00902006.

Rose, Jerzy E., John F. Brugge, David J. Anderson & Joseph E. Hind. 1967. Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *Journal of neurophysiology* 30(4). 769–793. DOI: 10.1152/jn.1967.30.4.769.

Rosén, Haim B. 1956. *Ha-ʿivrit shelanu: Demutah be-or shiṭot ha-balshanut*. Tel-Aviv: Am Oved [in Hebrew].

Rosen, Stuart. 1992. Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions: Biological Sciences* 336(1278). 367–373.

Rouse, Andrew A., Peter F. Cook, Edward W. Large & Colleen Reichmuth. 2016. Beat keeping in a sea lion as coupled oscillation: Implications for comparative understanding of human rhythm. *Frontiers in Neuroscience* 10(257). DOI: 10.3389/fnins.2016.00257.

Rubach, Jerzy & Geert E. Booij. 1990. Edge of constituent effects in Polish. *Natural Language & Linguistic Theory* 8(3). 427–463.

Rumelhart, David E., James L. McClelland & the PDP Research Group. 1986a. *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations*. Cambridge, Mass.: The MIT Press.

Rumelhart, David E., James L. McClelland & the PDP Research Group. 1986b. *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2: Psychological and biological models*. Cambridge, Mass.: The MIT Press.

Savino, Michelina, Caterina Ventura, Simona Sbranna, Aviad Albert & Martine Grice. 2021. Native variety interference when imitating intonation in a non-native variety of Italian. In *The 4th Phonetics and Phonology in Europe (PaPE). Barcelona, Spain [poster presentation]*.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Sbranna, Simona, Caterina Ventura, Aviad Albert & Martine Grice. 2021a. Developing prosodic competence: Marking information status in L2 German. In *The 1st International Conference on Tone and Intonation (TAI). Sonderborg, Denmark [poster presentation]*.

Sbranna, Simona, Caterina Ventura, Aviad Albert & Martine Grice. 2021b. Prosodic marking of information status: An exploratory study on L1 Italian and L2 German. In *TiPToP: Trends in pedagogical transmission of prosody. Konstanz, Germany [oral presentation]*.

Scheer, Tobias. 2007. On the status of word-initial clusters in Slavic (and elsewhere). In Richard U. Compton, Magdalena Goledzinowska & Ulyana M. Savchenko (eds.), *Proceedings of FASL*, vol. 15, 346–364. Ann Arbor: Michigan Slavic Publications.

Schielzeth, Holger & Wolfgang Forstmeier. 2009. Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology* 20(2). 416–420. DOI: 10.1093/beheco/arn145. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2657178%5C&tool=pmcentrez%5C&rendertype=abstract.

Schouten, Jan Frederik. 1938. The perception of subjective tones. *Proceedings of the Koninklijke Akademie van Wetenschappen* 41. 1086–1093.

Schwarzwald, Ora Rodrigue. 2005. Modern Hebrew consonant clusters. In Dorit Diskin Ravid & Hava Bat-Zeev Shyldkrot (eds.), *Perspectives on language and language development: Essays in honor of Ruth A. Berman*, 45–60. Dordrecht; Boston; London: Kluwer Academic Press.

Searle, John R. 1980. Minds, brains, and programs. *Behavioral and Brain Sciences* 3(3). 417–424. DOI: 10.1017/S0140525X00005756.

Searle, John R. 1990. Cognitive science and the computer metaphor. In Karamjit S. Gill, Bo Göranzon & Magnus Florin (eds.), *Artificial intelligence, culture and language: On education and work*, 23–34. London: Springer.

Seebeck, August. 1841. Beobachtungen über einige Bedingungen der Entstehung von Tönen. *Annalen der Physik* 129(7). 417–436. DOI: 10.1002/andp.18411290702.

Selkirk, Elisabeth O. 1984. On the major class features and syllable theory. In Mark Aronoff, Richard T. Oehrle & Moris Halle (eds.), *Language sound and structure: Studies in phonology presented to Morris Halle by his teacher and students*, 107–136. Cambridge, Mass.: The MIT Press.

Seshadri, Guruprasad & B. Yegnanarayana. 2009. Perceived loudness of speech based on the characteristics of glottal excitation source. *J Acoust Soc Am* 126(4). 2061–71. DOI: 10.1121/1.3203668.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Sharma, Bidisha & S. R. Mahadeva Prasanna. 2018. Significance of sonority information for voiced/unvoiced decision in speech synthesis. *Speech Communication* 99. 201–210. DOI: 10.1016/j.specom.2018.04.002.

Shaw, Jason, Adamantios I. Gafos, Philip Hoole & Chakir Zeroual. 2009. Syllabification in Moroccan Arabic: Evidence from patterns of temporal stability in articulation. *Phonology* 26(01). 187–215.

Shepard, Roger. 2001. Pitch perception and measurement. In Perry R. Cook (ed.), *Music, cognition, and computerized sound: An introduction to psychoacoustics*, 149–165. Cambridge, Mass.: MIT Press.

Shiffrin, Richard, Michael Lee, Woojae Kim & Eric-Jan Wagenmakers. 2008. A Survey of Model Evaluation Approaches With a Tutorial on Hierarchical Bayesian Methods. *Cognitive Science: A Multidisciplinary Journal* 32(8). 1248–1284. DOI: 10.1080/03640210802414826. (17 May, 2019).

Sievers, Eduard. 1893. *Grundzüge der Phonetik zur Einführung in das Studium der Lautlehere der indogermanischen Sprachen*. 4th. Leipzig: Breitkopf & Härtel.

Sigurd, Bengt. 1955. Rank order of consonants established by distributional criteria. *Studia Linguistica* 9(1-2). 8–20.

Skovenborg, Esben. 2012. Loudness range (LRA) – design and evaluation. In *Audio engineering society convention 132*.

Slowikowski, Kamil. 2019. *ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'*. R package version 0.8.1. https://CRAN.R-project.org/package=ggrepel.

Smolensky, Paul, Matthew Goldrick & Donald Mathis. 2014. Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science* 38(6). 1102–1138. DOI: 10.1111/cogs.12047.

Smolensky, Paul & Géraldine Legendre. 2006. *The harmonic mind: From neural computation to optimality-theoretic grammar*. Cambridge, Mass.: MIT Press.

Spivey, Michael. 2007. *The continuity of mind*. New York: Oxford University Press.

Stan Development Team. 2018a. *RStan: The R interface to Stan*. R package version 2.18.2. http://mc-stan.org/.

Stan Development Team. 2018b. *Stan: A C++ library for probability and sampling, Version 2.18.2*. http://mc-stan.org/.

Stan Development Team. 2018c. *StanHeaders: Headers for the R interface to Stan*. R package version 2.18.0. http://mc-stan.org/.

Steriade, Donca. 1982. *Greek prosodies and the nature of syllabification*. Massachusetts Institute of Technology. (Doctoral dissertation).

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Steriade, Donca. 1999. Alternatives To syllable-based accounts of consonantal phonotactics. In *Proceedings of the 1998 Linguistics and Phonetics Conference*, 205–242. Prague: Karolinum Press.

Stockhausen, Karlheinz & Cornelius Cardew. 1959. How time passes by. *Die Reihe* 3(English). 10–40.

Sueur, Jerome, Thierry Aubin, Caroline Simonis, Laurent Lellouch, Ethan C. Brown, Marion Depraetere, Camille Desjonqueres, Francois Fabianek, Amandine Gasc, Eric Kasten, Stefanie LaZerte, Jonathan Lees, Jean Marchal, Andre Mikulec, Sandrine Pavoine, David Pinaud, Alicia Stotz, Luis J. Villanueva-Rivera, Zev Ross, Carl G. Witthoft & Hristo Zhivomirov. 2020. *seewave: Sound Analysis and Synthesis*. R package version 2.1.6. https://CRAN.R-project.org/package=seewave.

Sung, Eunkyung. 2016. Perception of onset clusters by English and Korean listeners: Universal markedness and L2 phonotactic knowledge. *Studies in Phonetics, Phonology and Morphology* 22(3). 477–498.

Surprenant, Aimée M. & Louis Goldstein. 1998. The perception of speech gestures. *The Journal of the Acoustical Society of America* 104(1). 518–529.

Suzuki, Yoiti & Hisashi Takeshima. 2004. Equal-loudness-level contours for pure tones. *The Journal of the Acoustical Society of America* 116(2). 918–933.

Tal, Idan, Edward W. Large, Eshed Rabinovitch, Yi Wei, Charles E. Schroeder, David Poeppel & Elana Zion Golumbic. 2017. Neural entrainment to the beat: The "missing-pulse" phenomenon. *Journal of Neuroscience* 37(26). 6331–6341. DOI: 10.1523/JNEUROSCI.2500-16.2017.

Tamási, Katalin & Iris Berent. 2014. Sensitivity to phonological universals: The case of stops and fricatives. *Journal of Psycholinguistic Research* 44(4). 359–381. DOI: 10.1007/s10936-014-9289-3.

Tang, Kevin, Mellissa MC DeMille, Jan C. Frijters & Jeffrey R. Gruen. 2020. DCDC2 READ1 regulatory element: How temporal processing differences may shape language. *Proceedings of the Royal Society B* 287(1928). 20192712. DOI: 10.1098/rspb.2019.2712.

Tilsen, Sam & Amalia Arvaniti. 2013. Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America* 134(1). 628–639.

Tremblay, P., M. Baroni & U. Hasson. 2013. Processing of speech and non-speech sounds in the supratemporal plane: Auditory input preference does not predict sensitivity to statistical structure. *Neuroimage* 66. 318–332.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Tuller, Betty. 2004. Categorization and learning in speech perception as dynamical processes. In Michael A. Riley & Guy C. van Orden (eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences*. Arlington, VA: National Science Foundation. www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp.

Tuller, Betty, Pamela Case, Mingzhou Ding & J. A. Kelso. 1994. The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human perception and performance* 20(1). 3.

Tuller, Betty, McNeel G. Jantzen & Viktor K. Jirsa. 2008. A dynamical approach to speech categorization: Two routes to learning. *New Ideas in Psychology* 26(2). 208–226. DOI: 10.1016/j.newideapsych.2007.07.002.

Tupper, Paul K. & Michael Fry. 2012. Sonority and syllabification in a connectionist network: An analysis of BrbrNet. In Stephen G. Parker (ed.), *The sonority controversy*, 385–409. Berlin; Boston: De Gruyter Mouton.

Turk, Alice & Stefanie Shattuck-Hufnagel. 2013. What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology* 4(1). 93–118.

Turk, Alice E. & James R. Sawusch. 1996. The processing of duration and intensity cues to prominence. *The Journal of the Acoustical Society of America* 99(6). 3782–3790.

Ultan, Russell. 1978. A typological view of metathesis. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language, volume 2: Phonology*, 367–402. Stanford: University Press Stanford.

Urbanek, Simon & Jeffrey Horner. 2020. *Cairo: R Graphics Device using Cairo Graphics Library for Creating High-Quality Bitmap (PNG, JPEG, TIFF), Vector (PDF, SVG, PostScript) and Display (X11 and Win32) Output*. R package version 1.5-12. https://CRAN.R-project.org/package=Cairo.

Van de Vijver, Ruben & Dinah Baer-Henney. 2012. Sonority intuitions are provided by the lexicon. In Stephen G. Parker (ed.), *The sonority controversy*, 195–218. Berlin; Boston: De Gruyter Mouton.

van de Weijer, Jeroen. 1996. *Segmental structure and complex segments*. Tübingen: Niemeyer.

Vasishth, Shravan, Bruno Nicenboim, Mary E. Beckman, Fangfang Li & Eunjong Kong. 2018. Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics* 71. 147–161. DOI: 10.1016/j.wocn.2018.07.008.

Vaux, Bert & Andrew Wolfe. 2009. The appendix. In Eric Raimy & Charles E. Cairns (eds.), *Contemporary views on architecture and representations in phonology*, vol. 48 (Current Studies in Linguistics), 101–135. Cambridge, Mass.; London, England: A Bradford Book The MIT Press.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Vehtari, Aki, Andrew Gelman & Jonah Gabry. 2015. Pareto smoothed importance sampling. *arXiv preprint* 1507.02646. http://arxiv.org/abs/1507.02646 (17 May, 2019).

Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian and Latin.* Berlin; New York; Amsterdam: Mouton de Gruyter.

Ventura, Caterina, Martine Grice, Michelina Savino, Aviad Albert & Petra B. Schumacher. 2019. Perceptual evaluation of post-focal prominence in Italian by L1 and L2 naïve listeners. In *The 3rd Phonetics and Phonology in Europe (PaPE). Bari and Lecce, Italy [poster presentation].*

Verschooten, Eric, Shihab Shamma, Andrew J. Oxenham, Brian CJ Moore, Philip X. Joris, Michael G. Heinz & Christopher J. Plack. 2019. The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints. *Hearing research* 377. 109–121.

Vishnubhotla, Srikanth. 2007. *Detection of irregular phonation in speech.* University of Maryland. (MA thesis).

Vitevitch, Michael S. & Paul A. Luce. 2004. A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers* 36(3). 481–487. DOI: 10.3758/BF03195594.

von Békésy, Georg & Ernest Glen Wever. 1960. *Experiments in hearing.* New York, N.Y.: McGraw-Hill.

Wang, Dagen & S. S. Narayanan. 2007. Robust speech rate estimation for spontaneous speech. *IEEE Transactions on Audio, Speech, and Language Processing* 15(8). 2190–2201. DOI: 10.1109/TASL.2007.905178.

Ward, W. Dixon. 1954. Subjective musical pitch. *The Journal of the Acoustical Society of America* 26(3). 369–380. DOI: 10.1121/1.1907344.

Warren, Richard M. 1982. *Auditory perception: A new synthesis.* New York: Pergamon Press.

Warren, Richard M. & James A. Bashford. 1981. Perception of acoustic iterance: Pitch and infrapitch. *Perception & Psychophysics* 29(4). 395–402. DOI: 10.3758/BF03207350.

Wever, Ernest Glen & Charles William Bray. 1930. The nature of acoustic response: The relation between sound frequency and frequency of impulses in the auditory nerve. *Journal of experimental psychology* 13(5). 373–387. DOI: 10.1037/h0075820.

Whitney, William Dwight. 1865. The relation of vowel and consonant. *Journal of the American Oriental Society* 8. 357–73.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Wickelgren, Wayne A. 1969. Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review* 76(1). 1. DOI: 10.1037/h0026823.

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer. https://ggplot2.tidyverse.org.

Wickham, Hadley. 2019. *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.4.0. https://CRAN.R-project.org/package=stringr.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani & Dewey Dunnington. 2020. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. R package version 3.3.0. https://CRAN.R-project.org/package=ggplot2.

Wickham, Hadley, Romain François, Lionel Henry & Kirill Müller. 2020. *dplyr: A Grammar of Data Manipulation*. R package version 0.8.5. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley & Lionel Henry. 2018. *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*. R package version 0.8.2. https://CRAN.R-project.org/package=tidyr.

Wickham, Hadley, Jim Hester & Romain Francois. 2018. *readr: Read Rectangular Text Data*. R package version 1.3.1. https://CRAN.R-project.org/package=readr.

Wiese, Richard. 2001. The phonology of /r/. In T. Alan Hall (ed.), *Distinctive feature theory*, vol. 2, 335. Berlin: Walter de Gruyter.

Wilson, Colin, Lisa Davidson & Sean Martin. 2014. Effects of acoustic–phonetic detail on cross-language speech production. *Journal of Memory and Language* 77. 1–24. DOI: 10.1016/j.jml.2014.08.001.

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 34–57. Pearson Education.

Yao, Yuling, Aki Vehtari, Daniel Simpson & Andrew Gelman. 2017. Using stacking to average Bayesian predictive distributions. *Bayesian Analysis* 13(3). 917–1007. DOI: 10.1214/17-BA1091.

Yavaş, Mehmet, Avivit Ben-David, Ellen Gerrits, Kristian E. Kristoffersen & Hanne G. Simonsen. 2008. Sonority and cross-linguistic acquisition of initial s-clusters. *Clinical linguistics & phonetics* 22(6). 421–441.

Yeverechyahu, Hadas. 2019. Consonant co-occurrence restrictions in Modern Hebrew. *Brill's Journal of Afroasiatic Languages and Linguistics* 11(1). 57–68. DOI: 10.1163/18776930-01101006.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Yeverechyahu, Hadas & Outi Bat-El. 2020. Biblical Hebrew segholates: Universal and language-specific effects. *Brill's Journal of Afroasiatic Languages and Linguistics* 12(1). 31–73. DOI: 10.1163/18776930-01201002.

Young, Mackenzie & Colin Wilson. 2017. Markedness effects in visual processing of non-native onset clusters. In *Proceedings of the 34th West Coast Conference on Formal Linguistics*.

Zec, Draga. 1995. Sonority constraints on syllable structure. *Phonology* 12(01). 85–129.

Zec, Draga. 2003. Prosodic weight. In Caroline Féry & Ruben van de Vijver (eds.), *The optimal syllable*, 123–143. Cambridge University Press.

Zeileis, Achim, Gabor Grothendieck & Jeffrey A. Ryan. 2020. *zoo: S3 Infrastructure for Regular and Irregular Time Series (Z's Ordered Observations)*. R package version 1.8-7. https://CRAN.R-project.org/package=zoo.

Zhang, Jie. 2001. *The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis*. University of California, Los Angeles. (Doctoral dissertation).

Zhao, Xu & Iris Berent. 2015. Universal restrictions on syllable structure: Evidence from Mandarin Chinese. *Journal of Psycholinguistic Research* 45(4). 795–811. DOI: 10.1007/s10936-015-9375-1.

Zwicky, Arnold. 1972. Note on a phonological hierarchy in English. In Robert P. Stockwell & Ronald K. S. Macaulay (eds.), *Linguistic change and generative theory*, 275–301. Bloomington, IN: Indiana University Press.

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

Proofreading version. Do not quote. Final version available from http://www.langsci-press.org

A model of sonority based on pitch intelligibility

Sonority is a central notion in phonetics and phonology and it is essential for generalizations related to syllabic organization. However, to date there is no clear consensus on the phonetic basis of sonority, neither in perception nor in production. The widely used *Sonority Sequencing Principle* (SSP) represents the speech signal as a sequence of discrete units, where phonological processes are modeled as symbol manipulating rules that lack a temporal dimension and are devoid of inherent links to perceptual, motoric or cognitive processes. The current work aims to change this by outlining a novel approach for the extraction of continuous entities from acoustic space in order to model dynamic aspects of phonological perception. It is used here to advance a functional understanding of sonority as a universal aspect of prosody that requires pitch-bearing syllables as the building blocks of speech.

This book argues that sonority is best understood as a measurement of *pitch intelligibility* in perception, which is closely linked to *periodic energy* in acoustics. It presents a novel principle for sonority-based determinations of well-formedness – the *Nucleus Attraction Principle* (NAP). Two complementary NAP models independently account for symbolic and continuous representations and they mostly outperform SSP-based models, demonstrated here with experimental perception studies and with a corpus study of Modern Hebrew nouns.

This work also includes a description of ProPer (*Prosodic Analysis with Periodic Energy*). The ProPer toolbox further exploits the proposal that periodic energy reflects sonority in order to cover major topics in prosodic research, such as prominence, intonation and speech rate. The book is finally concluded with brief discussions on selected topics: (i) the phonotactic division of labor with respect to /s/-stop clusters; (ii) the debate about the universality of sonority; and (iii) the fate of the classic phonetics–phonology dichotomy as it relates to continuity and dynamics in phonology.