# A STORY OF /v/: VOICED SPIRANTS IN THE OBSTRUENT-SONORANT DIVIDE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Christina Jean Bjorndahl

August 2018

© 2018 Christina Jean Bjorndahl
ALL RIGHTS RESERVED

A STORY OF /v/: VOICED SPIRANTS
IN THE OBSTRUENT-SONORANT DIVIDE

Christina Jean Bjorndahl, Ph.D.
Cornell University 2018

The patterning of /v/ in Russian and other languages has long posed a problem for phonological theory because of its ambiguous classification with respect to the feature [sonorant]: like obstruents, /v/ participates as a target of final devoicing and regressive voicing assimilation, but like sonorants, fails to trigger regressive voicing assimilation. In this dissertation I tackle the problem of Russian /v/ by situating it in a broader cross-linguistic landscape, in terms of its acoustic properties, its relationship to other non-sibilant voiced fricatives, or *voiced spirants* /β, ð, ɣ/, and its phonological typology. The empirical results of the dissertation challenge the binary division between obstruents and sonorants, and suggest that a finer distinction is required.

Chapter 2 and Chapter 3 focus on the phonetic characteristics of /v/. Chapter 2 investigates the relationship between voicing and frication type in four languages—English, Greek, Serbian, and Russian—using measures that distinguish between voicing and frication type: duration, harmonicity, and spectral centroid. Two environments, word-initial stressed and word-medial unstressed, are used in order to assess the effect of word-internal prosodic factors. I argue based on these results that although /v/ has the highest harmonicity in each language, the relationship between voicing and frication type is revealed by spectral centroid, and that the relationship is independent in English and Greek, but not in Serbian; in Russian the results depend on environment. Chapter 3 tests whether there exists a correlation between phonological and phonetic identity, using Greek, in which /v/ patterns

as an obstruent, Serbian, in which /v/ patterns as a sonorant, and Russian, in which /v/ is ambiguous. A partial correlation between phonetic and phonological identity was found, with the relativized centroid values for Greek and Serbian /v/ being distinguished in both environments. Russian and Greek do not differ significantly in the WIS environment, while Russian and Serbian do not differ significantly in the WMU environment.

Chapter 4 turns to investigating the distribution in consonant inventories and implicational relations of voicing in spirants and sibilants. The results of this investigation show that voiced spirants and voiced sibilants do not share the same cross-linguistic distribution: voiced spirants frequently surface in small inventories, but only if unpaired, while the correlation between the presence of voiced sibilants (which only occur with a voiceless counterpart) and inventory size is linear. Voiced spirants also tend to appear in inventories without a voicing contrast in the stops, unlike voiced sibilants.

Chapter 5 argues that Russian /v/ is featurally specified as [+sonorant, +obstruent], and that regressive voicing assimilation is inherently asymmetric: all segments that are [+obstruent] are targets of devoicing under regressive voicing assimilation and final devoicing, but only segments that are *exclusively obstruents* trigger RVA. A typology of phonological patterning of /v/ under [obstruent] and [sonorant] is proposed.

Chapter 6 considers the quantal nature of [obstruent] and [sonorant], and sketches a model that combines Quantal Theory (Stevens, 1989) and Emergent Feature Theory (Mielke, 2008). It is argued that by expanding the set of quantal domains, both clear cases of feature assignment and ambiguous cases can be understood within the same framework.

## BIOGRAPHICAL SKETCH

Christina Bjorndahl was born in Toronto, Ontario, where she attended elementary and high school. She attended the University of Toronto for her undergraduate degree, pursuing a double major in Linguistics and Mathematics with a minor in French Studies, graduating with honours. She pursued a master's degree in Linguistics, also at the University of Toronto. She is currently living in Pittsburgh, teaching linguistics courses at Carnegie Mellon University in the Philosophy Department.

Για τη μαμά μου, η πρώτη μου δασκάλα.

## ACKNOWLEDGEMENTS

### Animals in Academia

One night when I was putting my then 3-year old son, Anagnostis, to bed, I had to explain to him why I couldn't hug him to sleep, and instead had to work on my laptop on the chair beside him. I explained that I was writing a special kind of book called a dissertation, and that sometimes I wasn't sure what to write. He had some sage advice: "If you get confused, I can help you, Mama. I can tell you what to write. You should tell them about animals. That's what you can write." I promised him that I would try, and so I begin these acknowledgements in an unusual way, with some musings on being an animal in academia.

Writing this dissertation while being a mammalian mother has been a greater challenge than I could have ever imagined. My pregnancies were, thankfully, uncomplicated from a medical point of view, but exhausting nonetheless, as they are for all women. With my son I had to recover not just from pregnancy and labour, but also from surgery, as he was born by cesarean section. I nursed both my children, and pumped when they were in the care of others. At every milestone I would feel like I'd finally "recovered" until I hit the next stage: when the postpartum period started to be measured in weeks, not days, and then in months; when I stopped needing to pump; when I stopped nursing; when they got out of diapers, etc. At each milestone I would look back and realize how tired I'd been trying to get to that point, and how compromised my energy had been.

One is not supposed to talk about being in the trenches of motherhood at the gates of the Ivory Tower. Do I betray my fellow academic mothers by publicly admitting how tired I was? How tired I still am? I didn't understand the myriad ways one could be tired until becoming a parent: physically exhausted, emotionally drained, spiritually weary, and even (shh!) intellectually lethargic at times. It is a kind of tiredness that coffee can help, but does

not eradicate. Extra sleep helps, but is often interrupted, and the aches you get from the funny positions you assume because of all the extra limbs poking you make you wonder if it was worth it. But if you speak of this, you are admonished to sleep train, and if you don't sleep train, then your commitment to academia is questioned, and regardless of what sleep philosophy you adhere to (for nowadays everything in parenting is a "philosophy"), you will be judged, and you will doubt yourself, and it turns out that you will lose sleep regardless, either to your children or to your worries.

But talking about the exhaustion is, I think, okay. Lots of people joke about it, and seasoned parents tease you that you'll sleep again in $18-x$ years, where $x$ is the age of your youngest. Being tired is some kind of badge of honour in our society: how busy you are, how tired you are, how much you can do with the resources of your brain, in spite of the limitations the mammalian body places on you.

It is less okay to talk about what it means to embrace the mammal and human selves, and truly reconcile those aspects of our identity with the aspiring-to-be-great identity. I was lucky in that I had no trouble breastfeeding my children, and my mammalian instincts (which only revealed themselves to me when my first was born) found it the easiest course. So I breastfed both children until they were toddlers, and it was not out of adherence to some inane parenting philosophy, but rather a pragmatic decision based on what was easiest at the time. Also, it was pretty much free, if you don't count all the extra food I had to consume. I guess I shouldn't have been shocked, but I was unsettled when other human mammals disapproved of this decision, pointing out that I could drink more coffee if I wasn't nursing. Anyway, it isn't important what fools say, so I won't waste more time on them.

In fact, the hardest part about writing a dissertation with children is that children provide an outlet for apathy. When writing isn't going well, and edits are boring, and you're sick of your topic, it can be difficult, if not impossible, to slog ahead. When faced with the choice of toiling through the seemingly interminable and isolating slog that is euphemistically called

"ABD", the beautiful faces of your children shine ever brighter, their laughter is that much more infectious, their little hands seem that much more in need of yours. And, unlike other temptations against productivity, spending time with your children never feels like a waste of time, because it *isn't* a waste of time. I cannot count the number of days where the only reason I had for dropping them off at daycare, instead of spending the day playing with them, was that daycare was paid for, and I couldn't justify throwing the money away.

Of course, apathy is one fault that academics Do Not Have. Academics are forgiven for being tardy, slovenly, rude, and worse, but apathetic? Never! Undying passion and hunger (meant, usually, I think, in the metaphoric sense) are the hallmarks of Proper Academics. But this is a lie. It is true that passion and hunger are our fuel, but we are human, and we get bogged down, and sometimes don't care about a Thing. Even if that Thing is our livelihood and our passion. My point is, the difference between a child-free human trying to live out their aspiring-to-be-great identity and a parent is not that one hungers and the other doesn't. It's that when the hunger falters, and it does for everyone, the parent always has a good reason to walk away; moreover the parent will always be fulfilled by that moment, and the decision will have felt like the right one, at least until the guilt and the longing for intellectual stimulation settles in. And if, by good fortune, non-metaphorical hunger does not threaten the children, or if another kind of employment promises more joyous time with the children for more money, then is it any wonder why parents choose not to finish?

In fact, it is. Being a parent is not inimical to academia, or research. What becoming a parent does is cast a harsh light on problems already there. And so, if I may be forgiven a Moralizing Aside in the middle of this story to give some advice: whether you are an advisor or an advisee, do not despair that the aspiring-to-be-great identity has been smothered under the diaper-changing, tear-wiping, possibly food-dispensing identity of human parenthood. As with all things, This Too Shall Pass. In the interim, working within the new constraints is much more productive than pretending they don't exist, and certainly more moral than acting as though they ought not exist.

This brings me to another interaction I had with my son a few months later, when he was almost 4. After a particularly rough period trying to make progress but failing, I was talking to my husband about wanting to quit, when my son came into the room and asked what we were talking about. I explained, “Do you remember how I’m writing a book? Well, sometimes writing it is very hard, and sometimes writing it is so hard that I don’t want to do it anymore. So sometimes I think of quitting. What do you think I should do?” He took the question seriously, thought for a minute, and then responded, “I think you should not stop writing the book, and you shouldn’t think about that, because that’s what work is all about”.

Which brings me to the best part of juggling everything that goes with being a mother in academia: In my children I see my values and goals reflected. I see the person they think I am, and strive to be that person. So for all those times their little hands and faces and laughter tempted me away from my work, and for all those times their tears hurt my heart when I dropped them off at daycare, the fact that they exist and that what I do shapes their view of the world is a constant reminder that who I am is much more than a hungering, passion-filled, aspiring-to-be-great brain, and that when that drive falters, I slog on because “that’s what work is all about”.

In the next bit you will find a proper acknowledgments section, where I thank all those people who have influenced and helped this dissertation, written in the conventional way. But here, in my story about Animals in Academia, I acknowledge my mammalian children, Anagnostis Sebastian and Athena Vasiliki, as the greatest hindrance to my progress, but the ultimate reason that this thing got done at all.

## Proper acknowledgments

When I had to find a final paper topic in Phonology II with Draga Zec, she suggested, “Why don’t you look at this little problem of Russian /v/?” Funny how “little problems” can be decidedly large. I am most grateful to Draga for the enormous amount of latitude that she entrusted me with in pursing the questions I was interested in, while still reigning me in when I needed it. Her feedback has been instrumental in making this dissertation much better than I could have made it myself, particularly in the final stages of writing.

Abby Cohn’s encyclopedic knowledge of phonology and phonetics has been incredibly valuable, and I am grateful for all her suggestions and detailed feedback. I always appreciated the unique and difficult questions that Mats Rooth brought to the table, and am only sorry that I couldn’t work more closely with him. I would like to also take this opportunity to thank other faculty members at Cornell for their support at various times, and specifically acknowledge Wayles Browne for his assistance in coming up with the Russian word lists I used for the phonetic studies. And of course, nothing at Cornell would happen without the excellent office staff, to whom I am incredibly grateful for all their help and patience with administrative matters.

Much of the writing of this text was done in the presence of and with the support of others in writing groups, both virtual and in real life. I owe it to them that the thoughts in my head and the statistical results on my screen got turned into actual sentences. The mantra of “shut up and write” is much easier to follow with company and a cheerleading squad, and I was lucky enough that, no matter how miserable the task, the company was always enjoyable and helpful. The statistical analyses in Chapters 2–4 benefited tremendously from consultations with statisticians at Carnegie Mellon University, Rafael Stern (now at Universidade Federal de São Carlos), Jerzy Wieczorek, and Harold Seltman. I am also indebted to my good friend Ewan Dunbar, not just for his invaluable help on mixed models, but also for his friendship,

which is of the truest kind.

My time at Cornell was sweetened by my very close friends, Nikola and Esra Predolac, Jiwon Yun and Seongyeon Ko, Sarah Murray and Will Starr, Todd Snider, Nan Li, and Robin Karlin. Since graduating, Jiwon has been like an academic big sister to me, checking in on me and providing invaluable support. Likewise, Sarah has been not only a great friend, but her informal role as mentor provided crucial support, and even luxuries, at key times. Robin and I have never really stopped talking: even at a distance, our chat windows are always open with conversations that interleave phonology, phonetics, cooking, politics, academic gossip, and just about everything else, and one of these days maybe we'll manage to turn some of these chat windows into an actual paper or three. She has been an incredible friend to talk shop with, and her help with various tasks, whether word lists (and thank you, Andrej Bjelaković for extra help on this front!), Praat scripting, or proofreading is sprinkled throughout this dissertation. In Ithaca, I spent countless hours with Esra and Nikola, and they were my main support system during my time at Cornell, whether by helping me come up with Serbian wordlists or settling some of the world's most pressing questions, such as whether Greek or Turkish baklava is better (answer: Greek). In Pittsburgh, Seth Wiener, Melinda Fricke, and Jevon Heathe have all made being a phonologist/phonetician in this town much less lonely, and I am particularly grateful for Melinda's encouragement, mentorship, and, especially, her friendship over these last few years. To all the many linguists who have responded to my queries, translated sections of grammars for me, or talked /v/ with me, thank you. I consider myself very fortunate to be part of such a friendly discipline.

When my husband got a job at Carnegie Mellon University, I had no idea what to expect of my new "department-in-law", but the Philosophy Department opened its arms to me in ways I could never have imagined. To all the faculty, staff, and graduate students at CMU, thank you, from the bottom of my heart. You have provided me with teaching, humoured my attempts at philosophizing with you, submitted to me teaching you phonology

and phonetics, welcomed me into your lunch room, celebrated our children, and most of all, given me a true home. I am grateful to have experienced the absolute best of academia in your company. Mara and Mandy: thank you for your friendship and guidance, whether academic or maternal.

Going back in time somewhat, I wish to thank the faculty—particularly the ph-faculty, Keren Rice, Elan Dresher, Yoonjung Kang, and Alexei Kotchetov—at the University of Toronto for my wonderful years as an undergraduate and master's student in Linguistics. I'd like to specifically acknowledge Alexei for generously allowing me to use the Phonetics Lab after I had left Toronto, and for helping me with the design of the acoustic study in Chapter 3. Keren was my MA advisor, and the reason I became a phonologist. I owe my foundations in phonology to UofT, and will forever hope that I'm doing right by their teachings.

During my time as an undergraduate at UofT, I worked with Yael Karshon in Mathematics on a paper in topology. I haven't worried about convexity in a while, but Yael's influence has been robust in the face of time. She was a true mentor, and never shied away from offering me humane and kind advice on how to navigate academia as a woman and, later, as a mother. I never realized the importance of role models until our weekly meetings stopped, and although I can't repay her for her kindness, I aspire to pay it forward.

I have been extraordinarily fortunate to have multiple support networks. To the village that helps us raise our children, namely, our children's caregivers, thank you. If it weren't for you, I would either be living life as a mother, or as an academic, but certainly not both. Your love and dedication to the children in your care are worth far more than you know, and our children are richer for having had you in their lives. My neighbours, friends, and all the baristas in my life have listened to me rant and whine about this dissertation for far more than anyone should have to endure, and are probably as relieved that this is over as I am. And of course, to my good friends and κουμπάροι, thank you for being so great. Scattered though we may be, I hope the end of this dissertation marks a new beginning in how often we see each other.

To my grandparents, thank you for your support and for setting a wonderful example. I lost my παππού and γιαγιά during graduate school; I can only imagine how proud, and also confused, they would be if they knew their sacrifices led to their granddaughter getting a PhD. . . on /v/. Μου λείπετε. Words cannot even begin to express my love and gratitude for my parents, siblings, brother-in-law, aunts, uncles, and cousins. Even if I were to be completely utilitarian and restrict myself just to those times that they have supported me in this dissertation, it would be impossible to describe the impact that their support has had. They have watched our children, cleaned and cooked for us, fixed our house, travelled with us, stuffed money in our hands as we said our goodbyes, and sent us back to our daily routines with food to stock the freezer with. They are, in short, everything that anyone could ever want in a family. Σας ευχαριστώ και σας αγαπώ.

As for you, Adam, you are my best friend, and building this life with you is the best life I could be leading. Thank you for linguisticating with me all these years, for making me laugh every day, for being a true partner in marriage, parenting, and life, for celebrating my successes, and for holding my hand through what I thought were failures, but that you rightly saw as mere setbacks. Fourteen years and counting, the conversation just gets better and better. I love you.

Anagnostis and Athena: You are still young, but you have felt the stress of this dissertation more than anyone. Αναγνωστάρα μου, thank you for your patience, for wishing me luck when I left for work, and for always being so helpful, kind, and thoughtful. You are a beautiful boy. Αθηνούλα μου, you are too young to have been anything but yourself—imperious, fierce, and hilarious. But thank you anyway, because your demands that I hug you and your antics that then made me laugh helped keep things in perspective. Ψυχές μου, I love you both with all my heart.

## CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

CHAPTER 1

**INTRODUCTION**

## 1.1 Introduction

One of the core tenets of segmental phonology is that consonants are divided into sonorants and obstruents. In this dissertation I propose that there is phonetic, typological, and phonological evidence to treat non-sibilant, voiced fricatives such as /β, v, ð, ɣ/, what I call the *voiced spirants*, as both obstruent and sonorant. Specifically, I argue that the class of voiced fricatives (that is, the spirants and sibilants together) do not form a unified class, and that this disunity is reflected in how these two classes distribute in consonant inventories, their phonetic realization, and in how they pattern with respect to laryngeal phenomena. The voiced spirants have long been recognized as more sonorous than other obstruents, but their intermediacy with respect to sonorants and obstruents has generally been treated idiosyncratically. This dissertation is an attempt to rectify this gap in the literature by presenting cross-linguistic investigations into various dimensions of the identity of the class of voiced spirants.

The inspiration for the dissertation comes from the anomalous patterning of /v/ in Russian, and as such, /v/ plays a special role, acting as a case study for the set of the voiced spirants. Like obstruents, Russian /v/ undergoes final devoicing and regressive voicing assimilation, yielding forms such as /prav/ → [praf] 'right (masc.)' and /v-supe/ → [fsupe] 'in the soup'; on the other hand, like sonorants, but unlike obstruents, /v/ does not trigger voicing assimilation, allowing contrasts such as [tver] 'Tver (place name)' vs. [dver] 'door'. If Russian /v/ is an obstruent then it should trigger voicing assimilation, and [tv] should not be a licit onset cluster; if Russian /v/ is a sonorant, then it should not undergo final devoicing or regressive voicing assimilation. In the words of Jakobson (1978), "...the Standard Russian *v* ...occupies an obviously intermediate position between the obstruents and the sonorants".

The puzzle of /v/'s patterning in Russian has a long history in the phonological literature. In spite of this, nearly all analyses of the voicing facts in Russian provide a general analysis of the voicing assimilation facts, and then require an additional rule, constraint, feature or representation that uniquely imbues Russian /v/ with special properties that account for its dual status as both an obstruent and sonorant. Interestingly, however, no analysis specifically argues for exactly this: that [v] patterns in this way precisely because it belongs to both classes. Nevertheless, a literal interpretation of Jakobson's comment suggests treating Russian /v/ neither as an exclusive obstruent, nor as an exclusive sonorant, but as a third category that is intermediate between these two major classes. In line with this, I propose that /v/ is special because it is a member of what I call the *voiced spirants*, the class of non-sibilant, voiced fricatives, e.g. /β, v, ð, ɣ/. This class of sounds has a set of unusual properties with respect to its phonetics, consonant inventory membership, and phonological patterning, exhibiting traits of both sonorants and obstruents. It has long been recognized that the voiced spirants can exhibit both obstruent-like and sonorant-like traits, but phonologists have resisted an explicit representation of their intermediate identity, particularly in featural terms. In contrast, building on work by Clements and Osu (2002) I propose in Chapter 5 that there is a feature [obstruent], and that in languages like Russian, /v/ is featurally specified as both [+sonorant] and [+obstruent].

In order to motivate the analysis in Chapter 5, I first situate Russian /v/ in a broader cross-linguistic landscape. Russian /v/ is considered anomalous not just because of its patterning within the Russian phonological system, but cross-linguistically as well: it is a general, if tacit, assumption in phonological theory that if an inventory has both /v/ and /f/, then they function as a voiced-voiceless pair within the obstruent system: in other words, the relationship between /v/ and /f/ parallels the relationship between /b/ and /p/, and between /z/ and /s/.

Broadly, I approach the problem of /v/ by examining the relationship between voicing and obstruency. In Chapter 2 I present an acoustic investigation into the relationship between frication and voicing in four languages, probing whether the relationship between /v/ and /f/ parallels that of /z/ to /s/, a standard assumption of feature theories. Chapter 3 tests the hypothesis that the realization of [v] tokens correlates with the phonological identity of /v/, by examining three languages in which the phonological identity of /v/ in relation to /f/ and the remaining obstruent class differs. Chapter 4 probes how voiced spirants pattern with respect to their distribution in consonant inventories, challenging the assumption that the voiced spirants and voiced sibilants form a unified class. These chapters paint a much more nuanced cross-linguistic picture of the phonetic and typological identity of /v/.

I return to the problem that Russian /v/ poses for phonological theory in Chapter 5, where I argue that /v/ in Russian and other languages with similar patterning is featurally specified as [+sonorant, +obstruent]. The use of [obstruent] in addition to [sonorant] has precedence in a series of papers by Nick Clements and Sylvestre Osu who argue that Ikwere non-explosive stops exhibit phonetic properties, typological identity, and phonological patterning that is neither sonorant nor obstruent, and hence are specified as [−sonorant, −obstruent] (Clements and Osu, 2002, 2003, 2005). Voiced spirants like Russian /v/ complete the typology, by exhibiting phonetic properties, typological identity, and phonological patterning that is *both* sonorant and obstruent. Further, I show that the full typology includes instances of /v/ that are featurally specified as obstruents, and instances of /v/ that are featurally specified as sonorants.

As a result of these investigations, I conclude that the class traditionally denoted as *voiced fricatives* is not a unified class, and that the voiced sibilants and voiced spirants are distinct classes, phonetically, phonologically, and with respect to their distributions in inventory systems. These distinctions can be traced back to the different aerodynamic tensions inherent to the production of voiced sibilants as opposed to voiced spirants, leading to different resolutions and hence a different phonological status between the two classes. These results cast doubt on the traditional division between voiced obstruents and sonorants deemed necessary for accounting for laryngeal phenomena.

**Phonological vs. typological identity** Throughout this dissertation I use the shorthand *typological identity* to refer to the distribution of consonants in inventory systems, and thus distinguish between *typological identity* and *phonological identity.* This might be confusing, given that the typology of segments with respect to their distribution across inventories is generally taken to be part of their phonological character; for example, discussions about markedness frequently refer to frequency of occurrence and whether such segments are the antecedent or consequent of an implicational relation. Certainly, understanding the cross-linguistic character of such segments is crucial for understanding their phonological identity within any given language, as the cross-linguistic character can suggest typical and atypical inventory configurations. Nevertheless, I distinguish between typology with respect to inventory structure, the subject of Chapter 4, and *phonological typology*—i.e., referring to patterning—which is the subject of Chapter 5. Phonological typology is a part of phonology, as it reflects the constrained variation in how speakers categorize segments into classes that pattern together and are productive, and thus may reflect a cognitive fact about the language system, unlike the typology of voiced spirants in inventory systems. It has long been known that databases of inventories are flawed because it is not clear at what level the inventory holds: is it surface true, or does it reflect abstract analyses of the linguists? To what extent are such inventory descriptions influenced by the native language of grammarian? To what extent are the discovered tendencies the result of geographic or geneological factors, as opposed to cognitive factors? Linguists are aware of these issues, and various steps can be taken to mitigate them, as I discuss in Section 4.2. Moreover, such investigations can be fruitful in terms of understanding how segments and classes relate to each other, as I show in Chapter 4. Nevertheless, insofar as phonology is a module of the cognitive system, I argue that typological statements about inventories derived from database studies are not phonological.

The remainder of this chapter is structured as follows. In Section 1.2 I briefly summarize the phonological phenomena of interest, and their relationship to the obstruent-sonorant divide. In Section 1.3 I introduce the term *voiced spirants* to distinguish the sibilant from non-sibilant fricatives, and provide a partial motivation for doing so; the remainder of the dissertation is an argument for distinguishing the voiced spirants from the voiced sibilants. In Section 1.4 I present an overview of the phonetics of frication and voicing, with particular emphasis on the aerodynamic tensions voicing and frication present, separately and together, and on the acoustic realizations of voicing vs. frication; exploring the relationship between these variables, and how they are realized in different languages is the subject of Chapter 2. In Section 1.5 I introduce the hypothesis tested in Chapter 3 on the relationship between phonological identity and phonetic realization, and I establish the phonological identity of /v/ in the three languages under consideration. I then turn to the structure of consonant inventories, specifically as related to frication and voicing in Section 1.6, setting the stage for the database study of Chapter 4. In Section 1.7 I summarize the background provided in this chapter and outline the structure of the dissertation.

## 1.2 The obstruent-sonorant divide

Phonologically, obstruents and sonorants exhibit a constellation of different properties that motivates their distinction. As summarized by Botma (2011), sonorants tend to be closer to the syllable nucleus than obstruents, and in languages that admit consonantal nuclei, typically only sonorants, or a subset thereof, can occupy the nucleic position, while obstruents typically cannot; in some quantity-sensitive languages, sonorants, or a subset thereof, can bear weight, while obstruents cannot (Zec, 1995); in languages where consonants can bear tone, sonorants can bear tone, while obstruents cannot (Yip, 2002). Syllabicity, properties of weight and tone typically distinguish between obstruents and sonorants, but such phenomena are also sensitive to the relative sonority of segments, and are thus related to the Sonority Hierarchy. Similarly, in nasal harmony systems, sonorant segments are the most likely to be nasalized, and obstruents are the least likely to be nasalized, but nasalizability is widely understood to obey the nasalizability hierarchy, which mirrors, though is not identical to, the sonority hierarchy quite closely (Cohn, 1993; Walker, 2000).

Perhaps the most robust phonological evidence for a *binary* division between obstruents and sonorants is their relationship to voicing. Indeed, in *Sound Pattern of English*, Chomsky and Halle (1968) specifically cite this distinction in the definition of [sonorant]: "Sonorants are sounds produced with a vocal tract cavity configuration in which spontaneous voicing is possible; obstruents are produced with a cavity configuration that makes spontaneous voicing impossible" (pg. 302). The close relationship between sonority and voicing is further entrenched in the representation of the feature [voice], which is available only on obstruents, and is redundant for sonorants. Section 1.4.2 outlines the aerodynamic tensions that are inherent to the production of voicing in obstruents, and which are lacking in sonorants. Here I focus only on the phonological patterning of *sonorant voicing* and *obstruent*, or *laryngeal* voicing.

### 1.2.1 Phenomena of interest

In many languages, obstruents pattern differently than sonorants, specifically with respect to their relationship to voicing. While there are many such processes, I concentrate here on those most relevant to the present investigations: *final devoicing* and *regressive voicing assimilation*; I briefly discuss *pre-sonorant voicing* due to its relationship with final devoicing, and differences it exhibits with regressive voicing assimilation.

#### 1.2.1.1 Final devoicing

In Russian, voiced obstruents devoice word finally, a process commonly referred to as *final devoicing*. For languages such as Russian, in which the laryngeal opposition is truly one of voicing (Ringen and Kulikov, 2012), the term *final devoicing* is appropriate. However, many languages that employ a laryngeal contrast in obstruents neutralize this contrast in syllable or word-final position, even in cases where the laryngeal contrast in stops is not one of voicing, and so the process cannot be properly called "devoicing". German is a famous example of an aspirating language that neutralizes the contrast in syllable final position (1).[1] In light of cases like this, some phonologists propose the term *final laryngeal neutralization* to capture both cases; see Iverson and Salmons (2011) for review. In cases like Russian, I will typically refer to the process as *final devoicing*; when discussing languages that do not have a voicing contrast in the stops, I will use the term *final neutralization*. It is important to note that even in aspirating languages such as German, final neutralization applied to /v,

[1] Example (a) is from Jessen and Ringen (2002); (b) from Beckman et al. (2006); (c) from Wiese (2000). For simplicity, I transcribe the lenis stop series as /b, d, g/ in German, in spite of the fact that this series is rarely voiced.

z/ is a devoicing process.[2] In Russian, the domain of final devoicing is the prosodic word (Padgett, 2011); in German, it is widely taken to be the syllable, though Wagner (2002) argues that the domain of final neutralization is also the prosodic word; the question of domains is beyond the scope of the present work, and in what follows, I consider "final" to refer to being final in some prosodic domain.[3]

(1) German FN

| | | | | |
|---|---|---|---|---|
| a) **stops:** | /lob/ | [lop] | | 'praise (nom)' |
| b) **sibilants:** | /graz/ | [gras] | | 'grass (sg)' |
| c) **sonorants:** | /viːl/ | [fiːl] | *[fiːl̥] | 'much' |

In general, sonorants are considered to be exempt from the kind of final devoicing found in Russian and German, although there are cases where sonorants are reported to devoice in final position. For example, Russian word final sonorants, particularly liquids, are sometimes reported to devoice in word final position, but as Padgett (2002) argues, such devoicing is generally understood to be gradient and optional, in contrast to the categorical and obligatory phenomenon of final devoicing. Iverson and Salmons (2011) notes that in some languages like Kaqchikel, sonorants are reported to devoice; these cases are left aside in the present work.[4]

[2] I take no position here on whether neutralization is complete or partial; for discussion, see Yu (2011) and references therein.

[3] Beckman et al. (2006) argue that German fricative voicing is better analysed as positional faithfulness in presonorant position, and is not in fact coda devoicing.

[4] Note that since sonorants rarely exhibit a voicing contrast between voiced and voiceless members, the process does not result in neutralization, either partial or complete.

#### 1.2.1.2 Regressive voicing assimilation

In general, voicing assimilation neutralizes laryngeal contrasts in tautosyllabic consonant clusters. Lombardi (1996) showed that progressive voicing assimilation, like in English plural suffixation, is generally morphologically restricted to certain suffixes. When no voicing assimilation takes place, such clusters are nearly always heterosyllabic; for example, English [pɪɡ.pɛn] 'pigpen'. Cross-linguistically, tautosyllabic consonant clusters nearly always agree in laryngeal specification (Kehrein and Golston, 2004).[5]

Phonologically conditioned voicing assimilation is regressive, as in English Level 1 alternations of the form [skrɑɪb] 'scribe' ∼ [skrɪptɹ] 'scripture', and the overarching generalization is that obstruent clusters agree with the voicing specification of the rightmost obstruent; sonorants do not trigger voicing assimilation. Russian examples of RVA are given in (2): (a) shows voicing of an underlyingly voiceless obstruent, and (b) shows devoicing of an underlyingly voiced obstruent; note that in both (a) and (b), the first example (i) shows that sonorants are inert with respect to RVA.

(2) Russian RVA (Padgett, 2011)

[5]Exceptions to this generalization generally include cases like /v/, and I return to these issues in Chapter 5. Kreitman (2008) argues that Hebrew obstruent onsets allow for mixed voicing clusters such as [kd] and [dk]. The acoustic evidence she presents clearly shows mixed voiced clusters, but also shows that there is a lack of overlap between the consonants in question, with the first consonant fully released before the second. These results suggest that gestural timing is key to understanding the licensing of voicing in clusters, but addressing these issues is beyond the scope of the present work.

| | | | | |
|---|---|---|---|---|
| a) | i. | /ot-ˈjexatʲ/ | [ɐt-ˈexətʲ] | ‘to ride off’ |
| | ii. | /ot-stuˈpʲitʲ/ | [ɐt-stuˈpʲitʲ] | ‘to step back’ |
| | iii. | /ot-ˈbrosʲitʲ/ | [ɐd-ˈbrosʲitʲ] | ‘to throw aside’ |
| b) | i. | /pod-nʲeˈstʲi/ | [pəd-nʲiˈstʲi] | ‘to bring (to)’ |
| | ii. | /pod-pʲiˈsatʲ/ | [pət-pʲiˈsatʲ] | ‘to sign’ |
| | iii. | /pod-ˈ zet͡ʃʲ/ | [pɐd-ˈ zet͡ʃʲ] | ‘to burn’ |

#### 1.2.1.3 Pre-sonorant voicing and sonorant devoicing

**Pre-sonorant voicing** Although in many languages only obstruents trigger voicing assimilation, other languages exhibit voicing assimilation triggered by sonorants, in a phenomenon known as pre-sonorant voicing (PSV). In these languages, what counts as a trigger minimally includes sonorant consonants, but can also include vowels and voiced obstruents. Regardless, the voicing triggered by sonorants is of a qualitatively different nature from RVA, in terms of degree, frequency, segments affected, and the prosodic boundary across which it happens (Strycharczuk, 2012).

> First, voicing applies in the environment of a segment that is not contrastively voiced. . . Second, voicing is positionally restricted to environments where laryngeal neutralisation is also observed. . . Third, pre-sonorant voicing displays an array of manner asymmetries, with fricatives being more prone to undergo the process than stops, even though there is no clear evidence for a preference for voiced fricatives over voiced stops in the languages of the world. (Strycharczuk, 2012, pg. 15)

Because of the differences between pre-sonorant voicing and those phenomena that traditionally fall under the rubric of laryngeal phonology—regressive voicing assimilation and final devoicing—I do not focus on pre-sonorant voicing in this dissertation. Nevertheless, as I show in Chapter 5, PSV plays a role in several languages or dialects with ambiguous patterning of /v/, and so cannot be ignored completely.

**Sonorant devoicing** Sonorants are also reported to devoice in some languages, as in English [pl̥iz] ‘please’, or in Icelandic [vɔr̥s] ‘spring gen.’, but is often partial and gradient. Sonorant devoicing is often linked to aspiration phenomena (Stevens and Hajek, 2004), and Hansson (2001) notes that for Scandinavian dialects, “Virtually all dialects with preaspiration proper also have preaspiration in the form of sonorant devoicing”. In general, sonorant devoicing does not bear upon the present concerns, and is generally left aside.

### 1.2.2 Phonological implications of the obstruent-sonorant divide

Within phonological theory, accounting for the different voicing phenomena exhibited by obstruents and sonorants has taken a number of forms. The general consensus is that the distinction between obstruents and sonorants is representationally encoded in the featural specification of [voice]. Specifically, obstruents are represented as [±voice], while sonorants either are not marked for [voice] underlyingly, or are endowed with a different feature that captures the kind of voicing on sonorants. For example, Cho (1990b) argues for a privative [voice] feature, which is borne only by obstruents, while Wetzels and Mascaró (2001) argue for binary [voice] feature, but which is again borne only by obstruents. Rice (1993) argues that obstruents and sonorants bear different voicing features: [voice] is a dependent of the Laryngeal node, while sonorants are specified underlyingly for a Sonorant Voice (SV) node. Obstruent devoicing is readily accounted for, either by changing the value of [+voice] to [−voice] or, in autosegmental accounts, as delinking of [voice] from the laryngeal node. Sonorants, not being specified for [voice] are predicted not to become devoiced. Likewise, the presence of [voice] on obstruents accounts for its propensity to trigger other obstruents to match in voicing specification. Again, because sonorants are not specified as [voice], they do not spread their voicing. These facts are made particularly salient in autosegmental accounts of assimilation via feature spreading, where [voice] can only spread if there is a Laryngeal node on which it can dock.[6]

Russian /v/ is a puzzle for phonological theory precisely because it violates the assumed divide between obstruent and sonorant voicing. Since /v/ undergoes devoicing, it must have a [voice] feature to change or lose; since /v/ does not trigger voicing assimilation, it must not have a feature [voice] to spread. For this reason, Russian /v/ has resurfaced time and time again in phonological theory, and each new feature theory or phonological architecture is tested against the patterning or Russian /v/.[7] In light of the vast literature on Russian /v/, Padgett (2002) suggests that "the Russian consonant [v] has a status in phonology out of proportion to its size". I submit that the status may be out of proportion to its size, but only because that is the wrong comparison to make: when compared to the size of the obstruent-sonorant divide in phonological theory, the status of Russian /v/ has been, if anything, not taken seriously enough.[8]

I return to questions having to do with the representation of voicing in Chapter 7. Until then, the perspective taken in this dissertation is to situate Russian /v/ in a cross-linguistic setting, phonetically, phonologically, and typologically. In the remainder of this chapter

[6]In more recent work, Kehrein and Golston (2004) argue that laryngeal features are properties not of segments, but of subsyllabic constituents such as the onset, nucleus, and coda. Nevertheless, the distinction between obstruents and sonorants is still crucial to their account: sonorants are not laryngeal, and so tautosyllabic clusters such as /pl/ are not counterexamples to their claim.

[7]A non-exhaustive list of some of the major works addressing /v/ in Russian and other languages in which /v/ also patterns ambiguously is: Halle (1959), Lightner (1965), Andersen (1969), Coats and Harshenin (1971), Daniels (1972), Barkai and Horvath (1978), Jakobson (1978), Vago (1980), Hayes (1984), Burton and Robblee (1997), Kavitskaya (1998), Padgett (2002), Petrova and Szentgyörgyi (2004) Lulich (2004), Kiss and Bárkányi (2006). This list does not include literature that is not in English.

[8]I take it as personally encouraging the following comment regarding /v/ in Hall and Żygis (2010): "We therefore regard language-specific studies on the phonetics and phonology of [v] as a profitable area of future research because these studies have the potential of shedding light on features (e.g. [sonorant]) as well as the general interaction of phonetics and phonology"(pg. 3).

I provide the requisite background information for chapters 2–4; I return to the featural specification of the obstruent-sonorant divide in Chapter 5, and broader implications for phonological theory in Chapter 7.

## 1.3 Terminology: spirants vs. sibilants

In this dissertation I depart from tradition and use the term *voiced spirants* to refer to voiced, non-sibilant fricatives such as /β, v, ð, ɣ/. The term *spirant* is often considered synonymous with the term *fricative*; Crystal (2008), for example, directs the reader to the entry for *fricative* under the entry for *spirant.* The term *fricative* is the more popular of the two terms: as one (albeit crude) assessment of the relative popularity of the two terms, as of August 2018, Google Ngram viewer shows that, especially after 1960, the use of *spirant* has steadily decreased to almost 0, while *fricative* after 1960s is at least twice as popular.[9] Additional support comes from the phonological process of *spirantization*: when voiced stops [b, d, g] become spirantized, as in Spanish, they are realized as [β, ð, ɣ], but not as [β, z, ɣ]. In languages where /d/ does not spirantize to [ð], it there is a tendency for it to be realized as a tap or flap, or a liquid, but crucially not as the voiced sibilant [z] (Gurevich, 2004).[10]

Chomsky and Halle (1968) grouped the sibilant fricatives [s, z, ʃ, ʒ] together with labiodental [f, v] under the definition of *strident* fricatives. This grouping was motivated by the presence of a baffle downstream from the locus of constriction; however, as discussed in Section 1.4.1, all buccal fricatives other than [ɸ, β] involve an additional source of noise generation from the airstream hitting a wall or obstacle, and since Chomsky and Halle (1968), few linguists employ [+strident] to distinguish labiodental [f, v, s, z, ʃ, ʒ] from other fricatives.

[9] I am vaguely aware of the term *spirant* already being used in certain subfields to refer to the non-sibilant fricatives.

[10] The same pattern may not be true for voiceless stops. For example, the voiceless stop /t/ spirantizes to [s] in Nez Perce (Gurevich, 2004, pg. 5).

An important fact regarding the phonetics and phonology of voiceless vs. voiced spirants and sibilants is that, while voiceless spirants and sibilants pattern uniformly as the class of voiceless fricatives, and generally share various characteristics in terms of their acoustic realization and their relationship to inventory structure, the voiced spirants and sibilants do not. Specifically, the voiced spirants systematically exhibit traits of sonorants, in terms of their acoustic realization, typological distribution in consonant inventories, and phonological inventories.

Although the divide between sonorants and obstruents is in many ways well entrenched, the strict partition of consonants according to a binary feature [sonorant] is challenged by voiced spirants on phonetic, phonological and typological grounds. Phonetically, the voiced spirants are particularly prone to either compromised voicing or frication; as Ohala (1983) remarks, "on nonsibilant voiced fricatives ([β, v, ð, j, ɣ, ʁ]) [the frication] is often so weak as to be barely detectable" (pg. 202). To the extent that such segments are voiced but have little to no frication, their realization may be quite sonorous.

Additionally, the voiced spirants are typologically anomalous compared to their stop and sibilant counterparts. While the implicational relation of voicing is robust for stops and sibilants, the voiced spirants violate the relation much more frequently, and in this respect are again similar to sonorants, which almost universally appear without a voiceless counterpart. For example, in the UPSID segmental database there are no languages that contain /z/ without /s/, while spirant pairs such as /f, v/ exhibit a violation rate of about 20%, and for the bilabial pair /ɸ, β/, more languages have /β/ without /ɸ/ than with /ɸ/, with a violation rate of 75% (Maddieson, 1984).

In terms of their phonological identity, the voiced spirants are known to pattern with [+sonorant] segments (Botma and van't Veer, 2013). For example, Serbian obstruent clusters must agree in voicing, yielding alternations such as [redak] 'a line' ∼ [retka] 'line, gen. sg.'; /v/ does not participate in such alternations, yielding surface forms such as [ovca] 'sheep' and [tvoj] 'your'. In Tiv, only /m, n, r, v, ɣ/ can occur at the end of a word, and these consonants are also the only ones that can bear tone, suggesting a [+sonorant] classification (Zec, 1995).

In short, the voiced spirants are already known to be uneasy members of the obstruent class, phonetically, phonology, and with respect to the relationship they exhibit with voiceless obstruents and consonant inventories as a whole. This dissertation takes as a starting point the hypothesis that the voiced spirants and sibilants do not form a coherent class. The conclusion of this dissertation is that voiced spirants are not a proper subset of the obstruents, and insofar as fricatives are obstruents, the term "voiced, non-sibilant fricatives" is inappropriate. That being said, for brevity I use the term "fricative" as a shorthand to refer to the spirants and sibilants together (rather than unwieldy descriptions like "sibilants and spirants", or "non-liquid, non-glide continuants"). The thesis that voiced sibilants and spirants are not a coherent class, and that voiced spirants are not a proper subset of obstruents raises the question of whether voiceless spirants and sibilants are a single class. The short answer to this appears to be "yes", but some complications arise regarding the status of voiceless fricatives depending on place. Excluding the class of laryngeal fricatives, there are some examples of voiceless velar fricatives patterning ambiguously. For example, Hungarian /x/ (or /h/)[11] exhibits patterning that is the mirror-image of the ambiguous patterning of /v/: like obstruents, /x/ triggers RVA, but like sonorants, does not alter its voicing value when it precedes a voiced obstruent (Szigetvári, 1998). As I discuss in Chapter 7, these facts suggest that in addition to distinguishing between sibilant and spirant manners of articulation and how they interact with the obstruent-sonorant divide, a more nuanced

[11] /h/ and /x/ are allophones of the same underlying segment, and I take no position here on which is underlying.

understanding of the role of place and its relationship to voicing is required. Nevertheless, a detailed examination of the patterning of voiceless spirants is beyond the scope of this work, and so for the purpose of this dissertation I take the position that voiceless spirants and sibilants pattern as a single class.

To summarize, I use the term *spirant* to refer to what are conventionally known as the non-sibilant fricatives, particularly those transcribed as /ɸ, β, f, v, θ, ð, x, ɣ/; the term *sibilant* to refer to sibilant fricatives such as /s, z, ʃ, ʒ/, and the term *fricative* as shorthand to refer to the sibilants and spirants together.

## 1.4 The phonetics of frication and voicing

There are two cross-linguistic phonetic investigations presented in this dissertation, in Chapter 2 and Chapter 3. The goal of Chapter 2 is to probe the relationship between voicing and the kind of frication (sibilant vs. non-sibilant) in four languages, English, Greek, Serbian, and Russian, based on the acoustic properties of four segments common to all four languages, [f, v, s, z]. It is well known that the simultaneous generation of voicing and frication are in conflict, but the tensions for voiced sibilants and voiced spirants are different. Nevertheless, the phonetic relationship between [f, v] is generally considered to be the same as the phonetic relationship between [s, z], with the tacit assumption that the mode of frication generation is independent of voicing. To my knowledge, there are no existing studies that investigate how frication and voicing interact in a cross-linguistic setting.

This section provides the relevant acoustic and aerodynamic background for understanding frication and voicing. I discuss frication generation in sibilants and non-sibilants, as well as the aerodynamic requirements for generating voicing. I then turn to an explanation of the aerodynamic tensions that underlie the simultaneous generation of frication and voicing, with an overview of previous research done in this area.

### 1.4.1 Generating frication

Fricatives are defined as sounds produced with a turbulent airstream resulting in audible frication (Ladefoged and Maddieson, 1996; Crystal, 2008; Ohala and Solé, 2008). Articulatorily, the generation of turbulent airstream is typically associated with a narrow, but not complete, stricture of the articulators, but the generation of turbulent airflow resulting in audible frication is considered paramount. For example, Crystal defines fricatives as "sounds made when two organs come so close together that the air moving between them produces audible *friction*, or *frication*" (Crystal, 2008, pg. 199) (emphasis in the original). More precisely, however, numerous factors are involved in generating a turbulent airstream required for audible frication.

Aerodynamically, turbulent flow of air through a channel depends on many factors, including factors such as the length, diameter, and shape of the channel and roughness of the channel walls (see Shadle (2010) for review). In the case of speech, the most important factor is the *particle velocity* $v$ of air as it flows through a constriction.[12] Particle velocity is determined in turn by two factors: (1) the *volume velocity* $U$, or the amount of air that passes through a given area per unit of time, and (2) the *cross-sectional area* $A$ of the constriction; the relationship between these variables is given by the equation $v = U/A$ (Ohala and Solé,

[12] The reason for this is that at higher velocities, inertial forces dominate over friction forces. Laminar (or smooth) flow is characterized by dominant friction forces, while turbulent flow is characterized by dominant inertial forces (Shadle, 2010). As Shadle explains, laminar flow through a pipe is characterized by particles flowing side by side without crossing paths, with particles in the middle of the pipe moving fastest, relatively unconstrained by the friction forces of the pipe walls; the particles on the outside thus "hug" the pipe, and this *still flow* is known as the *boundary layer*. As particle velocity increases, the inertial forces involved in particle motion "tend to overshoot" and can disrupt the boundary layer; with even greater particle velocity, inertial forces dominate over the friction forces, and paths of fluid particles cross each other unpredictably, and hence "the flow as a whole has a random fluctuating component superimposed on the mean flow" (Shadle, 2010, pg. 44).

2008). This relationship dictates that the greater the volume of air passing through the constriction, the greater the particle velocity; or, the smaller the cross-sectional area, the greater the particle velocity.[13]

The conditions required to cross the threshold from laminar to turbulent flow are quantified by a term known as the *Reynolds number Re*.[14] While the Reynolds number provides a precise quantification of the conditions required for turbulent flow in general, the application of the Reynolds number is less clear for the aerodynamic conditions that characterize the irregularly shaped vocal tract (Ohala and Solé, 2008). Thus, as Ohala and Solé explain, for the purposes of speech, the relationship between air velocity and noise generation are adequately captured qualitatively: "the intensity (i.e., loudness) and center frequency (i.e., pitch) of frication noise varies monotonically with the particle velocity of the air flow. . . a variant of this relation [is that the] "intensity of frication at a supraglottal constriction increases with increasing oral pressure and decreasing aperture of constriction" (Ohala and Solé, 2008, pg. 300).

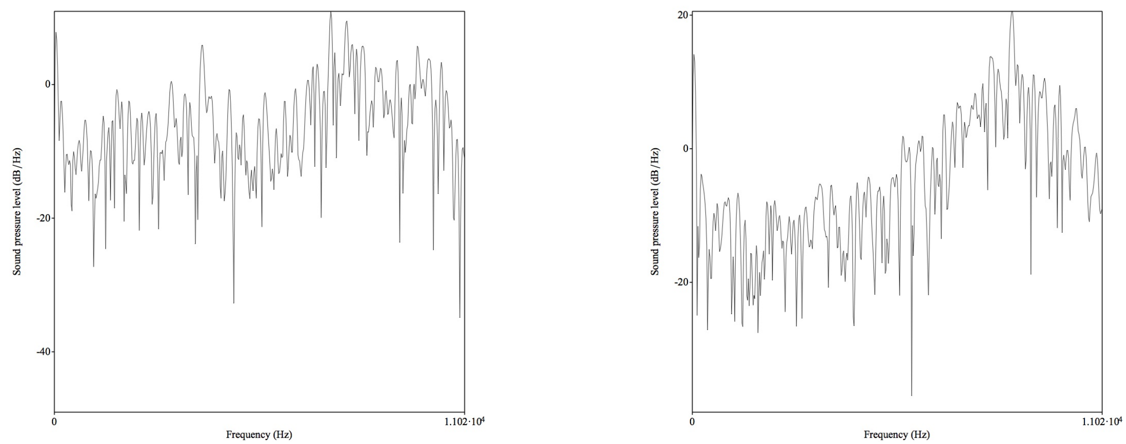
In speech, sounds characterized by turbulent noise generated strictly at the locus of constriction (*channel turbulence*) is only found at the bilabial place of articulation for fricatives [ɸ, β], and possibly at the glottis for [h] (Johnson, 2012, pg. 156). All other fricatives involve the jet of air hitting a wall or an obstacle. For back fricatives such as [ç, x], the jet of air hits

[13] The relationship between airflow and constriction degree involved in the generation of noise can be heard by starting with the articulatory configuration for [u], and steadily increasing the degree of labial constriction while maintaining a steady airflow; this quickly results in something akin to [β] being produced. Conversely, for a given constriction, a greater volume velocity will result in increased turbulence. This can be heard by starting with the articulatory configuration required for [f], and holding it until one runs out of air; over time, as the volume velocity decreases, the frication does as well, until what is being produced is a (voiceless) frictionless, labiodental continuant, (e.g., [f̞] or [ʋ̥]).

[14] "The reynolds number is defined as the ratio of inertial to viscous forces, and can be determined by: $Re = VD/v$, where $V$ = a characteristic velocity, $D$ = a characteristic dimension, and $v$ = the kinematic viscosity. For pipe flow, the $V$ normally used is the average particle velocity in the center of the pipe (and, because of the averaging, is therefore typically capitalized in the literature, confusing it with volume) and $D$ is the pipe diameter" (Shadle, 2010, pg. 43).

the palate at an oblique angle, and so the noise source is distributed along the hard palate; for [s, ʃ], the jet stream hits the teeth at approximately a right angle, and the noise source is localized at the teeth (Shadle, 2010, pg. 52-53). For labiodental fricatives such as [f, v], the jet of air hits the upper lip; manually pulling the upper lip away during articulation of [f] decreases the amplitude considerably (Johnson, 2012, pg. 156). Similarly, pulling the lips away during articulation of interdentals [θ, ð] illustrates the importance of the lips as a source of noise generation. As Shadle (2010) and Johnson (2012) explain, the distinction between fricatives such as [ç, x] and [s, ʃ] is not that the former involve only *channel turbulence* while the latter also involve *obstacle turbulence*; rather what is at issue is the angle at which the jet stream strikes the obstacle, and the resulting distribution or localization of the noise source.

Acoustically, the consequence of turbulent airflow is aperiodic energy, or noise, and place of articulation effects are manifested in how the aperiodic energy is distributed throughout the frequency domain. For example, [f] is characterized by comparatively weak turbulence in comparison to [s], manifested as lower amplitude in the waveform. Further, because [f] is articulated so far forward in the mouth, the noise source is subject to very little vocal tract filtering, yielding the fairly flat distribution of noise evidenced in the spectrum seen in Figure 1.1a; in contrast, [s] the effect of the vocal tract as filter can be seen in the spectral peak of [s] (at approximately 8000Hz), in Figure 1.1b.




(a) Spectrum of [f]    (b) Spectrum of [s]

Figure 1.1: 40ms spectra taken from middle of consonant

### 1.4.2 Generating voicing

In order for voicing to occur, two conditions must be met: first, there must be a pressure gradient across the glottis, thereby ensuring sufficient flow of air; second, the vocal folds must be held in the appropriate configuration to allow for vibration (Ohala, 1983). In modal voicing, as air flows from the lungs and up the trachea, pressure builds up below the lightly adducted vocal folds until such pressure is sufficient to push apart the vocal folds. Due to the narrowing of the passage, the air passes through the glottis at a higher velocity (the *Venturi effect*), causing a pressure drop perpendicular to direction of flow; this drop in pressure, together with the natural elasticity of the vocal folds, causes the folds to snap back into position, an effect known as the *Bernoulli effect*. This is called the *Aerodynamic Myoelastic Theory of Phonation*.

Halle and Stevens (1971) model vocal fold configuration as differing along two linguistically relevant parameters: glottal width, and vocal fold tension. For a given glottal width, voicing can only occur if the tension of the vocal folds is such that it allows for the maintenance of a pressure gradient across the glottis. The lines in Figure 1.2 represent the boundaries of vocal fold oscillations for vocal fold tension under a given glottal width and pressure gradient. Provided that the point specified by glottal width and pressure gradient lies above the boundary curve, vocal fold vibration can occur for the vocal fold tension indicated by the boundary curve. As seen in Figure 1.2, slack vocal folds allow for a greater range of glottal widths, and in general, stiff vocal folds require a smaller glottal width in order for voicing to occur.

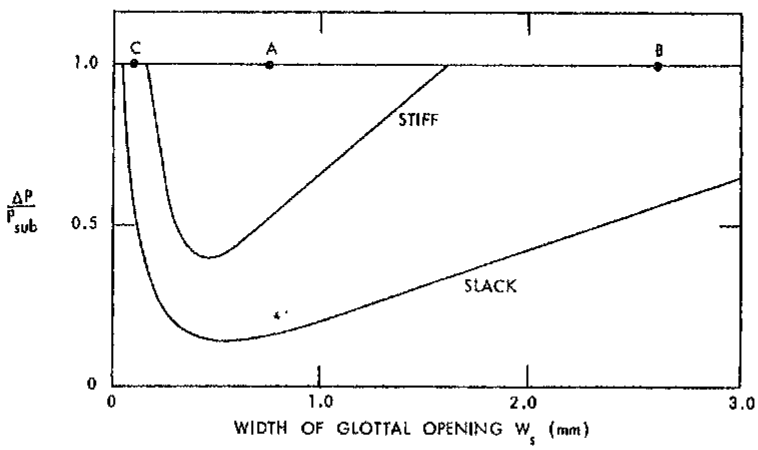



**Fig. 2.** Sketch showing approximate ranges of conditions under which vocal-cord vibration occurs. $\Delta P$ is the pressure across the glottis, $P_{sub}$ is the subglottal pressure, and $w_s$ is the static width that would be assumed by the glottis if there were no vibrations. If the values of $\Delta P$ and $w_s$ give rise to a point above the curve labeled "slack," then vocal-cord vibration is initiated when the vocal-cord stiffness is small. Likewise the curve labeled "stiff" represents the boundary of vocal-cord oscillation for relatively stiff vocal cords. Below these lines, the vocal cords remain in a static position with no oscillations. Points A, B, and C represent glottal widths that lie within regions of "normal" glottal vibration, spread glottis and constricted glottis for nonobstruents, for which $\Delta P = P_{sub}$. The portion of the chart corresponding to obstruent configurations is well below the line $\Delta P/P_{sub} = 1$. The regions are based on an assumed subglottal pressure of $\sim 8$ cm $H_2O$. The shapes of the curves for $w_s > 0.5$ mm are derived from theoretical analysis of a two-mass model of the vocal cords (Ishizaka and Matsudaira[4] and Stevens[2]). For smaller values of $w_s$ the curves are estimated.

Figure 1.2: Conditions for voicing based on pressure gradient across glottis and glottal width (Halle and Stevens, 1971, pg. 47)

Of relevance to the present work is the horizontal line (denoted 1.0 on the $y$-axis of Figure

1.2), where the pressure gradient across the glottis is equal to the subglottal pressure and the supraglottal pressure is equal to 0; this is the aerodynamic state achieved for non-obstruent vowels and glides. Sonorant consonants and vowels differ from obstruents because they do not impede airflow sufficiently to raise the supraglottal pressure: for vowels, glides and liquids, laminar air flow is unimpeded through the vocal tract, and for nasal stop consonants, air flow is vented through the nasal cavity.

Halle and Stevens (1971) propose four binary laryngeal features encoding glottal width and vocal fold tension. For glottal width, the features [spread glottis] and [constricted glottis]; for vocal fold tension, the features [stiff] and [slack]. Excluding the physiologically impossible combinations [+ spread, + constricted] and [+ stiff, + slack], these features yield nine combinations of features that are proposed to account for the range of voicing contrasts attested cross-linguistically, shown in Figure 1.3.

**Table 1. Classification of obstruent, glide, and vowel features in terms of proposed glottalized features.**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| obstruents | $b_l$ | b | p | $p_k$ | $b^h$ | $p^h$ | ɓ | ʔb | $p^ʔ$ |
| glides | w, y | | | | ɦ | h, W, Y | | ʔ | ʔ, ʔw, ʔy |
| vowels | V | V̀ | V́ | voiceless vowels | breathy vowels | | | creaky voice vowels | glottalized vowels |
| spread glottis | − | − | − | + | + | + | − | − | − |
| constricted glottis | − | − | − | − | − | − | + | + | + |
| stiff vocal cords | − | − | + | − | − | + | − | − | + |
| slack vocal cords | − | + | − | − | + | − | − | + | − |

Figure 1.3: Features of laryngeal adjustment (Halle and Stevens, 1971, pg. 51)

Because the focus of the present work is on modal phonation, I focus on the first three columns, which represent, in order, a "lax voiceless stop" [$b_l$], a voiced plosive [b], and a voiceless unaspirated plosive [p]. In terms of glottal width, all three are [−spread] and [−constricted], meaning that the vocal folds are lightly adducted, as appropriate for modal phonation. The voiceless unaspirated stop [p] has [+stiff] (and [−slack]) vocal folds, resulting in voicelessness. Both [$b_l$] and [b] have [−stiff] vocal folds; the fully voiced [b] is specified as having slack vocal folds ([+slack]), and the "lax voiceless stop" is specified to have vocal fold tension that is neither stiff nor slack. Halle and Stevens (1971) note that this segment is reported to occur in Danish, and likely occurs as the configuration for stops in initial position for many speakers of English (pg. 54).

With respect to fricatives, all voiceless fricatives are specified as [+stiff]: aspirated [$s^h$] is [+spread] ([−constricted]), ejective [s'] is [+constricted] ([−spread]), and the plain voiceless [s] is [−spread] and [−constricted], like [p]. However, as pointed out by Vaux (1998), later phonetic work, such as that by Kingston (1990) and Stevens (1991), has shown that at least voiceless fricatives appear to be produced with a spread glottis.

Halle and Stevens (1971) describe voiced fricatives as follows:

> Voiced fricatives are generated with vocal-cord vibration, as well as turbulence noise at a supraglottal constriction, the intraoral pressure being maintained at a value less than the subglottal pressure. Thus implosive fricatives with reduced intraoral pressure, having the features [+ constricted glottis, − stiff], are not possible. The combination [− constricted glottis, + slack] is, however, a possibility, and gives rise to the voiced fricative [z]. To our knowledge, there are no languages that contrast an unaspirated and an aspirated voiced fricative, i.e., columns 2 and 5 in Table 1. The feature combinations in columns 1 and 4 for fricatives are apparently either not utilized or, what is more likely, do not give acoustic outputs that are significantly different from those represented by columns 3 and 6.

With respect to non-obstruents, only two categories are of potential relevance: the glides are specified as [−stiff, −slack] for vocal fold tension, and their glottal width is specified as

Table 1.1: Feature specification for modal voiced and voiceless stops, sibilants, spirants, and approximants based on Halle and Stevens (1971)

| | vocal fold tension | | glottal width | |
|---|---|---|---|---|
| | [stiff] | [slack] | [constricted] | [spread] |
| [f, s, p] | + | − | − | − |
| [v, z, b] | − | + | − | − |
| [ʋ, w, j, $b_l$] | − | − | − | − |
| [ʋ̥, w̥, j̥] | + | − | − | + |

for modal phonation, [− constricted, − spread]; voiceless glides differ only in being specified as [+spread].

Assuming that [ʋ] is classified with the glides [j, w] and that [ʋ̥] is classified with the voiceless glides, the laryngeal features associated with the segments relevant to the present concerns are given Table 1.1. Under this framework, the distinction between [v], [ʋ] is that the production of the voiced fricative [v] involves additional slackening of the vocal folds, and the distinction between [v] and [f] is that [f] involves additional stiffening of the vocal folds. On the other hand, the distinction between the voiced approximant [ʋ] and the voiceless approximant [ʋ̥] is that in addition to the stiffening of the vocal folds required for voicelessness, the vocal folds are spread.

The framework of Halle and Stevens (1971) specifies all voicing (across both obstruents and nonobstruents) in terms of laryngeal configuration. For voiced fricatives, however, voicing or frication can be compromised if there is insufficient airflow, either through the glottis (compromising voicing) or across the supralaryngeal constriction (compromising frication). I address these issues in the following section (Section 1.4.3).

### 1.4.3 Production of voiced fricatives

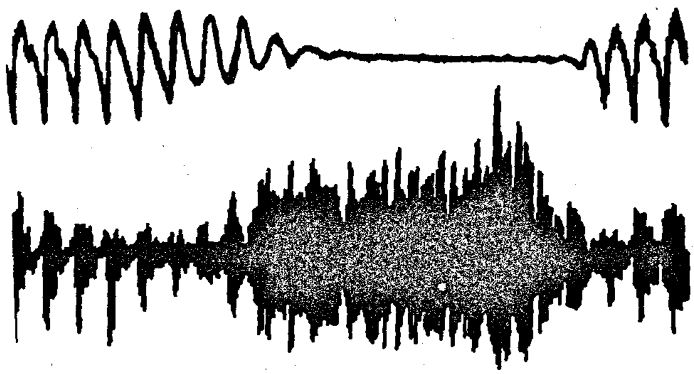
*A priori*, producing a sound with simultaneous voicing and frication presents significant aerodynamic challenges: the transglottal pressure drop must be maintained for voicing to occur, and the pressure drop across the constriction must be maintained in order to produce frication (Ohala, 1983). As a result of this delicate balancing act, small changes in various parameters can have drastic effects on the outcome. For example, the articulatory configuration for voiced approximants, laterals, and nasals can result in a high degree of frication if the airflow is not modulated at the glottis through voicing: in other words, if the vocal folds are positioned for voicelessness, then oral articulatory configurations that would normally result in voiced sonorants will yield frication if produced with sufficiently high airflow (Ohala and Solé, 2008). Similarly, for a given articulatory configuration, a decrease in airflow across the location of constriction can result in a loss of frication. The moral of these phonetic facts is that the divisions between [v] and [f], and [v] and [ʋ] can be due to laryngeal configuration, articulatory configuration, or airflow.

Within the class of voiced fricatives, the voiced spirants and voiced sibilants have different aerodynamic tensions to resolve, and because of this, the failure to achieve both voicing and frication tends to be realized differently between the two manners of articulation. Voiced spirants can easily be realized as sonorant approximants, because the generation of noise is less efficient. In sibilants the primary locus of noise generation is at the teeth because the airstream is focussed into a directed jet that hits the teeth at a right angle; for spirants, the obstruction downstream the locus of constriction is not as efficient at generating noise because the airstream is spread more diffusely. The result is that for voiced spirants, for a given constriction degree, if air flow is sufficiently reduced, it will be laminar, not turbulent, and so not generate noise. For a given rate of flow, slight widening of the constriction also leads to laminar flow, again resulting in an approximant realization. In contrast, for voiced sibilants such as [z, ʒ], reducing the flow of air across the constriction does not appear to lead to laminar flow, presumably due to the channelling of the airstream. In other words, voiced sibilants tend to devoice, rather than than sonorantize, as discussed by Smith (1997) (presumably the last sentence refers to the simplification of voiced sibilants, not spirants):

> At least for English, it appears that simplification of voiced fricatives into voiced approximants is limited to non-sibilant fricatives. There do not seem to be any reports in the literature of voiced sibilants being simplified into voiced approximants, but simplification in the form of a loss of voicing is quite common (Haggard, 1978; Docherty, 1992). This asymmetry suggests that speakers are simplifying voiced fricatives by allowing oral pressure to rise and/or subglottal pressure to fall, not by widening the oral constriction and allowing oral pressure to fall. (pg. 472)

In fact, not only is the loss of voicing quite common, but Haggard (1978) discusses the *dynamic* nature of voiced fricatives. The standard conception of a voiced fricative is the simultaneous generation of voicing and frication, as I laid out above; in this model, frication is modulated by voicing. However, as Haggard (1978) discusses, this conception is a simplification that disregards the dynamic and temporal nature of the interaction between voicing and frication, shown in the oscillogram and accelerometer output in Figure 1.4.

Importantly, as can be read in the caption, the base line of the accelerometer output raises during the voiceless portion (the flat line), which is interpreted to be a raising of the larynx. That voiceless sounds are produced with a raised larynx and voiced sounds with a lowered larynx, known as the *vertical tension hypothesis* (in the context of tonogenesis and CV interactions of consonant phonation and F0 on the following vowel), is discussed in Hombert et al. (1979). The implication is that voiced [z] does not simply devoice because of aerodynamic tensions, but rather because it has two phases, voiced and fricated, that are synchronised to a greater or lesser degree:



**Oscillogram of the intervocalic voiced fricative in /a z a/ (lower trace). The upper trace is the output of an accelerometer on the body wall below the larynx. Voicing ceases during the most intense friction and the larynx moves as witnessed by the shift in baseline of the accelerometer trace. Friction is only voice modulated near onset and offset.**

Figure 1.4: Oscillogram and accelerometer output of [z] (Haggard, 1978, pg. 96)

> In the lower (microphone) trace amplitude modulation of the friction is only appreciable for a few cycles at its onset and offset. In this example, most of the friction is *not* voice-modulated because there is no delicately maintained balance of constriction with sub-glottal pressure to maintain the pressure-drops at *both* glottis and fricative constrictions. The upper trace recording was taken below the thyroid cartilage with an accelerometer, and shows that glottal vibration and friction can be complementary in time and amplitude. The point where the glottal trace ceases is here defined as the point of de-voicing, and the major friction comes *after* this point. This example suggests that in fluent speech, voiced fricative production is a dynamic trajectory through a voiced phase and a fricative phase, with little effort given to the synchronization of these phases as might occur in the imitation of the sound of a bee.

Haggard (1978) investigates the percentage devoicing of voiced fricatives in British English, and the results are shown in Figure 1.5; I focus on [v, z]. In word final position and before a voiceless obstruent, [v, z] are produced as basically voiceless, with upwards of 90% devoicing. In all remaining positions, [z] exhibits more devoicing than [v], and Haggard notes that "it has to be noted that /v/ and /ð/ often do not devoice or produce any appreciable friction" (pg. 97), suggesting that even in English, /v/ can be produced as an approximant.

**Table I Results of experiment 1–32 subjects**

| Mean percentage devoicing in various conditions: | v | d | z | ʒ | dʒ |
|---|---|---|---|---|---|
| *Word-initial* | | | | | |
| Before stressed vowel | 21 | 17 | 34 | | 80 |
| Before unstressed vowel | 24 | 14 | | | 89 |
| *Word-final* | | | | | |
| After stressed vowel | 95 | | 99 | 100 | 99 |
| After unstressed vowel | 92 | | 99 | | 100 |
| *Intervocalic* | | | | | |
| Before stressed vowel | 23 | 9 | 30 | | 68 |
| Before unstressed vowel | 8 | 18 | 39 | 29 | 61 |
| Between stressed vowels | 37 | 27 | 90 | | |
| *After voiced stop* | | | | | |
| Before unstressed vowel | 40 | 40 | 83 | | |
| *Before voiceless stop* | | | | | |
| After stressed vowel | 100 | | 100 | | |
| After unstressed vowel | 100 | | 100 | | |

Figure 1.5: Percentage devoicing of English voiced fricatives; 'd' represents [ð] (Haggard, 1978, pg. 97)

The generation of simultaneous voicing and frication is indeed aerodynamically challenging, but these results show that the aerodynamic tensions are resolved differently for voiced sibilants and spirants. Voiced sibilants consistently devoice for some portion of the fricative, but this is a result of the dynamic interplay between voicing and sibilant frication. Voiced spirants /v, ð/ do not tend to devoice as much, either in terms of frequency or proportion, and even in English can be realized with little to no frication.[15]

### 1.4.4 Exploring the relationship between voicing and frication

The above discussion on the relationship between voicing and the kind of frication generated (sibilant vs. non-sibilant), was based on aerodynamic considerations, and on results of investigations on English voiced fricatives. Within the acoustic phonetic literature on

[15] It is not clear whether voiced spirants, when they do devoice, exhibit the same larynx raising effect. This is left for future research.

fricatives, much of the work has focussed on English; some recent exceptions relevant to the present work are Jesus and Shadle (2002) on Portuguese and Nirgianaki (2014) on Greek. Various measures have been shown to reliably discriminate fricatives within a given language. For example, duration reliably distinguishes voiced from voiceless fricatives, with voiced fricatives being shorter than their voiceless counterparts, and spectral moments reliably distinguish place of articulation.

Although the acoustic phonetic literature on fricatives is vast, it has primarily been concerned with fricative discrimination, namely identifying the acoustic parameters that reliably distinguish fricatives within a given language. To my knowledge, none of these studies have considered these questions from a cross-linguistic perspective, nor have they considered how measures that discriminate fricatives within a language compare across languages. Chapter 2 explores the relationship between frication type and voicing in English, Greek, Russian, and Serbian. The goal of Chapter 2 is to explore how phonetic parameters that distinguish voiced from voiceless, and sibilant from spirant frication are instantiated across different languages. Four segments common to all four languages are included in the analysis of each language, [f, v, s, z], which represent the phonetic parameters of voicing (voiced vs. voiceless), and frication type (spirants vs. sibilants). For each language I consider two environments, one that is prosodically strong—word-initial, onset of a stressed syllable—and another that is prosodically weak—intervocalic word-medial, onset of an unstressed syllable. These two environments are used to assess the effect of environment on fricative production, particularly voiced fricatives [v, z].

## 1.5 Cross-linguistic acoustic study of [v]

Following the exploratory acoustic analysis of [f, v, s, z] in Chapter 2, I turn in Chapter 3 to testing the hypothesis that the phonological identity of /v/ correlates with its phonetic identity. I use the same data set as in Chapter 2, but restrict my attention to Greek, Russian, and Serbian.

The study is motivated by the hypothesis articulated in Padgett (2002), that the phonological intermediacy of Russian /v/ is due to its phonetic intermediacy. In particular, Padgett (2002) hypothesizes that in a language where /v/ patterns as an obstruent, it will be realized as a true fricative [v], and in a language where /v/ patterns as a sonorant, it will be realized as a true approximant [ʋ]. I detail the operationalization of this hypothesis in Chapter 3, in terms of phonetic measures, environments tested, and statistical model. In the remainder of this section, I provide evidence for the phonological classification of /v/ in Greek, Russian, and Serbian. I present evidence for the classification of /v/ as either [−sonorant] or [+sonorant] in each language by considering its involvement in phonological processes and its static distribution.

### 1.5.1 Classification of Greek /v/

Greek has a five vowel system /i, e, a, o, u/; vowel length is not contrastive. Only vowels may be nucleic. The consonant inventory of Greek is given in Table 1.2. Voiceless stops are unaspirated, and voiced stops are fully voiced in all positions; Greek does not have final devoicing. Intervocalically, voiced stops are sometimes prenasalized as [$^{m}$b, $^{n}$d, $^{ŋ}$g], but Arvaniti (2007) describes prenasalization as depending on sociolinguistic factors like age, dialect, rate of speech and formality; Arvaniti also notes that in fast and casual speech, the voiced stops can lenite to voiced fricatives.

Table 1.2: Greek consonant inventory

| | Labial | Interdental | Alveolar | Velar |
|---|---|---|---|---|
| Stop | p b | | t d | k g |
| Fricative | f v | θ ð | s z | x ɣ |
| Affricates | | | t͡s d͡z | |
| Nasal | m | | n | |
| Lateral | | | l | |
| Rhotic | | | r | |

Table 1.3 shows the distribution of Greek consonants in word-initial two-consonant clusters; rows show the first consonant, and columns show the second consonant.[16] Greek allows sequences of obstruent-obstruent, with stop-stop, stop-fricative, fricative-stop, and fricative-fricative clusters all attested; the sonorant cluster [mn] is also allowed. Clusters with obstruents followed by sonorants are allowed, but clusters of the form sonorant-obstruent (i.e., sonority reversals) are not. Clusters with two obstruents must agree in voicing. Greek /v/ clearly distributes with the other obstruents, and exhibits parallel distribution to /f/.

[16]I treat [ts, dz] as single segments and not as clusters, but see Arvaniti (2007) for discussion on this point. Neither can enter into clusters, and so are omitted from Table 1.3.

Table 1.3: Word initial CC clusters in Greek

| | p | b | t | d | k | g | f | v | θ | ð | s | z | x | ɣ | m | n | l | r |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p | | | pt | | | | | | | | ps | | | | | pn | pl | pr |
| b | | | | | | | | | | | | | | | | | bl | br |
| t | | | | | | | | | | | | | | | tm | | | tr |
| d | | | | | | | | | | | | | | | | | | dr |
| k | | | kt | | | | | | | ks | | | | | | kn | kl | kr |
| g | | | | | | | | | | | | | | | | | (gl) | gr |
| f | | | ft | | | | | | fθ | | fx | | | | | | fl | fr |
| v | | | | | | | | | | vð | | | | vɣ | | | vl | vr |
| θ | | | | | | | | | | | | | | | | θn | θl | θr |
| ð | | | | | | | | | | | | | | | | | | ðr |
| x | | | xt | | | | | | xθ | | | | | | | xn | xl | xr |
| ɣ | | | | | | | | | | ɣð | | | | | | ɣn | ɣl | ɣr |
| s | sp | | st | | sk | | sf | | | | sx | | | | | | | |
| z | | | | | | | | zv | | | | | | zɣ | zm | | | |
| m | | | | | | | | | | | | | | | | mn | | |
| n | | | | | | | | | | | | | | | | | | |
| l | | | | | | | | | | | | | | | | | | |
| r | | | | | | | | | | | | | | | | | | |

**A note about Greek consonant clusters** The existence of stop-stop and fricative-fricative clusters require some comment, specifically with respect to clusters such as /fθ/ and /vð/, and whether they have the same status in Modern Greek phonology.

Until the late 1970s there were two varieties of Greek: a high variety, *katharevousa* (literally, "puristic") and a low variety, *dimotiki* ("colloquial", or "demotic") (Joseph and Philippaki-Warburton, 1987). The two varieties coexisted until language reforms in 1976, when dimotiki was declared the official language. Dimotiki disallows consonant clusters in which both segments share the same manner of articulation; hence φϑάνω [fθanɔ] (katharevousa) becomes φτάνω [ftanɔ] 'I complete an action'. Manner dissimilation did not exist in the katharevousa variety, and so clusters such as [fθ, kt, xθ,. . . ] are typically found in words of katharevousa origin (Joseph and Philippaki-Warburton, 1987). Furthermore, many words have two forms, one dimotiki and one katharevousa, as shown in (3).

(3) Examples of manner dissimilation in dimotiki

| Dimotiki | Katharevousa | Gloss |
|---|---|---|
| [ftinos] | [fθinos] | ‘cheap’ |
| [xtes] | [xθes] | ‘yesterday’ |
| [efta] | [epta] | ‘seven’ |
| [xtima] | [ktima] | ‘farm’ |
| [pextis] | [pektis] | ‘player’ |
| [askimos] | [asximos] | ‘ugly’ |
| [pistike] | [pisθike] | ‘was persuaded’ |

In response to this situation, some linguists have treated the two forms of the language separately; for example, Morelli (1999) considers all fricative-fricative clusters as being part of a katharevousa stratum, and peripheral to the core. This analysis faces a number of problems. First, there exists at least one minimal pair that contrasts [ft] with [fθ]: [ftino] ‘I spit’ vs. [fθino] ‘I wither’.[17] Second, manner dissimilation applies within the root and in inflectional morphology, but not in the derivational forms such as /ef-θanasia/ → [efθanasia], *[eftanasia] ‘euthanasia’ Holton et al. (2004).[18] Third, voiced fricative clusters are not subject to manner dissimilation, thus clusters /vð, vɣ, ɣð, zv, zɣ/ are all attested and part of the dimotiki lexicon.[19] An interesting example comes from the two forms for the word

[17]Thanks to Effi Georgala for alerting me to this example.

[18]The preservation of consonant clusters with identical manner specifications may derive from the possible (though not necessary) heterosyllabicity of such consonant clusters over morpheme boundaries.

[19]Cypriot Greek and some southeastern dialects are a counterexample, where sequences /vɣ, vð/ are realized as [vg, vd] (Newton, 1972).

'week': in dimotiki, the word is [vðomaða], while the katharevousa form is [evðomaða]; the dimotiki form has the voiced fricative consonant cluster word-initially, while the katharevousa form allows for a possible heterosyllabification of the cluster.[20]

More generally, the problem with treating dimotiki and katharevousa separately is that they are not in a true diglossic situation. In true cases of diglossia, such as Egyptian vs. Classical Arabic, the two varieties are relatively stable, established both diachronically and synchronically Ferguson (1959). As Alexiou (1982) argues, this is not the case for Greek:

> [K]atharevousa has no standardised form because there are so many degrees of archaism and purism. Its grammars therefore... reflect neither consistent linguistic rules nor common usage. It was never a spoken language of any historical period, nor is it the vehicle of a large and respected body of written literature, either past or present. It is an artificial compromise between archaism and colloquialism.... As the writer Nikos Kazantzakis said at the beginning of [last] century, 'Greece has acquired not one but hundreds of katharevousas.' pg. 158

Greek has distinguished between high and low varieties of the language since at least Medieval times: "This coexistence has led to considerable intermingling of the two varieties, so that present-day dimotiki as used by most Greeks contains a number of katharevousa elements" (Joseph and Philippaki-Warburton, 1987, 2). The katharevousa forms are so integrated that a child has words such as [ktima, fθinoporo, periptero] ("land, autumn, kiosk") alongside words such as [ftero, eftixos, xteni] ("wing, lucky, comb") as part of their primary linguistic data (Philippaki-Warburton, 1980). To summarize,

[20]Thanks to Brian Joseph for pointing this example out to me, and for confirming my intuitions that there is nothing inherently "katharevousa" about these clusters.

In terms of general phonotactics, Greek does not exhibit long strings of obstruent consonants. Many syllables are open, and codas are generally restricted to sonorants; native words can only end in vowels, [s, n]. Greek exhibits regressive voicing of the final /s/ in (4). Voicing is triggered by all voiced consonants (both obstruents and sonorants), but Baltazani (2007) shows that voicing triggered by voiced stops is much greater than voicing triggered by sonorants; voiced fricatives were not included in the study, so it is not known if they exhibit the same degree of voicing as stops or sibilants, or an intermediate degree.

(4) Greek /s/-voicing

| | | |
|---|---|---|
| a) /tous barbaðes/ | [touz barbaðes] | 'the uncles (acc.)' |
| b) /tis ðino/ | [tiz ðino] | 'I give her' |
| c) /tis varvaras/ | [tiz varvaras] | 'Barbara's' |
| d) /tis mamas/ | [tiz mamas] | 'the mother's' |

Within words, Greek exhibits voicing agreement, as in [evɣlotos] 'eloquent' and [efstaθia] 'steadiness' (same prefix). In sum, although sonorants also trigger voicing assimilation, the overall facts about Greek consonant distributions shows that Greek /v/ is classified as an obstruent.

### 1.5.2 Classification of Serbian /v/

Serbian has a five vowel system /i, e, a, o, u/; vowel length is contrastive. The rhotic /r/ may be nucleic. Voiced stops are fully voiced in all positions; Serbian does not have final devoicing. Voiceless stops are unaspirated.

Table 1.4 shows the distribution of Serbian consonants in word-initial two-consonant

Table 1.4: Word initial CC clusters in Serbian

| | Labial | Alveolar | Palatal | | Velar |
|---|---|---|---|---|---|
| Stop | p b | t d | | | k g |
| Fricative | f | s z | | ʃ̺ ʒ̺ | x |
| Affricates | | t͡s̻ | t͡ʃ̻ d͡ʒ̻ | t͡ʃ̺ d͡ʒ̺ | |
| Nasal | m | n | | ɲ | |
| Lateral | | l | | ʎ | |
| Rhotic | | r | | | |
| Approximant | v | | | j | |

clusters.[21] To save space, I do not include the affricates or the palatals /ɲ, ʎ/ in the table.[22]

Serbian allows sequences of obstruent-obstruent, with stop-stop, stop-fricative, fricative-stop, and fricative-fricative clusters all attested; sonorant-sonorant clusters [mn, ml, mr] are also allowed. Clusters with obstruents followed by sonorants are allowed, but clusters of the form sonorant-obstruent (i.e., sonority reversals) are not. Clusters with two obstruents must agree in voicing. Serbian /v/ can appear after voiced and voiceless obstruents, like sonorants; it cannot only come first in a cluster when followed by sonorants (liquids and, marginally, the glide /j/).

Obstruent clusters in Serbian must agree in voicing, so that both voiced and voiceless obstruents trigger RVA; /v/ does not trigger RVA, nor is it a target of RVA.

(5) Serbian RVA across morpheme boundaries

[21] Thanks to Draga Zec for providing me with this chart.

[22] Affricates, like stops and fricatives, must agree in voicing with other obstruents, though voiceless affricates can precede /v/; affricates can appear as the second member after /s/, and some can appear before /m, l, r/.

| | p | b | t | d | k | g | f | v | s | z | ʃ | ʒ | x | m | n | l | r | j |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p | | | pt | | | | | | ps | | pʃ | | | | pn | pl | pr | |
| b | | | | bd | | | | | | | | | | | | bl | br | |
| t | | | | | tk | | | tv | | | | | | tm | | tl | tr | |
| d | | | | | | | | dv | | | | | | | dn | dl | dr | |
| k | | | kt | | | | | kv | | | | | | km | kn | kl | kr | |
| g | | | | gd | | | | gv | | | | | | gm | gn | gl | gr | |
| f | fp | | ft | | | | | | | | | | | | | fl | fr | fj |
| v | | vb | | | | | | | | | | | | | | vl | vr | (vj) |
| s | sp | | st | | | | sf | sv | | | | | sx | sm | sn | sl | sr | sj |
| z | | zb | | zd | | | | zv | | | | | | zm | zn | zl | zr | zj |
| ʃ | ʃp | | ʃt | | | | | | | | | | | ʃm | ʃn | ʃl | ʃr | |
| ʒ | | ʒb | | ʒd | | | | | | | | | | ʒm | | ʒl | | |
| x | | | xt | | | | | xv | | | | | | xm | | xl | xr | |
| m | | | | | | | | | | | | | | | mn | ml | mr | (mj) |
| n | | | | | | | | | | | | | | | | | | |
| l | | | | | | | | | | | | | | | | | | |
| r | | | | | | | | | | | | | | | | | | |
| j | | | | | | | | | | | | | | | | | | |

a) /s ložiti/ [složiti] ‘to assemble’

b) /s gaziti/ [zgaziti] ‘to trample’

(6) Serbian RVA within morphemes

a) [redak] ‘a line (nom.sg.)’

b) [retka] ‘a line (gen.sg.)’

(7) Non-participation of /v/ in RVA

a) [ovca] ‘sheep’

b) [tvoj] ‘your’

In sum, based on phonotactic distribution and its lack of participation in regressive voicing assimilation, Serbian /v/ patterns with sonorants.

### 1.5.3 Classification of Russian /v/

Russian has a five vowel system /i, e, a, o, u/; vowel length is not contrastive. The consonant inventory of Russian is provided in Table 1.5.

Table 1.5: Consonant inventory of Russian (Kavitskaya, 1998)

| | Labial | Dental | Palato-Alveolar | Velar |
|---|---|---|---|---|
| Stop | p b<br>$p^j$ $b^j$ | t d<br>$t^j$ $d^j$ | | k g<br>($k^j$) ($g^j$) |
| Affricate | | ts | $tʃ^j$ | |
| Fricative | f v<br>$f^j$ $v^j$ | s z<br>$s^j$ $z^j$ | ʃ ʒ<br>$ʃ^j$ | x<br>($x^j$) |
| Nasal | m<br>$m^j$ | n<br>$n^j$ | | |
| Lateral | | l<br>$l^j$ | | |
| Rhotic | | r<br>$r^j$ | | |
| Glide | | | j | |

The possible word-initial two-consonant clusters are provided in Table 1.6; in order to save space, the palatalized consonants are omitted from this chart, as well as the affricates. In general, nasals and liquids can only appear as the second member of a cluster, with the exception of /m/, which can appear in clusters before /l, r, j/. Stops and fricatives can appear in first or second position, but can only appear in second position if followed by a stop or fricative, and must agree in voicing. Stops and fricatives readily appear as first member if a nasal, liquid, or glide is the second member. /v/ can appear in both first and second position, and exhibits the distribution of both an obstruent and a sonorant. When it is in first position, it only precedes voiced stops and fricatives, and all sonorants; when it is in second position, it does not have to agree in voicing with the preceding obstruent. With respect to word-initial consonant clusters, /v/ distributes both as an obstruent, and as a sonorant.

Table 1.6: Word-initial clusters in Russian

| | p | b | t | d | k | g | f | v | s | z | ʃ | ʒ | x | m | n | l | r | j |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p | | | pt | | | | | | ps | | pʃ | | | | pn | pl | pr | pj |
| b | | | | bd | | | | | | | | | | | | bl | br | bj |
| t | | | | | tk | | | tv | | | | | | | | tl | tr | tj |
| d | | | | | | | | dv | | | | | | | dn | dl | dr | dj |
| k | kp | | kt | | | | | kv | ks | | | | | km | kn | kl | kr | |
| g | | gb | | gd | | | | gv | | gz | | | | | gn | gl | gr | |
| f | | | ft | | fk | | | | fs | | | | | | | fl | fr | fj |
| v | | | | vd | | | | | | vz | | vʒ | | vm | vn | vl | vr | vj |
| s | sp | | st | | | | sf | sv | | | | | sx | sm | sn | sl | sr | sj |
| z | | zb | | zd | | zg | | zv | | | | | | zm | zn | zl | zr | zj |
| ʃ | | | | | | | | | | | | | | ʃm | | | | |
| ʒ | | | | | | | | | | | | | | | | | | |
| x | | | | | | | | xv | | | | | | | | xl | xr | |
| m | | | | | | | | | | | | | | | | ml | mr | mj |
| n | | | | | | | | | | | | | | | | | | nj |
| l | | | | | | | | | | | | | | | | | | lj |
| r | | | | | | | | | | | | | | | | | | |
| j | | | | | | | | | | | | | | | | | | |

Padgett (2002) notes that clusters with [v] as second member are quite common, identifying "common" clusters as those that appeared in a Russian dictionary (Ozhegov, 1984) at least three times. In longer clusters of three consonants, "the third member was *always* [n, l, r, j] or [v], e.g., [skvaʒina] 'chink'" (pg. 17). The common clusters are given in (8). The significance of listing the common clusters is that clusters with /v/ as second member are not just attested, but a core part of the system, and so cannot be explained away as marginal.

(8) Common word-initial two-consonant clusters; adapted from Padgett (2002, 16)[23]

| | | | | | |
|---|---|---|---|---|---|
| pr | tr | kr | br | dr | gr |
| pl | | kl | bl | dl | gl |
| | tm | | | | |
| pn | | kn | | | gn |
| | tv | kv | | dv | gv |
| pj | | | | dj | |
| ps | tʃʲ | ks | | | |

Russian is famous for its typologically unusual phonotactics, allowing up to four consonants in the onset ([fskrɨl] 'he opened') and coda ([tʃorstf] 'stale'), sonorant-obstruent sequences in onsets ([rtutʲ] 'mercury') and obstruent-sonorant sequences in codas ([ʒezl] 'staff') (Proctor, 2009, pg. 126).[24]

The labiodentals /f, v/ play a special role in licensing three- and four-consonant initial clusters (Proctor, 2009). For three-consonant clusters, the only segments that can begin a three-consonant cluster are /s, z, v, f, m, t, k/; however, /t/ is only found in the cluster [tkn] (6 instances) and /k/ in [kst] (1 instance). Four-consonant clusters are even more constrained: they cannot contain nasals, and they must be of the form fricative-fricative-stop-liquid ([vzbros] 'upthrust', [fsplesk] 'splash') (Proctor, 2009, pg. 131).

While the voicing assimilation patterning of Russian /v/ is the most famous, other data bears on its partial sonorant classification. For example, (9) shows that the final consonant of verb stems is deleted before another consonant if that consonant is [j, v, n, m].

[23]The symbol tʃʲ does not represent the affricate.

[24]Such clusters have been been traditionally viewed as counterexamples to the Sonority Sequencing Principle (SSP), but Proctor (2009) argues that sonority violations are only permitted at word-edges, and argues that word-medially, Russian syllable structure generally follows the SSP. Based on a corpus study based on Sharoff (2002) of Russian two-consonant onset clusters, Proctor further concludes that these clusters are dispreferred and that "the modern lexicon of Russian demonstrates an overwhelming preference for syllable structures in which onsets, and to a lesser extent also codas, conform to typologically-standard sonority sequencing principles" (Proctor, 2009, pg. 130).

(9) Sonorant truncation in Russian (Padgett, 2002, 17)

| | | | |
|---|---|---|---|
| delaj-u-t | stan-u-t | ʒiv-u-t | (3rd plural) |
| dela-l-a | sta-l-a | ʒi-l-a | (Past fem.) |
| ‘do’ | ‘become’ | ‘live’ | |

Given /v/’s robust patterning with sonorants thus far, its participation as a target of obstruent voicing assimilation is all the more surprising. The following description of voicing processes are based primarily on Kavitskaya (1998), Padgett (2002) and Padgett (2011). There are various intricacies with respect to the domain of application of voicing processes; these are mentioned in the examples that follow, but for further details, I refer the reader to Padgett (2011). Further, I adopt Padgett’s position on the non-participation of sonorants in the voicing processes, as whatever devoicing occurs is optional and, when it does occur, gradient at best; I therefore only present those facts which are uncontroversially part of the phonology.[25] Since the palatalized and unpalatalized consonants pattern identically with respect to voicing processes, in what follows I describe the data only for the plain consonants, but all descriptions are to be interpreted to apply also to their palatalized counterparts.[26]

Russian obstruents, but not sonorants, devoice word-finally (10).

(10) Russian final devoicing

| | | |
|---|---|---|
| a) [sled-a] | [slet] | ‘track (gen./nom.sg)’ |
| b) [mil] | *[mil̥] | ‘dear’ |

Devoicing occurs at the end of the prosodic word; in (11), (a) illustrates devoicing of a

[25] According to Maria Gouskova (personal communication), Padgett’s presentation of the data is a reliable representation of the facts of the standard dialect spoken by Russians in Moscow.

[26] All examples in this section are from Padgett (2002).

lexical word before another lexical word, while (b) shows that prepositions do not devoice before lexical words, as they are considered to be part of the same phonological phrase. Example (c) shows that clitics are not incorporated into the prosodic word, as final devoicing occurs.

(11) Domain of final devoicing

| | | |
|---|---|---|
| a) /sad mixaila/ | [sat] [mixaila] | 'Mikhail's garden' |
| b) /pod moskvoj/ | [pod moskvoj] | 'near Moscow' |
| c) /sad li/ | [sat] [li] | 'garden (int.)' |

Regressive voicing assimilation is triggered by all obstruents except /v/; all obstruents including /v/ are targets of RVA.

(12) Russian regressive voicing assimilation

| | | |
|---|---|---|
| a) /ot brosit$^j$/ | [od-brosit$^j$] | 'to throw aside' |
| b) /pod pisat$^j$/ | [pot-pisat$^j$] | 'to sign' |

Padgett (2002) shows that while final devoicing operates at the level of the prosodic word, regressive voicing assimilation crosses prosodic word boundaries, thus /sad/ + /to/ → [sat][to]garden (topic.). Finally, we see in (13) that final devoicing feeds regressive voicing assimilation.

(13) [pojezd-a] [pojest] train (gen./nom.sg)

**Patterning of** /v/ Like obstruents, word-final /v/ devoices, as seen in (14).

(14) [prav-a] [praf] right (fem./masc.)

Unlike obstruents, but like sonorants, /v/ does not trigger regressive voicing assimilation, as seen by the contrast in (15). Interestingly, unlike sonorants, but like obstruents, /v/ is a target for regressive voicing assimilation, devoicing to [f], as in (16).

(15) Russian /v/ is not a trigger for RVA

a. [tver$^j$] 'Tver'

b. [dver$^j$] 'door'

(16) Russian /v/ is a target of RVA

a. [v gorode] 'in the city'

b. [f supe] 'in the soup'

The data in this section show that with respect to its participation in voicing processes, Russian /v/ is intermediate between sonorants and obstruents.

## 1.6 Inventory structure of voicing and frication

One of the criteria for the selection of Greek, Russian, and Serbian as languages of study in the acoustic investigations was that they all have both /v/ and /f/ in their consonant inventories. This criterion was motivated not only by the desire for uniformity across consonant inventories, but also by implicational relations of voicing. Maddieson (1984) identifies three implicational universals with respect to voicing in consonants, given in (17). A universal such as (17a) is read as “the presence of a voiceless sonorant in an inventory implies the presence of a voiced sonorant”.

(17) Implicational universals of voicing in consonants

a. Voiceless sonorants $\Longrightarrow$ voiced sonorants

b. Voiced stops $\Longrightarrow$ voiceless stops

c. Voiced fricatives $\Longrightarrow$ voiceless fricatives

Additionally, the voicing of fricatives has been claimed to follow from the voicing of stops; i.e., the presence of a voicing contrast in the fricatives implies a voicing contrast in the stops (Maddieson, 2010). The articulation of implicational relations between voiced and voiceless stops and fricatives, and between voicing contrasts in fricatives and stops, encodes, in terms of inventory structure, the relationship between voicing and frication, and is the starting point of the investigation in Chapter 4.

The relationship between voicing and consonants can be summarized by saying that the default laryngeal setting for obstruents is voicelessness and the default laryngeal setting for sonorants is voicing. The tension between obstruency and voicing is phonetically grounded, as the aerodynamic conditions required to maintain voicing are compromised by the articulatory configurations required to produce oral stops and fricatives (Ohala, 1983). Furthermore, within the obstruents, voicing in fricatives is viewed as an additional level of complexity over the voicing in stops (Lindblom and Maddieson, 1988).

The generalizations in (17) are derived from the relative frequencies of voiced and voiceless segments in inventories in the Maddieson (1984) version of the UCLA Phonological Segment Inventory Database (UPSID), and the implicational relations have varying degrees of robustness.[27] The relationship between voicing and sonorants articulated in (17a) appears to be a true universal: no language has voiceless sonorants without voiced sonorants, and indeed the contrastive status of voiceless sonorants is controversial (Blevins, 2018).

With respect to stops, the implicational relation is nearly exceptionless. All languages have at least one series of stops, and the most common type of stop series is plain voiceless, with an overall frequency of 91.8% in UPSID. Table 1.7, replicated from Maddieson (1984), displays the prevalence of stop series by the number of stop series in an inventory. Languages have an overwhelming tendency to have plain voiceless stops, regardless of the number of stop series, as seen from the top row. The leftmost column shows that plain voiceless stops are in some sense the "default" stop series, since if a language has only one series of stops, it is almost always the voiceless plain series, attested in 49 of the 50 languages in Maddieson (1984) with a single series.

Table 1.7: Frequency of stop series by number of series; from Maddieson (1984, pg. 28)

| | Number of series | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Plain voiceless | 98.0% | 90.1% | 89.5% | 96.0% |
| Plain voiced | 2.0% | 81.5% | 69.7% | 88.0% |
| Aspirated voiceless | 0.0% | 16.0% | 63.2% | 52.0% |
| Voiceless ejective or voiceless laryngealized | 0.0% | 3.7% | 42.1% | 56.0% |
| Voiced implosive or voiced laryngealized | 0.0% | 1.2% | 27.6% | 48.0% |

Among languages with two series of stops, plain voiced stops are attested in 81.5% of languages, and the majority contrast plain voiceless and voiced stops (72.2%); other contrasts include plain vs. aspirated voiceless, and plain voiced vs. aspirated voiceless, with these two

[27]Maddieson (1984) is based on the UPSID database of 317 languages; the familiar online version, accessible at `http://web.phonetik.uni-frankfurt.de/upsid`, represents an updated version with an additional 134 languages, bringing the total to 451 languages (Maddieson and Precoda, 1990). In this introduction all tables and reported numbers are based on Maddieson (1984) (hence the older version of UPSID).

contrasts providing an additional 16.7% of contrast types for two series. The remainder of the contrasts are between voiceless stops (either plain or aspirated) and one of the other phonation types listed. For languages with three or more series of stops, the pattern is the same: all such languages have at least one series that is voiceless, whether plain or aspirated, and elaborate from there. In sum, the implicational relation that if a language has voiced stops it has voiceless stops is robust: no language with two or more series of stops lack voiceless stops, and Maddieson (1984) identifies only one language in UPSID that is claimed to have a voiced series of stops alone.

The robustness of the implicational relation for voicing in stops is due, in part, to interpretation: namely, that the absence of a single stop at a particular place does not constitute a violation. When single stop consonants are missing from a stop series, they are considered *gaps*. Such is the case in exhibited by the consonant inventory of Chuave (`cjv`), shown in Table 1.8 which constitutes a violation of (17b) if interpreted in the strongest sense, since it has voiced /b/ without its voiceless counterpart /p/. However, (17b) is generally interpreted in a weaker sense: framed in terms of (binary) features, [−voice] is present on /t, d/, [+voice] is present on /b, d, g/, and thus the absence of /p/ is viewed as a gap in the [−voice] inventory.[28]

The implicational relation for fricative voicing articulated in (17c) is not nearly as robust as that for stops and sonorants. Table 1.9, from Maddieson (1984), lists the frequency with which various fricative pairs violate (17c). Specifically, for each voicing pair (e.g., /s, z/, /ɸ, β/), the middle column shows the number of languages that have a given voiced fricative but lack its voiceless counterpart; the final column expresses this proportion as a percentage. There is a remarkable range across fricative pairs, with the number of exceptions ranging

[28] Greenberg (1966) discusses the strength relations of various interpretations of implicational relations, a point taken up again in Section 4.6.

Table 1.8: Chuave (`cjv`) consonant inventory

| | Labial | Coronal | Palatal | Velar | Labio-velar |
|---|---|---|---|---|---|
| Stop | b | t d | | k g | |
| Fricatives | f | s | | | |
| Nasal | m | n | | | |
| Rhotic | | ɾ | | | |
| Approximant | | | j | | w |

from 0%, for /s, z/ to 75% for /ɸ, β/.

Table 1.9: Voiced fricatives without voiceless fricatives, adapted from Maddieson (1984) (some symbols changed to standard IPA); /s, z, ɬ, ɮ/ refer to dental, alveolar or unspecified place of articulation.

| Fricative pair | Unpaired voiced fricative / total voiced fricative | Exceptions as % of cases |
|---|---|---|
| /s, z/ | 0/96 | 0.0% |
| /ħ, ʕ/ | 0/9 | 0.0% |
| /ʃ, ʒ/ | 2/51 | 3.9% |
| /ɬ, ɮ/ | 1/8 | 12.5% |
| /f, v/ | 11/51 | 21.5% |
| /ʂ, ʐ/ | 1/3 | 33.3% |
| /χ, ʁ/ | 5/14 | 35.7% |
| /x, ɣ/ | 15/40 | 37.5% |
| /θ, ð/ | 12/21 | 57.1% |
| /ç, ʝ/ | 5/7 | 71.4% |
| /ɸ, β/ | 24/32 | 75.0% |

In spite of such heterogeneity, there appears to be a clear distinction between sibilant and non-sibilant, i.e., *spirant*, pairs: Table 1.9 shows that sibilants pattern similarly to stops with respect to the implicational relations of voicing, while spirants incur higher rates of violation. Of the 96 languages that have /z/, none lack /s/; of the 51 languages with /ʒ/, only two of them lack /ʃ/; the apparent high rate of violation for the retroflex sibilants /ʂ, ʐ/ at 33.3% is almost certainly due to the sparsity of data, since there are only three languages that have voiced retroflex sibilants. In contrast to the sibilants, the voiced spirants appear without their voiceless counterparts quite frequently. For example, of the 51 inventories with /v/, 11 of them lack /f/; thus the *rate of violation* is 21.5%. The rates of violation for the remaining spirant pairs range from 37% (velars) to 75% (bilabials). Based on Table 1.9, there appears to be a distinction between sibilants and spirants with respect to their adherence to the implicational relations of voicing, where sibilants pattern similarly to stops, and spirants exhibit a much greater tendency to appear *unpaired*.

Unpaired voiced spirants present a challenge to our phonological understanding of segmental inventories. Voicing in obstruents is generally (though at times tacitly) interpreted as an elaboration over the voiceless default state; for example, the terminology in Lindblom and Maddieson (1988) of *basic* and *elaborated* articulations (of which the voiceless and voiced fricatives belong to the former and latter, respectively) is suggestive of this kind of thinking, as are phonological representations that specify voiced segments featurally, and leave voiceless segments unspecified.

Such a perspective is impossible to maintain for a language such as Bannoni (`bcm`), which has unpaired /v, ɣ/: it is not possible to interpret /v, ɣ/ as an elaboration (in terms of voicing) of voiceless /f, x/, since /f, x/ are not in the inventory.

Matters are even worse in Rapanui (`rap`), shown in Table 1.11: since there are no other voiced fricatives or stops, it is not possible to interpret /v/ as sharing the [+voice] feature of other voiced obstruents (with “normal” markedness relations) as was possible for Bannoni, and thereby interpreting the absence of /f/ merely as a gap within the [−voice] inventory.

Table 1.10: Consonant inventory of Bannoni `(bcm)` (Mielke, 2008)

| | Labial | Coronal | Velar |
|---|---|---|---|
| Stop | p b | t d | k g |
| Fricative | v | s | ɣ |
| Nasal | m | n | ŋ |
| Rhotic | | r | |

Table 1.11: Consonant inventory of Rapanui `(rap)` (Mielke, 2008; Du Feu, 1996)

| | Labial | Dental/Alveolar | Velar | Labio-velar | Glottal |
|---|---|---|---|---|---|
| Stop | p | t | k | | ʔ |
| Fricative | v | | | | h |
| Nasal | m | n | ŋ | | |
| Rhotic | | r | | | |

Taking the work of Maddieson (1984) as a starting point, Chapter 4 presents a nuanced picture of the cross-linguistic distribution of voiced spirants, and discusses the findings in light of previous work on segmental inventories. In particular, I explore additional parameters of inventory structure that correlate with the presence of voiced spirants in an inventory, and whether they appear paired or unpaired.

## 1.7 Summary and structure of the dissertation

To summarize, the fundamental difference between obstruents and sonorants is, arguably, their relationship to voicing, a difference that is phonetically grounded, typologically instantiated, and phonologically active. The unequal status of voicing across consonant classes is grounded in aerodynamic constraints: the aerodynamic conditions that favour voicing are inherently at odds with those required for obstruency, namely the generation of plosion or frication (Ohala, 1983). Phonologically, voiced obstruents are often the target of processes or constraints that resolve the tension between obstruency and voicing, exemplified, for example, by word-final obstruent devoicing. Typologically, voiced obstruents are dispreferred in comparison to their voiceless counterparts, a fact typically stated in the form of an implicational relation: if an inventory has a voiced obstruent, it has its voiceless counterpart (Maddieson, 1984).

While there is significant literature supporting the binary partition of segments into those that are [−sonorant] and those that are [+sonorant], the patterning of Russian /v/, and the typological characteristics of the voiced spirants with respect to their distribution in consonant inventories challenge this divide. In the remainder of this dissertation I present a multi-pronged approach to understanding the cross-linguistic identity of /v/.

The dissertation is organized as follows. In Chapter 2 I present the results of a cross-linguistic acoustic study designed to probe the relationship between voicing and type of frication, focussing on segments /f, v, s, z/ in English, Greek, Serbian, and Russian. In Chapter 3 I explicitly test the hypothesis that the phonetic realization of /v/ correlates with phonological identity, using Greek, Serbian, and Russian as an instantiation of a typology of /v/'s phonological identity. In Chapter 4 I present the results of a database investigation of the distribution of /v/ and the other voiced spirants /β, ð, ɣ/ in consonant inventories, specifically probing the relationship between the presence of voiced spirants and (1) inventory size, and (2) the presence of other voiced obstruents. Chapter 5 returns to the problem of /v/ in Russian, in which I propose that the voicing assimilation facts are straightforwardly accounted for by analysing Russian /v/ as [+sonorant, +obstruent]. I also present a typology of laryngeal phenomena from the perspective of /v/'s patterning, illustrating the extent to which traditional perspectives of laryngeal phonology have failed to adequately account for all the data. In Chapter 6 I explore the quantal nature of [sonorant] and [obstruent], and propose a model that combines Quantal Theory (Stevens, 1989) with Emergent Feature Theory (Mielke, 2008). Chapter 7 concludes the dissertation.

CHAPTER 2

# CROSS-LINGUISTIC ACOUSTIC CHARACTERISTICS OF THE RELATIONSHIP BETWEEN FRICATION AND VOICING

## 2.1 Introduction

This chapter presents the results of acoustic investigations of the spirants [f, v] and the sibilants [s, z] in four languages: English, Greek, Russian, and Serbian. These languages were chosen with phonological motivations in mind: the phonological identity of /v/ within each language differs, either in terms of patterning, phonotactics, or local inventory structure. Each language is investigated independently; an explicit cross-linguistic comparison follows in the next chapter.

It is well known that the maintenance of voicing and the generation of turbulent noise, or frication, are antagonistic for aerodynamic reasons (see Section 1.4 for details). Less well understood is how the relationship between voicing and the *type* of frication, whether sibilant or spirant, manifests acoustically, particularly across different languages. The primary goal of these investigations is to determine how [v] is related to other segments in the fricative inventory with respect to phonetic parameters related to type of frication and voicing, as well as the interaction between them. In particular, the investigations presented in this chapter seek to understand how [v] is related to [f], and how this relationship compares to the relationship between [z] and [s].

A tacit assumption in phonological theory is that if a consonant inventory has both [f] and [v], they form a voiceless-voiced pair within the obstruent system, much like sibilants [s, z], or voiceless and voiced stop pairs, such as [p, b], [t, d], and [k, g]. As I demonstrated in Section 1.5.1, this is the case for Greek, where the distribution and patterning of /v/

follows that of other voiced obstruents, and the relationship between /v/ and /f/ parallels that of other voiced-voiceless pairs. In Serbian, however, /v/ is not the phonological voiced counterpart to /f/, either in terms of its patterning or its distribution, as detailed in Section 1.5.2. Russian, detailed in Section 1.5.3, presents an intermediate case, where /v/ patterns with both obstruents and sonorants, thereby exhibiting a partial pairing with /f/. Greek, Russian, and Serbian therefore instantiate a typology with respect to how tightly paired /v/ is with /f/.

The intuition that /v/ is *paired* with /f/ (to a greater or lesser degree) is translated into phonetic terms by assessing whether the acoustic relationship between [v] and [f] mirrors that of [z] and [s], within a given language. In each language, /s, z/ are uncontroversially phonological obstruents, and /z/ patterns as the voiced counterpart of /s/. Phonetically, [s] and [z] are both realized with high frequency noise, though the noise component of [z] is modulated due to the presence of voicing.

An important feature of the consonant inventories of Greek, Russian, and Serbian is that none of these inventories include /w/ or any other labial approximant against which /v/ contrasts. English is included because it differs in precisely this way, where in addition to /f, v/, English has /w/. In Greek, the phonological identity of /v/ as an obstruent is justified strictly by analogy to other voiced obstruents, but the phonological identity of English /v/ as an obstruent is justified at least in part by its contrast with the sonorant approximant /w/, though this intuition is rarely made explicit.

The relationship between [f, v, s, z] in each language is assessed using three measures: duration, harmonicity, and the spectral centroid calculated over the high-pass filtered signal. These measures, detailed further in Section 2.4, have been used to distinguish voiced from voiceless fricatives, as well as the degree and type of frication. The relationship between segments is assessed by implementing a hierarchical mixed model, fit over factors that encode the voicing value (voiced vs. voiceless) and the type of frication (sibilant vs. spirant).

An additional factor of interest is whether the relationship between voicing and the type of frication is stable across prosodically strong and weak environments, and whether the effect of environment differs across segments, particularly with respect to the realization of /v/. In languages as varied as Hindi and German, /v/ has allophonic variants [v, ʋ, w], conditioned by prosodic factors (see Pierrehumbert and Nair (1996) for discussion of Hindi /v/ allophony, and Wiese (2000) for German). To my knowledge, allophony of this magnitude is not reported for any of the languages in the present study, but these cases suggest that /v/ may be particularly prone to prosodic conditioning.

To summarize, the question that this chapter seeks to answer is as follows: Is the relationship between [v] and [f] parallel to the relationship between [z] and [s]? In other words, modulo whatever effects sibilance has on a given phonetic parameter, are differences [v] between [f] the same kinds of differences between [z] and [s]?

The remainder of this chapter is structured as follows. In Section 2.2, I discuss in more depth the selection of languages; in Section 2.3, I discuss the methodology used, including speaker selection (Section 2.3.1), word lists (Section 2.3.2), recording setup (Section 2.3.3), and segmentation (Section 2.3.4). I detail the acoustic measures in Section 2.4, and the statistical analysis in Section 2.5. I turn to the results in Section 2.6 – Section 2.9, dealing with each language independently, and discuss the results in Section 2.10.

## 2.2 Language Selection

Russian, Greek, and Serbian were selected as the languages of study because they exhibit a partial typology of patterning with respect to the patterning of /v/. Additionally, they were selected with the following principles in mind: First, the local surface inventory structure regarding /v/ is the same in all three languages: all three have /f/, and none have /w/ or any other labial approximant. Second, the vowel inventories are sufficiently similar, as all are five vowel systems that are transcribed roughly as /i, e, a, o, u/, meaning that it was easy to design a word list that maintained a similar vowel environment. Third, all three languages had at least one voiceless-voiced sibilant pair against which /f, v/ could be compared. Fourth and finally, all three languages are true voicing languages, rather than aspirating languages.

English differs in a couple of crucial ways from the other three languages. First, English has /w/ in addition to having /f, v/, meaning that in English [v] tokens contrast not just with [f], but also with [w]. Second, the English voicing contrast in initial and stressed position is aspirating rather than voicing (Halle and Stevens, 1971), though the fricative contrast is between voiced and voiceless members. These properties of the English inventory and contrast system provide a different setting in which to assess the relationship between [v] and [f, s, z]. On a practical level, English differs from Greek, Russian, and Serbian in that the English vowel inventory is much larger, and crucially lacks a central low vowel /a/, which was used in most of the words for the other languages; for this reason, nonce words with a low back [ɑ] were used.

## 2.3 Methodology

### 2.3.1 Speakers

For Greek, Russian, and Serbian, seven native speakers of each language were recorded: Greek (4F, 3M; aged 26-32), Serbian (3F, 4M; aged 29-47), Russian (4F, 3M; aged 22-73).[1] None reported any hearing loss or showed any evidence of a speech impairment. With the exception of RuF2, who immigrated to the US at 16, all had left their home country after the age of 18 and the majority had been in North America for 10 years or fewer at the time of recording (with the aforementioned exception of RuF4 noted above).[2] Dialect was partially controlled for, in that certain dialects were avoided, but it was not possible to find enough speakers from only one dialect. Regardless, the segments under study are not expected to display much, if any, dialectal variation, and none of the (potential) dialects of the speakers indicate any differences in the consonantal phonology of /f/ or /v/; for all languages, a linguist native speaker was consulted regarding the dialect situation.

For Greek, dialectal variation is not as pronounced in the educated, urbanized, younger generation, the group available for study (graduate students at either Cornell or the University of Toronto). Of the Greek speakers, two (GrF4, GrM1) were born in Northern Greece, one (GrM2) was born in Crete, and the remaining four were born and raised in or near Athens (GrF1, GrF2, GrF3, GrM3). For Serbian, three participants (SeF1, SeM1, SeM3) were from the south (SeF1, SeM1 from Niš; SeM3 from Kruševac) and four from northern

[1] The age range for Russian is significantly skewed by RuF4, who was 73 years old at the time of recording; the next oldest Russian participant was 29. Though RuF4 had been in North America for over 20 years, she works as a Russian language instructor, and maintains strong ties to the Russian language and culture.

[2] RuF2 maintained strong ties to Russia and the Russian community in the US, and speaks Russian daily.

and central Serbia (SeF2, SeF3 from Novi Sad; SeM2 from Belgrade; SeM4 from Paraćin).[3] For Russian, speakers from neighbouring countries were excluded to control for language contact effects of closely related languages (e.g., Ukrainian), and speakers from the south were also excluded.[4] Three Russian speakers (RuF1, RuF2, RuM4) were from in or near Moscow (RuM4 is from Bryansk), three from St. Petersburg (RuF4, RuF6, RuM2) and one (RuM1) from Yekaterinburg.

English speakers were recruited from the graduate student population at Cornell University, and were between the ages of 22 and 32 at the time of recording. All eight speakers were native speakers of either American or Canadian English, and none reported any hearing loss or showed any evidence of a speech impairment.

All speakers of the study, except for GrF1, SeF1, and SeM3, who are linguists, were naïve as to the purpose of the study; GrF1 and SeM3 helped design the word lists for Greek and Serbian, respectively. A few of the participants had some linguistics background, but when they were queried as to what they thought the recording session was about, none guessed the purpose of the study correctly. Most of the English speakers knew that the general study was on consonant articulations, but did not have further details.

### 2.3.2 Word list

#### 2.3.2.1 Greek, Russian, Serbian

The complete word list included stops, fricatives, and sonorants in various positions (word initial, word medial, and word final), in words with a varying number of syllables (up

[3]Although the northern and southern dialects are known to differ in their prosodic patterns, no relevant differences in the segmental phonology are reported (Draga Zec, personal communication).

[4]Excluding speakers from the south was based on the advice of two Russian linguists (Alexei Kotchetov and Maria Gouskova) due to dialectal differences in the segmental phonology of the southern dialects: /g/ is frequently realized as [ɣ] and vowels are much longer than their northern counterparts.

to three), in stressed and unstressed position, as applicable given the phonotactic and stress restrictions of the language in question. Only a subset of the collected data were selected for analysis; the remaining words therefore acted as detractors to the present study. A total of 910 tokens were elicited over all words, repetitions and speakers, for all languages combined, for each environment; the number of tokens is lower than expected since many words could be used for more than one environment).

The word list for each language was devised with the following principles in mind. Target segments /f, v, s, z/ were elicited in two environments, word-initial, stressed (WIS) position, and word-medial unstressed (WMU) position.[5] Word size was generally limited to bisyllabic words of the form CVCV(C), with stress on the first syllable, though Serbian [ˈzapeta] was included.[6] In order to avoid consonant palatalization triggered by high and front vowels, adjacent vowels were limited to underlying /a/ and /o/.[7] In all three languages, /o/ is realized somewhat intermediate between [o] and [ɔ], and in Russian, unstressed /o/ is realized as [ɑ]. The word list (transcribed in IPA) is given in Table 2.1.

For the WIS environment, no tokens were removed. For the WMU environment, some tokens were removed either because the participant did not know the word, or as a result of

[5] Due to an oversight in coming up with the word lists, words with [s] in WMU position for Russian were not elicited. In their place, I use /faˈsad/, and /zaˈsov/, which have stress on the second syllable. In general, this should not have an outsized effect on the results, but should nevertheless be kept in mind, particular with respect to duration, which is expected to be sensitive to position and stress. In an acoustic study of Russian voiceless sibilants, Kochetov (2017) found that in word-initial position, fricatives are longer in stressed syllables over unstressed syllables, and that sibilants in word-initial position are longer than those in word-medial position. However, a full comparison of stress and position was not done, and so the results of Kochetov (2017) do not provide a more precise estimate of the durational difference expected between stressed and unstressed [s] in word-medial position. Moreover, as duration is sensitive to word duration, syllable complexity, and other extra-linguistic factors (e.g., rate of speech and experimental conditions), results cannot be compared across different experiments. Nevertheless, while duration may be affected by the use of WMS [s] rather than WMU [s], other measures are not expected to differ, as found in Kochetov (2017).

[6] Visual inspection of the data for [ˈzapeta] and [ˈzadar] were done prior to all analyses to ensure that the words were roughly comparable; although [ˈzapeta] was slightly shorter in duration, the differences were slight, and no appreciable differences were detected in harmonicity or spectral measures.

[7] Greek /ˈpezo/ is an exception to this, but [z] is not palatalized in this environment.

Table 2.1: Words with target segments: word-initial, stressed (WIS) syllables vs. word-medial unstressed (WMU) syllables.

| | WIS environment | | | | WMU environment | | | |
|---|---|---|---|---|---|---|---|---|
| Greek | ˈfovos | ˈvolos | ˈsoma | ˈzori | ˈtafos | ˈfovos | ˈðasos | ˈvrazo |
| | ˈfaros | ˈvaθos | ˈsakos | ˈzali | | ˈkava | ˈposa | ˈpezo |
| Russian | ˈfaza | ˈvata | ˈsaxar | ˈzapax | ˈpafos | ˈslovo | faˈsad | ˈbaza |
| | ˈfara | ˈvozit | ˈsokol | | ˈtrefa | ˈpravo | zaˈsov | ˈfaza |
| Serbian | ˈfotsa | ˈvoda | ˈsaxat | ˈzadar | ˈkafa | ˈlava | ˈkasa | ˈbaaza |
| | ˈfatsa | ˈvakuf | ˈsada | | ˈrafal | ˈovaj | ˈtasa | ˈkoza |

oddities in the recording.[8]

Words were read in a frame sentence (1) – (3), and in all cases, the vowel preceding the target word ended in /a/.

(1) Greek
eɣrapsa ______ tris fores
I wrote ______ three times

(2) Russian
sveta skazala ______ odin ras
Sveta said ______ one time

(3) Serbian
kaʒe jetsa ______ opet
says, Jeca ______ again

[8]Twenty-five tokens were not recorded because the participant did not know the word: /trefa/ (RuM1, RuM4); /slovo/ (RuF2); /tasa/ (SeF3); /kama/ (SeF1). All five tokens of /ðasos/ for GrF1 were removed due to file corruption. The remaining three tokens were removed due to oddities in the recording or file corruption: /posa/ (GrF2: Rep2); /vrazo/ (GrF2: Rep1); /pafos/ (RuF2: Rep1).

#### 2.3.2.2 English

As noted in Section 2.2, English differs from Greek, Russian, and Serbian in multiple ways. In order to maximize the comparability of English to the other languages, nonce words were used for English, of the form [ˈCɑCɑ], rendered orthographically as <*CahCa*>, as in, for example, <*mahfa*>. Although using nonce words carries a risk of analysing unnatural speech, this liability is minimized because of the extensive literature on English fricatives. As with the word lists for Greek, Serbian, and Russian, many additional segments were included in the data collection as detractor segments, including stops and sonorants [l, m, w], but only [f, v, s, z] were analysed.

Each nonce word recorded had one of the five target segments in one of the two positions, and thus each word served as a token for the WIS and WMU environments, and in words selected for analysis, no word had the same segment twice. Thus for each segment, in each environment, 10 words were used for analysis.[9]

### 2.3.3 Recording setup

The recording session consisted of reading a word list in a frame sentence, with five randomized repetitions. Recordings for Greek, Russian, and Serbian took place in a sound attenuated chamber in the Phonetics Laboratory at either Cornell University or the University of Toronto. Recordings for English took place only at Cornell. The recording device in both locations was a Sony SD722 digital recorder, and recordings of Greek, Russian, and Serbian were sampled at 44,100 Hz with 16-bit quantization; recordings of English were sampled at 22,050 Hz. Recordings for Greek, Russian, and Serbian were downsampled to

[9] The following forms were omitted from the WMU environment: <*sahfa*> (EnF2: Rep 2), and <*vahma*> (EnM3: Rep 1 and 2) due to anomalies produced after the first syllable, and <*zahsa*> (EnM1: Rep 1) due to an unnatural pronunciation.

22,050 Hz. At Cornell, the microphone used was an Electrovoice RE20 dynamic cardioid microphone; at Toronto, a DPA 4011 cardioid shotgun microphone for all speakers except SeM1, for which a Shure SM10−A headworn microphone was used. For Study 1, the word list was read from paper; the word lists were randomized five times (using Excel), and presented as separate lists. For English the words were presented on a monitor in the sound attenuated chamber, with a lag of two seconds between each word using the E-Prime 2.0 software (`www.pstnet.com`); the word lists were randomized twice by the EPrime software.

Participants were asked to read at a comfortable, conversational pace, to skip words they did not know, and were asked to repeat themselves if some disturbance occurred during a particular token (e.g., a cough, rustling of papers, etc.). Participants were given breaks between reading each list, and each recording session took between 40 and 60 minutes. All speakers of Russian, Greek, and Serbian were remunerated $10 for their participation; speakers of English were not remunerated.

### 2.3.4 Segmentation

The recorded signals were hand segmented in PRAAT, based on visual inspection of the waveform and spectrogram. In word-initial conditions, target consonants and the following vowel were segmented; in word-medial conditions, the preceding vowel was segmented as well. Jesus and Shadle (2002) note that the segmentation of fricatives, voiced fricatives in particular, can be somewhat subjective.

> During the vowel-fricative transition, there is a decrease in amplitude, voicing ceases (for unvoiced fricatives) and frication noise starts. During the fricative-vowel transition, there is an increase in amplitude, voicing starts (for unvoiced fricatives) and frication noise ceases. These events do not occur simultaneously or always in the same order, making the segmentation a somewhat subjective process.

Segmentation of [f, v, s, z], depended in part on position. For word-initial consonants, the onset of the consonant was determined what appeared to be the onset of constriction, gauged by a low amplitude waveform. The offset was determined to be the onset of clear formant structure (first three formants) of the following vowel. For word-medial consonants, the onset was determined to be the end of clear formant structure of the preceding vowel; I did not use this method for word-initial consonants, as speakers sometimes produced a pause before the target word. Some articulations of Serbian [v] in the WMU environment were so sonorous as to be glide-like, and thus the onset and offset were marked approximately part way through the formant transition from the previous and following vowels; stability of the third formant was often used as a heuristic in segmentation as well.

## 2.4 Measures

Acoustically, noise corresponds to the presence of energy in the high frequency domain, while voicing corresponds to energy in the low frequency domain. Assessing whether the fricatives [v] and [z] are realized as the voiced counterparts to [f] and [s], respectively, involves assessing the relative contribution of voicing and frication to the acoustic signal. Three point measures are discussed—duration, spectral centroid, and harmonicity—each of which is described in detail below.

### 2.4.1 Duration

Duration is reported for all segments tested in all languages, reported in milliseconds. In the WIS environment, all syllables were open; in the WMU environment, some syllables were closed, and some open. All else being equal, voiced fricatives are shorter in duration than voiceless fricatives, and so we predict that tokens of [v] and [z] will be shorter than [f] and

[s], respectively, and that sibilants will be longer than non-sibilants, as found in Jongman et al. (2000) for English, and Nirgianaki (2014) for Greek.[10]

### 2.4.2 Harmonicity

*Harmonicity* calculates the relative contribution of voicing and frication in the acoustic signal, and can be interpreted as a measure of the degree of acoustic periodicity (Boersma, 1993).

Harmonicity was computed over the middle 80% of the sound, highlighted in blue in the illustration in Figure 2.1, because the presence of vowel transitions (due to the segmentation method) could have an undue effect on the mean harmonicity. Harmonicity was calculated automatically as part of a Praat script using the `To Harmonicity (cc)...` function in Praat, with default parameters: time step = .01 s; minimum pitch = 75 Hz; silence threshold = .1; 1 period per window. This procedure generates a matrix of values at several time steps through the duration of the segment; I report on the mean harmonicity, using the `Get mean...` function in Praat.

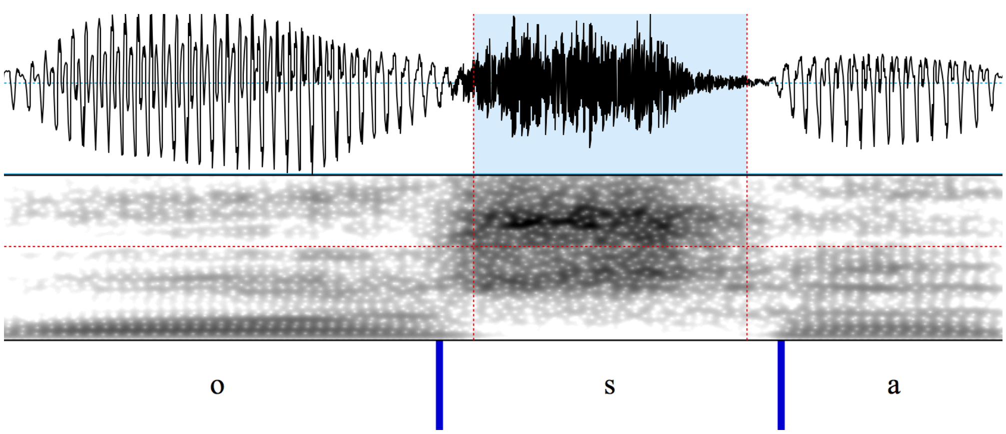



Figure 2.1: Middle 80% of segment [s] in Greek [pɔsa]

[10]Both Jongman et al. (2000) and Nirgianaki (2014) used *normalized duration*, defined as the ratio of the duration of the fricative to the duration of the word.

As Hamann and Sennema (2005) explain, "A harmonicity median of 0 dB means that there is equal energy in the harmonics and in the noise of a signal, and a harmonicity median of 20 dB that there is almost 100% of the energy of the signal in the periodic part (Boersma 1993)." Voiceless fricatives such as [f, s] are predicted to have lower harmonicity values than voiced fricatives such as [v, z]. Furthermore, since sibilants are characterized by a energy in the high frequency range, while labiodental spirants have a more diffuse spectrum of noise distribution, we predict that [v] will have a higher harmonicity than [z].

### 2.4.3 Spectral centroid

It is well known that fricatives are distinguished by their statistical distribution over the frequency domain. This can be assessed by comparing power spectra over the whole range, or by summary statistics, known in the literature as *spectral moments* (Jongman et al., 2000; Jesus and Shadle, 2002; Maniwa et al., 2009). As this literature has primarily focussed on distinguishing place of articulation within a given language, not all spectral measures are applicable to the goals of the present study.[11] I limit my attention to the *spectral centroid* (or *centre of gravity*).

The spectral centroid measures the average frequency of the spectrum, weighted by the energy; the variance represents the spread of the spectral energy. For a complex spectrum $S(f)$, where $f$ is the frequency, the centroid is given by (2.1) (Boersma and Weenink, 2011).[12] The centroid is a measure of the location of energy concentration in the frequency domain,

[11]For example, *skewness* is a measure of the asymmetry of a distribution about its mean, by measuring the asymmetry in the tails of a distribution. A distribution with positive skewness has a longer *or* fatter tail on the right side than on the left side, but skewness does not distinguish the distribution's shape more finely than this, making it difficult to assign an interpretation to skewness measures over the frequency domain. Measures such as *skewnes* and *kurtosis* are therefore useful for distinguishing major differences that one might expect from place of articulation, particularly for voiceless fricatives, but are less useful for exploring how sibilant and non-sibilant frication interacts with voicing.

[12]Changing the exponent from 2 to 1 would correspond to weighting by the absolute spectrum rather than by the power spectrum, as given here.

or "how high the frequencies in a spectrum are on average" (Boersma and Weenink, 2011).

$$\frac{\int_0^\infty f|S(f)|^2\,df}{\int_0^\infty |S(f)|^2\,df} \tag{2.1}$$

All else being equal, voicing and noise will have opposite effects on the value of the centroid. Since noise is energy in the high-frequency range, higher centroid values correspond to noisier sounds. In contrast, voicing introduces energy in the low-frequency range, and thus voiced sounds will have lower centroid values than their voiceless counterparts. Nevertheless, as a statistical measure over a distribution, the centroid only makes sense for a roughly monomodal distribution. For voiceless sounds, calculating the spectral moments over the whole distribution is unproblematic, as the distributions are monomodal to begin with, but for voiced fricatives with one peak in the lower frequency range and another peak in the high frequency range, this criterion is not met. Because of the effect of voicing on the spectral centroid, and because the goal is to quantitatively assess the concentration of noise in the frequency domain, the centroid is computed after the signal has been high-pass filtered at 1500 Hz to remove the effects of voicing and the first several harmonics.[13] As a result, the centroid values are interpreted as the concentration of energy *modulo* the effect of voicing.

The spectral centroid cannot, in general, be interpreted as a measure of the degree of frication, because the distribution of energy is subject to the downstream filtering of the air stream. For example, one would not say that [s] is "more fricated" than [ʃ], even though [s] has a higher centroid measure than [ʃ] (Jongman et al., 2000); the difference in centroid measures is a result of the different articulations which have different effect on the filtering of the airstream downstream from the locus of constriction. A similar result can also be seen in

[13]This methodology is similar to that of Hamann and Sennema (2005), though they used a high-pass filter of 500 Hz, rather than 1500 Hz.

the results here: Greek sibilants [s, z] have much lower spectral centroid measures than those of Russian and Serbian, but this is due to the fact that Greek [s, z] are produced posterior to Russian and Serbian [s, z], and not because they are produced with "less frication".

Variability in the spectral centroid of [v] should not be subject to such filtering effects due to place of articulation, since [v] is labiodental in all languages considered. Nonetheless, there could be subtleties in the articulation of the labiodentals that results in different centroid results that do not correspond to the degree of constriction, but rather to slight place differences. For this reason, centroid values of voiced segments [v, z] are interpreted in comparison to the centroid values of their voiceless counterparts [f, s], under the assumption that [f, v] and [s, z] share the same place of articulation, respectively.

#### 2.4.3.1 Time-averaging of spectra

To the extent that sounds such as [f, v, s, z] are produced with turbulent airflow, the signal is marked by small, random perturbations characteristic of random noise. There are several ways to account for this, as explained by Jesus and Shadle (2002):

> Stochastic signals require some form of averaging for their spectra to be both consistent and low-error estimates of the underlying distribution (Bendat & Piersol, 2000, p. 423−442). For a stationary signal, time-averaging can be used; for non-stationary signals where an ensemble exists, ensemble-averaging can be used; or a single spectrum can be smoothed, and the averaging achieved at the expense of frequency resolution. (pg. 443)

In keeping with previous literature on fricative spectral analysis, I use time averaging for the calculation of the spectral centroid (Shadle et al., 1996; Jesus and Shadle, 2002; Ladefoged, 2003). Time-averaging was done over spectra calculated over the middle 80% of the fricative, removing transitions.[14] Following the methodology laid out by Jesus and Shadle

[14]Analysis was done based on the script "Spectral moments of fricative spectra script in Praat", written

(2002), time-averaged spectra were computed by averaging seven 10 ms Hann windows spread out over the analysis duration, illustrated in Figure 2.2. In longer fricatives, windows do not overlap, but in shorter fricatives, the windows can overlap considerably; parameters were chosen to have the greatest number of windows possible while still ensuring that windows did not overlap more than 50% in the shortest segments. An illustration of the time-averaging method used is provided in Figure 2.2.

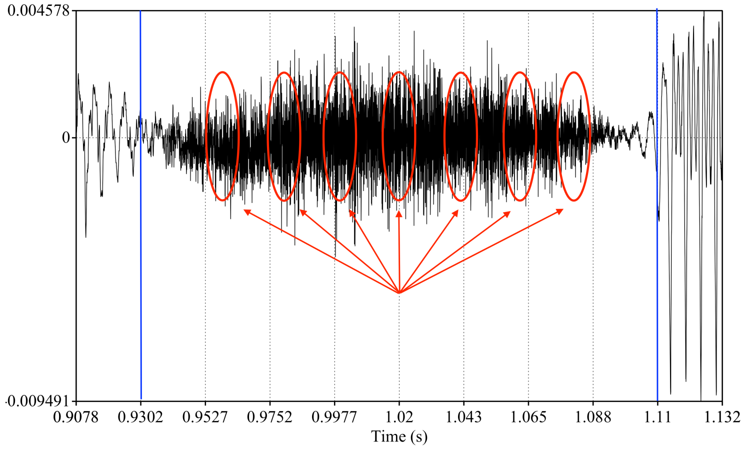



Figure 2.2: Time averaging

## 2.5 Statistical analysis

Before undertaking any statistical analysis, data were inspected visually. Potential outliers were investigated by going back to the original sound files and textgrids, in order to determine if the outlier measurement was due to either experiment error or mispronunciation; errors that arose from segmentation or mislabelling were corrected and measurement scripts rerun, but errors due to equipment problems or mispronunciation were removed from the dataset. No outliers were otherwise removed.

by Christian DiCanio, available here at `http://www.acsu.buffalo.edu/~cdicanio/scripts.html`, though the script was heavily modified, both in structure and in analysis parameters.

Table 2.2: Specification of all models in R. Dataframe and REML specification omitted; REML was set to FALSE. Env = Environment; Vc = Voicing; FricType = Frication Type; Spkr = Speaker.

```
m1  ← lmer(X ∼ 1 + (1|Spkr))
m2  ← lmer(X ∼ Env + (1|Spkr))
m3  ← lmer(X ∼ Vc + (1|Spkr))
m4  ← lmer(X ∼ FricType + (1|Spkr))
m5  ← lmer(X ∼ Env + Vc + (1|Spkr))
m6  ← lmer(X ∼ Env∗Vc + (1|Spkr))
m7  ← lmer(X ∼ Env + FricType + (1|Spkr))
m8  ← lmer(X ∼ Env∗FricType + (1|Spkr))
m9  ← lmer(X ∼ Vc + FricType + (1|Spkr))
m10 ← lmer(X ∼ Vc∗FricType + (1|Spkr))
m11 ← lmer(X ∼ Env + Vc + FricType + (1|Spkr))
m12 ← lmer(X ∼ Env + Vc∗FricType + (1|Spkr)))
m13 ← lmer(X ∼ Env∗Vc + FricType + (1|Spkr))
m14 ← lmer(X ∼ Env∗FricType + Vc + (1|Spkr))
m15 ← lmer(X ∼ Env∗Vc∗FricType + (1|Spkr))
```

Statistical analyses were carried out in R (R Core Team, 2017). Linear mixed models were fit using the `lme4` package (Bates et al., 2015b). The following model selection approach was used: all possible models were fit, using the fixed effects `Environment` (values: WIS, WMU), `Voicing` (values: 0 = voiceless, 1 = voiced), and `FricationType` (values: 0 = spirant, 1 = sibilant), and specifying `Speaker` as a random intercept, yielding a total of 15 models, shown in Table 2.2.

The model with the lowest Bayesian Information Criterion (BIC) value was selected provided that it differed from the next lowest BIC by at least 2 points. In cases where the BIC was not definitive for choosing between two models (a difference of less than 2 points), the model with the lower value for the Akaike Information Criterion (AIC) was selected. Finally, model fit was assessed graphically, testing the residuals for normality, independence and homoskedasticity.[15] Models were fit with maximum likelihood (`REML = FALSE`).

[15]This model selection procedure was chosen in order to avoid the problem of local minima in incremental model fitting procedures. Baayen (2008) recommends incrementally fitting models, beginning with the null

Reported $p$-values are based on Satterthwaite approximation using the package `lmerTest` Kuznetsova et al. (2016). Significance was assessed at the $\alpha = .5$ level; cases where factors were significant at the $\alpha = .1$ level are noted. Confidence intervals of the fixed and random effects were computed using the `confint.merMod` function part of the `lme4` package with the default profile method (Bates, 2010). Model results are presented in tabular form, and confidence intervals are plotted alongside.

The `lsmeans` package was used to understand the relationship between the particular levels of the fixed effects (i.e., voiced vs. voiceless; sibilant vs. spirant; WIS vs. WMU) (Lenth, 2016). Least squares means and associated confidence intervals of the fixed effect levels were computed with a confidence level of 0.95, and degrees of freedom were computed using Satterthwaite approximation. Least squares means are presented as interaction plots showing the least squares means and confidence intervals.

### 2.5.1 Coding fixed effects

Fixed effects were coded using the default treatment (or dummy) coding in R, shown in Table 2.3. Intercepts in language specific analyses represent the value of `Environment` = 0, `Voicing` = 0, `Frication Type` = 0; in other words, the value of the voiceless, non-sibilant (i.e., [f]) in the word-initial stressed environment.

In the presentation of the model, these factors are referred to as `Environment`, `Frication Type`, and `Voicing`, with the following abbreviations used in tables and plots: Env. (WMU), Frication (Sib), and Voicing (Vcd). For example, if type of frication is a significant effect

model, and using the `anova` function that is part of the `lmerTest` package to do model comparison (i.e., a likelihood ratio test). The two model selection methods are, in general, equivalent. A difficulty with the incremental approach is that sometimes a more complex model is not justified by the procedure, even if it is a better fit to the data. In the present dataset, this typically happened when an interaction was significant, but the main effect of one of the factors did not improve model fit.

Table 2.3: Treatment coding of fixed effects

| Factor | Level = 0 | Level = 1 |
|---|---|---|
| `Environment` | WIS | WMU |
| `Voicing` | [f, s] | [v, z] |
| `Frication Type` | [f, v] | [s, z] |

in the model, then the estimate associated with Frication (Sib) refers to the effect that the sibilant type of frication has on the measure in question.

### 2.5.2 Random effects structure

I did not include word as a random effect. For English, words were nonce-words of the same phonological form, and thus there is no prediction that certain combinations might be more or less representative of English pronunciations. For Greek, Russian, and Serbian, only two words were used per segment (and sometimes only one), thus precluding the possibility of using word as a random effect. I did not include repetition as a random effect due to the design of the study: word lists were randomized, and participants were given ample time between each word list reading, so that the effect of repetition would be minimized. Moreover, unlike, say, reaction time, in which either habituation or fatigue would be expected to play a serious confounding role within the time course of the study, there is no reason why these factors would be expected to play a significant role in the variables under study. Nevertheless, the data were inspected visually to ensure that there were no overall trends over the time course of the recording session.

Mixed-effect models with speaker as a random intercept encodes the speaker-specific means within each fixed-effect grouping; for example, different speakers may exhibit across-the-board effects with respect to duration, and this effect is taken into account by modelling

`Speaker` as a random effect. Including random slopes models the speaker-specific variation in the degree to which they respond to the different factor levels; to use the same example, some speakers may show more sensitivity to the effect of voicing on duration than others, and incorporating random slopes models varying sensitivity in addition to the across-the board effects.

Determining the appropriate random effects structure in (psycho)linguistics is an ongoing topic of debate in the literature of mixed-effect modelling, specifically with respect to the inclusion of random slopes. In many cases, including random slopes led to convergence problems; in other cases the inclusion of random slopes led to extremely high correlations in the interaction term, suggesting overparametrization (Baayen et al., 2008). Thus, in spite of the extra modelling potential that random slopes afford, I follow the recommendations of Baayen et al. (2008); Bates et al. (2015a) in fitting the simplest model possible (c.f. Barr et al. (2013)), and generally do not include random slopes for speaker in the specification of the model, making sure that the omission did not change the qualitative results of the model.

## 2.6 English: Results

### 2.6.1 Duration

Violin plots of duration in each environment for English are shown in Figure 2.3.

The most parsimonious model for *duration* included `Environment`, `Voicing`, and `Frication Type` as fixed effects, but none of the interactions between them; `Speaker` was specified as a random effect. Fricatives in the WMU environment were shorter than fricatives in the WIS environment, voiced fricatives [v, z] were shorter than voiceless fricatives [f, s],

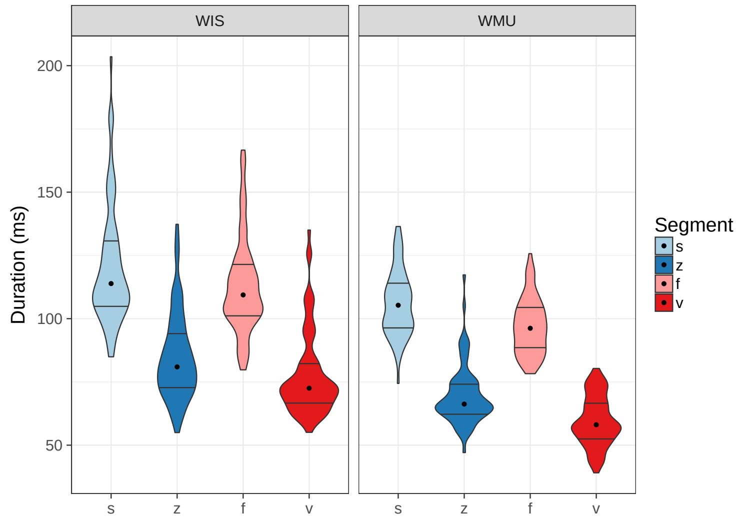


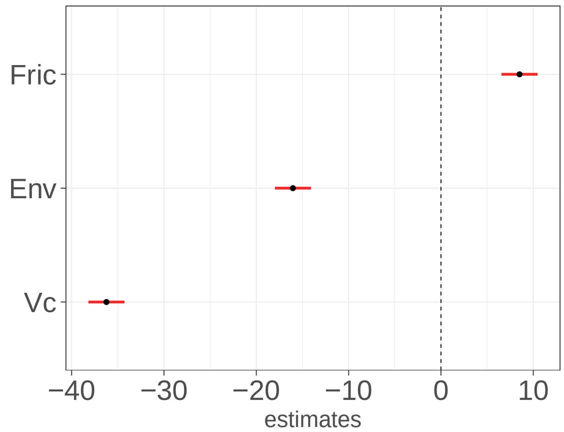

Figure 2.3: Violin plots of duration in English

and spirants [f, v] were shorter than sibilants [s, z]. Confidence intervals for the fitted model are shown in Figure 2.4a.

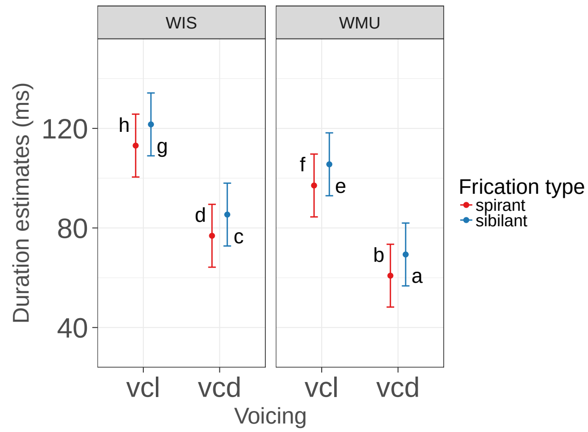
Table 2.4: Fitted model of duration (ms) in English: Estimates shown with standard errors in parentheses; lower and upper bounds of 95% confidence intervals shown in parentheses underneath estimate and standard error.

| | $\beta$ (SE)<br>95% CI | t-value | p-value |
|---|---|---|---|
| (Intercept) | 113.1 (3.58)<br>(105.24, 120.96) | 31.61 | $<$ **.0001** |
| Env. (WMU) | -16.03 (1)<br>(-17.99, -14.08) | -16.11 | $<$ **.0001** |
| Frication (Sib) | 8.51 (1)<br>(6.56, 10.47) | 8.55 | $<$ **.0001** |
| Voicing (Vcd) | -36.23 (1)<br>(-38.19,-34.28) | -36.39 | $<$ **.0001** |

Figure 2.4b shows the least squares means estimates for each factor level. Since each




(a) Estimates and CIs




(b) Interaction plot of LS means

Figure 2.4: Estimates and profile confidence intervals of fixed effects for fitted model of duration (ms) in English (2.4a). Interaction plot of least squares means with 95% Tukey-adjusted comparison confidence intervals; means sharing a letter are not significantly different (2.4b)

factor has only two levels, these plots can be interpreted as estimates for each segment by environment: i.e., the estimate that corresponds to "voiceless" and "spirant" in the WIS panel (beside the letter 'h') corresponds to the duration estimate of [f] in word-initial stressed position. As seen in Figure 2.4b, duration estimates for each segment differ from every other, since no two confidence intervals share a letter. The lack of interaction between the fixed effects shows an additive relationship between voicing, frication type, and environment. In particular, the relationship between [v] and [f] parallels that between [z] and [s], and the relationship between [v] and [z] parallels that between [f] and [s]. The effect of environment is also additive, evident from the fact that between the WIS and WMU panel, the relationship between the four segments stays the same, only their duration values are all lower.

## 2.6.2 Harmonicity

Figure 2.5 shows violin plots for harmonicity by segment in each environment in English.

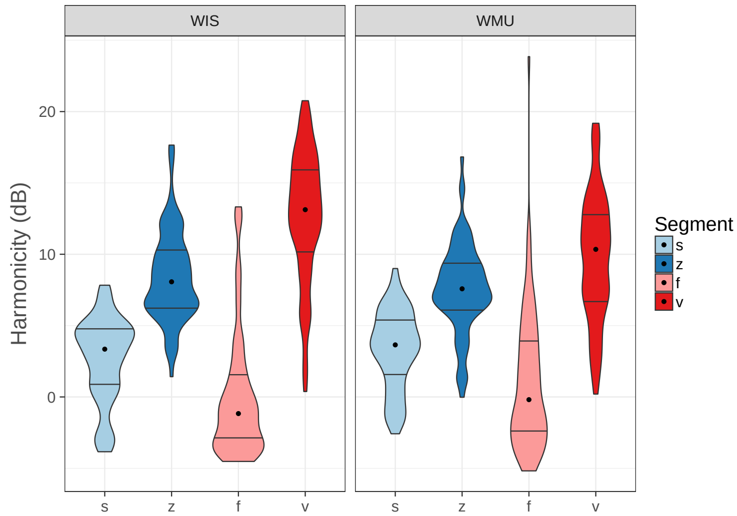


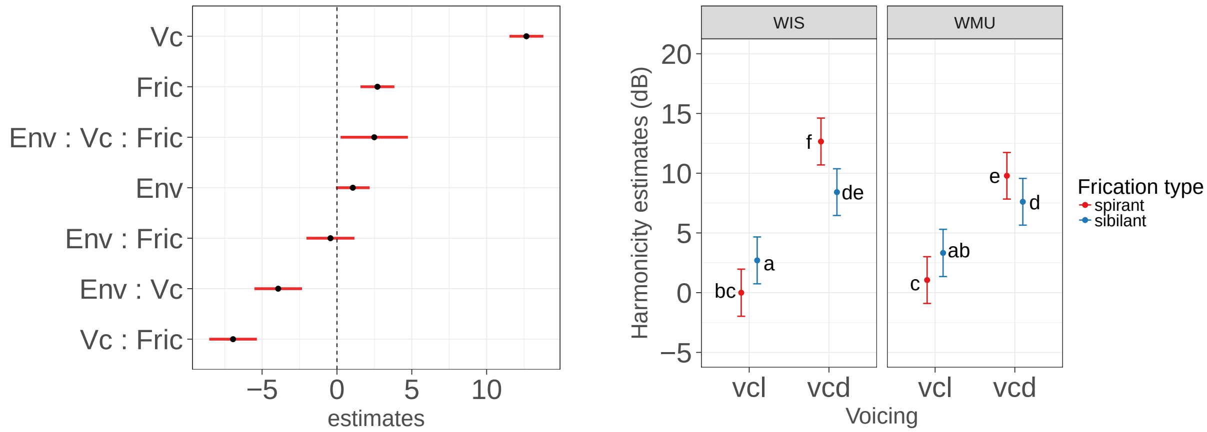

Figure 2.5: Violin plots of harmonicity in English by environment

The final fitted model for *harmonicity* included `Environment`, `Voicing`, and `Frication Type` as fixed effects, and the three-way interaction between them; `Speaker` was specified as a random effect. Not all factors and interactions were significant in the final model. Both `Voicing` and `Frication Type` significantly increased harmonicity, so voiced segments have higher harmonicity than voiceless, and sibilants have a higher harmonicity than spirants. The interaction between `Voicing` and `Frication Type` was negative, meaning that [z] has lower harmonicity than [v]. `Environment` had a positive effect on harmonicity, but was only significant at the $\alpha = .1$ level, meaning that segments produced in the word-medial unstressed position were slightly more harmonic than when in the word-initial stressed position. The interaction between `Environment` and `Frication Type` was not significant, but the interaction between `Environment` and `Voicing` was significant and had a negative effect on harmonicity. This means that voiceless fricatives were more harmonic in the word-medial position, but voiced fricatives were less harmonic in word-medial position. Finally, the the three-way interaction between `Frication Type`, `Voicing`, and `Environment` was significant and positive, meaning that the harmonicity of voiced sibilant [z] did not decrease as much as the harmonicity of [v] between the two environment.

Table 2.5: Fitted model of harmonicity (dB) in English: Estimates shown with standard errors in parentheses; lower and upper bounds of 95% confidence intervals shown in parentheses underneath estimate and standard error.

| | $\beta$ (SE)<br>95% CI | t-value | p-value |
|---|---|---|---|
| (Intercept) | 0 (0.64)<br>(-1.34, 1.34) | -0.01 | .996 |
| Env. (WMU) | 1.06 (0.57)<br>(-0.07, 2.19) | 1.84 | .0662 |
| Voicing (Vcd) | 12.66 (0.58)<br>(11.52, 13.79) | 21.95 | $<$ **.0001** |
| Frication (Sib) | 2.71 (0.58)<br>(1.57, 3.84) | 4.69 | $<$ **.0001** |
| Env.:Voicing | -3.93 (0.81)<br>(-5.51, -2.34) | -4.85 | $<$ **.0001** |
| Env.:Frication | -0.43 (0.82)<br>(-2.04, 1.17) | -0.53 | .5953 |
| Voicing:Frication | -6.94 (0.81)<br>(-8.53, -5.35) | -8.55 | $<$ **.0001** |
| Env.:Voicing:Frication | 2.49 (1.15)<br>(0.24, 4.74) | 2.17 | **.03** |

Figure 2.6b shows the least squares means estimates for each factor level. Within each environment, the order from most to least harmonic is: [v] > [z] > [s] > [f]. The effect of the WMU environment is to shrink the total range of harmonicity values, increasing the harmonicity of [f, s], and decreasing the harmonicity of [v, z]. The relative harmonicity of [f] and [s] is preserved across environments, but the difference in harmonicity between [v]




(a) Estimates and CIs

(b) Interaction plot of LS means

Figure 2.6: Estimates and profile confidence intervals of fixed effects for fitted model of harmonicity (dB) in English (2.6a). Interaction plot of least squares means with 95% Tukey-adjusted comparison confidence intervals; means sharing a letter are not significantly different (2.6b)

and [z] decreases between the WIS and WMU environments.

### 2.6.3 Spectral Centroid

Figure 2.7 shows violin plots for spectral centroid on the 1500 Hz high pass filtered signal, by segment, in each environment in English.

The final fitted model for *spectral centroid* included `Voicing` and `Frication Type` and the interaction between them; `Environment` was not included, and `Speaker` was specified as a random effect. The main effect of voicing was to decrease the spectral centroid, while sibilance increased the spectral centroid. The interaction between `Voicing` and `Frication Type` was positive, indicating a discrepancy between [v] and [z].[16]

Figure 2.8b shows the least squares means estimates for each factor level; `Environment` was not included in the model, so there is only one panel. The estimates for [s] and [z] do

[16]Recall that because the spectral centroid is computed on the high-pass filtered signal, the effect of voicing on the calculated distribution of energy in the frequency domain is mitigated.

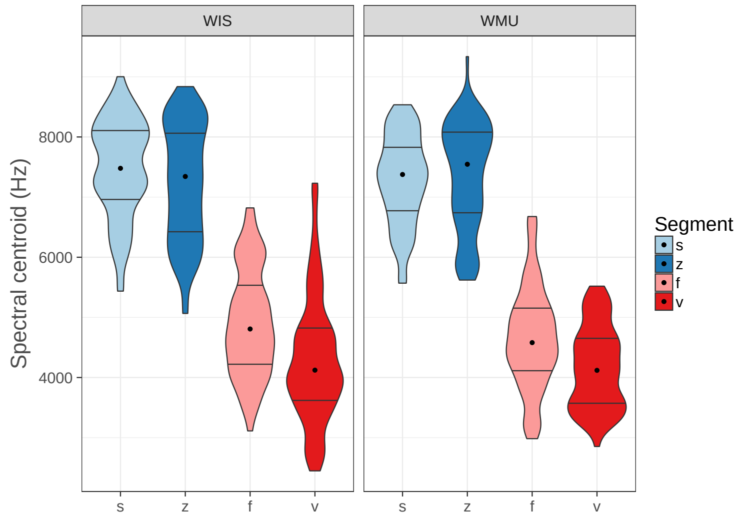
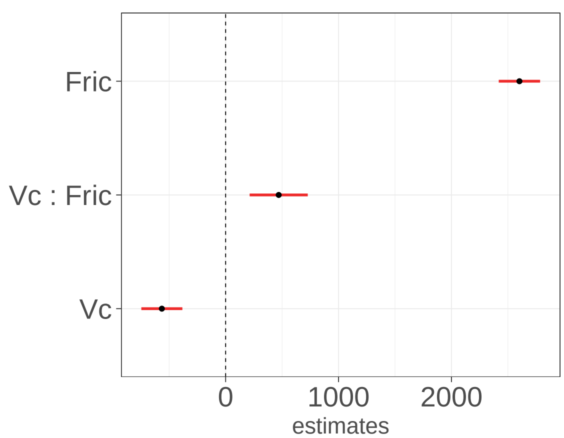



Figure 2.7: Violin plots of spectral centroid (Hz) on the 1500 Hz high pass filtered signal in English by environment

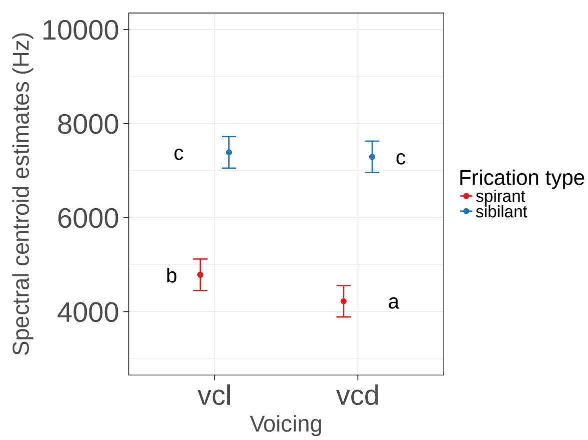
not differ significantly from each other, indicating that, with the effect of voicing and first several harmonics removed due to the 1500 Hz high pass filter, the distribution of noise in the high frequency range is the same. The estimates for [f] and [v], on the other hand, differ significantly, and [v] has a lower spectral centroid than [f]. The lower centroid value for [v] indicates that, even with the high pass filter in place, the distribution of noise in the high frequency range for spirants is modulated to a greater degree for English spirant [v] than English sibilant [z].

Table 2.6: Fitted model of of spectral centroid (Hz) on the 1500 Hz high-pass filtered signal in English: Estimates shown with standard errors in parentheses; lower and upper bounds of 95% confidence intervals shown in parentheses underneath estimate and standard error.

| | $\beta$ (SE)<br>95% CI | t-value | p-value |
|---|---|---|---|
| (Intercept) | 4785.38 (116.9)<br>(4537.26, 5033.73) | 40.94 | $<$ **.0001** |
| Voicing (Vcd) | -564.78 (92.57)<br>(-746.49, -383.07) | -6.1 | $<$ **.0001** |
| Frication (Sib.) | 2602.03 (93.3)<br>(2418.88, 2785.19) | 27.89 | $<$ **.0001** |
| Voicing:Frication | 470.09 (131.11)<br>(212.72, 727.46) | 3.59 | **.00004** |




(a) Estimates and CIs




(b) Interaction plot of LS means

Figure 2.8: Estimates and profile confidence intervals of fixed effects for fitted model of spectral centroid on the 1500 Hz high pass filter in English (2.8a). Interaction plot of least squares means with 95% Tukey-adjusted comparison confidence intervals; means sharing a letter are not significantly different (2.8b)

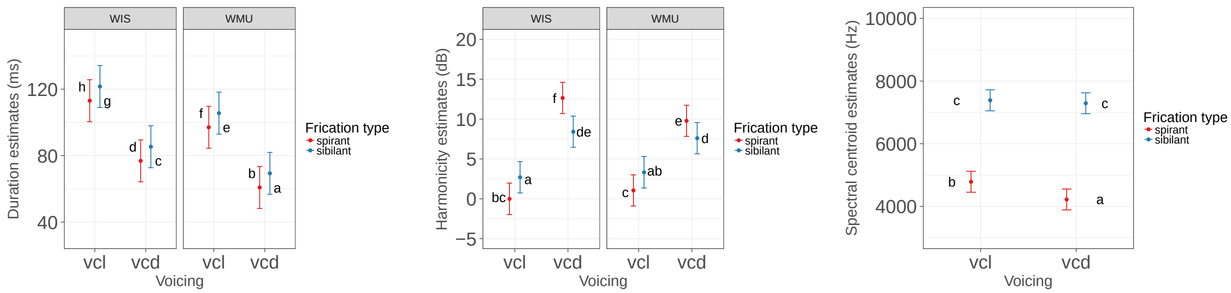
### 2.6.4 English: summary of point measures

Figure 2.9 repeats the three interaction plots of duration Figure 2.9a, harmonicity Figure 2.9b, and spectral centroid Figure 2.9c for ease of reference.

The results presented in this section are consistent with previous results of the acoustic effects of the manifestation of voicing and type of frication in English: with respect to duration, English voiced [v, z] are shorter than voiceless [f, s], and English spirants [f, v] are shorter than sibilants [s, z]; with respect to spectral centroid, sibilants have higher spectral centroid than non-sibilants (Jongman et al., 1989, 2000).

Voicing shortens fricative duration, increases harmonicity, and lowers spectral centroid. With respect to frication type, sibilants are longer than spirants, have higher harmonicity, and higher spectral centroid values. The interaction of voicing and frication type was not significant for duration, suggesting that the effect of voicing and frication type is additive, but was significant for both harmonicity and spectral centroid. Specifically, [v] had highest harmonicity and the lowest spectral centroid.

The effect of environment was not significant for spectral centroid, but was significant for the other two measures. For duration, there was no interaction effect, so segments produced in the prosodically weak word-medial unstressed position were generally shorter than when in the prosodically strong word-initial stressed position. For harmonicity, the three-way interaction between environment, frication type, and voicing was significant. As seen in Figure 2.9b, the effect of the WMU environment is to reduce the total range of values, particularly between [v] and [z].




(a) Duration (ms)



(b) Harmonicity (dB)



(c) Spectral centroid (Hz)

Figure 2.9: Interaction plots of least squares means with 95% Tukey-adjusted comparison confidence intervals for duration, harmonicity, and spectral centroid on the 1500 Hz high pass filtered signal in English (repeated). Means sharing a letter are not significantly different

## 2.7 Greek: Results

### 2.7.1 Duration

Violin plots of duration for Greek are shown in Figure 2.10.

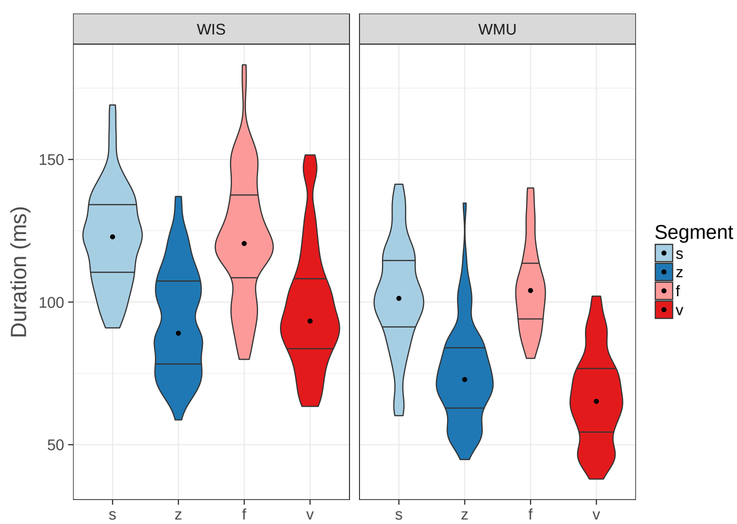



Figure 2.10: Violin plots of duration in Greek

The final fitted model for duration in Greek included `Environment` and `Voicing`, as well as the interaction between them; `Frication Type` was not included, and `Speaker` was specified as a random effect. This model had a slightly higher BIC than the one that did not include the interaction term, but the more complex model was chosen because the BIC between the two differed by less than 2, and so other model selection criteria were invoked: the AIC was lower for the more complex model, and a comparison of the two models using anova showed that the interaction was significant ($\chi^2 = 5.0541$, $df = 1$, $p = 0.02457$).[17]

Both `Environment` and `Voicing` had a negative effect on duration, meaning that fricatives were shorter in word-medial unstressed position than in word-initial stressed position, and voiced fricatives were shorter than voiceless fricatives. The interaction between `Environment` and `Voicing` was negative, indicating that voiced fricatives are more prone to shortening in the WMU position than voiceless fricatives.

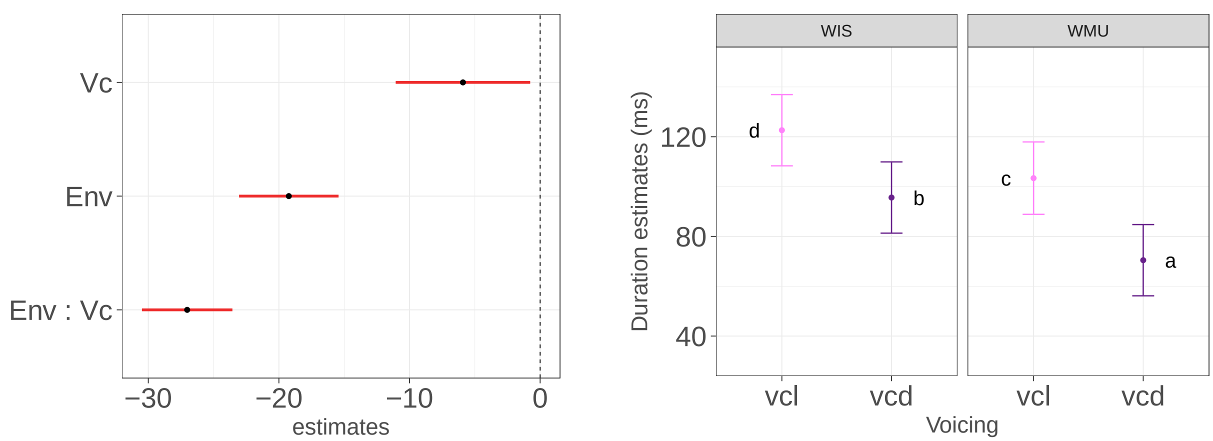
Figure 2.11b shows the least squares means estimates for each factor level; since `Frication Type` was not included in the model, the least squares means and confidence intervals refer to voiced fricatives [v, z] (in purple) and voiceless fricatives (in pink). The interaction between environment and voicing is evident in the greater decrease in duration for the voiced fricatives in the WMU panel than for the duration of voiceless fricatives.

[17]Although the information criteria typically agree in model selection, they may disagree, as they do here, precisely in cases where the AIC selects the more complex model. Briefly, this is because the BIC penalizes model complexity more heavily than the AIC; see `https://methodology.psu.edu/AIC-vs-BIC` and the linked tech report for discussion of this. Output of model comparison and information criteria scores are provided here for reference:

| Model | Df | AIC | BIC | logLik | deviance | Chisq | Df | Pr(>Chisq) |
|---|---|---|---|---|---|---|---|---|
| Environment + Voiced | 5 | 4308.7 | 4330.0 | -2149.3 | 4298.7 | | | |
| Environment ∗ Voiced | 6 | 4305.7 | 4331.2 | -2146.8 | 4293.7 | 5.0541 | 1 | .02457 |

Table 2.7: Fitted model of duration (ms) in Greek: Estimates shown with standard errors in parentheses; lower and upper bounds of 95% confidence intervals shown in parentheses underneath estimate and standard error.

| | $\beta$ (SE)<br>95% CI | t-value | p-value |
|---|---|---|---|
| (Intercept) | 122.63 (4.46)<br>(112.66, 132.59) | 27.5 | $<$ **.0001** |
| Env. (WMU) | -19.24 (1.94)<br>(-23.05, -15.43) | -9.92 | $<$ **.0001** |
| Voicing (Vcd) | -27.02 (1.76 )<br>(-30.49, -23.56) | -15.32 | $<$ **.0001** |
| Env.:Voicing | -5.91 (2.62)<br>(-11.06, -0.76) | -2.25 | **.0246** |




(a) Estimates and CIs

(b) Interaction plot of LS means

Figure 2.11: Estimates and profile confidence intervals of fixed effects for fitted model of duration (ms) in Greek (2.11a). Interaction plot of least squares means with 95% Tukey-adjusted comparison confidence intervals; means sharing a letter are not significantly different (2.11b)

### 2.7.2 Harmonicity

Figure 2.12 shows violin plots for harmonicity by segment in each environment in Greek.

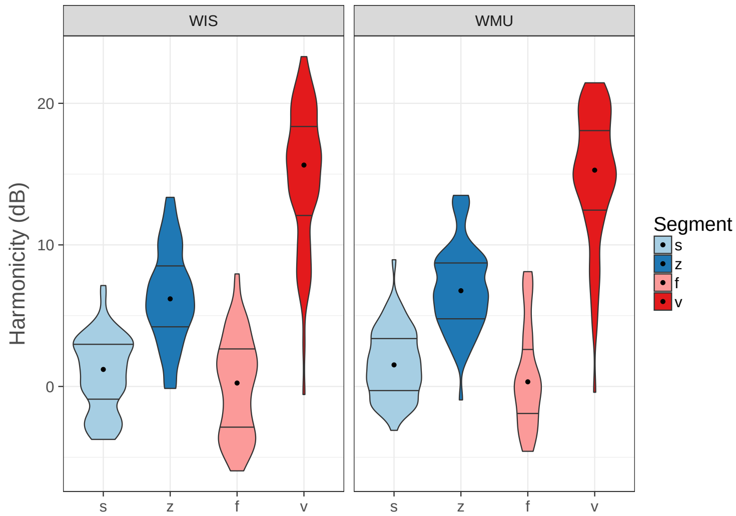



Figure 2.12: Violin plots of harmonicity (dB) in Greek by environment

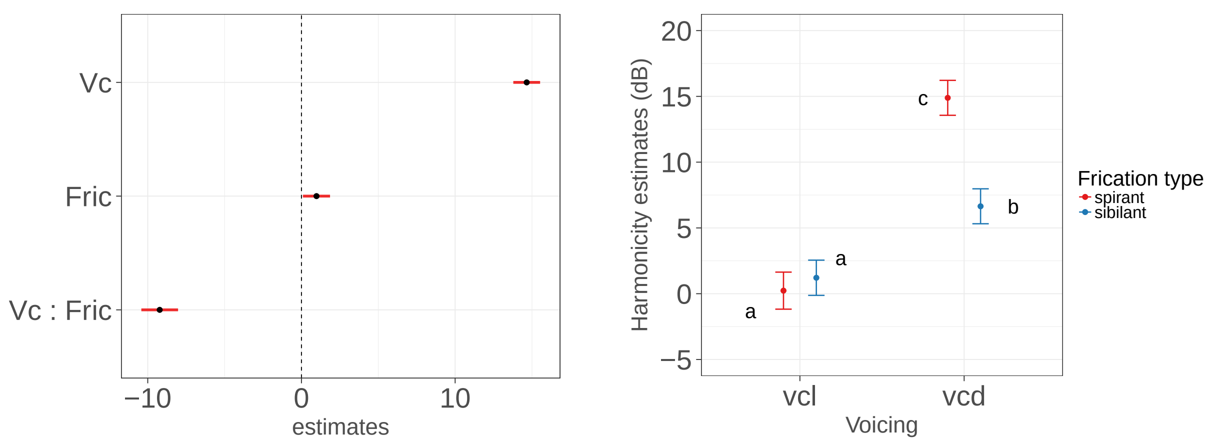
The final model fit for *harmonicity* included `Voicing`, `Frication Type` and the interaction between them; . `Environment` was not included, and `Speaker` was specified as a random effect. `Voicing` and `Frication Type` both had a positive effect on harmonicity, so voiced segments have higher harmonicity than voiceless, and sibilants have a higher harmonicity than spirants. The interaction between `Voicing` and `Frication Type` was negative, meaning that [z] had a lower harmonicity than [v].

As can be seen from the least squares means plot Figure 2.13b, the voiced fricatives are more harmonic than the voiceless fricatives [f, s], which do not differ from each other, but the voiced spirant [v] is much more harmonic than the voiced sibilant [z].

Figure 2.13b shows the least squares means estimates for each factor level. The harmonicity of [s] and [f] do not differ significantly; [v] is the most harmonic, followed by [z].

Table 2.8: Fitted model of harmonicity (dB) in Greek: Estimates shown with standard errors in parentheses; lower and upper bounds of 95% confidence intervals shown in parentheses underneath estimate and standard error.

| | $\beta$ (SE)<br>95% CI | t-value | p-value |
|---|---|---|---|
| (Intercept) | 0.23 (0.5)<br>(-0.83, 1.29) | .46 | .6503 |
| Voicing (Vcd) | 14.66 (0.44)<br>(13.78, 15.53) | 33 | $<$ **.0001** |
| Frication (Sib) | 0.98 (0.45 )<br>(0.10, 1.86) | 2.18 | **.0295** |
| Voicing : Frication | -9.22 (0.61)<br>(-10.42, -8.03) | -15.16 | $<$ **.0001** |




(a) Estimates and CIs

(b) Interaction plot of LS means

Figure 2.13: Estimates and profile confidence intervals of fixed effects for fitted model of harmonicity (dB) in Greek (2.13a). Interaction plot of least squares means with 95% Tukey-adjusted comparison confidence intervals; means sharing a letter are not significantly different (2.13b)

### 2.7.3 Spectral centroid

Figure 2.14 shows violin plots for spectral centroid on the 1500 Hz high-pass filtered signal by segment in each environment in Greek.

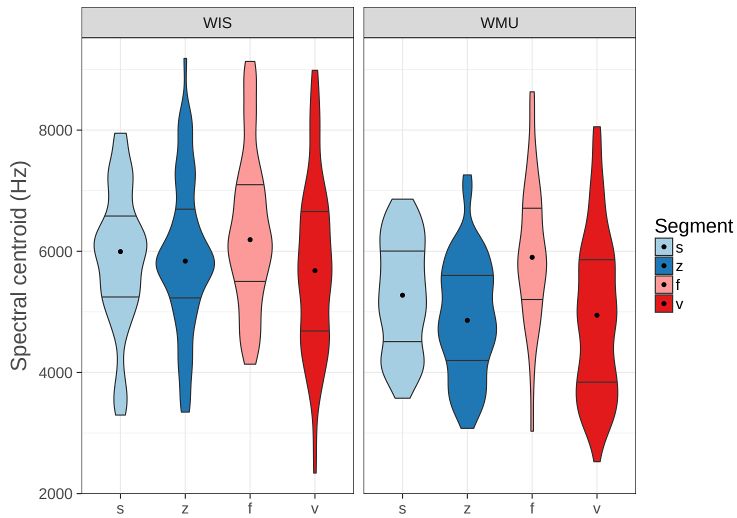



Figure 2.14: Violin plots of spectral centroid (Hz) on the 1500 Hz high-pass filtered signal in Greek by environment

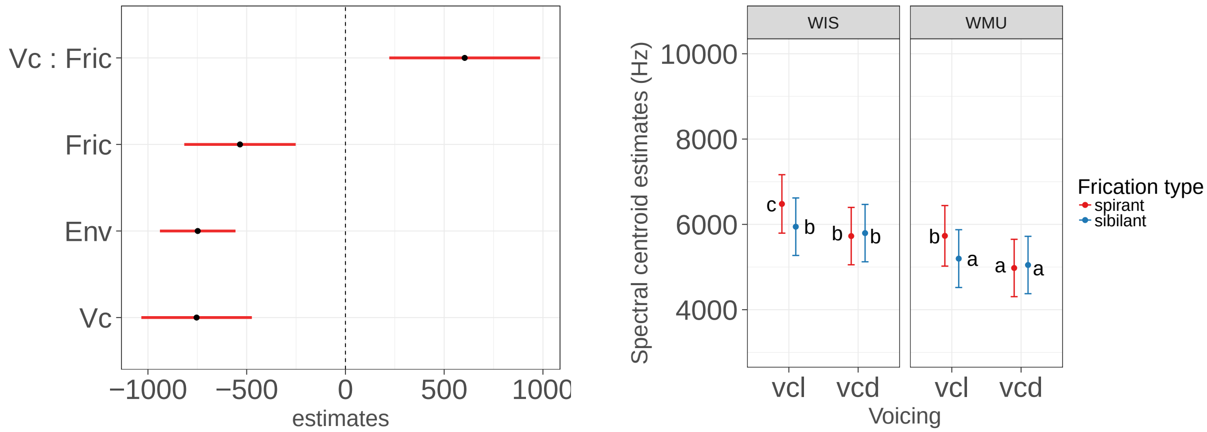
The final fitted model for Greek *spectral centroid* included `Environment`, `Voicing`, `Frication Type`, and the interaction between `Voicing` and `Frication Type`; `Speaker` was specified as a random effect. All three main effects had the effect of reducing the spectral centroid, meaning that, compared to the reference level of the voiceless spirant [f] in WIS position, all other segments had lower spectral centroid. The interaction between `Voicing` and `Frication Type` is positive, indicating that [z] has higher spectral centroid than [v]. `Environment` had a negative effect on spectral centroid, so segments had a lower spectral centroid in the WMU position than in the WIS position.

Figure 2.15b shows the least squares means interaction plot for Greek spectral centroid. The estimates for [s] and [z] do not differ significantly from each other in either environment,

Table 2.9: Fitted model of of spectral centroid (Hz) on the 1500 Hz high-pass filtered signal in Greek: Estimates shown with standard errors in parentheses; lower and upper bounds of 95% confidence intervals shown in parentheses underneath estimate and standard error.

| | $\beta$ (SE)<br>95% CI | t-value | p-value |
|---|---|---|---|
| (Intercept) | 6479.89 (207.76)<br>(6032.80, 6927.07) | 31.19 | < **.0001** |
| Env (WMU) | -747.91 (97.43)<br>(-939.22, -556.59) | -7.68 | < **.0001** |
| Voicing (Vcd) | -753.6 (142.5)<br>(-1033.42, -473.78) | -5.29 | < **.0001** |
| Frication (Sib) | -533.93 (143.67)<br>(-816.04, -251.81) | -3.72 | < **.0002** |
| Voicing:Frication | 603.83 (194.47)<br>(221.95, 985.70) | 3.1 | **.002** |

indicating that, with the effect of voicing and first several harmonics removed due to the 1500 Hz high pass filter, the distribution of noise in the high frequency range is the same. The estimates for [f] and [v], on the other hand, differ significantly, and [v] has a lower spectral centroid than [f]. Of note is that the centroid measure for [v] does not differ from sibilant centroid measures, but the centroid for [f] is higher than all other segments. The discrepancy in centroid values between [v] and [f] indicates that, even with the high pass filter in place, the distribution of noise in the high frequency range for spirants is modulated to a greater degree for Greek spirant [v] than Greek sibilant [z].




(a) Estimates and CIs

(b) Interaction plot of LS means

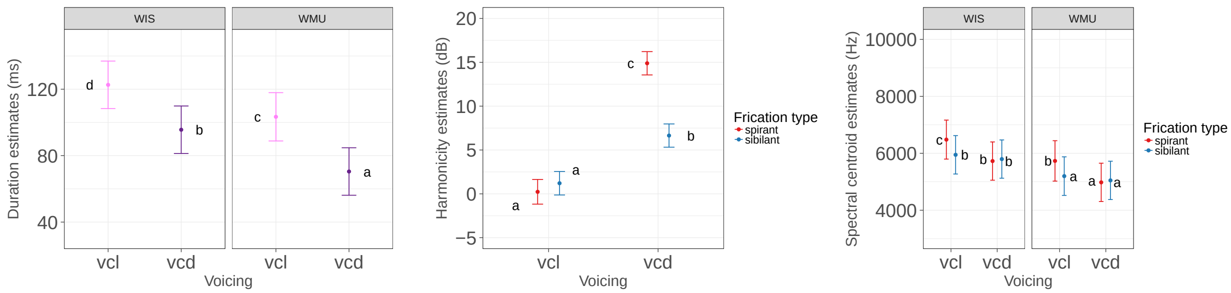
Figure 2.15: Estimates and profile confidence intervals of fixed effects for fitted model of spectral centroid on the 1500 Hz high pass filter in Greek (2.15a). Interaction plot of least squares means with 95% Tukey-adjusted comparison confidence intervals; means sharing a letter are not significantly different (2.15b)

### 2.7.4 Greek: summary of point measures

Figure 2.16 repeats the three interaction plots of duration Figure 2.16a, harmonicity Figure 2.16b, and spectral centroid Figure 2.16c for ease of reference.

The results of this section showed that Greek voiced fricatives are shorter than their voiceless counterparts, and all fricatives are shorter in the WMU position compared to the WIS stressed.[18] Type of frication was not found to be significant for duration. Both voicing and sibilance were found to increase harmonicity, but environment did not have an effect. Figure 2.16b shows that there is no difference in harmonicity between the voiceless fricatives, and that the voiced spirant [v] has the highest harmonicity. With respect to spectral centroid calculated on the 1500 Hz high pass filter, all three factors were significant, as well as the interaction between voicing and frication. As seen in Figure 2.16c, there is no difference in

[18] Nirgianaki (2014) found that sibilants were longer than non-sibilants using a measure of normalized duration, rather than the raw duration measure used here. Since **?** used a different measure and also included the segments [θ, ð, ɣ, x], it is difficult to compare results explicitly.




(a) Duration (ms) (b) Harmonicity (dB) (c) Spectral centroid (Hz)

Figure 2.16: Interaction plots of least squares means with 95% Tukey-adjusted comparison confidence intervals for duration, harmonicity, and spectral centroid on the 1500 Hz high pass filtered signal in Greek (repeated). Means sharing a letter are not significantly different

centroid values between the voiced and voiceless sibilants within a given environment, but there is a difference between [f] and [v] within each environment, with [v] having a lower centroid measure.

## 2.8 Results: Serbian

### 2.8.1 Duration

Figure 2.17 shows violin plots for duration by segment in each environment in Serbian.

The final fitted model for *duration* included `Environment`, `Voicing`, and `Frication Type` as fixed effects, and the three-way interaction between them; `Speaker` was specified as a random effect. `Voicing` and `Environment` had a negative effect, meaning that voiced segments were shorter than voiceless segments, and all segments were shorter in the WMU environment than in the WIS environment. The two-way interaction between `Environment` and `Voicing` shows that the duration of voiced fricatives is reduced in the word-medial unstressed environment more than that of voiceless fricatives. `Frication Type` had a positive

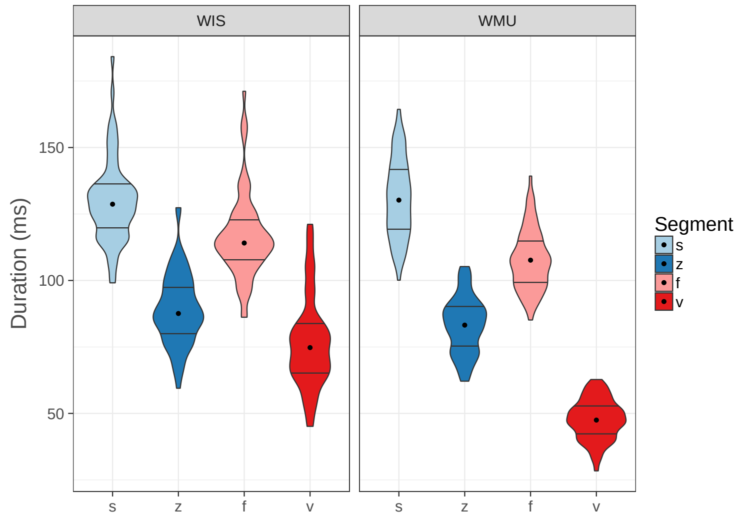



Figure 2.17: Violin plots of duration (ms) in Serbian by environment

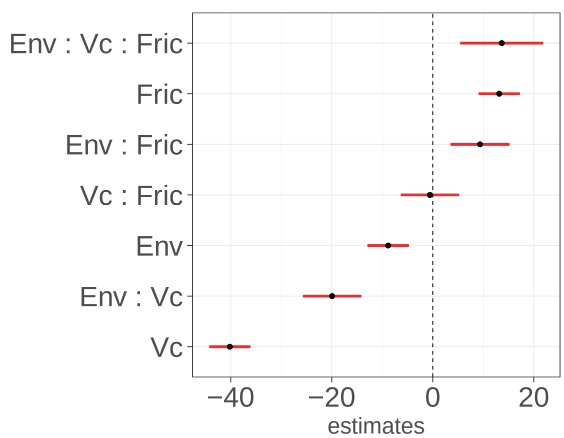
effect on duration, meaning that Serbian sibilants were longer than spirants. The two-way interaction between `Frication Type` and `Environment` is positive, indicating that sibilants do not shorten as much as spirants do in the WMU environment. The three-way interaction between environment, voicing and type of frication is positive, a result of the different effect that environment has on [z] vs. [v]; specifically, [z] is longer in the word-medial environment than [v].

Figure 2.18b shows the least squares means estimates and confidence intervals for each factor level. Within each environment, duration estimates for each segment differ from every other.[19] Voiced segments are shorter than voiceless segments; the interaction with environment means that [v, z] are more prone to shortening in WMU position than voiceless fricatives. The three-way interaction term manifests in the extreme shortening of [v] in the WMU environment compared to [z].

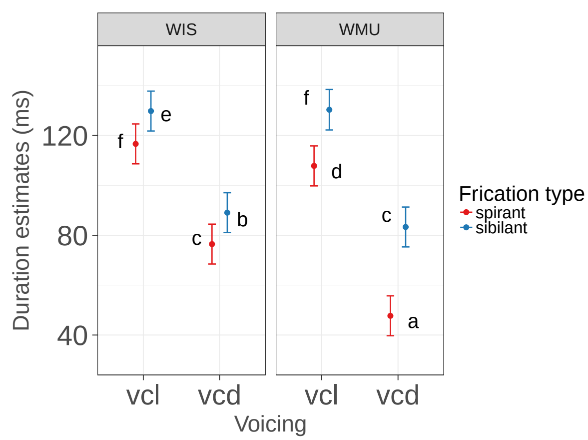
[19]The fact that, say, the duration of [v] in WIS position is equal to the duration of [z] in WMU environment is not relevant.

Table 2.10: Fitted model of duration (ms) in Serbian: Estimates shown with standard errors in parentheses; lower and upper bounds of 95% confidence intervals shown in parentheses underneath estimate and standard error.

| | $\beta$ (SE)<br>95% CI | t-value | p-value |
|---|---|---|---|
| (Intercept) | 116.64 (2.51)<br>(-12.94, 121.96) | 46.53 | $<$ **.0001** |
| Env. (WMU) | -8.83 (2.09)<br>(-23.05, -4.72) | -4.22 | $<$ **.0001** |
| Voicing (Vcd) | -40.18 (2.08)<br>(-44.27, -36.08) | -19.27 | $<$ **.0001** |
| Frication (Sib) | 13.16 (2.08)<br>(9.07, 17.25) | 6.31 | $<$ **.0001** |
| Env.:Voicing | -19.93 (2.95)<br>(-25.73, -14.13) | -6.75 | $<$ **.0001** |
| Env.:Frication | 9.36 (2.98)<br>(3.50, 15.22) | 3.14 | **.0018** |
| Voicing:Frication | -0.57 (2.95)<br>(-6.36, 5.22) | -0.19 | .8478 |
| Env.:Voicing:Frication | 13.66 (4.19)<br>(5.42, 21.89) | 3.26 | **.0012** |




(a) Estimates and CIs




(b) Interaction plot of LS means

Figure 2.18: Estimates and profile confidence intervals of fixed effects for fitted model of duration (ms) in Serbian (2.18a). Interaction plot of least squares means with 95% Tukey-adjusted comparison confidence intervals; means sharing a letter are not significantly different (2.18b)

### 2.8.2 Harmonicity

Figure 2.19 shows violin plots for harmonicity by segment in each environment in Serbian.

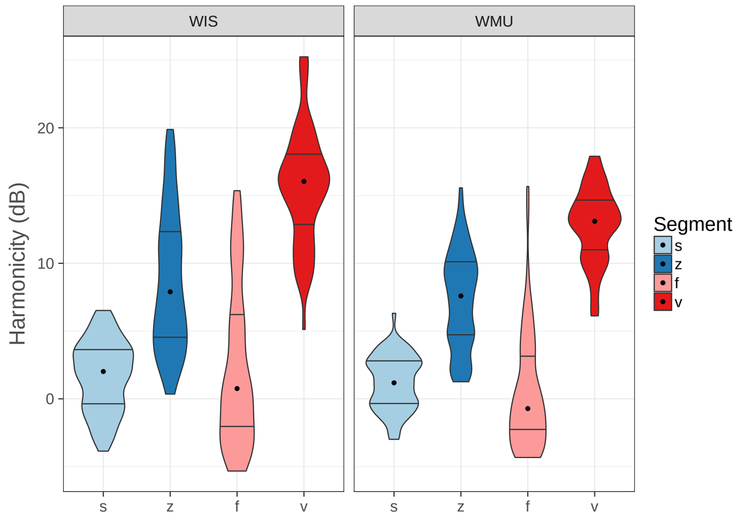


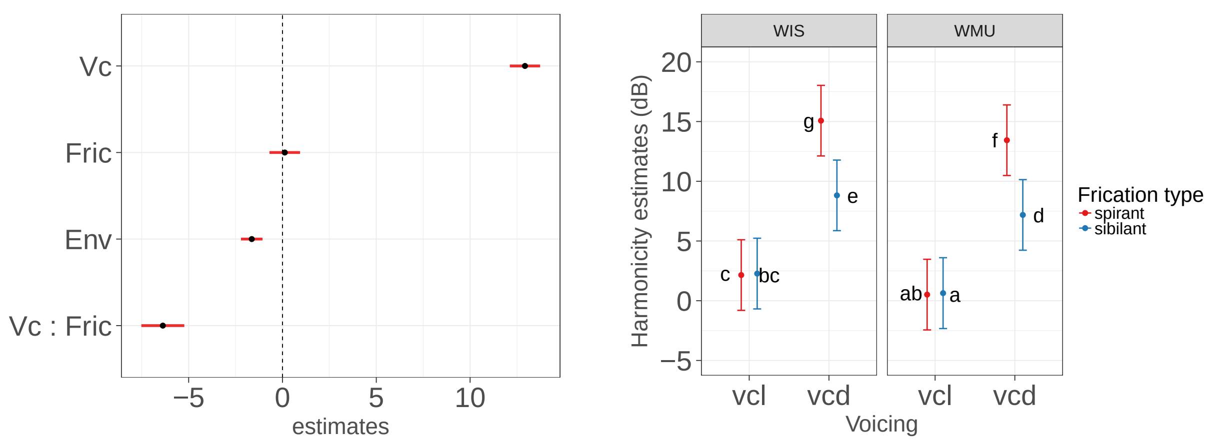

Figure 2.19: Violin plots of harmonicity (dB) in Serbian by environment

The final fitted model for *harmonicity* in Serbian was fit with the `Environment`, `Voicing`, and `Frication Type` and the interaction between `Voicing` and `Frication Type`; `Speaker` was specified as a random effect. `Frication Type` was not significant in the final model, but was retained because of its interaction with `Voicing`. `Voicing` had a positive effect on harmonicity, hence voiced [v, z] were more harmonic than voiceless [f, s]. The interaction between `Voicing` and `Frication Type` was negative, reflecting the lower harmonicity of sibilant [z] compared to spirant [v]. `Environment` had a negative effect on harmonicity, meaning fricatives in the WMU environment had lower harmonicity than in the WIS environment.

Figure 2.20b shows the least squares means estimates and confidence intervals for harmonicity in Serbian. The voiced spirant [v] has the highest harmonicity, followed by [z], followed by [f, s]. The harmonicity of [f, s] is not distinguished within environments. All segments have lower harmonicity in the WMU environment over the WIS environment.

Table 2.11: Fitted model of harmonicity (dB) in Serbian: Estimates shown with standard errors in parentheses; lower and upper bounds of 95% confidence intervals shown in parentheses underneath estimate and standard error.

| | $\beta$ (SE)<br>95% CI | t-value | p-value |
|---|---|---|---|
| (Intercept) | 2.15 (0.84)<br>(0.30, 4.00) | 2.57 | **0.0303** |
| Env (WMU) | -1.64 (0.29)<br>(-2.21, -1.06) | -5.61 | $<$ **.0001** |
| Voicing (Vcd) | 12.92 (0.41)<br>(12.12, 13.73) | 31.44 | $<$ **.0001** |
| Frication (Sib) | 0.12 (0.42)<br>(-0.69, 0.94) | 0.3 | 0.7676 |
| Voicing:Frication | -6.38 (0.58)<br>(-7.52, -5.23) | -10.92 | $<$ **.0001** |




(a) Estimates and CIs

(b) Interaction plot of LS means

Figure 2.20: Estimates and profile confidence intervals of fixed effects for fitted model of harmonicity (dB) in Serbian (2.20a). Interaction plot of least squares means with 95% Tukey-adjusted comparison confidence intervals; means sharing a letter are not significantly different (2.20b)

### 2.8.3 Spectral centroid

Figure 2.21 shows violin plots for spectral centroid on the 1500 Hz high-pass filtered signal by segment in each environment in Serbian.

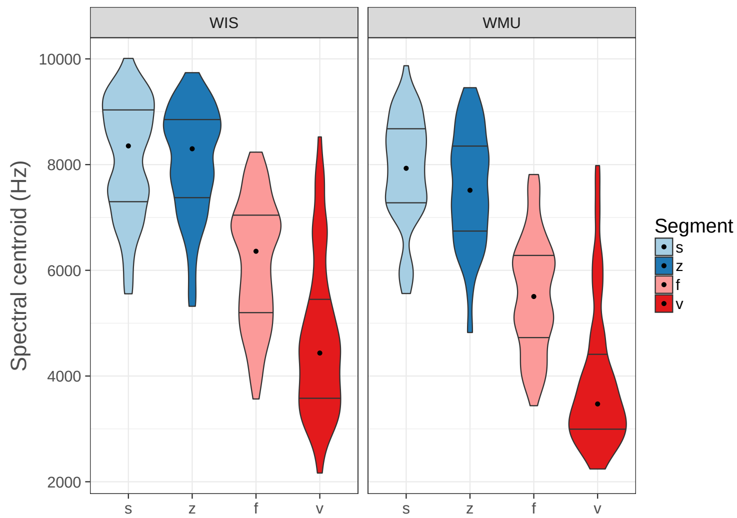



Figure 2.21: Violin plots of spectral centroid (Hz) on the 1500 Hz high pass filtered signal in Serbian by environment

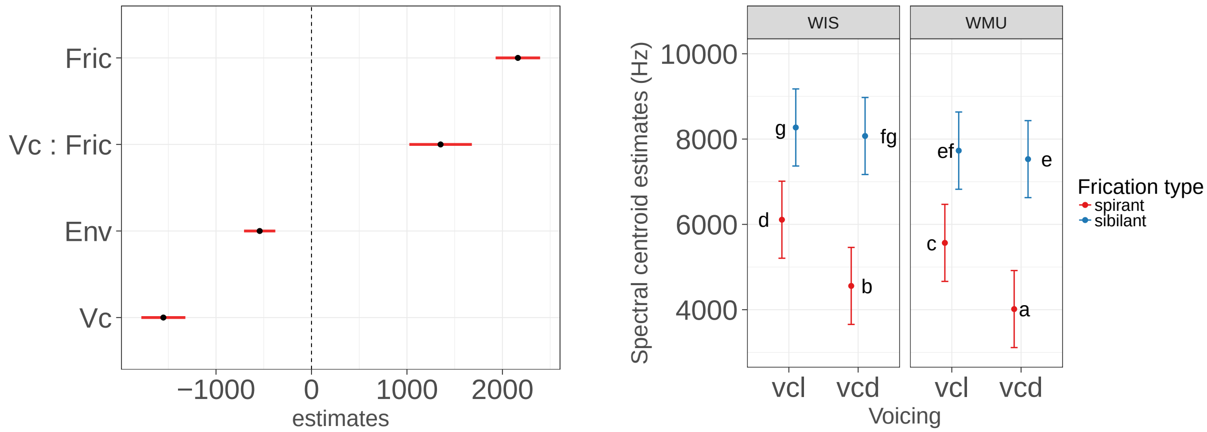
The final fitted model was fit with `Frication Type`, `Voicing`, `Environment`, as well as the interaction between `Frication Type` and `Voicing`; `Speaker` was specified as a random effect. The effect of `Frication Type` was positive, so Serbian sibilants have higher spectral centroids than Serbian spirants. The effect of `Voicing` was negative, meaning that voiced fricatives have lower centroids than voiceless fricatives. The interaction between `Frication Type` and `Voicing` was positive, indicating that [z] has a higher centroid than [v]. The effect of `Environment` was negative, thus fricatives in the WMU position had lower spectral centroid values than those in WIS position.

The least squares means and confidence intervals of Serbian spectral centroid for each factor level is shown in Figure 2.22b. The estimates for [s] and [z] do not differ significantly

Table 2.12: Fitted model of of spectral centroid (Hz) on the 1500 Hz high-pass filtered signal in Serbian: Estimates shown with standard errors in parentheses; lower and upper bounds of 95% confidence intervals shown in parentheses underneath estimate and standard error.

| | $\beta$ (SE)<br>95% CI | t-value | p-value |
|---|---|---|---|
| (Intercept) | 6109.98 (253.76)<br>(5548.70, 6671.11) | 24.08 | < **.0001** |
| Env (WMU) | -542.74 (83.39)<br>(-706.47, -379.01) | -6.51 | < **.0001** |
| Voicing (Vcd) | -1551.83 (117.46)<br>(-1782.45, -1321.20) | -13.21 | < **.0001** |
| Frication (Sib) | 2161.96 (118.61)<br>(1929.07, 2394.84) | 18.23 | < **.0001** |
| Voicing:Frication | 1352.36 (166.78)<br>(1024.90, 1679.82) | 8.11 | < **.0001** |

from each other in either environment, indicating that, with the effect of voicing and first several harmonics removed due to the 1500 Hz high pass filter, the distribution of noise in the high frequency range is the same. The estimates for [f] and [v], on the other hand, differ significantly, and [v] has a lower spectral centroid than [f] in both environments.




(a) Estimates and CIs

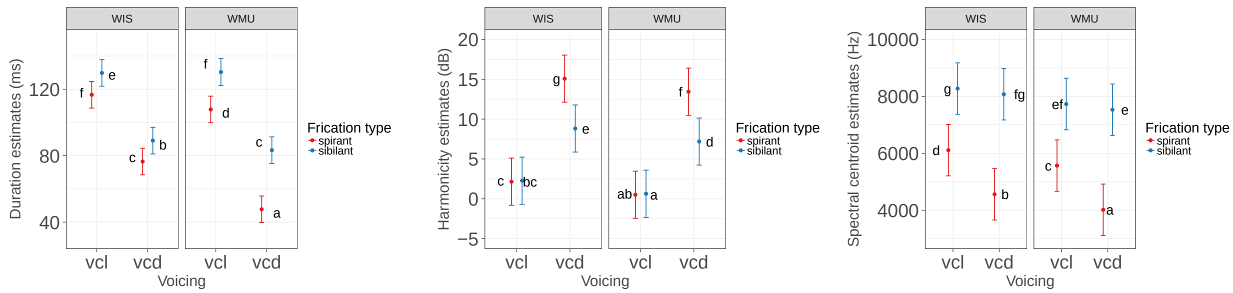
(b) Interaction plot of LS means

Figure 2.22: Estimates and profile confidence intervals of fixed effects for fitted model of spectral centroid on the 1500 Hz high pass filter in Serbian (2.22a). Interaction plot of least squares means with 95% Tukey-adjusted comparison confidence intervals; means sharing a letter are not significantly different (2.22b)

### 2.8.4 Serbian: summary of point measures

Figure 2.23 repeats the three interaction plots of duration Figure 2.23a, harmonicity Figure 2.23b, and spectral centroid Figure 2.23c for ease of reference.

The results of this section show that in Serbian, voiced fricatives are shorter in duration, have higher harmonicity, and have lower spectral centroids than voiceless fricatives. Sibilant fricatives were longer and had higher spectral centroids than spirant fricatives. Environment was a significant factor for all three measures, but its effect differed by measure. With respect to harmonicity and spectral centroid, environment had a negative effect, but did not interact with other measures. For duration, the effect of environment was significant, as were the two-way interactions between it and the other two measures, as well as the three-way interaction term. The effect is most readily seen in how much shorter [v] is in the WMU environment compared to when in the WIS environment; this difference is not seen in the other segments. The two-way interaction between voicing and frication type was significant




(a) Duration (ms) (b) Harmonicity (dB) (c) Spectral centroid (Hz)

Figure 2.23: Interaction plots of least squares means with 95% Tukey-adjusted comparison confidence intervals for duration, harmonicity, and spectral centroid on the 1500 Hz high pass filtered signal in Serbian (repeated). Means sharing a letter are not significantly different

for both harmonicity and spectral centroid, and in both cases, the driving force of the effect is the discrepancy between [v] and [z]. Specifically, [v] has higher harmonicity than all other segments and it has a much lower spectral centroid than all other segments.

## 2.9 Results: Russian

### 2.9.1 Duration

Figure 2.24 shows violin plots for duration by segment in each environment in Russian; recall that WMU [s] is in fact stressed.

The final model selected was fit with `Voicing`, `Frication Type`, `Environment`, and the three-way interaction between them; `Speaker` was specified as a random effect. Note that `Frication Type` and the two-way interaction between `Environment` and `Voicing` were significant at the $\alpha = .1$ level, rather than the usual threshold of $\alpha = .05$; they were retained in the model because both are involved in higher order interactions. The final fitted model had a slightly higher BIC than the model that specified `Voicing`, `Frication Type`,

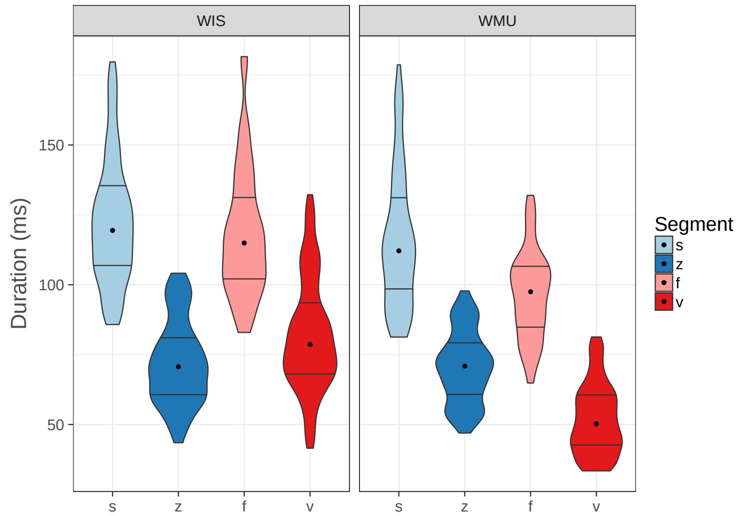



Figure 2.24: Violin plots of duration (ms) in Russian by environment

`Environment`, and the two-way interaction between `Environment` and `Frication Type`. The more complex model was chosen between the BIC between the two differed by only 0.601547, and the AIC of the full model was lower than the AIC of the model without the three-way interaction.[20]

`Voicing` had a negative effect on duration, meaning that voiced fricatives were shorter than voiceless fricatives. `Frication Type` was significant at the $\alpha = .1$ level, and showed that sibilants were slightly longer than spirants. `Environment` had a negative effect, indicating that fricatives were shorter in the WMU environment.

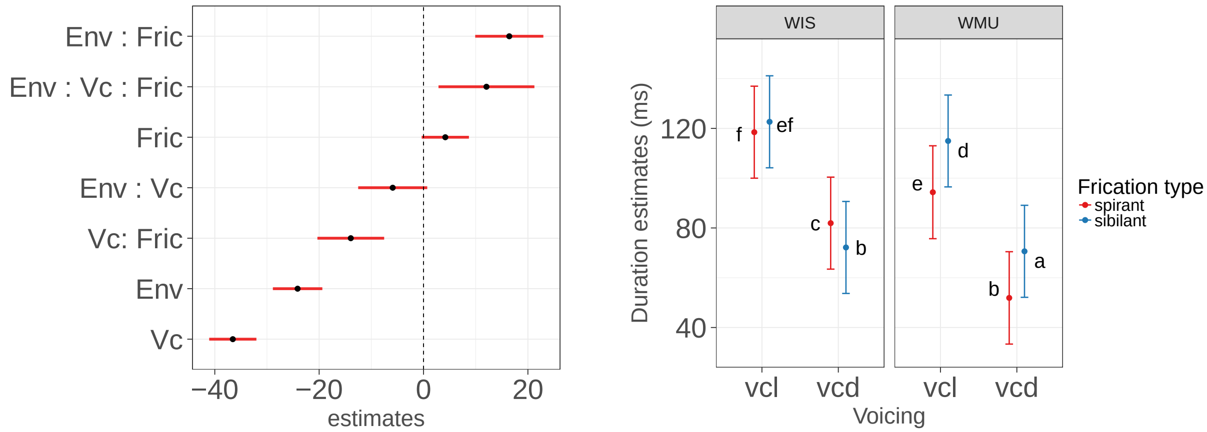
The interpretation of some of the interactions is somewhat subtle, but facilitated by referring to Figure 2.25b. The two-way interaction between `Voicing` and `Frication Type` was negative, meaning that voiced sibilants were shorter than voiced spirants, but the three-way interaction that includes `Environment` was positive, meaning that [z] was longer than

[20] Direct model comparison using anova is not appropriate in this situation because the three-way interaction introduces two extra degrees of freedom, specifically the two-way interactions between (1) `Voicing` and `Frication Type`, and (2) `Voicing` and `Environment`. Since the AIC of the full model was 12.30631 points lower than the AIC of the less complex model, this was deemed a sufficient improvement in model fit.

Table 2.13: Fitted model of duration (ms) in Russian: Estimates shown with standard errors in parentheses; lower and upper bounds of 95% confidence intervals shown in parentheses underneath estimate and standard error.

| | $\beta$ (SE)<br>95% CI | t-value | p-value |
|---|---|---|---|
| (Intercept) | 118.48 (5.13)<br>(107.09, 129.87) | 23.1 | < **.0001** |
| Env. (WMU) | -24.13 (2.41)<br>(-28.87, -19.39) | -10 | < **.0001** |
| Voicing (Vcd) | -36.55 (2.3)<br>(-41.07, -32.02) | -15.86 | < **.0001** |
| Frication (Sib) | 4.17 (2.3)<br>(-0.35, 8.70) | 1.81 | .0706 |
| Env.:Voicing | -5.91 (3.37)<br>(-12.52, 0.71) | -1.75 | .0802 |
| Env.:Frication | 16.43 (3.33)<br>(9.90, 22.96) | 4.94 | < **.0001** |
| Voicing:Frication | -13.94 (3.26)<br>(-20.34, -7.54) | -4.28 | < **.0001** |
| Env.:Voicing:Frication | 12.06 (4.68)<br>(2.87, 21.24) | 2.58 | **.0102** |

[v] in the WMU position. Indeed, as seen in Figure 2.25b, [z] is shorter than [v] in the WIS environment, but longer than [v] in the WMU environment. Similarly, while [f] has a duration value that does not differ significantly from [s] in the WIS environment, [f] is shorter than [s] in the WMU environment.




(a) Estimates and CIs (b) Interaction plot of LS means

Figure 2.25: Estimates and profile confidence intervals of fixed effects for fitted model of duration (ms) in Serbian (2.25a). Interaction plot of least squares means with 95% Tukey-adjusted comparison confidence intervals; means sharing a letter are not significantly different (2.25b)

### 2.9.2 Harmonicity

Figure 2.26 shows violin plots for harmonicity by segment in each environment in Russian; recall that WMU [s] is in fact stressed.

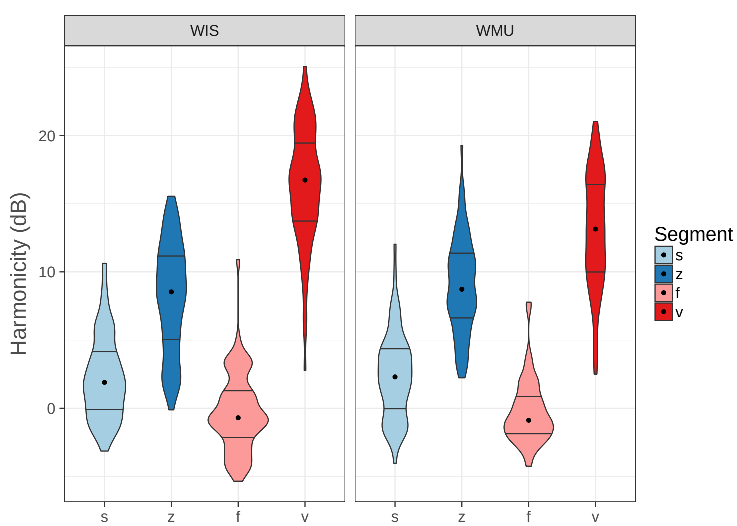


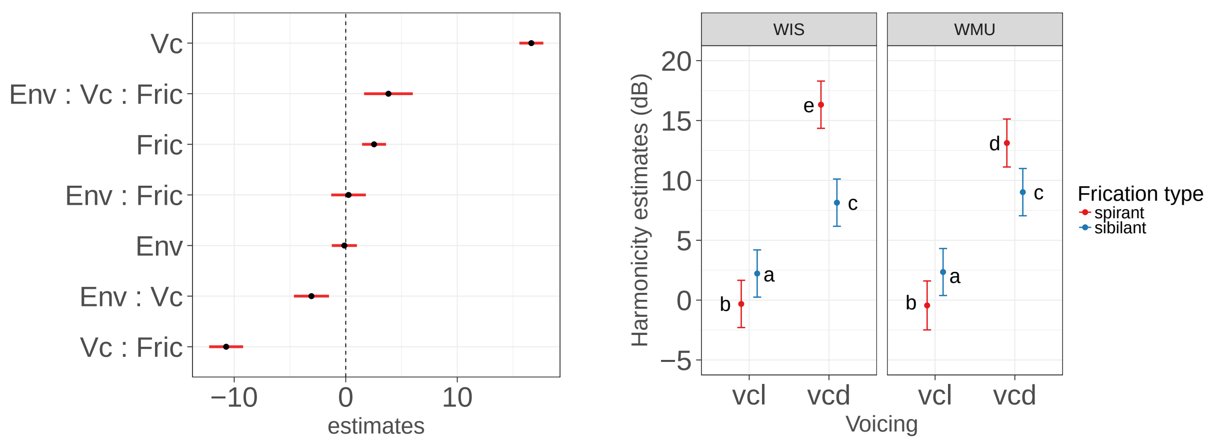

Figure 2.26: Violin plots of harmonicity (dB) in Russian by environment

The final fitted model for *harmonicity* included `Voicing`, `Frication Type`, `Environment` and the three-way interaction between them; `Speaker` was specified as a random effect. `Voicing` and `Frication Type` had positive effects on harmonicity, meaning that voiced fricatives were more harmonic than voiceless segments, and sibilants were more harmonic than spirants. The interaction between `Voicing` and `Frication Type` was negative, indicating that the voiced sibilant [z] had lower harmonicity than the voiced spirant [v]. `Environment` was not significant in the final model, nor was the interaction between `Environment` and `Frication Type`. The interaction between `Environment` and `Voicing` was negative, thus voiced segments had lowered harmonicity in the WMU environment compared to the WIS environment. The three-way interaction between `Voicing`, `Frication Type`, and `Environment` was positive, meaning that the harmonicity of [z] was higher in WMU position than the harmonicity of [v].

The interpretation of model effects is facilitated by considering the least squares means plot of estimates and confidence intervals for factor levels. Within each environment, the order of segments from most harmonic to least harmonic is [v] > [z] > [s] > [f]. Across both environments, the harmonicity of [f, z] do not change, but the harmonicity of [v] is unique among the segments in that it differs between the two environments, and is lower in the WMU environment compared to the WIS environment.

Table 2.14: Fitted model of harmonicity (dB) in Russian: Estimates shown with standard errors in parentheses; lower and upper bounds of 95% confidence intervals shown in parentheses underneath estimate and standard error.

| | $\beta$ (SE)<br>95% CI | t-value | p-value |
|---|---|---|---|
| (Intercept) | -0.32 (0.63)<br>(-1.64, 1.00) | -0.51 | .6184 |
| Env. (WMU) | -0.13 (0.57)<br>(-1.25, 1.00) | -0.22 | .8265 |
| Voicing (Vcd) | 16.64 (0.55)<br>(15.57, 17.72) | 30.39 | < **.0001** |
| Frication (Sib) | 2.54 (0.55)<br>(1.46, 3.61) | 4.64 | < **.0001** |
| Env.:Voicing | -3.08 (0.8)<br>(-4.65, -1.50) | -3.84 | **.0001** |
| Env.:Frication | 0.25 (0.79)<br>(-1.30, 1.80) | 0.32 | 0.7513 |
| Voicing:Frication | -10.72 (0.77)<br>(-12.24, -9.20) | -13.84 | < **.0001** |
| Env.:Voicing:Frication | 3.83 (1.11)<br>(1.64, 6.01) | 3.44 | **.0006** |




(a) Estimates and CIs

(b) Interaction plot of LS means

Figure 2.27: Estimates and profile confidence intervals of fixed effects for fitted model of harmonicity (dB) in Russian (2.27a). Interaction plot of least squares means with 95% Tukey-adjusted comparison confidence intervals; means sharing a letter are not significantly different (2.27b)

### 2.9.3 Spectral centroid

Figure 2.28 shows violin plots for spectral centroid on the 1500 Hz high-pass filtered signal, by segment in each environment in Russian; recall that WMU [s] is in fact stressed.

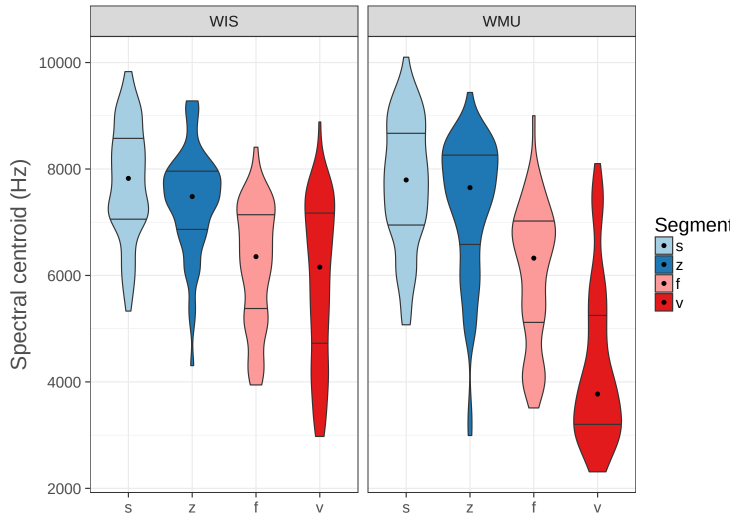
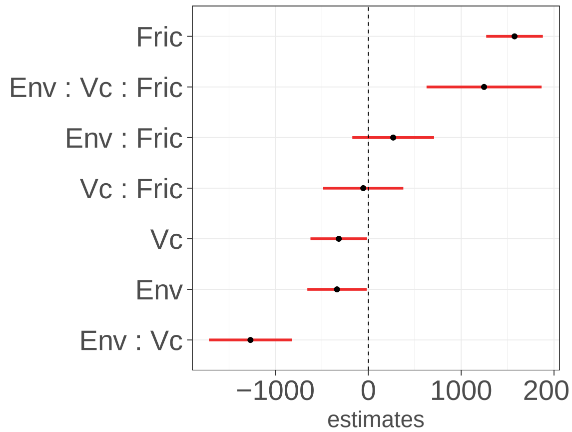



Figure 2.28: Violin plots of spectral centroid (Hz) in Russian by environment

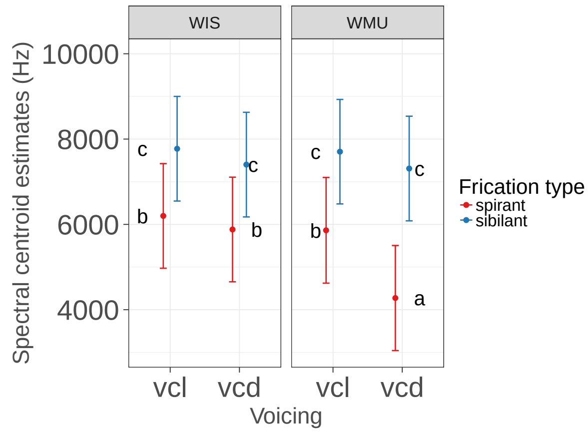
The final fitted model for *spectral centroid* on the 1500 Hz high pass filtered signal included the fixed effects of `Environment`, `Voicing`, and `Frication Type`, as well as the three-way interaction between them; `Speaker` was specified as a random effect. `Frication Type` had a positive effect, thus sibilants had higher centroid measures than spirants. `Voicing` had a negative effect, thus voiced segments had lower centroid measures than voiceless segments. The two-way interaction between `Frication Type` and `Voicing` was not significant. `Environment` had a negative effect, meaning that all segments had lower centroid measures in the WMU position compared to the WIS position. The interaction between `Frication Type` and `Environment` was not significant. The two-way interaction between `Environment` and `Voicing` was negative, while the three-way interaction was positive. The interpretation of these interactions is facilitated by the least squares means plot in Figure 2.29b. Voiced and voiceless sibilants do not differ significantly from each other with respect to spectral centroid, not just within a given environment, but across environments as well. In WIS position, [f] and [v] have lower centroids than the sibilants, but do not differ from each other; [f] remains constant in the WMU position, and like the sibilants shows no effect of environment. This contrasts with [v], which has a lower spectral centroid in the WMU environment, and differs significantly from [f].

Table 2.15: Fitted model of of spectral centroid (Hz) on the 1500 Hz high-pass filtered signal in Russian: Estimates shown with standard errors in parentheses; lower and upper bounds of 95% confidence intervals shown in parentheses underneath estimate and standard error.

| | $\beta$ (SE)<br>95% CI | t-value | p-value |
|---|---|---|---|
| (Intercept) | 6198.91 (340.92)<br>(5442.28, 6955.54) | 18.18 | < **.0001** |
| Env. (WMU) | -337.26 (162.78)<br>(-656.88, -17.61) | -2.07 | **.0388** |
| Voicing (Vcd) | -317.99 (155.38)<br>(-623.07, -12.91) | -2.05 | **0.0412** |
| Frication (Sib) | 1574.9 (155.38)<br>(1269.83, 1879.98) | 10.14 | < **.0001** |
| Env.:Voicing | -1269.34 (227.25)<br>(-1715.57, -823.15) | -5.59 | **.0001** |
| Env.:Frication | 268.1 (224.27)<br>(-172.26, 708.43) | 1.2 | 0.2324 |
| Voicing:Frication | -54.35 (219.74)<br>(-485.79, 377.09) | -0.25 | 0.8047 |
| Env.:Voicing:Frication | 1246.77 (315.58)<br>(627.16, 1866.42) | 3.95 | **.0004** |




(a) Estimates and CIs



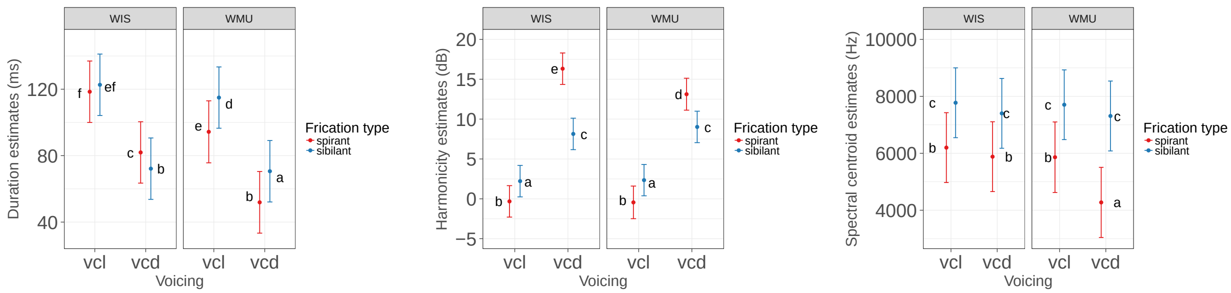

(b) Interaction plot of LS means

Figure 2.29: Estimates and profile confidence intervals of fixed effects for fitted model of spectral centroid on the 1500 Hz high pass filter in Russian (2.29a). Interaction plot of least squares means with 95% Tukey-adjusted comparison confidence intervals; means sharing a letter are not significantly different (2.29b)

### 2.9.4 Russian: summary of point measures

Figure 2.30 repeats the three interaction plots of duration Figure 2.30a, harmonicity Figure 2.30b, and spectral centroid Figure 2.30c for ease of reference.

This section showed that in Russian, voiced fricatives are shorter in duration, have higher harmonicity, and have lower spectral centroids than voiceless fricatives. Sibilant fricatives are somewhat longer, have higher harmonicity, and have higher spectral centroids than spirant fricatives. Environment had a negative effect on duration and harmonicity, meaning that segments were shorter, and had lower harmonicity when in the WMU environment compared to the WIS environment. For spectral centroid, the effect of environment was driven entirely by [v]. For all three measures, the full model was fit, meaning that the three-way interaction between voicing, frication, and environment was significant. In all cases, this interaction showed that measures of [v] were more affected by the environment than were the measures of other segments.




(a) Duration (ms) (b) Harmonicity (dB) (c) Spectral centroid (Hz)

Figure 2.30: Interaction plots of least squares means with 95% Tukey-adjusted comparison confidence intervals for duration, harmonicity, and spectral centroid on the 1500 Hz high pass filtered signal in Russian (repeated). Means sharing a letter are not significantly different

## 2.10 Discussion

This chapter investigated the relationship between voicing and the type of frication in English, Greek, Russian, and Serbian using the four segments typically transcribed as [f, v, s, z]. It also investigated the effect of word-internal prosody by comparing the realization of the voiced and voiceless sibilant and spirant fricatives in two environments: prosodically strong word-initial stressed (WIS) position, and prosodically weak word-medial unstressed (WMU) position. Three measures that are known to discriminate fricatives were used, *duration*, *harmonicity*, and the *spectral centroid* calculated on the 1500 Hz high pass filtered signal.

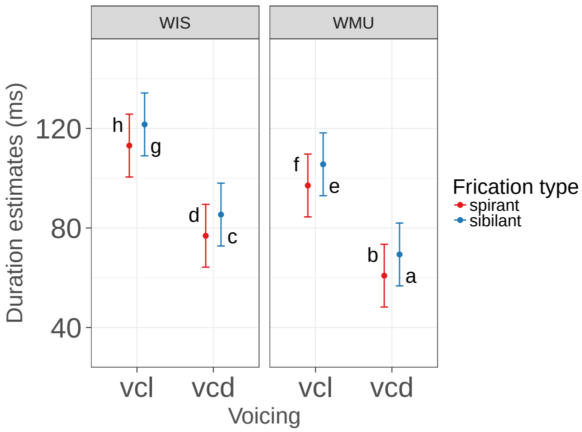
Various cross-linguistic generalizations can be made: in every language, voiced fricatives [v, z] were shorter than voiceless fricatives [f, s], and segments were longer in the prosodically strong WIS environment than in the weak WMU environment. Harmonicity results were generally consistent across languages: [v] is the most harmonic segment, followed by [z], followed by the voiceless fricatives. In every language, the spectral centroid of the voiceless and voiced sibilants was the same, suggesting that, modulo the effect of voicing, the distribution of noise in the sibilants is the same. Such generalizations are consistent with previous results (e.g., Jongman et al. (2000) and Maniwa et al. (2009) report similar findings for fricatives in English, Jesus and Shadle (2002) for Portuguese, Nirgianaki (2014) for Greek).

How do these results bear on the *relationship* between voicing and frication type, as well as the role of environment in mediating this relationship? Consider, by way of example, the results of duration in English; for convenience, the least squares means interaction plot is shown below in Figure 2.31a. Recall that in English, the most parsimonious model for duration included `Voicing`, `Frication Type`, and `Environment`, but none of the interactions between them. Voiced fricatives [v, z] were shorter than voiceless fricatives [f, s], sibilants [s, z] were longer than spirants [f, v], and all fricatives were shorter in the WMU environment compared to the WIS environment. Because there is no interaction between voicing and frication type, we can think of them as independent with respect to duration; in this respect, we can say that [v] is the voiced counterpart to [f] much as [z] is the voiced counterpart to [s]. In fact, we could likewise say that [s] is to [f] as [z] is to [v], although this interpretation is disfavoured since it is not clear how to interpret [s] as the "sibilant counterpart of f", whatever that might mean. With respect to environment, because it does not interact with either voicing or frication type, it does not change the relationship between voicing and frication. That is, the independence between voicing and frication type obtains, whether in the prosodically strong WIS environment or the prosodically weak WMU environment.

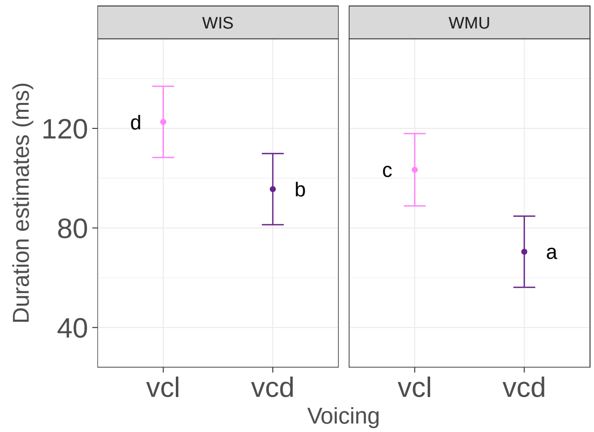
In contrast, the relationship between voicing and frication type for English harmonicity is not independent. Recall that the most parsimonious model for English harmonicity was the full model that included `Voicing`, `Frication Type`, `Environment`, and all interactions between them, though not all factors were significant in the final model. Voiced [v, z] had higher harmonicity than voiceless [f, s], but the significant interaction between voicing and frication indicates that the relationship between these factors is not independent. Specifically, while the voiceless spirant [f] had a lower harmonicity than voiceless sibilant [s], voiced spirant [v] had a higher harmonicity than voiced sibilant [z]. With respect to harmonicity then, we cannot say that [v] is to [f] as [z] is to [s]. Furthermore, the significant three-way interaction between `Voicing`, `Frication Type`, and `Environment` shows that environment mediates the relationship between voicing and frication type. In this case, environment had the effect of shrinking the harmonicity space, though it did not change the overall relationship between segments.

The relationship between voicing and frication thus depends on the measure in question. Regardless of the relationship between voicing and frication type for duration and spectral centroid, harmonicity exhibited a non-independent relationship in every language, evidenced by the interaction between voicing and frication that pointed to the discrepancy between [v] and [z] vs. [f] and [s]. Acoustically, this is because, regardless of subtle variations in place and/or manner that exist within and between languages, the broad-strokes differences between segments are the same: labiodental spirants have a relatively diffuse spectrum of noise, due to the fact that there is little filtering action downstream the place of constriction. Sibilants, on the other hand, are produced in such a way that the noise generated is amplified by the downstream filter. Harmonicity is therefore a global measure that quantifies the relative contribution of high and low frequency energy, which is affected by the overall distribution of energy in the frequency domain.

In spite of the cross-linguistic similarities, harmonicity interacted with environment differently across languages. In Greek, environment had no effect on harmonicity. In Serbian, the WMU environment had a global effect of lowering harmonicity, but with no interactions. Finally, in Russian, the effect of the WMU environment was to decrease the harmonicity of voiced segments, but it had no effect on voiceless segments. In spite of these differences, however, the relative harmonicity was preserved across environments. As I show in Section




(a) English




(b) Greek

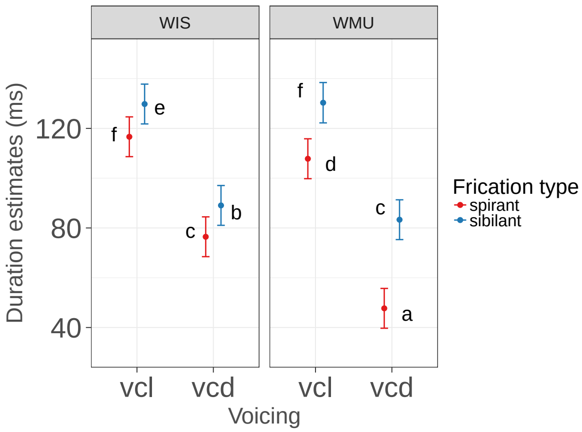
Figure 2.31: Interaction plots of least squares means with 95% Tukey-adjusted comparison confidence intervals for English and Greek duration; means sharing a letter are not significantly different

3.5.1, some of these results can be explained by examining the relationship between harmonicity and devoicing/voicing of fricatives across the two environments. I therefore put aside harmonicity for the remainder of this discussion.

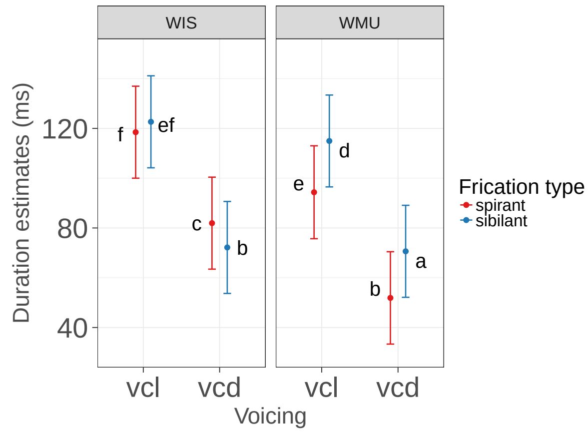
Returning to the relationship between voicing and frication type, we might ask whether independence is always manifested as the absence of an interaction, as it is in English duration. First, consider the case of Greek duration, in which voicing and frication type do not interact for the trivial reason that frication type was not included in the final model at all. Both English and Greek exhibit independence of frication type and voicing, but for different reasons; the least squares means interaction plots are repeated below for reference. In both languages voiced [v, z] are shorter than voiceless [f, s], and in both languages segments are shorter in the WMU environment than in the WIS environment. As discussed previously, in English, sibilants are longer than spirants, but as there is no interaction between voicing and frication type, these two parameters are independent; in Greek, sibilance does not increase raw duration, and so independence follows.

In Serbian, the two-way interaction between `Voicing` and `Frication Type` was not significant ($p = .8474$), but the three-way interaction that included `Environment` was significant. As seen in Figure 2.32a, this amounts to independence in the WIS environment, where sibilants are longer than spirants, and voiced segments are shorter than voiceless, and the relationship between [v] and [f] parallels that between [z] and [s]. In the WMU environment, however, the relationship is not independent: voiced segments are shorter than voiceless segments by more than they were in the WIS environment, but the shortened duration is realized on [v] to a greater extent than [z].

Interpreting the Russian model is more subtle. As seen in Figure 2.32b, voiced segments are shorter than voiceless segments across both environments, but the role of frication type changes. In the WMU environment, spirants are shorter than sibilants. The final fitted model included the three-way interaction term, pointing to a discrepancy between [v] and [z] in the WMU environment, indicating that [v] is shorter than [f] by more than [z] is shorter than [s]; nevertheless, Figure 2.32b does not reveal a large discrepancy between [v] and [z]. I would therefore argue that voicing and frication type are (mostly) independent in the WMU environment with respect to Russian duration. It is not clear what to make of the WIS environment: voiceless [f] and [s] do not differ with respect to each other, but [z] is apparently *shorter* than [v]; nevertheless, the difference, though significant, is slight. I interpret these results as pointing to the fact that frication type is not a factor for duration in Russian in the WIS environment, meaning that voicing and frication type are, in fact, independent in the WIS environment as well. To summarize, duration measures for Russian (mostly) exhibit independence in voicing and frication type in both environments, but the nature of the independence differs: in the WIS environment, the independence is of the kind seen in Greek, where frication type is not relevant; in the WMU environment, the independence is of the kind seen in English, where both voicing and frication type affect




(a) Serbian




(b) Russian

Figure 2.32: Interaction plots of least squares means with 95% Tukey-adjusted comparison confidence intervals for Serbian and Russian duration; means sharing a letter are not significantly different

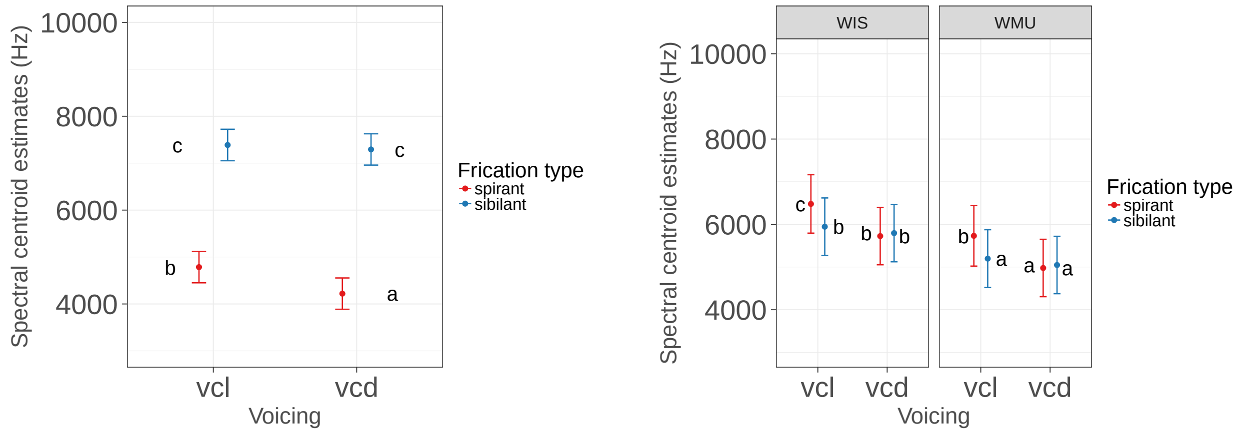
duration, but the parameters are independent.

We now turn to spectral centroid measured on the 1500 Hz high pass filtered signal, beginning with English. First, as seen in Figure 2.33a, the centroid values for [s] and [z] do not differ from each other, indicating that, modulo the effect of voicing, the distribution of energy in the high frequency range is the same. The spirants have generally lower centroid measures, consistent with previous literature on English (Forrest et al., 1988; Jongman et al., 2000; Maniwa et al., 2009), and [v] has a lower spectral centroid than [f]. The interaction between voicing and frication type suggests that, in spite of the high pass filter, the distribution of energy in the high frequency range differs between the voiced and voiceless spirants, and therefore that the relationship between voicing and frication type in English is not independent with respect to spectral centroid. Nevertheless, as I will argue further on, I reject this interpretation and will suggest instead that voicing and frication type exhibit independence for English centroid, particularly when viewed in light of the Serbian and Russian results.

Greek differs from English in that `Environment` was included in the final model; the effect of the WMU environment was to lower the centroid, but as no higher order interactions involving `Environment` were included in the model, the relationship between voicing and frication type is not affected by environment. Like English, voiced and voiceless sibilants do not differ in their centroid values, but in contrast to English, frication type does not affect the spectral centroid values in Greek, and the centroid values for Greek sibilants are lower than the centroid values for English sibilants. Although in both languages the voiced spirant [v] has a lower centroid than [f], in English it is because [v] has a lower centroid value than all other fricatives, while in Greek it is because [f] has a higher centroid value than all other fricatives. Without further articulatory investigations, it is not clear why the voiceless labiodental fricative [f] has a higher spectral centroid than all other fricatives in my data, particularly since I am not aware of any other studies that report a high centroid values for Greek [f].[21] With respect to what these results mean for the relationship between voicing and frication type, I propose to treat Greek like English, and consider the relationship between voicing and frication type to be independent with respect to spectral centroid. Although the voiced and voiceless labiodentals differ from each other, this difference appears to follow from the unusually high centroid value for [f].

Turning now to Serbian and Russian, a very different picture of the relationship between voicing and frication type emerges. In Serbian, like in English and Greek, the centroid measures of the voiced and voiceless sibilants do not differ significantly from each other, indicating that, modulo the effect of voicing, the distribution of noise in the high frequency domain is the same. The significant interaction between `Voicing` and `Frication Type` is evident in the low centroid of [v]. Unlike in Greek and English, there is no sense in which one would interpret these results as independence between the factors: voicing and frication type

[21]For example, Nirgianaki (2014) reports on the spectral moments of Greek fricatives, but does not indicate that [f] is anomalous in any way.

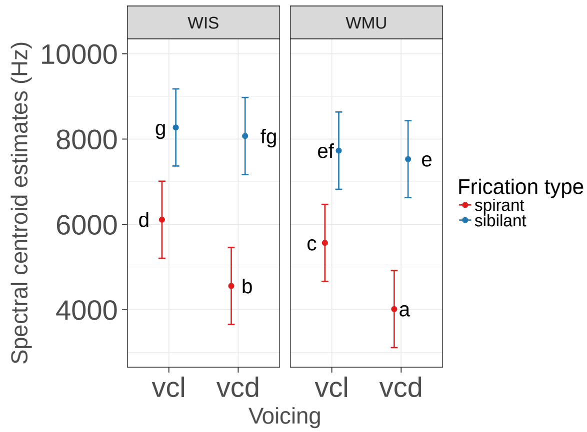



(a) English (b) Greek

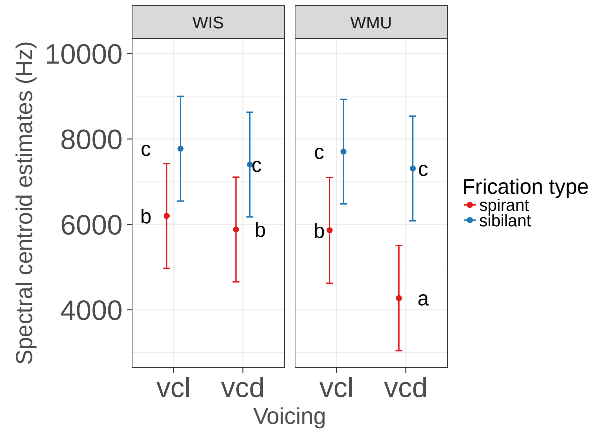
Figure 2.33: Interaction plots of least squares means with 95% Tukey-adjusted comparison confidence intervals for English and Greek; means sharing a letter are not significantly different

interact in Serbian so that the relationship between [f] and [v] does not parallel that between [s] and [z]. The WMU environment had the effect of lowering the centroid of all segments, but did not interact significantly with either of the other factors, and so environment did not change the relationship between voicing and frication type.

The Russian spectral centroid results suggest a different relationship between voicing and frication type, as well as the role of environment. In both environments, like the other three languages considered so far, voiced and voiceless sibilants do not significantly differ from each other, suggesting that modulo the effect of voicing, the distribution of energy in the high frequency range is the same. Unlike the other three languages, Figure 2.34b shows that the centroid values for [v] and [f] also do not differ significantly from each other in the WIS environment. This indicates that voicing and frication type are independent parameters with respect to spectral centroid in the WIS environment. The main effect of environment was not significant, meaning that overall, centroid values did not differ between the WIS and WMU environments, but the significant three-way interaction between is very evident in Figure




(a) Serbian




(b) Russian

Figure 2.34: Interaction plots of least squares means with 95% Tukey-adjusted comparison confidence intervals for English and Serbian; means sharing a letter are not significantly different

2.34b: in the WMU environment, the centroid measure of [v] differs significantly from [f], and is much lower. This indicates that the relationship between voicing and frication type is non-independent in the WMU environment, and so, although `Environment` as a main effect was not significant, environment changes the relationship between voicing and frication type as phonetic parameters of the acoustic space of fricatives.

In broad strokes, the results of this chapter can be summarized by stating that the relationship between voicing and frication type is independent in Greek and English, but non-independent in Serbian and Russian. Greek and English exhibit different kinds of independence, but in both we can interpret [v] as the voiced counterpart to [f] much as [z] is the voiced counterpart to [s]. In Serbian and Russian, voicing and frication type are non-independent: in Serbian, [v] is not the voiced counterpart to [f]; in Russian, [v] is the counterpart to [f] in the WIS position, but not in the WMU position. In the next chapter I turn to testing the hypothesis that the phonological identity of /v/ correlates with its phonetic realization, explicitly testing the cross-linguistic differences suggested here.

CHAPTER 3

# CROSS-LINGUISTIC ACOUSTIC STUDY OF [v]

## 3.1 Introduction

Chapter 2 investigated the relationship between voicing and frication in English, Greek, Russian, and Serbian. I showed that frication and voicing operate as independent factors in English and Greek, though independence is instantiated differently in each; in Serbian and Russian, voicing and frication exhibit interaction effects that depend on the measure in question, and on environment. The present chapter uses the same data, but restricting attention to Greek, Russian, and Serbian, presents the results of an acoustic study designed to test the hypothesis that tokens of [v] are produced with acoustic correlates that reflect their phonological identity.[1] Following the presentation of these results, I discuss how they compare to English.

In most descriptions of Greek, Russian, and Serbian, the voiced, labiodental continuant is transcribed as [v], and so the null hypothesis is that, for any given measure, tokens of [v] in Greek, Russian, and Serbian will not differ significantly from each other.[2] Against this we formulate the alternative hypothesis that, for a given measure, not only will tokens of [v] in each language differ from tokens of [v] in every other, but crucially that the differences will track the phonological identity of /v/ in each language. In the remainder of this section I make this hypothesis precise and explain its motivation.

[1]The study was designed and the data collected for Greek, Russian, and Serbian well before I decided to also collect English data. In the interests of good statistical practice, I test the hypothesis as it was originally articulated.

[2]Morén (2003) notes that Serbian /v/ is not the phonological voiced counterpart to /f/, but but still places /v/ in the row for fricatives; in a later paper, Morén (2006) places /v/ in the 'Glide' row of the consonant chart, but notes that he does this in spite of its "surface fricative property".

If we assume that there is a one-to-one relationship between phonological identity and phonetic realization, then we might expect that the realization of /v/ as more or less of a an obstruent correlates with its phonological patterning. Such an approach is adopted by functionalist analyses of “ambiguous /v/”, which have hypothesized that the intermediate phonological patterning of /v/, namely with respect to obstruents and sonorants, is due to the fact that it is phonetically intermediate with respect to obstruents and sonorants. For example, Padgett (2002) hypothesizes that Russian /v/ is neither a true fricative [v], nor a true approximant [ʋ], but a *narrow approximant*, transcribed as [ʋ̝].

As I showed in Section 1.5, Greek, Russian, and Serbian instantiate such a typology of phonological identity with respect to distribution and patterning: Greek /v/ is categorized as an obstruent, Serbian /v/ is categorized as a sonorant, and in Russian, /v/ patterns with both obstruents and sonorants.

**The notation <*v*>** Throughout the remainder of this chapter I use the notation <*v*> to refer to realizations of /v/. The question at stake is whether the realization of the voiced labiodental spirant is more of an approximant or fricative, and so I use <*v*> instead of [v] or [ʋ], so as not to evoke the standard assumptions on the phonetic realization of a segment transcribed as [v] or [ʋ].

In Section 3.2 I articulate hypotheses for each measure, but first introduce them here in general terms. Throughout, each hypothesis is formulated *within* a particular environment. That is, for a given measure, tokens of <*v*> are compared within either the word-initial stressed (WIS) or the word-medial unstressed (WMU) environment. For ease of exposition, I articulate each hypothesis in general terms, and then include the environment-specific hypotheses below for reference.

**Null Hypothesis** The null hypothesis is that <*v*> in Greek, Russian, and Serbian will not differ in terms of acoustic measures that assess degree of frication and noise distribution. That is, for a given measure, within a given environment, the measure in question will not be statistically different.

With three languages and two environments, there are a number of potential hypotheses, but I consider only those that are phonologically motivated. The null hypothesis is that there is only one phonetic category, and there are two main alternative hypotheses: (1) there are three categories, meaning that each language is distinguished from each other; (2) there are two categories, meaning that two of the languages are not distinguished from each other, but are distinguished from the third.

**Alternative Hypothesis #1** The first alternative hypothesis is that <*v*> will differ with respect to these measures, and that there will be three categories. The nature of the difference hypothesized is phonologically motivated, and the hypothesis is that differences will track phonological identity: Greek <*v*> will be realized with frication characteristic of an obstruent fricative [v]; Serbian <*v*> will be realized without significant frication, characteristic of a sonorant approximant [ʋ]; finally, Russian <*v*> will be realized with frication intermediate to that of Greek and Serbian, as [ʋ̝]. Table 3.1 summarizes the phonological classification of <*v*> in Greek, Russian, and Serbian from Section 1.5, and provides a transcription of the hypothesized realization under this alternative hypothesis.

The hypothesis given in Table 3.1 has not, to by knowledge, been tested.[3] Padgett (2002) discusses the realization of each kind of voiced labiodental continuant in terms of sonority,

[3]Padgett (2002) does not include a phonetic investigation. Kiss and Bárkányi (2006) present results of an acoustic investigation of <*v*> in Hungarian but because they do not include other languages to act as phonological controls, it is not clear whether results that suggest phonetic intermediacy are due to idiosyncracies of Hungarian, or a true correlation with the phonological intermediacy of Hungarian /v/.

Table 3.1: Hypothesized realization of voiced labiodental continuants (based on Padgett (2002))

| | Greek | Russian | Serbian |
|---|---|---|---|
| Phonotactics: | obs. | obs. & son. | son. |
| Undergoes final devoicing? | N/A | ✓ | N/A |
| Undergoes regressive voicing assimilation? | ✓ | ✓ | ✗ |
| Triggers regressive voicing assimilation? | ✓ | ✗ | ✗ |
| Phonetic realization: | [v] | [ʋ̞] | [ʋ] |

where [v] < [ʋ̞] < [ʋ] follows a cline from least sonorous to most sonorous. The primary difference between the three labiodental continuants is hypothesized to be constriction degree: [v] characterized by the narrowest constriction, [ʋ] by the widest constriction, and [ʋ̞] having intermediate constriction. Since, all else being equal, increased constriction degree roughly correlates with increased turbulent airflow, the greater the degree of constriction, the more acoustic frication will characterize the sound (Stevens, 1987). Therefore a more direct assessment of the hypothesis relies not on assessing sonority, for which robust acoustic measures have not been forthcoming, but rather on assessing the degree of frication. For voiced fricatives, assessing frication degree must be done by controlling for the effect of voicing; I approach this in two ways.

**Alternative Hypothesis #1: WIS** Within the WIS environment, tokens of Greek <*v*> will be realized with frication characteristic of an obstruent fricative [v]; Serbian <*v*> will be realized without significant frication, characteristic of a sonorant approximant [ʋ]; Russian <*v*> will be realized with frication intermediate to that of Greek and Serbian, as [ʋ̞].

**Alternative Hypothesis #1: WMU** Within the WMU environment, tokens of Greek <*v*> will be realized with frication characteristic of an obstruent fricative [v]; Serbian <*v*> will be realized without significant frication, characteristic of a sonorant approximant [ʋ]; Russian <*v*> will be realized with frication intermediate to that of Greek and Serbian, as [ʋ̞].

The second alternative hypothesis is that, within a given environment, two of the three languages cluster together. There are three logical possibilities to the two languages that could cluster together, listed in (1) with the potential motivation for the cluster.

(1) Alternative hypothesis #2 possibilities

a. Russian and Serbian group together to the exclusion of Greek
*Motivation:* Both Russian and Serbian are Slavic languages; <*v*> derives from a sonorant in both languages; <*v*> patterns at least partially as a sonorant in both languages. Greek is not a Slavic language; Greek <*v*> derives from the obstruent **b*; <*v*> patterns strictly as an obstruent.

b. Russian and Greek group together to the exclusion of Serbian
*Motivation:* Russian and Greek both exhibit at least partial phonological patterning with obstruents.

c. Greek and Serbian group together to the exclusion of Russian
*Motivation:* ???

There is no phonological reason to hypothesize that Greek and Serbian will pattern together to the exclusion of Russian, so I do not consider this hypothesis going forward. In terms of motivation, there is clearly more motivation to hypothesize that Russian and Serbian will pattern together to the exclusion of Greek, based on the relationship between languages and the common source of <*v*>. Nevertheless, the obstruent patterning of Russian <*v*> does mean that it is possible that it will be realized as an obstruent, together with Greek. Notice that this hypothesis carries with it some additional assumptions of the realization: note that (1a) aligns most naturally with the assumption that Russian and Serbian <*v*> are realized as sonorants, hence [ʋ], while (1c) aligns most naturally with the assumption that Russian and Greek <*v*> are realized as obstruents, hence [v]. These are summarized below:

**Alternative Hypothesis #2a** For a given environment, Russian and Serbian will pattern together to the exclusion of Greek. Specifically, <*v*> in Russian and Serbian will be realized with the lack of frication characteristic of a sonorant [ʋ], while Greek <*v*> will be realized with the frication characteristic of an obstruent [v].

**Alternative Hypothesis #2b** For a given environment, Russian and Greek will pattern together to the exclusion of Serbian. Specifically, <*v*> in Russian and Greek will be realized with the frication characteristic of an obstruent [v], while Serbian <*v*> will be realized with the lack of frication characteristic of a sonorant [ʋ].

Two measures are used to investigate the preceding hypotheses. The first approach directly compares the *harmonicity*, or *harmonics-to-noise ratio*, of <*v*> across languages. Recall that harmonicity measures the relative contribution of low frequency energy and high frequency energy, and therefore controls the effect of voicing relationally. In Chapter 2, <*v*> had the highest harmonicity across all languages, regardless of environment. The question is whether the harmonicity of <*v*> differs between languages. Using harmonicity as a measure of obstruency is motivated by a study by Hamann and Sennema (2005), who show that harmonicity distinguishes fricative and approximant realizations of voiced labiodental continuants in German and Dutch.

The second approach uses the *spectral centroid* calculated over the 1500 Hz high pass filtered signal. The high pass filter has the effect of removing the effect of voicing and first several harmonics, so that voicing is controlled for explicitly, and only the contribution of noise is assessed. Spectral centroid is not, in and of itself, a measure of frication degree, because it is subject to the downstream filter on the airstream. This is obvious, for example, when we consider the sibilants in Greek vs. Russian and Serbian: Greek sibilants [s, z] have a much lower spectral centroid than [s, z] in Russian and Serbian, due to the difference in place of articulation.[4] For spirants [f, v], the effect of place of articulation is not predicted to have as large an effect, since there is not as much available variability with respect to different possible articulatory configurations involving the lips. Nevertheless, in order to control for possible slight differences in the place of articulation, the centroid of <*v*> is assessed relative to [f], in a manner detailed in Section 3.3.2. As I discuss in Section 3.5, this approach has the effect of assessing the degree to which <*v*> is the voiced counterpart of [f], acoustically speaking.

The remainder of this chapter is structured as follows: in Section 3.2 I articulate specific null and alternative hypotheses with respect to *harmonicity*, and *spectral centroid*.[5] In Section 3.3, I detail the statistical models used. Data collection, segmentation, and acoustic measures can be found in Section 2.3. I present the results in Section 3.4, and a discussion in Section 3.5.

[4]This can be seen from comparing mean centroid values of the sibilants in Chapter 2, and by looking at Figure 3.3.

[5]For fuller descriptions of these measures, see Section 2.4.

## 3.2 Hypotheses

A <*v*> produced with both voicing and frication (i.e., as [v]) will have two main peaks in the spectrum: a peak in the low frequency range, due to voicing, and a peak in the high frequency range, due to frication. A <*v*> produced with voicing but little to no frication (i.e., as [ʋ]) will not have a peak in the high frequency range; instead, the peak in the low frequency range will drop off steadily, interrupted by formants which will appear as peaks in the spectrum.

### 3.2.1 Hypothesis: harmonicity

*Harmonicity* (or *harmonics to noise ratio*) is a measure of acoustic periodicity that calculates the relative contribution of low frequency and high frequency energy in the signal. Higher harmonicity corresponds to more of the energy coming from the low frequency domain, and low harmonicity corresponds to more energy coming from the high frequency domain, and so harmonicity is expected to correlate with sonority (see Section 2.4.2 for more details). As Hamann and Sennema (2005) explain, “A harmonicity median of 0 dB means that there is equal energy in the harmonics and in the noise of a signal, and a harmonicity median of 20 dB that there is almost 100% of the energy of the signal in the periodic part (Boersma 1993).”

The alternative hypotheses for harmonicity are:

**Hypothesis 1** (Alternative hypothesis 1: harmonicity)**.**
For a given environment, Greek <*v*> will be characterized by the lowest harmonicity values, Serbian <*v*> will be characterized by the highest harmonicity values, and Russian <*v*> will be characterized by intermediate harmonicity.

i. In the WIS position, Greek $<v>$ will be characterized by the lowest harmonicity values, Serbian $<v>$ will be characterized by the highest harmonicity values, and Russian $<v>$ will be characterized by intermediate harmonicity.

ii. In the WMU position, Greek $<v>$ will be characterized by the lowest harmonicity values, Serbian $<v>$ will be characterized by the highest harmonicity values, and Russian $<v>$ will be characterized by intermediate harmonicity.

**Hypothesis 2** (Alternative hypothesis 2a: harmonicity)**.**

For a given environment, the difference between harmonicity values for Russian and Serbian $<v>$ will not be significant; harmonicity values for Greek $<v>$ will be significantly different than both Russian and Serbian. Furthermore, Greek $<v>$ will have lower harmonicity values than Russian and Serbian $<v>$.

i. In the WIS position, the harmonicity values for Greek $<v>$ will be lower than the harmonicity values for both Russian and Serbian $<v>$, which will not differ from each other.

ii. In the WMU position, the harmonicity values for Greek $<v>$ will be lower than the harmonicity values for both Russian and Serbian $<v>$, which will not differ from each other.

**Hypothesis 3** (Alternative hypothesis 2b: harmonicity)**.**

For a given environment, the difference between harmonicity values for Russian and Greek $<v>$ will not be significant; harmonicity values for Serbian $<v>$ will be significantly different than both Russian and Serbian. Furthermore, Serbian $<v>$ will have higher harmonicity values than Russian and Greek $<v>$.

i. In the WIS position, the harmonicity values for Serbian $<v>$ will be higher than the harmonicity values for both Russian and Greek $<v>$, which will not differ from each other.

ii. In the WMU position, the harmonicity values for Serbian <*v*> will be higher than the harmonicity values for both Russian and Greek <*v*>, which will not differ from each other.

### 3.2.2 Hypothesis: relative spectral centroid

The *spectral centroid* is a measure of the average frequency weighted by the energy; because of the high pass filter, the average is calculated only on the frequency range over 1500 Hz. Because <*v*> is assessed relative to the baseline of [f], comparisons are between $v - f$, not <*v*> alone; henceforth I will refer to this measure as the *relative spectral centroid* of <*v*>. Tokens of <*v*> produced with high frication (as [v]) will be more similar to [f] than tokens of <*v*> produced with low frication (as [ʋ]). As such this measure provides a notion of the similarity of the labiodental spirant pair, modulo the effect of voicing.

The alternative hypotheses for spectral centroid are:

**Hypothesis 4** (Alternative hypothesis 1: spectral centroid)**.**
For a given environment, Greek <*v*> will be characterized by the highest relative spectral centroid values, Serbian <*v*> will be characterized by the lowest relative spectral centroid values, and Russian <*v*> will be characterized by intermediate relative spectral centroid values.

i. In the WIS position, Greek <*v*> will be characterized by the highest relative spectral centroid values, Serbian <*v*> will be characterized by the lowest relative spectral centroid values, and Russian <*v*> will be characterized by intermediate relative spectral centroid .

ii. In the WMU position, Greek <*v*> will be characterized by the highest relative spectral centroid values, Serbian <*v*> will be characterized by the lowest relative spectral centroid values, and Russian <*v*> will be characterized by intermediate relative spectral centroid.

**Hypothesis 5** (Alternative hypothesis 2a: relative spectral centroid)**.**
For a given environment, the difference between relative spectral centroid values for Russian and Serbian <*v*> will not be significant; relative spectral centroid values for Greek <*v*> will be significantly different than both Russian and Serbian. Furthermore, Greek <*v*> will have higher relative spectral centroid values than Russian and Serbian <*v*>.

i. In the WIS position, the relative spectral centroid values for Greek <*v*> will be higher than the relative spectral centroid values for both Russian and Serbian <*v*>, which will not differ from each other.

ii. In the WMU position, the relative spectral centroid values for Greek <*v*> will be higher than the relative spectral centroid values for both Russian and Serbian <*v*>, which will not differ from each other.

**Hypothesis 6** (Alternative hypothesis 2b: relative spectral centroid)**.**
For a given environment, the difference between relative spectral centroid values for Russian and Greek <*v*> will not be significant; relative spectral centroid values for Serbian <*v*> will be significantly different than both Russian and Serbian. Furthermore, Serbian <*v*> will have lower relative spectral centroid values than Russian and Greek <*v*>.

i. In the WIS position, the relative spectral centroid values for Serbian <*v*> will be lower than the relative spectral centroid values for both Russian and Greek <*v*>, which will not differ from each other.

ii. In the WMU position, the relative spectral centroid values for Serbian <*v*> will be lower than the relative spectral centroid values for both Russian and Greek <*v*>, which will not differ from each other.

### 3.2.3 Effect of environment

Recall that segments were elicited in two kinds of words: in word-initial stressed (WIS) position, and in word-medial (intervocalic) unstressed (WMU) position. The WIS environment is expected to favour more fortis realizations, while the WMU environment favours more lenis realizations. With respect to $<v>$, the WIS environment is predicted to be produced with greater airflow and smaller constriction, resulting in more fricative-like articulations, while the WMU environment is predicted to be produced with less airflow and a more lax constriction, resulting in more approximant-like articulations. Environment is used as a controlling factor, and the previous hypotheses are predicted to hold within a given environment. Because I do not explicitly compare languages differences across environments, I do not articulate specific hypotheses for the effect of environment.

## 3.3 Statistical analysis

As in Section 2.5, outliers were only removed due to experiment error or mispronunciation. The statistical analysis was performed using the `lme4` package (Bates et al., 2015b) in R (R Core Team, 2017).[6] For each measure, I tested the hypothesis separately within each environment, word-initial stressed (WIS) and word-medial unstressed (WMU).

### 3.3.1 Harmonicity analysis

Only $<v>$ tokens were included in the harmonicity analysis, rather than the full set of segments. As I show in Figure 3.8, $<v>$ tokens were not devoiced in any language, and so percentage unvoiced was not included in the model, given in 3.1; `Speaker` was included as

[6]Because the goal of this investigation is to test a specific hypothesis, I did not follow an incremental model fitting approach (as in Chapter 2).

a random effect. Post-hoc Tukey tests were implemented with the `glht` function from the `multcomp` package Hothorn et al. (2008).[7]

$$\texttt{lmer(Harmonicity} \sim \texttt{Language + (1|Sp))} \tag{3.1}$$

### 3.3.2 Relative spectral centroid analysis

All segments were included in the analysis of relative spectral centroid. The full model was fit testing the interaction between `Language`, `Voicing`, `Frication Type`, and all interactions between them; `Speaker` was specified as a random effect; the fitted model is given in 3.2.[8]

$$\texttt{lmer(CoG} \sim \texttt{Seg*Lang + (1|Sp))} \tag{3.2}$$

The two categorical predictors are coded differently. For `Language`, I used the default *dummy coding* used in R, with the default ordering of the factor levels, of `Greek`, `Russian`, `Serbian`; the intercept with respect to `Language` is therefore `Greek`.

For `Segment`, I implemented *deviation coding*, in order to ensure that differences between <*v*> tokens were not due to language-specific properties of the dependent variable (either *spectral centroid* or *harmonicity*). Deviation coding compares the mean of the dependent variable for a given level (say, [f, v, s, z]) to the overall mean of the dependent variable.[9] The implementation of deviation coding involves setting a reference level, which here is [f];

[7]For a fitted model *fm*, the code to to implement the post hoc tests comparing `Language` is: `glht(fm, linfct=mcp(Language="Tukey"))`.

[8]The following abbreviations are used: `Seg` = Segment, `Lang` = Language.

[9]Different coding schemes are explained here: `https://stats.idre.ucla.edu/r/library/r-library-contrast-coding-systems-for-categorical-variables/#SIMPLE`.

as a result, the dependent variable estimates encode the following deviations: `v - f`, `s - f`, `z - f`.

Post-hoc comparisons tests were carried out to compare the `v - f` level across languages using the `multcomp` package in R (Hothorn et al., 2008). This was done with a manually coded contrast matrix specifying the levels to compare, using the `glht` command.[10] Because the question of interest is how the languages compare to each other, I only present the results of the post-hoc comparison tests.

## 3.4 Results

### 3.4.1 Harmonicity

Chapter 2 showed that, for every language, the harmonicity of [v] was highest of all the segments, regardless of environment. Figure 3.1 shows density plots of the harmonicity of <*v*> in Greek, Russian, and Serbian, by environment. In both environments, the distributions of harmonicity values overlap entirely, though in the WMU environment Serbian shows less variability in realization.

The fitted model showed that `Language` was not significant for either the WIS environment, $F(2, 18) = 0.26331, p = .7714$ or the WMU environment, $F(2, 17.733) = 0.75487, p = .4846$. Furthermore, post-hoc Tukey tests revealed that no language differed from any other with respect to harmonicity values of [v], as seen in Table 3.2; Figure 3.2 plots confidence intervals on the post-hoc Tukey tests.

[10]For information on how to manually encode a contrast matrix, see this entry in the R-lang mailing list: `https://mailman.ucsd.edu/pipermail/ling-r-lang-l/2012-November/000393.html`. For a fitted model *fm* and a manually coded contrast matrix *cm*, the code to implement post-hoc comparisons using a single-step method is: `glht(fm, cm)`.

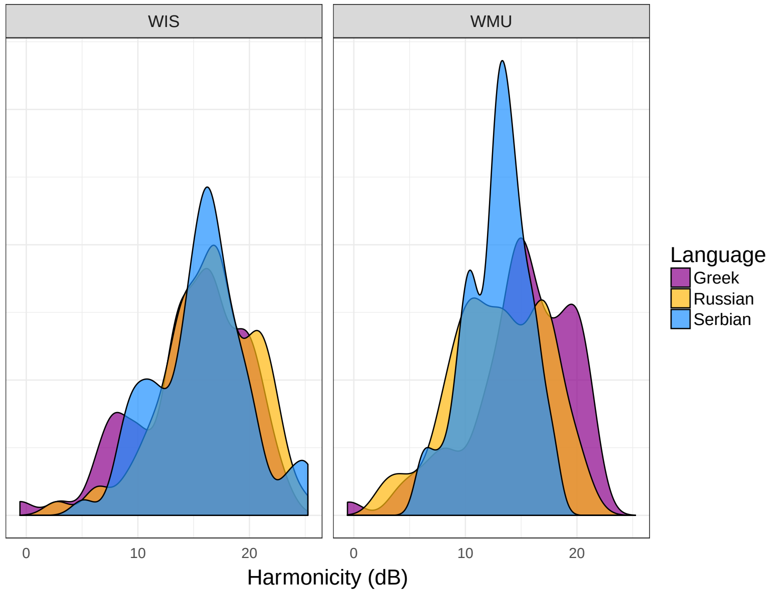
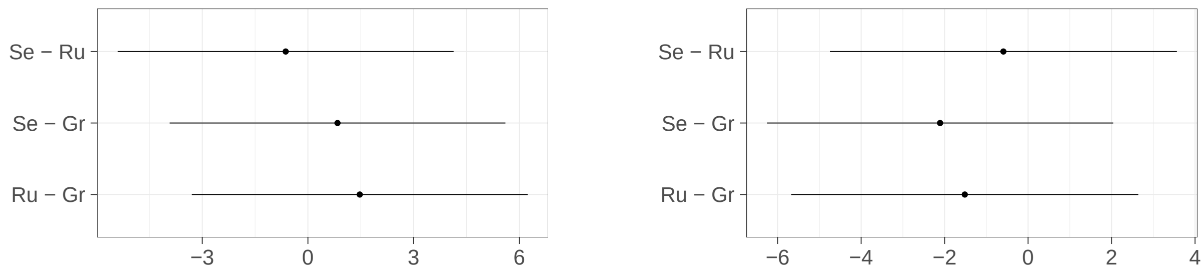



Figure 3.1: Density plots of harmonicity (dB) of <$v$> by language

Table 3.2: Post-hoc Tukey tests comparing harmonicity values of <$v$> between languages

| | WIS | | | | WMU | | | |
|---|---|---|---|---|---|---|---|---|
| | $\beta$ | SE | $t$-value | $p$-value | $\beta$ | SE | $t$-value | $p$-value |
| Se − Ru | -0.63 | 2.03 | -0.31 | .95 | -0.59 | 1.77 | -0.33 | .94 |
| Se − Gr | 0.84 | 2.03 | 0.41 | .91 | -2.11 | 1.77 | -1.19 | .46 |
| Ru − Gr | 1.47 | 2.03 | 0.72 | .75 | -1.52 | 1.77 | 0.86 | .67 |

These results show that harmonicity does not distinguish between <$v$> tokens across languages. Specifically, we cannot reject the null hypothesis that the harmonicity of <$v$> tokens does not differ between languages. I return to a discussion of these results in Section 3.5.




(a) WIS environment

(b) WMU environment

Figure 3.2: Confidence intervals of post-hoc Tukey tests testing whether harmonicity values of <*v*> differed between languages

### 3.4.2 Relative spectral centroid

Figure 3.3 shows violin plots of the spectral centroid calculated on the 1500 Hz high pass filtered signal; this is the same data as seen in Chapter 2, but organized to compare segments across languages.

Recall that because spectral centroid is sensitive to small differences in place of articulation, the model is fit to compare the *relative spectral centroid* of <*v*>, using deviation coding that takes into account the difference between <*v*> and [f].

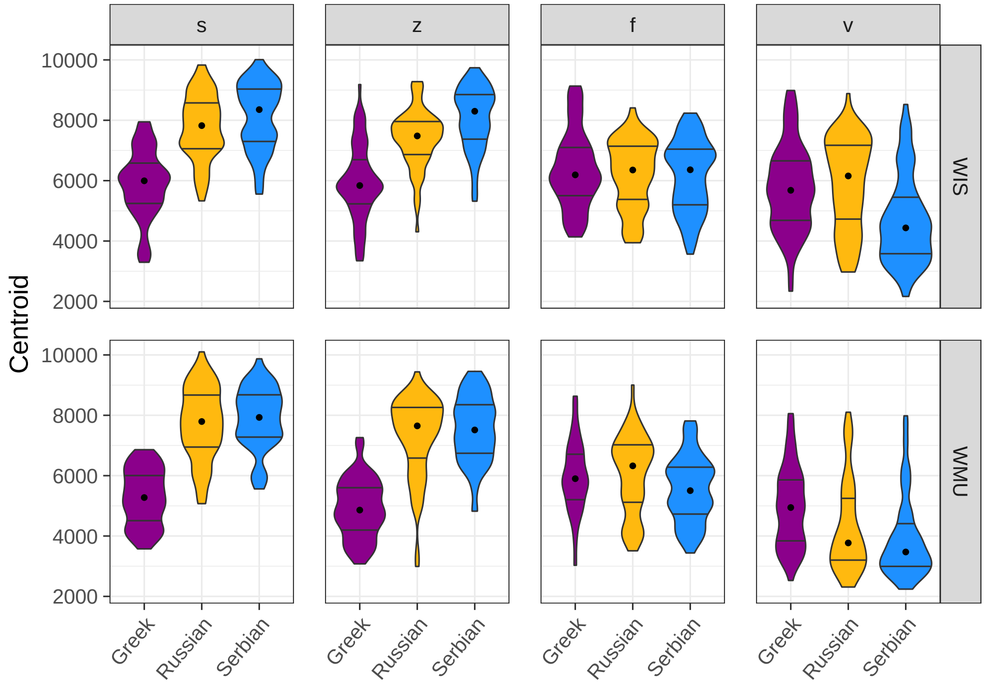



Figure 3.3: Violin plots of spectral centroid (Hz) on 1500 Hz high pass filtered signal, by segment and language

#### 3.4.2.1 Relative spectral centroid in WIS position

The fitted model for the WIS position is given in Table 3.3; the rows highlighted in grey show the estimates for $v - f$ within each language, which is the contrast of interest. The model shows that in all language, <*v*> does differ significantly from [f] in all languages. This is in line with the results of Chapter 2, since in every language, [v] and [f] were distinguished from each other in post-hoc analyses.

The hypothesis of interest, however, is whether the relative spectral centroid, that is, the *difference* <*v* – *f*>, differs across languages. To test this, post-hoc tests were carried out, and showed that in the WIS position, there was no significant difference between Greek and Russian, but that Serbian differed from both Greek and Russian. The results of the post-hoc analysis are given in Table 3.4, and are visualized in Figure 3.4.

Table 3.3: Fitted model for spectral centroid in the WIS environment; (I) represents the intercept of the fitted model

| | $\beta$ | SE | $t$-value | $p$-value |
|---|---|---|---|---|
| Greek (I) | 5987.08 | 278.37 | 21.51 | $< .0001$ |
| v - f | -232.65 | 105.48 | -2.21 | .0277 |
| z - f | -47.77 | 105.48 | -0.45 | .6508 |
| s - f | -94.78 | 105.48 | -0.9 | .3692 |
| Russian | 826.7 | 393.68 | 2.1 | .0501 |
| v - f | -700.21 | 149.17 | -4.69 | $< .0001$ |
| z - f | 635.46 | 149.17 | 4.26 | $< .0001$ |
| s - f | 1054.81 | 149.17 | 7.07 | $< .0001$ |
| Serbian | 766.05 | 393.68 | 1.95 | .0675 |
| v - f | -1844.95 | 149.17 | -12.37 | $< .0001$ |
| z - f | 1357.99 | 149.17 | 9.1 | $< .0001$ |
| s - f | 1462.21 | 149.17 | 9.8 | $< .0001$ |

Table 3.4: Post-hoc test results for the relative spectral centroid on the 1500 Hz high pass filter, compared across languages; WIS environment; Bonferroni correction applied

| | $\beta$ | SE | $z$-value | $p$-value |
|---|---|---|---|---|
| Ru − Se | 1144.7 | 149.2 | 7.674 | $< .0001$ |
| Gr − Se | 1612.3 | 235.9 | 6.836 | $< .0001$ |
| Gr − Ru | 467.6 | 235.9 | 1.982 | .142 |

These results suggest that we reject the null hypothesis that there is no difference in relative spectral centroid values of <*v*> across Greek, Russian, and Serbian in the WIS position. The data do not support Alternative Hypothesis #1, namely that all three languages are distinguished from each other. The data also do not support Alternative Hypothesis #2a, namely that Russian and Serbian pattern together to the exclusion of Greek. The data support Alternative Hypothesis #2b, namely that Russian and Greek pattern together to

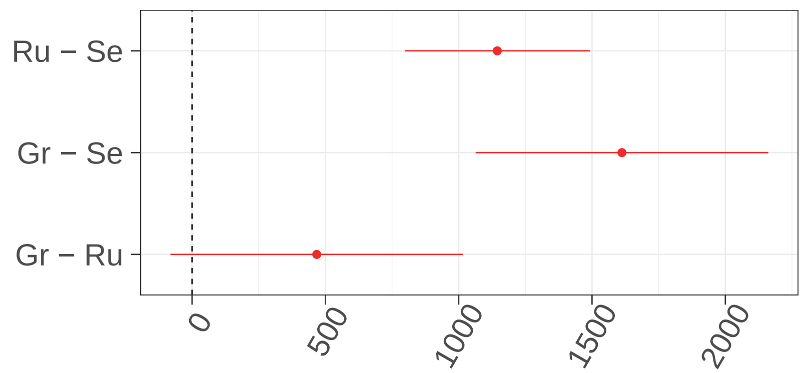



Figure 3.4: Confidence intervals of post-hoc tests (WIS)

the exclusion of Serbian.

#### 3.4.2.2 Spectral centroid in WMU position

The fitted model for the WMU position is given in Table 3.5; the rows highlighted in grey show the estimates for $v - f$ within each language, which is the contrast of interest. The model shows that in all language, <$v$> does differ significantly from [f] in all languages. This is in line with the results of Chapter 2, since in every language, [v] and [f] were distinguished from each other in post-hoc analyses.

Post-hoc tests verifying whether the difference <$v$ – $f$> differed between languages showed that there was no significant difference between Serbian and Russian, but that Greek differed from both Serbian and Russian. The results of the post-hoc analysis are given in Table 3.6, and are visualized in Figure 3.5.

These results suggest that we reject the null hypothesis that there is no difference in relative spectral centroid values of <$v$> across Greek, Russian, and Serbian in the WMU position. The data do not support Alternative Hypothesis #1, namely that all three languages are distinguished from each other. The data also do not support Alternative Hypothesis

Table 3.5: Fitted model for spectral centroid in the WMU environment; (I) represents the intercept of the fitted model

| | $\beta$ | SE | $t$-value | $p$-value |
|---|---|---|---|---|
| Greek (I) | 5270.79 | 286.52 | 18.4 | $< .0001$ |
| v - f | -320.55 | 105.54 | -3.04 | .0025 |
| z - f | -365.64 | 105.54 | -3.46 | .0006 |
| s - f | -10.23 | 108.87 | -0.09 | .9252 |
| Russian | 1010.75 | 404.32 | 2.5 | .0222 |
| v - f | -1690.67 | 148.87 | -11.36 | $< .0001$ |
| z - f | 1393.84 | 146.85 | 9.49 | $< .0001$ |
| s - f | 1430.1 | 148.6 | 9.62 | $< .0001$ |
| Serbian | 942.07 | 404.14 | 2.33 | .0315 |
| v - f | -1994.27 | 146.37 | -13.62 | $< .0001$ |
| z - f | 1691.62 | 146.37 | 11.56 | $< .0001$ |
| s - f | 1686.35 | 150.76 | 11.19 | $< .0001$ |

Table 3.6: Post-hoc test results for spectral centroid on the 1500 Hz high pass filter, compared across languages; WMU environment; Bonferroni correction applied

| | $\beta$ | SE | $z$-value | $p$-value |
|---|---|---|---|---|
| Ru − Se | 303.6 | 146.0 | 2.080 | 0.113 |
| Gr − Se | 1673.7 | 234.2 | 7.147 | $< .0001$ |
| Gr − Ru | 1370.1 | 235.8 | 5.812 | $< .0001$ |

#2b, namely that Russian and Greek pattern together to the exclusion of Serbian. The data support Alternative Hypothesis #2a, namely that Russian and Serbian pattern together to the exclusion of Greek.

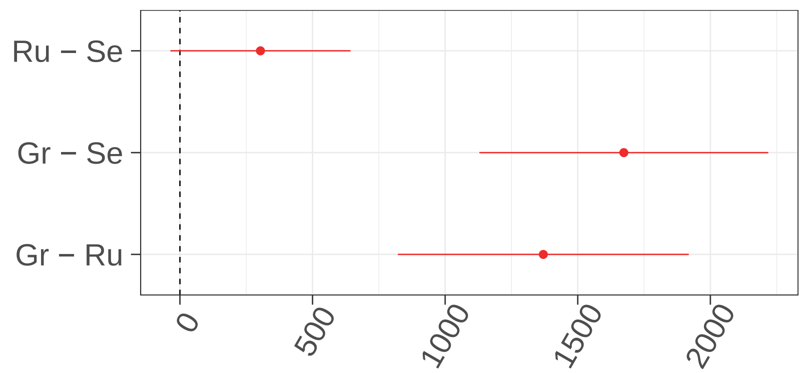



Figure 3.5: Confidence intervals of post-hoc tests (WIS)

#### 3.4.2.3 Spectral centroid: summary

In all languages, the difference between <*v*> and [f] is statistically significant, the relative spectral centroid of <*v*> does not differ between Greek and Russian in the WIS environment, and does not differ between Russian and Serbian in the WMU environment. In both environments, the relative spectral centroid of <*v*> differs between Greek and Serbian.

## 3.5 Discussion

Whether the segment phonologically transcribed as /v/ is realized as a fricative or as an approximant is, in articulatory terms, a question of labiodental aperture and the gradient between oral and atmospheric pressure. Acoustically, this is most likely to manifest in the presence of high frequency energy, which results from the turbulence generated both at the place of stricture between the upper teeth and lower lip, and where the airstream is impeded downstream from the stricture at the upper lip (Stevens, 1988). This was assessed using the spectral centroid on the 1500 Hz high-pass filtered signal. Harmonicity (or harmonics-to-noise ratio) was used to assess the relative contribution of high frequency energy (or noise) and low frequency energy (contributed by voicing).

Overall, the results did not support the Alternative Hypothesis #1, that the phonetic realization of <*v*> differs between languages and tracks phonological identity. I discuss the results of each measure separately.

### 3.5.1 Discussion: harmonicity

The results for harmonicity did not provide evidence to reject the null hypothesis: harmonicity values of <*v*> are not distinguished in Greek, Russian, and Serbian. The motivation to test harmonicity came from Hamann and Sennema (2005), who used harmonicity as a measure to assess differences between fricative and approximant realizations of voiced labiodental continuants in German and Dutch. Summary results from Hamann and Sennema (2005) are schematized in Figure 3.6.

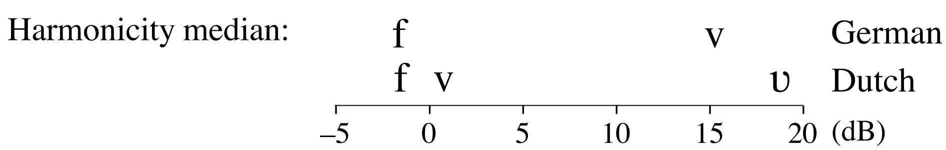



Figure 3.6: Harmonicity values of German and Dutch labiodentals, reproduced from Hamann and Sennema (2005)

The hypotheses articulated in Section 3.2.1 were motivated by Hamann and Sennema's results: if Serbian /v/ is a phonological approximant sonorant, it might be realized in a manner similar to Dutch [ʋ], and if Greek /v/ is a phonological fricative obstruent, it would be realized in a manner similar to Dutch [v].[11] In light of the results in Hamann and Sennema (2005), it is perhaps surprising that no harmonicity differences were found in Greek, Russian,

[11]In fact, many linguists, including Silke Hamann and Paul Boersma, suggested that I consider harmonicity as a way to test the hypothesis.

and Serbian.

I hypothesize that this discrepancy is due to the fact that Dutch [v] is reported to be frequently devoiced, and that in fact, the harmonicity results of Hamann and Sennema (2005) are tracking devoicing, which would lower harmonicity values significantly. Devoicing of [v] or [z] would affect the distribution of energy in the frequency range, specifically by reducing or eliminating the low-frequency energy associated with voicing. Support for this hypothesis draws from the fact that the harmonicity value of Dutch [v] tokens is near 0, which suggests that most of the energy comes from the high frequency range; this would be surprising if Dutch [v] tokens were voiced, since this would put the Dutch harmonicity values as lower than even the harmonicity values of [z] in Greek, Russian, Serbian, and English, as can be seen in the results of harmonicity in Chapter 2. Harmonicity values near 0 suggest that the tokens of Dutch [v] are mostly voiceless. Although Hamann and Sennema (2005) do not present devoicing results, this is consistent with literature that reports that Dutch voiced fricatives are nearly always devoiced.[12].

Additional support for the hypothesis that harmonicity is confounded by devoicing can be seen by looking at scatterplots that compare *percentage unvoiced* to harmonicity, as seen in Figure 3.7, which clearly suggests a correlation between harmonicity values and percentage unvoiced.[13] Specifically, the higher the percentage unvoiced, the lower the harmonicity value, which would justify the inclusion of `Unvoiced` as a covariate.[14]

[12]The distinction between /v/ and /f/ is manifested as length, not as voicing (Gussenhoven and Bremmer Jr., 1983)

[13]*Percentage unvoiced* was determined automatically with a Praat script to calculate the number of unvoiced frames in a labelled interval using the voice report function in Praat. Parameters were set as follows: for women, minimum pitch = 100 Hz, maximum pitch = 300 Hz; for men, minimum pitch = 75 Hz, maximum pitch = 250 Hz; all other parameters were set to the defaults. These parameters were selected following Eager (2015).

[14]The correlation between harmonicity and percentage unvoiced was calculated using Kendall's tau-b within each environment and language, yielding values between -.70 and -.74.

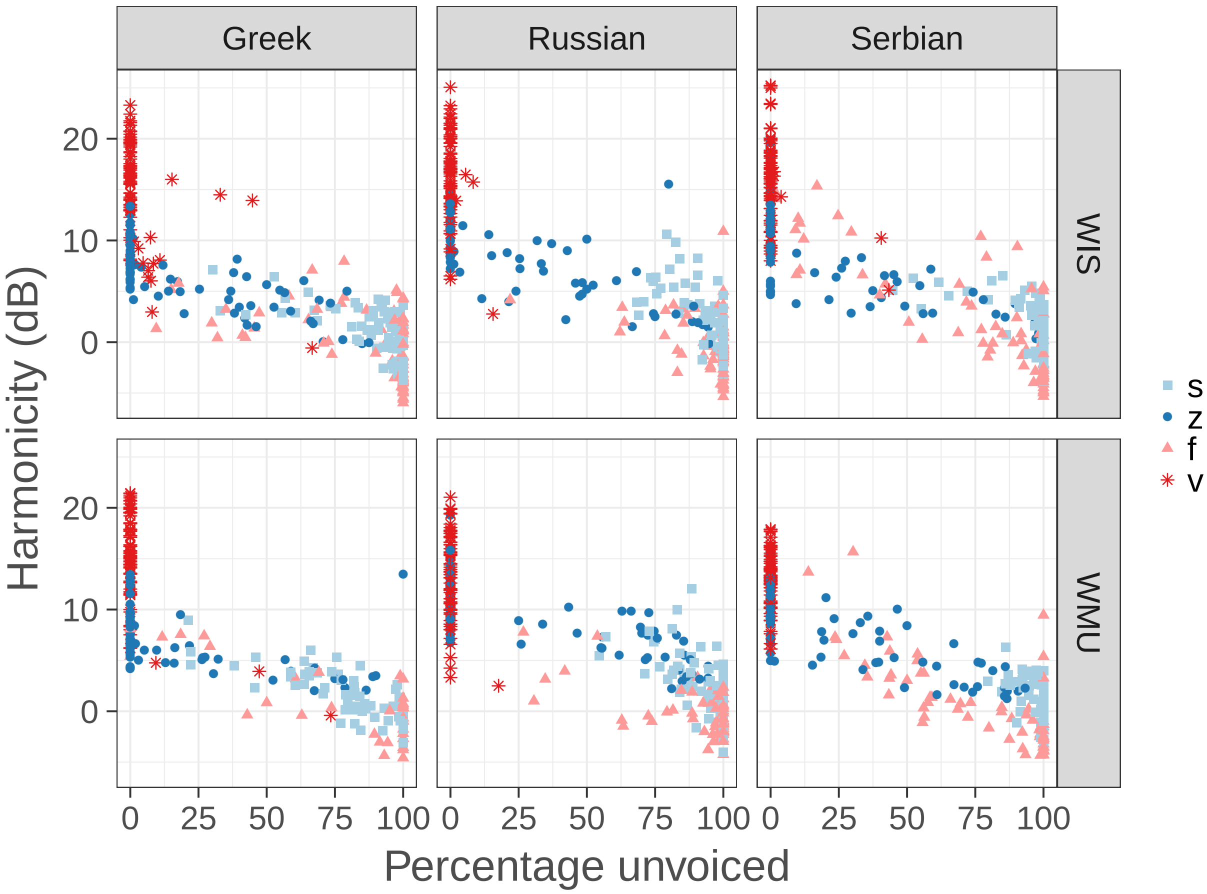



Figure 3.7: Scatter plot of harmonicity (dB) by percentage unvoiced, by segment and language

Percentage unvoiced was not included in the present analysis of harmonicity because <*v*> is not devoiced in any of the three languages. This is more clearly seen in Figure 3.8, which shows the *percentage unvoiced* for each token, by language; the black dot and error bars refer to the mean and one standard deviation. Segments /f, s/, which are underlyingly unvoiced, are predicted to have a strong skew to the right of the graph, with most tokens 100% unvoiced; segments /v, z/, which are underlyingly voiced, are predicted to have a strong skew to the left of the graph, with most tokens 0% unvoiced, we can expect that even voiceless [f, s] will not be entirely devoiced.

As expected, [f] and [s] are generally voiceless, with the data skewed to the right; interest-

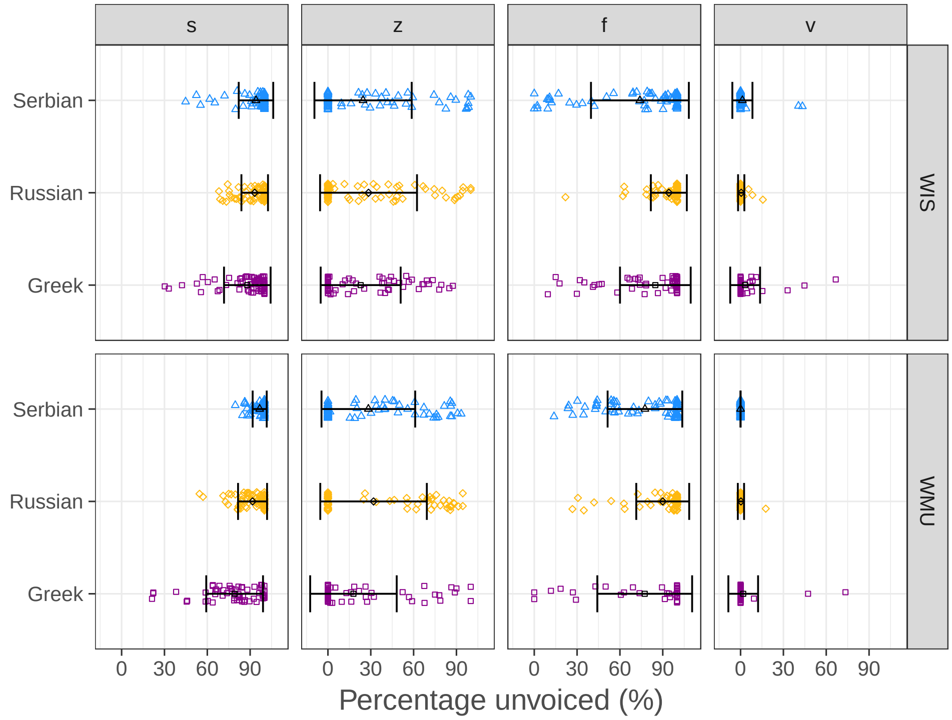



Figure 3.8: Voicing percentage

ingly, the degree of voicing in [f, s] is much greater than might otherwise be expected. This is particularly salient for Greek [f, s], and Serbian [f]. For Greek, this may be related to the fact that voiceless stops /p, t, k/ have been shown to be produced with significant voicing as carryover effects from neighbouring vowels (Nicolaidis, 2002), though Nicolaidis did not find this result applied to the voiceless fricatives. For Serbian [f], I noticed during segmentation that a fair number of female speakers produced [f] as weakly voiced, and indeed there seem to be some general effects of gender, as seen in Figure 3.9.

For all languages, women seem to produce more deviation from the underlying voicing of /f, s, z/ than men. Although this could be due to the parameters used to determine pitch, I do not believe these results are spurious, based on my observations during segmentation. That said, understanding why there is an apparent gender-based difference in voicing percentage is beyond the scope of this work, and left for future research.

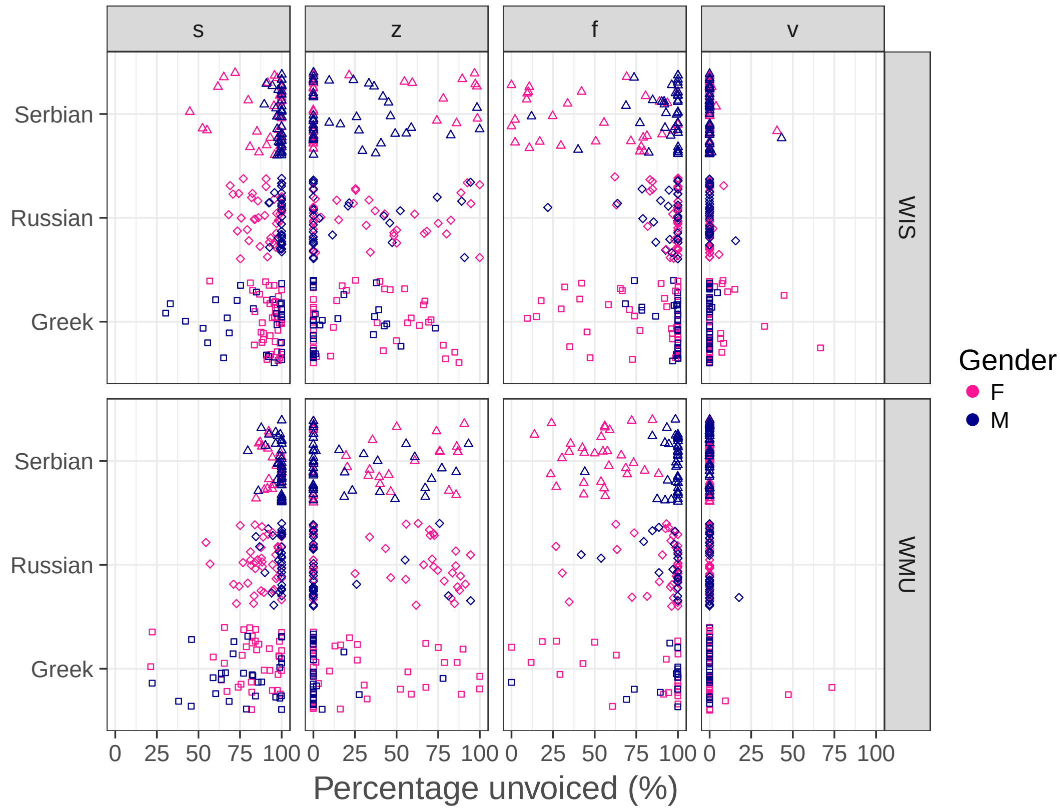



Figure 3.9: Voicing percentage

Further support for the argument that harmonicity is tracking percentage unvoiced comes from considering the harmonicity and devoicing values in English [v] and [w]. Figure 3.10 reproduces the harmonicity data for Greek, Russian, and Serbian from Figure 3.1, and includes English [v] and [w] for comparison.

The distribution of harmonicity values for English [v] is generally lower than the distribution of harmonicity values of English [w], though the distributions overlap a fair bit, and paired *t*-tests did not reveal significant differences in harmonicity between [v] and [w] tokens, in either environment. English [v] is mostly voiced, though it is devoiced more frequently than <*v*> in Greek, Russian, and Serbian, as seen in Figure 3.11a; interestingly, English [w] exhibits some devoicing as well. Like in Greek, Russian, and Serbian, English harmonicity

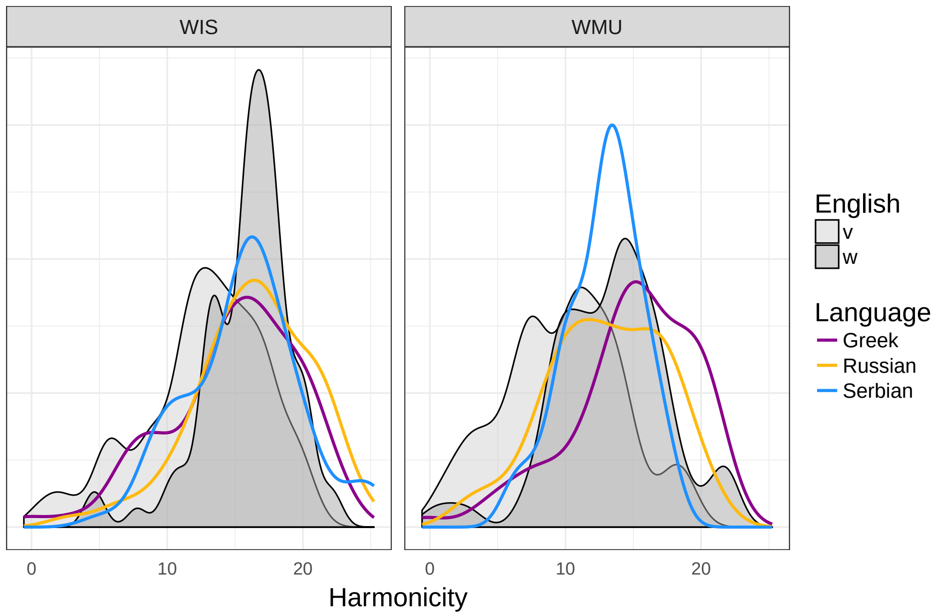
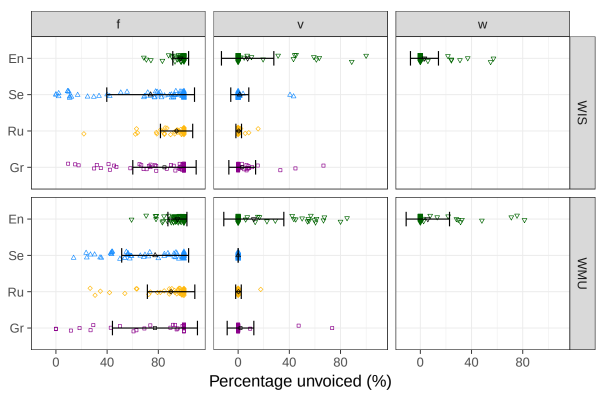
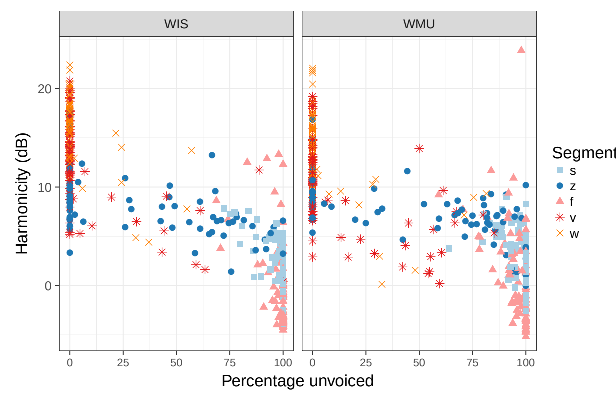



Figure 3.10: Harmonicity values of Greek, Russian, and Serbian, compared to English [v] and [w]




(a) Voicing percentage of [f, v, w]




(b) Scatterplot of harmonicity (dB) by percentage unvoiced in English

Figure 3.11: Comparison of English devoicing and harmonicity to Greek, Russian, and Serbian

exhibits some correlation with percentage unvoiced, as seen in Figure 3.11b.

Returning to the null results of harmonicity presented in Section 3.4.1, these facts suggest that the realization of <*v*> in terms of its harmonics-to-noise ratio is not influenced by its phonological status. Rather, the discrepancy between Hamann and Sennema (2005) and the harmonicity results presented here suggest that devoicing is a significant confound for harmonicity.

### 3.5.2 Discussion: spectral centroid

The results for relative spectral centroid did not support Alternative Hypothesis #1, namely that the realization of <*v*> tokens tracks phonological identity. Specifically, Russian <*v*> did not exhibit values for the phonetic parameters that were consistently intermediate between the values found in Greek and Serbian. Although tokens of Greek and Serbian <*v*> were distinguished by spectral centroid in both environments, suggesting a partial correlation between phonological identity and phonetic realization, Russian <*v*> tokens did not fall consistently in the middle: in the WIS environment, Russian <*v*> tokens patterned as in Greek, and in the WMU environment, Russian <*v*> tokens patterns as in Serbian. In terms of the hypotheses articulated in Section 3.2.2, the data support Alternative Hypothesis #2b for the WIS position, but Alternative Hypothesis #2a for the WMU position.

The results for relative spectral centroid presented in this chapter are in line with the results of spectral centroid presented in Chapter 2. For all languages, the spectral centroid of <*v*> was lower than all other segments, and more importantly, the interaction effect between frication type and voicing was significant in all languages. This suggests that, regardless of how much of a fricative <*v*> is, the aerodynamic consequences of maintaining non-sibilant frication with voicing are such that the distribution of noise in the articulation of [v] will have the relationship to [f] that we see between [z] and [s]. In other words, modulo the effect of voicing, [z] is indistinguishable from [s] (in all languages, in all environments), but <*v*> can always be distinguished from [f] (in all languages, in all environments).

In spite of this potentially universal, low-level phonetic effect, the relationship between

<*v*> and [f] differs across languages, though not to the extent hypothesized by functionalist accounts of Russian /v/. Padgett (2002) hypothesizes that Russian <*v*> is inherently intermediate between obstruent [v] and [ʋ], offering the transcription [ʋ̞]. The inherent intermediacy of [ʋ̞] between fricative [v] and sonorant [ʋ] is intended to capture the phonological intermediacy of Russian /v/ between obstruents and sonorants. Based on the results of this study, I reject Padgett's proposal that there exists a class of *narrow approximants* that is distinguished in articulatory terms from both fricatives on the one hand and approximants on the other. This class is purported to be highly unstable in its phonetic realization, prone to lose either its frication or voicing; nevertheless, in phonetically favourable environments, such as in pre-sonorant position, it is expected to surface. This study was designed in part to test this, by measuring the phonetic parameters of <*v*> in not just pre-sonorant position, but pre-vocalic position, which ought to be the most phonetically favourable position to see such a contrast. I did not find a three-way distinction in the degree of frication, and thus no acoustic evidence supporting the existence of a third phonetic category of voiced labiodental continuant [ʋ̞].

Nevertheless, a particular kind of intermediacy was discovered. Greek and Serbian both showed an effect of environment, namely with higher centroid values in the WIS environment, suggesting a more fricative realization, and lower centroid values in the WMU environment, suggesting a more approximant realization. Recall from Section 2.7.3 that in Greek, only `Environment` was a factor for spectral centroid, and that all fricatives had lower spectral centroid values in the WMU environment. The same effect of environment was seen in Serbian, where the effect of environment operated additively on the general interaction of `Voiced` and `Sibilant` (Section 2.8.3). For Russian, the effect of lenition (or fortition), was much more drastic, and as a result, patterned with Greek in the WIS environment, and with Serbian in the WMU environment.

The phonological interpretation of the centroid results is suggested by the statistical coding, which assessed the deviation of spectral centroid of all segments from [f]. Because of the high-pass filter, the spectral centroid assessed the distribution of high frequency energy, and calculating the deviation from [f] essentially compares the similarity between <*v*> and [f] in the high frequency range. For Greek, modulo the effect of voicing, the distribution of high frequency energy differs from [f], but to a much smaller extent than is seen in either Russian and Serbian, and suggests that the transcription [v] is appropriate for the voiced labiodental spirant in Greek. For Serbian, <*v*> is not characterized by energy in the high frequency range of the spectrum, supporting a transcription of [ʋ]. These results support a partial correlation between the phonological identity of <*v*> and its realization: Greek /v/ is [v], but Serbian has /ʋ/ which is realized as [ʋ]. Russian does not exhibit the intermediate realization hypothesized by functionalist analyses of /v/'s patterning, as in Padgett (2002). Nevertheless, Russian exhibits variability in its realization conditioned by environment.

### 3.5.3 General discussion

The implications of the present study extend beyond the purported intermediacy of Russian <*v*>. Specifically, this study was designed to address a particular case study that assumes a one-to-one relationship between phonological identity and phonetic realization. By hypothesizing that Russian <*v*> is realized as neither [v] nor [ʋ], Padgett (2002) hypothesizes that phonology and phonetics are tightly linked, and that in fact the phonology can be "read off of" the phonetics, so to speak.

As I discuss further in Chapter 5, the contrast between /v/ and /ʋ/ is, perhaps unsurprisingly, very rare, and exemplified by Dutch and several closely related languages spoken in Nigeria and Cameroon. All other things being equal, the difference between [v] and [ʋ] is constriction degree, but of course it is no trivial task to keep all other things equal. For example, as mentioned above, the contrast between Dutch [v] and [ʋ] appears to be maintained in part by devoicing, and the contrast between [v] and [f] by length. Positing a third category intermediate between these two that speakers are able to pick up on is improbable.

The intermediacy that Russian *does* display, however, is suggestive. Russian <*v*> is arguably [v] when in a simple onset in WIS position, and arguably [ʋ] when in intervocalic position as the onset of an unstressed syllable. These environments do not play a role in voicing assimilation, but we might imagine that if phonological patterning tracked phonetic identity, then we might expect that, for example, Russian <*v*> triggers voicing assimilation in WIS position. This is not what is reported in the literature, suggesting that phonological representation does not exhibit high fidelity to phonetic realization. Nevertheless, the range of realizations of Russian <*v*> may account for why it is partially paired with [f], and partially not. Although this study only considered two environments, it seems reasonable to conclude that Russian <*v*> is characterized by a greater range of realizations than either Greek [v], an obstruent, or Serbian [ʋ], a sonorant. Since the realizations of Russian <*v*> include both obstruent-like and sonorant-like realizations, this may facilitate its interpretation by speakers as both an obstruent and sonorant.

In the following chapter I turn to an investigation of voiced spirants and their relationship to inventory structure. I return to the question of how Russian <*v*> ought to be represented in Chapter 5.

CHAPTER 4

**THE TYPOLOGICAL IDENTITY OF VOICED SPIRANTS**

## 4.1 Introduction

The focus of the previous two chapters was on the acoustic properties of [v] and its relationship to its voiceless counterpart [f], and the voiced-voiceless sibilant pair [z, s]. Understanding the phonetic properties of [v] is important, but is only part of the story. What is the typological identity of /v/? Does /v/ exhibit the typological properties of obstruents or sonorants? Do all the voiced spirants share the same typological properties? This chapter shifts focus to the cross-linguistic typological identity of /v/ and the voiced spirants /β, ð, ɣ/ in an effort to answer these questions, focussing specifically on inventory size and implicational relation violations.

As discussed in Section 1.6, the voiced spirants challenge the canonical divide between obstruents and sonorants, from multiple perspectives: they are often described as “frictionless” or “weak”, and they exhibit ambivalent and ambiguous patterning with respect to the [sonorant] feature. The focus of the present chapter is their typological character in consonant inventories; I address their participation in phonological processes and their phonotactics in Chapter 5.

Recall from Section 1.4 that the aerodynamic requirements of voicing and frication are in conflict. This fact is widely regarded as explaining the *typological markedness* of voiced fricatives, frequently articulated as the typological observation that if a language has a voiced obstruent, it has its voiceless counterpart, but not vice versa. This observation is articulated as an implicational relation, as in (1) (Maddieson, 1984). Recall from Section 1.6 that Maddieson (1984) shows that, unlike for stops and sonorants, the fricative consonants exhibit a fair degree of heterogeneity in how well they adhere to (1c), with voiced spirants /β, v, ð, ɣ/ exhibiting much higher violation rates than other consonant classes.[1] That is, while the implicational relations (1a) and (1b) are robust, the implicational relation (1c) is not homogenous with respect to the fricative class. Specifically, while voiced plosives and sibilants rarely appear in inventories without their voiceless counterparts, voiced spirants appear unpaired with a frequency of 20% to 75%.

(1) Implicational relations of voicing in consonants

a. Voiceless sonorants $\Longrightarrow$ voiced sonorants

b. Voiced stops $\Longrightarrow$ voiceless stops

c. Voiced fricatives $\Longrightarrow$ voiceless fricatives

In spite of the heterogeneity in violation rates across fricative voicing pairs, the voiced fricatives are nonetheless often treated as a uniform class. In light of the established differences, both phonetic and typological, between voiced sibilants and spirants, this chapter seeks to clarify the typological character of voiced spirants with respect to claims made about voiced obstruents as a class, taking the work of Maddieson (1984) described in Section 1.6 as a starting point.

### 4.1.1 Typological claims about voiced fricatives

I begin with two typological investigations motivated by claims in the literature regarding the typological character of voiced fricatives as a class: in Section 4.3 I consider the status of

[1] The relationship between voiced and voiceless sonorants on the one hand, and voiced and voiceless obstruents on the other, are arguably different kinds of relationships, due in part to the diachronic origins of voiceless sonorants as opposed to voiced obstruents (see Blevins (2018) and references therein for discussion of voiceless sonorants). Several researchers have shown that voiceless sonorants exhibit patterning that is more akin to aspiration rather than voicelessness; see Botma (2011) and references therein.

voicing within the class of voiced obstruents, and in Section 4.4 I investigate the relationship of voiced fricatives to consonant inventory size. I show that, as seen in the violation rates of (1c), the voiced fricatives do not pattern uniformly: in both cases, the voiced spirants exhibit a different typological character than the voiced sibilants.

Maddieson (2010) considers the relationship of voicing within the class of voiced obstruents, specifically between voiced fricatives and voiced plosives, claiming that, "if a language has a fricative voicing contrast, then it is highly likely that it also has a plosive voicing contrast" (pg. 536). With respect to inventory size, Lindblom and Maddieson (1988) propose that the articulatory complexity of segments tracks inventory size, meaning that inventories first make use of basic articulations, then elaborated articulations, and finally complex articulations. Thus small inventories will comprise segments that are considered basic, rather than a random set of possible segments, and only large inventories will comprise complex segments. Voiced fricatives are classified as elaborated articulations, and so are predicted to surface in inventories only after more basic articulations have been exploited.

Both claims crucially rely on the interaction between voicing and frication, with the assumption that fricative voicing is more complex than either sonorant or stop voicing, an assumption grounded in the particular aerodynamic challenges required to maintain both voicing and frication. However, given the heterogeneity in violation rates of (1c), one might reasonably wonder if this assumption is warranted for all kinds of fricatives. This chapter shows that while sibilant voicing does adhere to the typological claims regarding obstruent voicing and inventory size, voiced spirants do not. Specifically, I show that voiced spirants are more likely than voiced sibilants to appear in inventories that do not have voiced stops, and that voiced spirants, but not voiced sibilants, tend to appear in small inventories, but only when *unpaired*.[2]

[2] The focus here is on contrastive inventories, not on allophonic relationships. For example, voiced spirants

### 4.1.2 Typological markedness: segments, features, or contrast?

Complementing the empirical goals laid out above, this chapter has as its secondary goal the conceptual task of probing our understanding of typological markedness, specifically with respect to the role that features play in circumscribing typological statements. These concerns are not purely philosophical, but rather have implications for how the empirical questions laid out above are probed.

By way of example, consider the consonant inventory of Chuave (`cjv`) in Table 4.1, which has voiceless stops /t, k/ and voiced stops /b, d, g/. The presence of voiced /b/ without its voiceless counterpart is an instantiation of a violation of (1b): the presence of a voiced stop implies the presence of its voiceless counterpart. Nevertheless, inventories like Chuave's are rarely interpreted that way; instead, the absence of /p/ is interpreted as an accidental gap.[3] This interpretation suggests that the implicational relations of voicing are over features and not segments.

Table 4.1: Consonant inventory of Chuave (`cjv`) (Maddieson et al., 2015)

| | Labial | Coronal | Palatal | Velar | Labio-velar |
|---|---|---|---|---|---|
| Stop | b | t d | | k g | |
| Fricative | f | s | | | |
| Nasal | m | n | | | |
| Rhotic | | ɾ | | | |
| Approximant | | | j | | w |

may arise from intervocalic lenition of voiced stops /b, d, g/, as in Spanish. LAPSyD does not include allophonic variants in its inventories, and so these are not considered here unless specifically remarked upon.

[3]It should be noted that the absence of /p/ from a voiceless inventory truly is an accident. Although the bilabial place of articulation is the most likely place to be missing from a voiceless stop inventory (Ohala, 1983) in terms of raw frequency, Maddieson (2013b) shows that this is due to geographical and geneological factors.

However, if we consider an inventory such as that of Rukai `(dru)`, as in Table 4.2, the absence of /f/ is typically interpreted as an instantiation of the violation (1c), an interpretation corroborated by the fact that, for example, Maddieson (1984) lists the violation rates for fricatives for individual pairs. Unlike in stops, the markedness facts for fricatives are discussed and interpreted over (pairwise) segments, not features.

Table 4.2: Consonant inventory of Rukai `(dru)` (Maddieson et al., 2015)

| | Labial | Dental | Alveolar | Retroflex | Velar | Glottal |
|---|---|---|---|---|---|---|
| Stop | p b | t d | | ɖ | k g | ʔ |
| Affricate | | | ts | | | |
| Fricative | v | θ ð | s | | | h |
| Nasal | m | | n | | ŋ | |
| Rhotic | | | ɾ | | | |
| Lateral | | | l | ɭ | | |
| Approximant | w | | | j | | |

The likely source of this discrepancy has to do with other differences in stops and fricatives, apparently unrelated to the voicing facts: stops typically appear in series, while fricatives do not, even though fricatives can be articulated at more places of articulation than stops. Maddieson (1984) observes that languages overwhelmingly use three places of articulation for plosives, with only two exceptions in UPSID. All languages contain plosives, and no language only uses one place of articulation for plosives. Hawaiian `(haw)` and Wichita `(wic)` utilize only two places of articulation: Hawaiian lacks a coronal stop, and Wichita has no labials. Among the remaining 449 languages that employ a three-way place distinction, all have “bilabial, dental or alveolar and velar stops” (Maddieson, 1984, pg. 31), with three exceptions: Hupa `(hup)` lacks labial and velar plosives, Aleut `(ale)` lacks labial plosives, and Kirghiz `(kir)` lacks velar plosives. Thus among the 451 languages in UPSID, 446 contrast plosives at the labial, coronal and velar places of articulation (98.9%).

The overwhelming tendency for languages to contrast plosives at three places of articulation within a given laryngeal configuration, e.g., "voiced", "aspirated", "palatalized", etc., gives rise to the notion of a stop *series*, and to the complementary notion of a *gap*. Although fricatives are, in theory, able to contrast more places of articulation than stops, as represented by the IPA chart in Figure 4.1, fricatives do not typically surface as a series, whether voiceless or voiced.[4] As described in Maddieson (1984), almost all languages have at least one fricative (93.4%), and Maddieson notes that this fricative is usually an 's'-like sound, occurring in 88.5% of languages. Beyond that, however, and the modal number of fricatives in a language is only 2: the second most common fricative after /s/ is either /f/ or /ʃ/; the third most common is a 'z'-like sound (Maddieson, 1984, pg. 45). In other words, fricatives do not tend to exploit the three major places of articulation the way that stops do, and hence even in systems with three fricatives, no more than two places are usually exploited.

CONSONANTS (PULMONIC) © 2005 IPA

| | Bilabial | Labiodental | Dental | Alveolar | Postalveolar | Retroflex | Palatal | Velar | Uvular | Pharyngeal | Glottal |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Plosive | p b | | | t d | | ʈ ɖ | c ɟ | k ɡ | q ɢ | | ʔ |
| Nasal | m | ɱ | | n | | ɳ | ɲ | ŋ | ɴ | | |
| Trill | ʙ | | | r | | | | | ʀ | | |
| Tap or Flap | | ⱱ | | ɾ | | ɽ | | | | | |
| Fricative | ɸ β | f v | θ ð | s z | ʃ ʒ | ʂ ʐ | ç ʝ | x ɣ | χ ʁ | ħ ʕ | h ɦ |
| Lateral fricative | | | | ɬ ɮ | | | | | | | |
| Approximant | | ʋ | | ɹ | | ɻ | j | ɰ | | | |
| Lateral approximant | | | | l | | ɭ | ʎ | ʟ | | | |

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Figure 4.1: Consonants identified by the IPA

Whether typological statements about markedness hold over segments or features is rel-

[4] IPA chart downloaded from `http://www.internationalphoneticassociation.org/content/ipa-chart` .

evant to another way that markedness might be expressed, namely in terms of contrast, as Maddieson (2010) does for the relationship between stop and fricative voicing: "if a language has a fricative voicing contrast, then it is highly likely that it also has a plosive voicing contrast" (pg. 536). The consonant inventory of Hopi `(hop)` (Table 4.3) does not count as having a fricative voicing contrast because, although there is both a voiceless fricative /s/ and a voiced fricative /v/, /s/ and /v/ do not share the same place of articulation.

Table 4.3: Consonant inventory of Hopi `(hop)` (Maddieson et al., 2015)

| | Labial | Alveolar | Palatal(ized) | Velar Plain | Velar Labialized | Uvular | Glottal |
|---|---|---|---|---|---|---|---|
| Stop | ʰp p | ʰt t | kʲ | ʰk k | ʰkʷ kʷ | ʰq q | ʔ |
| Affricate | | ʰts ts | | | | | |
| Fricative | v | s | | | | | h |
| Nasal | m̥ m | n̥ n | ɲ | ŋ̊ ŋ | ŋʷ | | |
| Lateral | | l̥ l | | | | | |
| Approximant | ʍ w | | j̊ j | | | | h |

As a more subtle example, consider the inventory of Tuva `(tyv)` in Table 4.4. The Tuva inventory includes the voiceless labiodental fricative /f/ and the voiced bilabial fricative /β/. Given that both /f/ and /β/ are both labial sounds, one might argue that their place of articulation is "close enough", and that Tuva has a fricative voicing contrast; on the other hand, this argument cannot always apply, since there exist languages such as Ewe `(ewe)` that contrast bilabial and labiodental fricatives. In this case it seems relevant to determine if the pair of sounds in question is *phonologically contrastive* (Dresher, 2009), but this is not a notion of contrast that can be assessed by large-scale typological studies of consonant inventories.

Empirically, the purpose of this chapter is to assess whether voiced fricatives pattern uniformly with respect to their cooccurrence with laryngeal contrasts in stops, and with respect to inventory size. The results of this chapter cannot inform phonological theory

Table 4.4: Consonant inventory of Tuva (`tyv`) (Maddieson et al., 2015)

| | Labial | Dental | Alveolar | Palatoalveolar | Palatal | Velar |
|---|---|---|---|---|---|---|
| Stop | p b | t̪ d̪ | | | | k g |
| Affricate | | t̪s̪ | dz | tʃ dʒ | | |
| Fricative | f β | s̪ | | ʃ | | |
| Nasal | m | n̪ | | | | ŋ |
| Rhotic | | | r | | | |
| Lateral | | l̪ | | | | |
| Approximant | | | | j | | |

directly, but they do illustrate that, at least typologically, voiced spirants do not exhibit the same tendencies as voiced sibilants.

## 4.2 Database

The present investigation is based on the updated UPSID-92 database (Maddieson and Precoda, 1990) with 451 languages, as it is designed for phonological typological research, (i.e. is relatively geneologically and geographically balanced). I used the downloadable text version that is freely available from the UPSID website (`http://web.phonetik.uni-frankfurt.de/upsid.html`), developed by Henning Reetz.[5] Although the data were downloaded from UPSID-92, I verified the data in the *Lyon-Albuquerque Phonological Systems Database* (LAPSyD), a new online phonological database created by Ian Maddieson in collaboration with the Laboratoire Dynamique de Langage (DDL) in Lyon. LAPSyD was used to verify data since it prioritizes reliability of the source material, and the linguistic descriptions are meant to maximize comparability between languages (Maddieson et al.,

[5] Following Clements and Osu (2003), I use the name UPSID-92 to distinguish the updated version of UPSID from the original version reported in Maddieson (1984), which I refer to as UPSID-84. UPSID-84 contains 317 languages.

2013); all discrepancies were resolved in favour of the LAPSyD description. In spite of the greater number of languages, I was not able to use LAPSyD for the present investigations because at the time of writing it had several bugs in its web interface that prevented efficient data analysis, and the server timed out when large datasets were accessed. Finally, for some languages I consulted primary sources, and occasionally made changes to the database as a result. To distinguish my database from the other two versions of UPSID, I refer to the database used here as UPSID-CB. Finally, all languages are referred to by their current accepted name, as some of the language names listed in UPSID-92 are no longer used by the linguistic communities in question; to facilitate comparisons, throughout the dissertation all languages are also identified by their ISO 639-3 code as on Ethnologue (e.g., Greek `(ell)`). For each language, meta-linguistic data was added: geneological classification (*family* and *genus*) came from WALS (Maddieson, 2013a), while geographical information (*country* and *area*) came from Ethnologue (Lewis et al., 2013); latitude and longitude came from LAPSyD, as did consonant inventory size.

In order to facilitate the investigations of interest, only contrasts and segments relevant to the current study are encoded. For example, the distinctions between various secondary parameters on voiceless plosives and sibilants (e.g., palatalization) were not encoded, but potentially relevant laryngeal contrasts were. For sibilants, this was reduced to only two categories, voiceless and voiced. For plosives, there were three categories within the voiceless and voiced categories; for voiceless, the categories were plain, aspirated, and either laryngealized or ejective; for voiced, the categories were plain, implosive, and either pre- or postnasalized. The only specific segments encoded were the voiced and voiceless spirants and the approximants at the same places of articulation, again ignoring secondary articulations such as palatalization. For some of the investigations, the categories of voiceless and voiced plosives are collapsed, and in these cases are simply referred to as voiceless and voiced, respectively. In the graphs and tables throughout this chapter plosives and sibilants are represented as `T` for the voiceless plosives, `D` for the voiced plosives, `S` for the voiceless sibilants, and `Z` for the voiced sibilants.

In order to see how the above methodological decisions affect the database, and in particular, how the current database differs from Maddieson (1984) in terms of the spirants, we can compare the present database with that of UPSID-84. Table 4.5 shows the frequency with which various fricative pairs violate the implicational relations with respect to voicing in UPSID-CB used here, with the violation rates from UPSID-84 (reported in Maddieson (1984)) repeated for reference from Table 1.9 in the third column. In general, the class behaviour of plosives and sibilants does not differ very much between the two versions; similarly, the violation rates for the bilabial and labiodental pairs also does not differ. Interestingly, the violation rates for both the dentals and the velars is higher in the present database than that of Maddieson (1984).

Table 4.5: Voiced fricatives without voiceless fricatives in current database, with comparison to violation rates in Maddieson (1984), repeated here for reference. Sibilants value for Maddieson (1984) calculated by summing number of unpaired voiced fricatives over total number of voiced fricatives for /s, z/, /ʃ, ʒ/ and /ʂ, ʐ/.

| | Unpaired voiced segment / total voiced segment | Exceptions as % of cases (current) | Exceptions as % of cases (Maddieson, 1984) |
|---|---|---|---|
| plosives | 5/324 | 1.5% | — |
| sibilants | 0/149 | 0% | 2% |
| /ɸ, β/ | 38/51 | 74.5% | 75% |
| /f, v/ | 19/93 | 20.4% | 21.5% |
| /θ, ð/ | 22/31 | 71% | 57.1% |
| /x, ɣ/ | 23/50 | 46% | 37.5% |

Finally, it is worth pointing out that the values given for plosives and sibilants somewhat underestimate the rate of violations because they group plosives and sibilants into classes. That is, as long as a language had a voiceless stop it was considered to have a voiceless series, and as long as it had a voiced stop, it was considered to have a voiced series. As a result, a stop inventory such as /t, k, b, d, g/ would not be considered to have a violation of the implicational relation of obstruent voicing, in spite of the fact that it has voiced /b/ without its voiceless counterpart [p]. Likewise, because the sibilants are treated as a class, an inventory with /s, ʒ/, but not /ʃ/ would not incur a violation.

## 4.3 Relationship to plosive voicing

Within the voiced obstruents, voiced fricatives are thought to be particularly disfavoured, or marked. The additional markedness of voiced fricatives relative to voiced stops is articulated by Maddieson with the implicational relation stated in (2):

(2) Implicational relation within voiced obstruents (Maddieson, 2010, pg. 536)
"[I]f a language has a fricative voicing contrast, then it is highly likely that it also has a plosive voicing contrast."

In this section I question whether the proposed asymmetry within voiced obstruents applies equally to sibilants and spirants. Maddieson (2010) cites the numbers in Table 4.6 to make the following argument:

> The fricative voicing contrast predominantly occurs in languages that also have a voicing contrast in plosives. Of the 221 languages with fricative voicing. . . 80% have both contrasts. Only 62% would be expected if fricative voicing was distributed independently of plosive voicing" (pg. 537).

The numbers in Table 4.6 are, presumably, derived from LAPSyD (Maddieson et al., 2013), an ongoing expansion of the UPSID-92 database, but that is neither geographically nor

Table 4.6: Plosive and fricative voicing contrasts (number of languages) (Maddieson, 2010)

| Plosive voicing | Fricative voicing | | |
|---|---|---|---|
| | Yes | No | Total |
| Yes | 177 | 218 | 395 |
| No | 44 | 198 | 242 |
| Total | 221 | 416 | 637 |

geneologically balanced.[6] Though Maddieson is careful not to claim statistical significance, the percentages that he provides (for expected counts) are derived from a chi-squared test of independence.

Table 4.7 shows contingency tables and results of chi-squared tests of independence assessing whether the presence of a voicing contrast in stops is independent of the presence of a voicing contrast in fricatives (sibilants and spirants together), sibilants, and spirants. In this section, a voicing contrast in plosives is interpreted in its most liberal sense: a language is considered to have a voicing contrast in plosives as long as there is one voiceless plosive (whether laryngealized, aspirated, or plain), and one voiced plosive (whether pre- or post-nasalized, implosive, or plain).

Grouping the sibilants and spirants into a single class of fricatives, Table 4.7 shows that the presence of voiced fricatives is not independent of the presence of voiced plosives ($\chi^2 = 15.523$, p-value $= 8.153 \times 10^{-5}$), replicating the results of Maddieson (2010). If, however, the sibilants and spirants are considered separately, we see that this effect appears to be driven by the sibilants: the presence of a voicing contrast in sibilants is not independent of

[6] I say "presumably" because Maddieson (2010) does not specifically cite the database that these numbers are derived from, but it seems reasonable given Maddieson's ongoing work with LAPSyD. Importantly, although LAPSyD contains more languages than UPSID-92 (637 in Maddieson (2010) as of 2013 compared to 451 in the expanded UPSID database), it is neither geographically nor geneologically balanced, and so it is not possible to assess independence in the LAPSyD database.

a voicing contrast in plosives ($\chi^2 = 17.376$, p-value $= 3.067 \times 10^{-5}$), but the presence of a voicing contrast in spirants is independent ($\chi^2 = 3.3576$, p-value $= 0.0669$).

Table 4.7: Contingency tables and Pearson's chi-squared test with Yates' continuity correction tests of independence for the presence of voicing contrast in plosives with (a) fricatives (sibilants and spirants); (b) sibilants only; (c) spirants only. Expected counts are below observed counts.

| voiced plosives | | all fricatives | | | sibilants | | | spirants | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | No | Yes | Total | No | Yes | Total | No | Yes | Total |
| | No | 93 | 26 | 119 | 99 | 20 | 119 | 100 | 19 | 119 |
| | | 74.67 | 44.33 | | 80.21 | 38.79 | | 92.35 | 26.65 | |
| | Yes | 190 | 142 | 332 | 205 | 127 | 332 | 250 | 82 | 332 |
| | | 208.33 | 123.67 | | 223.79 | 108.21 | | 257.65 | 74.35 | |
| | Total | 283 | 168 | 451 | 304 | 147 | 451 | 350 | 101 | 451 |
| | | $\chi^2 = 15.523$, df $= 1$ | | | $\chi^2 = 17.376$, df $= 1$ | | | $\chi^2 = 3.3576$, df $= 1$ | | |
| | | p-value $= 8.153 \times 10^{-5}$ | | | p-value $= 3.067 \times 10^{-5}$ | | | p-value $= 0.0669$ | | |

These results suggest that there is a distinction between spirants and sibilants with respect to their relationship to stop voicing. Within the class of spirants, it is not possible to reliably test independence of individual pairs of segments, because the counts are too small; for example, there are only 13 languages with both /β, ɸ/, and of these, only 3 lack a voicing contrast in the stops, meaning that the cell counts of the relevant contingency tables are too low in order to run a chi-squared test.

In what follows I use a Bayesian approach to analysing violation rates, for two reasons. First, the model used is robust in the face of small numbers, allowing me to probe whether there are differences between, say, bilabial /ɸ, β/ and velar /x, ɣ/. Second, and most importantly, if the focus of inquiry is implicational relations, then the appropriate approach to assessing the strength of these relations is by looking at *violation rates*, which directly test the strength of an implicational relation.

### 4.3.1 Assessing violations

Contingency tables, as in Table 4.6, show how two variables are associated. One of their main uses, and one possible approach to analysing the relationship of voicing within fricative pairs, is to assess independence of two segments, using a chi-squared test of independence. Non-independence is a weaker notion than an implicational relation, and assessing whether two segments are independent does not address the question of violations with respect to the hypothesized implicational relations. To illustrate the difference between the two variables being correlated as opposed to being in an implicational relationship, consider the contingency tables associated with the relationship of voiced and voiceless fricative pairs: Table 4.8 shows contingency tables for velar spirants /x, ɣ/; Table 4.9 shows contingency tables for sibilants.

| | no /ɣ/ | /ɣ/ | Total |
|---|---|---|---|
| no /x/ | 324 | 23 | 347 |
| /x/ | 77 | 27 | 104 |
| Total | 401 | 50 | 451 |

Table 4.8: Velar spirants

| | no Z | Z | Total |
|---|---|---|---|
| no S | 54 | 1 | 55 |
| S | 249 | 147 | 396 |
| Total | 302 | 149 | 451 |

Table 4.9: Sibilants

Applying the Pearson's chi-squared test with Yates' continuity correction test to the spirant pair /x, ɣ/, we obtain a chi-squared value of 28.412, and a highly significant p-value of $9.806 \times 10^{-8}$; thus we conclude the presence of /x/ and /ɣ/ in inventories is not independent. If the cooccurrence of /x/ and /ɣ/ were independent, then of the 104 languages with /x/, we would expect only 11% (approximately 12 languages) to also have /ɣ/; instead of the 104 languages with /x/, we have 27 languages with /ɣ/, or 26%. Likewise, of the 347 languages that lack /x/, we would expect 11%, or approximately 38 languages, to contain /ɣ/ if voicing in the velar spirant pair were not independent; instead of the 347 languages that lack /x/, we have only 23 languages that contain /ɣ/, or 6.6%.

Applying the Pearson's chi-squared test with Yates' continuity correction test to the sibilants, we obtain a chi-squared value of 25.722, and a highly significant p-value of $3.994 \times 10^{-7}$; thus we conclude the presence of voiceless and voiced sibilant classes in inventories is not independent. If voicing in sibilants were independent, then we would predict that approximately 33% (approximately 18 languages) would have voiced sibilants without voiceless sibilants; instead, we get only one.

The chi-squred test of independence shows that /x/ and /ɣ/ are not independent, and nor are the voiceless and voiced sibilants. However, despite apparently similar chi-squared values, the results are not directly comparable, and although we can reject the null hypothesis of independence for both the velar spirants and sibilants, doing so does not address the very different rates of violation for the implicational relation stated in (1c), that if a language has a voiced fricative, it has a voiceless counterpart. In particular, 46% of languages have /ɣ/ without /x/, but only one language has a voiced sibilant without a voiceless sibilant. The velar spirants exhibit both non-independence as well as high rates of violation for the implicational relation, while the relationship between voiceless and voiced sibilants is an example of non-independence with only one instance of a violation of the implicational relation.

In sum, in order to both address the issues that arise from the relative rarity of voiced spirants as well as more directly investigate violations of the implicational relation that voicing in fricatives implies a voicing contrast in the stops, I use a Bayesian approach to analysing violation rates. Specifically, I illustrate violations using credible intervals derived from a Binomial model with a beta prior distribution. Details of the model and model parameters are provided in Appendix 4.A, but interpreting the graphs is straightforward, and does not require an in-depth understanding of the model.

By way of example, consider Figure 4.2, which shows the rates of violations for the implicational relations of voicing within obstruent pairs for the present database, as listed in Table 4.5 and repeated here as Table 4.10 for convenience.

Table 4.10: Voiced fricatives without voiceless fricatives in current database, with comparison to violation rates in Maddieson (1984), repeated here for reference. Sibilants value for Maddieson (1984) calculated by summing number of unpaired voiced fricatives over total number of voiced fricatives for /s, z/, /ʃ, ʒ/ and /ʂ, ʐ/.

| | Unpaired voiced segment / total voiced segment | Exceptions as % of cases (current) | Exceptions as % of cases (Maddieson, 1984) |
|---|---|---|---|
| plosives | 5/324 | 1.5% | — |
| sibilants | 0/149 | 0% | 2% |
| /ɸ, β/ | 38/51 | 74.5% | 75% |
| /f, v/ | 19/93 | 20.4% | 21.5% |
| /θ, ð/ | 22/31 | 71% | 57.1% |
| /x, ɣ/ | 23/50 | 46% | 37.5% |

For each fricative pair, the violation rate is shown as an interval, where the midpoint of the interval corresponds, approximately, to the violation rate listed; I say "approximately" because the midpoint of the interval is a function of both the violation rate and the model parameters (*prior probabilities*). The graph shows intervals (as opposed to point estimates), because Bayesian estimates are distributions; the credible interval is interpreted as the probability that the interval contains the true value, based on the data and the model. In this chapter the thick interval line represents a 90% credible interval, while the thin interval line represents a 95% credible interval.

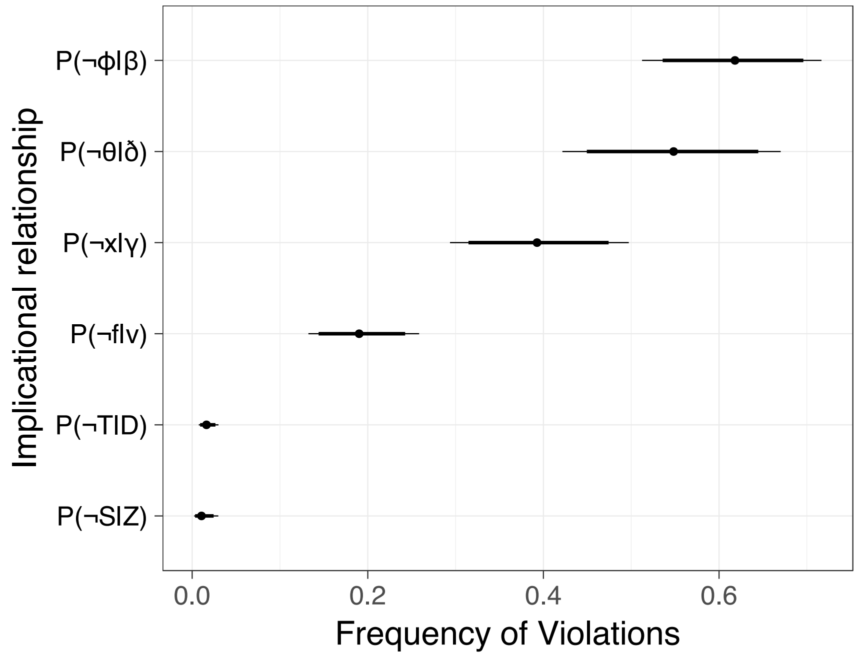



Figure 4.2: Credible intervals (90%/95%) showing violation rates for implicational relations of voicing for obstruent pairs; `/T, D/` refers to plosives, `/S, Z/` refers to sibilants.

### 4.3.2 Voicing relationships within obstruents: types of violations

From the perspective of *violations*, Maddieson's claim about the relationship between voicing contrasts in fricatives vs. stops fails if a language has a voicing contrast in the fricatives, but does not have a voicing contrast in the stops. There are two primary ways that this could happen. First, if the inventory has only one stop series, either voiceless or voiced, any voicing contrast in the fricatives will incur a violation. Second, a language can incur a violation if an inventory has multiple stop series, but all are voiceless, say, the contrast is between unaspirated /p, t, k/ and aspirated /p$^{h}$, t$^{h}$, k$^{h}$/ or ejective /p', t', k'/.

In principle, another possibility arises under a more restricted interpretation of the implicational relation. Specifically, violations of a more local sense can also occur in inventories that have a voicing contrast at at least one place of articulation, but lack the contrast where there is a fricative voicing contrast. For example, if an inventory with a voicing contrast in the stops has both /ɸ, β/ but is missing /p/, then this could count as an instance of the violation with respect to voicing contrasts in labial obstruents.

Maddieson (2010) frames the relationship between voicing in fricatives and stops in terms of *contrast*, by which he simply means that both members of the voicing pair are present.[7] The voicing relationship between fricatives and stops can also be formulated more weakly, strictly in terms of the *presence* (or absence) of voiced fricatives and stops. This alternative formulation is not likely to make a difference for the relationship between stops and sibilants, since voiced sibilants almost never appear in an inventory without a voiceless sibilant. The two formulations are not equivalent for spirants, which often appear unpaired in inventories.

In order to facilitate the discussion, I introduce a terminological distinction for the two formulations of the voicing relationship between fricatives and stops, and articulate the implicational relations that go with each in (3). (3a) is the implicational relation articulated by Maddieson (2010); (3b) articulates the weaker implicational relation, referring only to whether the voiced fricative is present, regardless of whether it is part of a voicing contrast.

(3) Implicational relations of voicing in obstruents
   a. **Cross-manner (voicing) contrast:** If a language has a fricative voicing contrast, then it is highly likely that it also has a plosive voicing contrast
   b. **Cross-manner (voicing) presence:** If a language has a voiced fricative, then it is highly likely that it also has voiced plosives

To illustrate these terms, consider the inventory of Luiseño `(lui)` in Table 4.11. Luiseño exhibits all three kinds of violations: *within-manner*, *cross-manner contrast*, and *cross-manner*

[7] This is distinct from the notion of contrast articulated by Dresher (2009).

*presence*. Since the Luiseño has /ð/ without /θ/, /ð/ incurs a *within-manner violation*, of the kind familiar from (1); /v/ does not incur a within-manner violation since /f/ is in the inventory. However, since there is not a voicing contrast in the plosives, /f, v/ incur a *cross-manner contrast violation*. Finally, both /v, ð/ incur *cross-manner presence violations* because they are in an inventory that lacks voiced plosives, independently of whether they incur within-manner or cross-manner contrast violations.

Table 4.11: Consonant inventory of Luiseño (`lui`), Maddieson et al. (2013)

| | Labial | Dental | Alveolar | Palatoalveolar | Retroflex | Palatal | Velar | Uvular | Glottal |
|---|---|---|---|---|---|---|---|---|---|
| Stop | p | t | | | | | k $k^w$ | q $q^w$ | ʔ |
| Fricative | f v | ð | s | ʃ | ʂ | | x $x^w$ | | h |
| Affricate | | | | tʃ | | | | | |
| Nasal | m | n | | | | | ŋ | | |
| Lateral | | l | | | | | | | |
| Rhotic | | | | ɾ | | | | | |
| Approximant | β̞ | | | | | j | | | |

Cross-manner violations can occur in very reasonable-looking inventories as well, as in Naga (`njo`) in Table 4.12. Because Naga has both /s, z/ it does not incur an within-manner violation, but the lack of voiced plosives means that /s, z/ incur a cross-manner contrast violation, and /z/ incurs a cross-manner presence violations.

Logically, it is possible to incur a cross-manner contrast violation without incurring a cross-manner presence violation, if an inventory has a voicing contrast in fricatives but only a voiced stop series; for example, an inventory that contains a voicing contrast in the fricatives, such as /f, v/, but that only has voiced plosives /b, d, g/. However, since it is rare for languages to have only a voiced stop series, this configuration rarely attains,

Table 4.12: Consonant inventory of Naga, Ao (`njo`), Maddieson et al. (2013)

| | Labial | Dental | Alveolar | Palatoalveolar | Velar | Glottal |
|---|---|---|---|---|---|---|
| Stop | p pʰ | t tʰ | | | k kʰ | ʔ |
| Fricative | | | s z | | | h |
| Affricate | | | ts tsʰ | tʃ tʃʰ | | |
| Nasal | m | n | | | ŋ | |
| Lateral | | | l | | | |
| Rhotic | | | ɻ | | | |
| Approximant | w | | | | | |

and, as mentioned above, all UPSID-CB languages that are reported to have only a voiced plosive series do not have voiced fricatives, so such a configuration is unattested. In practice then, cross-manner contrast violations typically arise as in Luiseño, namely when the voiced fricative does not incur an within-manner violation (i.e., is paired), and there is not a voiced stop series.

In the remainder of this section I assess the relationship between voiced fricatives and voiced stops with three investigations. First, I consider the narrowest interpretation of (3a), restricted to place of articulation; namely, languages that have a voicing contrast in the fricatives, but lack a voicing contrast in the the stops at that place of articulation. Second, I assess violation rates of (3a) in its broader interpretation, where a language has a voicing contrast in the fricatives, but does not have a voicing contrast in the stops, either because there is only one series, or all series are voiceless. Finally, I assess violation rates of the related implicational relation (3b), where a language may have voiced fricatives (regardless of contrast), but lacks voiced plosives.

### 4.3.3 Violations: pairwise segments

A language is considered to have a voicing contrast in the stops if it has at least one voiceless (plain, aspirated, ejective/laryngealized) stop and one voiced (plain, implosive, pre/post-nasalized) stop. Typically this results in a voicing contrast at at least one place of articulation (labial, coronal, velar), but some inventories do not conform to this. For example, Warao (`wba`) is listed in LAPSyD as having /b, t, k, kʷ/, but this is because "Osborn (1966) notes that /p/ is usually pronounced [b] so it is treated as such here".[8] Another example is the inventory of Yagaria (`ygr`), which has a voiceless series /p, t, k/, voiced /g/ and implosive /ɓ, ɗ/, of which Maddieson notes, "/ɓ, ɗ/ occur phonetically as plain voiced stops in word-initial position, but their medial variant is taken as unmotivated unless they are underlyingly glottalized".[9] Similarly, Rotokas (`roo`) has only six consonants in its inventory, listed as /p, t, k, β, ð, g/, and it is likely that a similar analysis is at play, though there is no commentary in LAPSyD for this inventory. None of these languages has a voicing contrast in the fricatives, and so the analysis of these inventories does not affect the present investigation.

Several languages have what might be called *degenerate voicing contrast*, where there is only a voicing contrast at one place of articulation, even though the typical three-place series exists otherwise. For example, four languages (Diyari (`dif`), Gadsup (`gaj`), Trumai (`tpy`), Yaathe (`fun`)) have only a coronal voicing contrast; none of these has a fricative voicing contrast, though Gadsup has unpaired /β/. Seven languages have only a labial voicing contrast (both /p, b/ are in the inventory, but no coronal or velar voiced stop): Alabama (`akz`), Japreria (`jru`), Lakkia (`lbc`), Pech (`pay`), Tzeltal (`tzh`), Yaqui (`yaq`), and Nenets

[8]Quoted from the field Consonant Notes, written by Ian Maddieson, at `http://www.lapsyd.ddl.ish-lyon.cnrs.fr/lapsyd/index.php?data=view&code=179`.

[9]From `http://www.lapsyd.ddl.ish-lyon.cnrs.fr/lapsyd/index.php?data=view&code=606`, only available with (free) registration to LAPSyD.

(yrk). Of these, only Nenets has voiced fricatives /β, ð, ðʲ/, but these are unpaired, and so do not instantiate a voicing contrast in the fricatives.

#### 4.3.3.1 Labials

There are eight languages in UPSID that lack both /p, b/: Aleut (ale), Cherokee (chr), Eyak (eya), Hupa (hup), Mor (mhz), Seneca (see), Tlingit (tli), and Wichita (wic). Only three of these languages have a voicing contrast in the stops: Eyak (coronals and velars), Tlingit (velars), and Mor (velars); only Mor has the voiced fricative /β/, but there is no voicing contrast in the fricatives because there is no voiceless labial fricative in Mor.

**Missing** /p/ There are 34 languages with a voicing contrast in the stops, have /b/, but are missing /p/, listed in Table 4.13. Most of the languages with a voicing contrast in the stops that lack /p/ either only have voiceless fricatives ("voiceless only"), or have a voicing contrast in the sibilants and perhaps some other places, as in Egyptian Arabic. Only three "missing /p/" languages have a voicing contrast in the labial fricatives: Fe'fe' (fmp) has /f, v/, Kwoma (kmo) has /ɸ, β/, and Yareba (yrb) has /ɸ, v/.

**Missing** /b/ There are only two languages with both a coronal and velar voicing contrast, but missing /b/. Khalkha Mongolian (khk) has /ɸ, β/, but only voiceless labial stops, and thus a violation of (2) within the labials.[10] Mixe (mto) has a voicing contrast in the sibilants (/ʃ, ʒ/), but only has unpaired /v/, and so does not incur a violation.

[10] In fact, Khalkha only has a voicing contrast in the palatal and velar stops, contrasting /cʰ, ɟ/ and /kʰ, g/. Khalkha also happens to have unpaired /βʲ/, but this is irrelevant for the present investigation.

Table 4.13: Languages with voicing contrast in stops but missing /p/ .
Fricative contrast: *voiceless only* means only voiceless fricatives are present; *different place* means that both voiceless and voiced fricatives are present but do not contrast at the same place of articulation.

| ISO | Language | Fricative Contrast | ISO | Language | Fricative Contrast |
|---|---|---|---|---|---|
| aey | Amele | different place | mde | Maba | sibilants |
| arz | Arabic, Egyptian | sibilants, uvular, pharyngeal | mdx | Dizi | sibilants |
| bej | Beja | voiceless only | nhu | Noni | voiceless only |
| cjv | Chuave | voiceless only | nrb | Nera | voiceless only |
| fia | Nobiin | voiceless only | nyi | Nyimang | voiceless only |
| fmp | Fe'fe' | f, v; sibilants | ses | Songhay | sibilants |
| fvr | Fur | voiceless only | shi | Tashlhiyt | sibilants |
| gmo | Kullo | sibilants | som | Somali | voiceless only |
| hau | Hausa | sibilants | sqt | Soqotri | sibilants |
| irh | Irarutu | voiceless only | taq | Tamasheq | sibilants |
| kbk | Koiari | voiceless only | tet | Tetun, Western | voiceless only |
| ket | Ket | different place | tig | Tigre | sibilants |
| kew | Kewa | voiceless only | tiy | Tiruray | voiceless only |
| kmo | Kwoma | ɸ, β | tma | Tama | voiceless only |
| knc | Kanuri | sibilants | wba | Warao | voiceless only |
| kun | Kunama | voiceless only | yrb | Yareba | ɸ, v |
| kwd | Kwaio | voiceless only | yss | Yessan-Mayo | voiceless only |

#### 4.3.3.2 Coronals and velars

None of the languages in the database has a voiced coronal stop without a voiceless counterpart among languages with a voicing contrast. Two languages, Tigak (`tgc`) and Pirahã (`myp`) have labial and velar voicing contrasts, but only voiceless /t/ without voiced /d/. Pirahã has /s/ but no voiced fricatives. Tigak has unpaired /β/ (irrelevant to the present investigation) and potentially has a voicing contrast in the sibilants between /s/ and the voiced lateral fricative /lʒ/.

Only two languages have a voiced velar stop without a voiceless counterpart: Kwaio (kwd) has a series of prenasalized voiced stops /mb, nd, ŋg, ŋgʷ/, but only has voiceless /t/ (and glottal stop /ʔ/); Usan (wnu) has both plain voiced stops /b, d, g/ and prenasalized voiced stops /mb, nd, ŋg/, but the voiceless series has /p, t, ʔ/, thus missing the voiceless velar /k/. Neither Kwaio nor Usan have a voicing contrast in the fricatives.

There are 28 languages that have a voicing contrast at both the labial and coronal places, have a voiceless velar stop such as /k/, but are missing /g/. Only one instantiates a violation of (2) at the velar place of articulation: Sui (swi) has velar fricatives /x, ɣ, ɣ̰/, voiceless velar stops, but no voiced velar stops. Of the remaining languages, 21 have either no fricatives at all, or only voiceless fricatives: Andoke (ano), Baré (bae), Bruu (brv), Bribri (bzd), Caddo (cad), Nyah Kur (cbn), Cham (cja), Asmat (cns), Efik (efi), Ekari (ekg), Hixkaryana (hix), Irantxe (irn), Itonama (ito), Khasi (kha), Naasioi (nas), Yay (pcc), Pomo (pom), Thai (tha), Wintu (wnw), Yagua (yad), Yawa (yva). Five languages have voiced fricatives (either paired or unpaired) at labial and coronal places: Cubeo (cub), Cayubaba (cyb), Guahibo (guh), Lu (khb), and Tacana (tna). Kawaiisu (xaw) has voiced velar fricatives /ɣ, ɣʷ/, but they are unpaired and so do not instantiate a violation of (2).

In sum, if we interpret the implicational relation that a voicing contrast in fricatives implies a voicing contrast in stops at the same place of articulation, there are very few violations overall, summarized in Table 4.14. Such languages warrant further study, as the fricative in question may pattern with the stops, and so phonologically function to "fill the gap".

In what follows, I interpret the voicing contrast of stops to refer to the existence of at least one member of a stop series, but I assess rates of violation separately for sibilants and

Table 4.14: Languages that incur a cross-manner contrast violation, narrowly interpreted.

| ISO | Language | Family | Area | Size | ɸ | β | f | v | θ | ð | x | ɣ | S | Z | Pairwise violation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| kmo | Kwoma | Sepik | Pacific | 24 | ✓ | ✓ | | | | | | | ✓ | | missing /p/ |
| mhz | Mor | Austronesian | Pacific | 14 | | ✓ | | | | | | | ✓ | | missing /p, b/ |
| mto | Mixe, North Highland | Mixe-Zoquean | Americas | 14 | | | | ✓ | | | | | ✓ | ✓ | missing /b/ |
| swi | Sui | Tai-Kadai | Eurasia | 47 | ✓ | | | | | | ✓ | ✓ | ✓ | ✓ | missing /ɣ/ |
| xaw | Kawaiisu | Uto-Aztecan | Americas | 25 | | | | ✓ | | | | ✓ | ✓ | ✓ | missing /g/ |
| yrb | Yareba | Trans-New Guinea | Pacific | 13 | ✓ | | | ✓ | | | | | ✓ | | missing /p/ |
| yrk | Nenets | Uralic | Eurasia | 27 | | ✓ | | | | ✓ | ✓ | | ✓ | | missing /d, g/ |

the individual spirants.

### 4.3.4 Cross-manner contrast violations

We return to the question of whether sibilants and spirants violate the implicational relationship to stop voicing, as articulated by Maddieson (2010), at the same rate. Violations occur when a language has a voicing contrast in the fricatives, but lack a voicing contrast in the plosives; as shown above in Table 4.7, the presence of a voicing contrast in sibilants is not independent of a voicing contrast in the plosives, but the presence of a voicing contrast in the spirants is.

Figure 4.3 shows that the spirants are not uniform with respect to cross-manner contrast violations, using 90(/95)% credibility intervals. For sibilants and all spirant pairs other than the velar /x, ɣ/, the rate of cross-manner contrast violations is low; for sibilants and /f, v/ the violation rate is between 10% and 20%, while for spirant pairs /θ, ð/ and /ɸ, β/ the violation rate hovers at just under 20%, but with much more uncertainty. The velars clearly standout, with a mode violation rate at near 40%. Note that the wide credibility intervals for the spirants, in particular the bilabial, dental and velar pairs, is due to the sparsity of data. Nonetheless, the velars stand out, suggesting that they are driving the independence between languages with a voicing contrast in the stops and spirants.

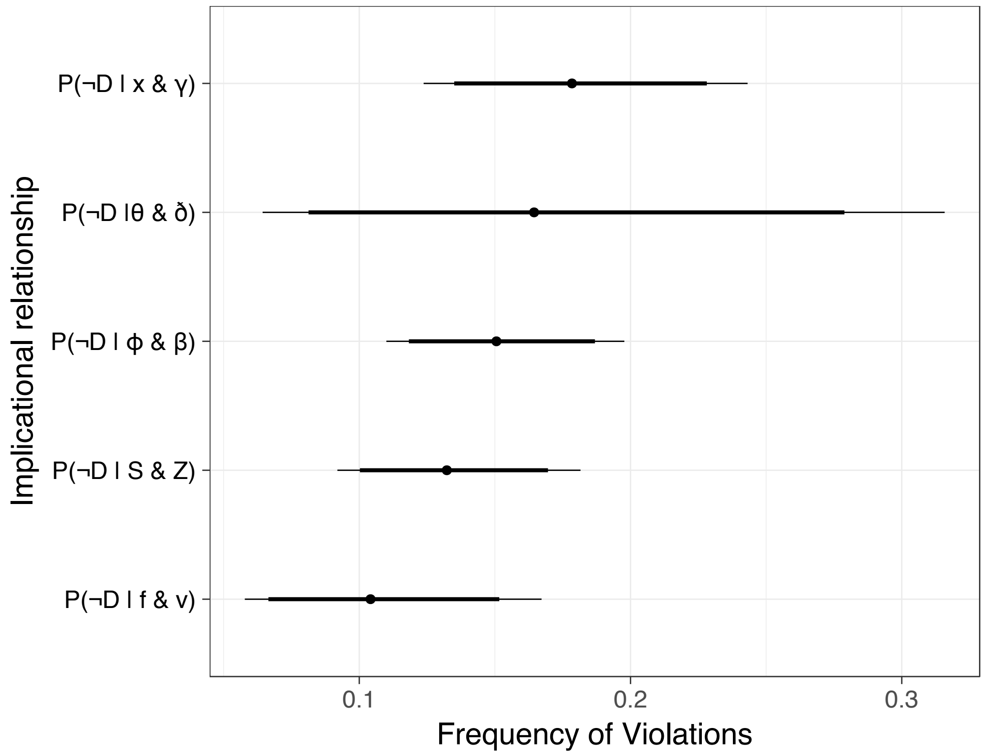



Figure 4.3: Proportion of cross-manner contrast violations by fricative pair; probability = 90(/95)%

#### 4.3.4.1 Stop systems in sibilant cross-manner contrast violations

Twenty languages have a voicing contrast within the sibilants but do not contrast voicing in the stops. Of these, six have a single (voiceless) stop series; the remaining 14 have a voiceless laryngeal contrast in the stops, as shown in Table 4.15. Of the six that have only a single stop series, Aleut (`ale`) and Chinanteq (`chq`) contrast voiced and voiceless sonorants; Atayal (`tay`), Yanesha' (`ame`) and Mari (`mhr`) have other voiced fricatives (voiced spirants) in the inventory; Guambiano (`gum`) reportedly has /ʒ/, but this "occurs only in one common diminutive formation" according to the consonant notes in LAPSyD, and thus the voicing contrast it has in the sibilants is marginal.[11]

[11] `http://www.lapsyd.ddl.ish-lyon.cnrs.fr/lapsyd/index.php?data=view&code=384`

Table 4.15: Languages that incur a cross-manner contrast violation in the sibilants.
Legend: T = plain; T' / T̰ = ejective/laryngealized; ʰT / Tʰ = pre/postaspirated

| | | | | | Voiceless plosives | | | |
|---|---|---|---|---|---|---|---|---|
| ISO | Language | Family | Area | Size | T | T' / T̰ | ʰT / Tʰ | Contrast? |
| aht | Ahtna | Eyak-Athabaskan | Americas | 27 | ✓ | ✓ | ✓ | ✓ |
| ale | Aleut, Eastern | Eskimo-Aleut | Americas | 24 | ✓ | | | |
| ame | Yanesha' | Maipurean | Americas | 24 | ✓ | | | |
| bca | Bai, Central | Sino-Tibetan | Eurasia | 18 | ✓ | | ✓ | ✓ |
| chp | Dene | Eyak-Athabaskan | Americas | 38 | ✓ | | ✓ | ✓ |
| chq | Chinantec, Quiotepec | Otomanguean | Americas | 33 | ✓ | | | |
| cmn | Mandarin | Sino-Tibetan | Eurasia | 25 | ✓ | | ✓ | ✓ |
| cnh | Haka Chin (Lai) | Sino-Tibetan | Eurasia | 45 | ✓ | | ✓ | ✓ |
| gqu | Gelao | Tai-Kadai | Eurasia | 31 | ✓ | | ✓ | ✓ |
| gum | Guambiano | Paezan | Americas | 19 | ✓ | | | |
| hnj | Hmong Njua | Hmong-Mien | Eurasia | 43 | ✓ | | ✓ | ✓ |
| hye | Armenian | Indo-European | Eurasia | 31 | ✓ | ✓ | ✓ | ✓ |
| itl | Itelmen | Chukotko-Kamchatkan | Eurasia | 27 | ✓ | ✓ | | ✓ |
| mhr | Mari, Meadow | Uralic | Eurasia | 25 | ✓ | | | |
| nav | Navajo | Eyak-Athabaskan | Americas | 35 | ✓ | ✓ | | ✓ |
| niv | Gilyak | Language Isolate | Eurasia | 28 | ✓ | | ✓ | ✓ |
| njo | Naga, Ao | Sino-Tibetan | Eurasia | 22 | ✓ | | ✓ | ✓ |
| quh | Quechua, Cochabamba | Quechuan | Americas | 28 | ✓ | ✓ | ✓ | ✓ |
| tay | Atayal | Austronesian | Pacific | 19 | ✓ | | | |
| wuu | Changzhou Wu | Sino-Tibetan | Eurasia | 25 | ✓ | | ✓ | ✓ |

#### 4.3.4.2 Stop systems in spirant cross-manner contrast violations

Nineteen languages have a voicing contrast in the spirants at at least one place of articulation but do not contrast voicing in the stops. Of these, six have a single (voiceless) stop series, as listed in Table 4.16. Thirteen of the languages in Table 4.16 are repeats from Table 4.15, since these languages contrast voicing in the sibilants. Of the remaining six, Beembe (`beq`), Fwai (Po-Ai) (`fwa`), and Khalkha Mongolian (`khk`) have a laryngeal contrast within the voiceless plosives; Inuktitut (West Greenlandic) `(kal)`, Luiseño `(lui)`, and Spanish `(spa)` all have a single series of voiceless plosives.

Table 4.16: Languages that incur a cross-manner contrast violation in the spirants.
Legend: T = plain; T' / T̰ = ejective/laryngealized; $^{h}$T / T$^{h}$ = pre/postaspirated

| | | | | | | | | | | | | | | | Voiceless plosives | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ISO | Language | Family | Area | Size | ɸ | β | f | v | θ | ð | x | ɣ | S | Z | T | T' / T̰ | $^{h}$T / T$^{h}$ | Contrast? |
| ale | Aleut, Eastern | Eskimo-Aleut | Americas | 24 | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| ame | Yanesha' | Maipurean | Americas | 24 | | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| bca | Bai | Sino-Tibetan | Eurasia | 18 | | | ✓ | ✓ | | | ✓ | | ✓ | ✓ | ✓ | | ✓ | ✓ |
| beq | Beembe | Niger-Congo | Africa | 16 | | | ✓ | ✓ | | | | | ✓ | | ✓ | | ✓ | ✓ |
| chp | Dene | Eyak-Athabaskan | Americas | 38 | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ |
| cnh | Haka Chin | Sino-Tibetan | Eurasia | 45 | | | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ |
| fwa | Fwai | Austronesian | Pacific | 19 | | | ✓ | ✓ | | | | | ✓ | | ✓ | | ✓ | ✓ |
| gqu | Gelao | Tai-Kadai | Eurasia | 31 | | | ✓ | ✓ | | | ✓ | | ✓ | ✓ | ✓ | | ✓ | ✓ |
| hnj | Hmong Njua | Hmong-Mien | Eurasia | 43 | | | ✓ | ✓ | | | | | ✓ | ✓ | ✓ | | ✓ | ✓ |
| itl | Itelmen | Chukotko-Kamchatkan | Eurasia | 27 | ✓ | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |
| kal | Inuktitut | Eskimo-Aleut | Americas | 19 | | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | | | |
| khk | Mongolian | Altaic | Eurasia | 35 | ✓ | ✓ | ✓ | | | | ✓ | | ✓ | | ✓ | | ✓ | ✓ |
| lui | Luiseno | Uto-Aztecan | Americas | 24 | | | ✓ | ✓ | | ✓ | ✓ | | ✓ | | ✓ | | | |
| mhr | Mari | Uralic | Eurasia | 25 | | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| nav | Navajo | Eyak-Athabaskan | Americas | 35 | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |
| niv | Gilyak | Language Isolate | Eurasia | 28 | ✓ | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ |
| spa | Spanish | Indo-European | Eurasia | 20 | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | | |
| tay | Atayal | Austronesian | Pacific | 19 | | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| wuu | Changzhou Wu | Sino-Tibetan | Eurasia | 25 | | | ✓ | ✓ | | | | | ✓ | ✓ | ✓ | | ✓ | ✓ |

To summarize, cross-manner contrast violations occur in approximately equal numbers for sibilants and spirants overall, but velar spirants incur higher rates of violation, and appear to be driving the non-independence between spirant voicing and plosive voicing contrasts. Thirteen languages that incur cross-manner contrast violations in the spirants also incur such violations in the sibilants. Overall, approximately two-thirds of such violations exhibit a laryngeal contrast within the voiceless plosives. Such cases warrant further study, since in these languages the voicing contrast in the fricatives might parallel—in terms of phonological patterning—the laryngeal contrast in the plosives. Conversely, the patterning of voiced fricatives in those languages that have only a single plosive series are interesting in light of their isolation within the obstruent system.

### 4.3.5 Cross-manner presence violations

Similar to what was shown in Table 4.7, Table 4.17 shows that when sibilants and spirants are grouped together as fricatives, the presence of a voiced fricatives is not independent from the presence of voiced plosives; when sibilants and spirants are treated separately, we see that the non-independence is driven by the sibilants, and not the spirants.

Table 4.17: Contingency tables and Pearson's chi-squared test with Yates' continuity correction tests of independence for the presence of voiced plosives with (a) all voiced fricatives (sibilants and spirants); (b) voiced sibilants only; (c) voiced spirants only. Expected counts are below observed counts.

| | | all fricatives | | | sibilants | | | spirants | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | No | Yes | Total | No | Yes | Total | No | Yes | Total |
| voiced plosives | No | 75 | 39 | 114 | 94 | 20 | 114 | 80 | 34 | 114 |
| | | 60.16 | 53.84 | | 76.59 | 37.41 | | 72.04 | 41.96 | |
| | Yes | 163 | 174 | 337 | 209 | 128 | 337 | 205 | 132 | 337 |
| | | 177.84 | 159.16 | | 226.41 | 110.59 | | 212.96 | 124.04 | |
| | Total | 238 | 213 | 451 | 303 | 148 | 451 | 285 | 166 | 451 |
| | | $\chi^2 = 9.6863$, df = 1 | | | $\chi^2 = 15.226$, df = 1 | | | $\chi^2 = 2.8089$, df = 1 | | |
| | | p-value = 0.001856 | | | p-value = 0.00009538 | | | p-value = 0.09375 | | |

Figure 4.4 shows that the spirants are not uniform with respect to cross-manner presence violations, using 90(/95)% credibility intervals. For sibilants and labiodental /v/, the rate of cross-manner presence violations are low: in other words, if an inventory has a voiced sibilant or /v/, the implicational relation that it is likely to have a voiced stop series appears to hold. For the spirants /β, ð, ɣ/, however, the rates of violation are higher; this is different than what was seen in Section 4.3.4, where the non-independence seemed to be driven primarily

by the velars. Note however that Figure 4.4 shows the rate of cross-class presence violations, which conflates cases of paired and unpaired voiced spirants.

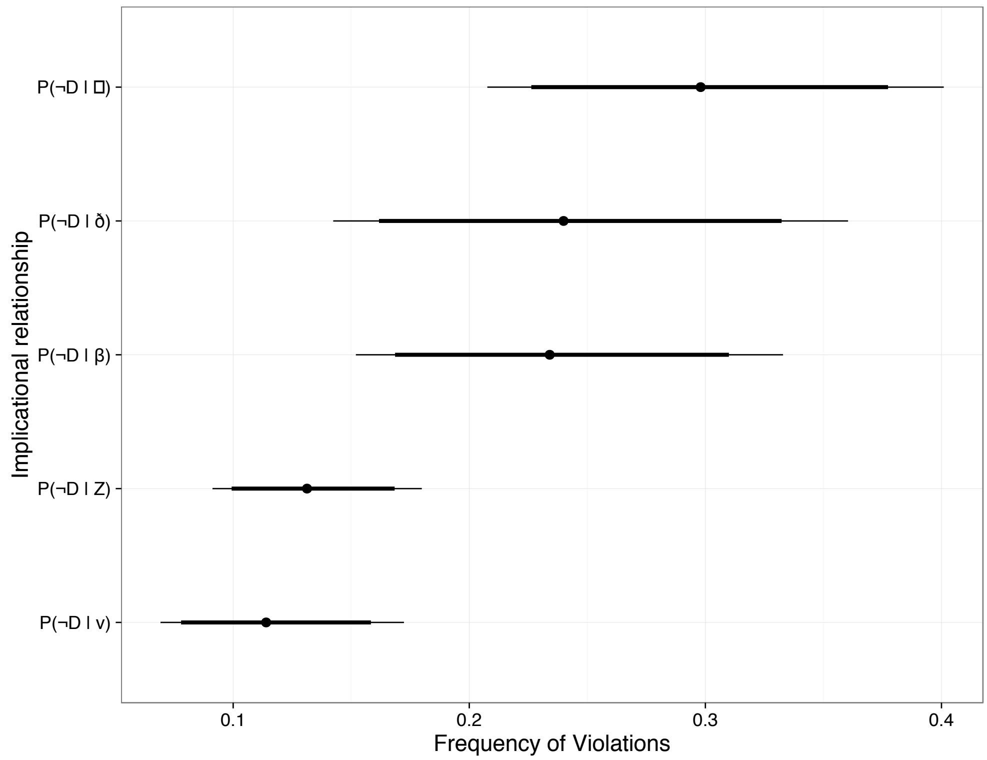



Figure 4.4: Proportion of cross-manner presence violations by fricative pair; probability = 90(/95)%

There are 22 languages that have unpaired voiced spirants that also lack voiced stops, listed below in Table 4.18. Of these languages, only five have a laryngeal contrast among voiceless plosives, less than 25%. Also of note, five languages do not have any sibilants at all.

Table 4.18: Languages that incur a cross-manner presence violation in the spirants; unpaired spirants.
Legend: T = plain; T’ / T̰ = ejective/laryngealized; $^{h}$T / T$^{h}$ = pre/postaspirated

| ISO | Language | Family | Area | Size | ɸ | β | f | v | θ | ð | x | ɣ | S | Z | Voiceless plosives | | | Contrast? |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | T | T’ / T̰ | $^{h}$T / T$^{h}$ | |
| ale | Aleut, Eastern | Eskimo-Aleut | Americas | 24 | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| ame | Yanesha’ | Maipurean | Americas | 24 | | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| arn | Mapudungun | Mapudungu | Americas | 20 | ✓ | | | | | ✓ | | | ✓ | | ✓ | | | |
| axb | Abipon | Guaykuruan | Americas | 15 | | | | | | | | ✓ | | | ✓ | | | |
| chq | Chinantec | Otomanguean | Americas | 33 | ✓ | | | | | ✓ | | ✓ | ✓ | ✓ | ✓ | | | |
| ckt | Chukchi | Chukotko-Kamchatkan | Eurasia | 14 | | | | | | ✓ | | ✓ | ✓ | | ✓ | | | |
| cni | Campa | Maipurean | Americas | 15 | | ✓ | | | | | | | ✓ | | ✓ | | ✓ | ✓ |
| dih | Kumeyaay | Cochimi-Yuman | Americas | 25 | | ✓ | | | | | ✓ | | ✓ | | ✓ | | | |
| hop | Hopi | Uto-Aztecan | Americas | 22 | | | | ✓ | | | | | ✓ | | ✓ | | ✓ | ✓ |
| hus | Huastec | Mayan | Americas | 21 | | ✓ | | | ✓ | | | | ✓ | | ✓ | ✓ | | ✓ |
| hye | Armenian | Indo-European | Eurasia | 31 | | | | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| jic | Tol | Jicaquean | Americas | 22 | | ✓ | | | | | | | ✓ | | ✓ | ✓ | ✓ | ✓ |
| kal | Inuktitut | Eskimo-Aleut | Americas | 19 | | ✓ | ✓ | | | | ✓ | ✓ | ✓ | | ✓ | | | |
| kpy | Koryak | Chukotko-Kamchatkan | Eurasia | 18 | | | | ✓ | | | | ✓ | | | ✓ | | | |
| lui | Luiseno | Uto-Aztecan | Americas | 24 | | | ✓ | ✓ | | ✓ | ✓ | | ✓ | | ✓ | | | |
| mhr | Mari, Meadow | Uralic | Eurasia | 25 | | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| mph | Maung | Australian | Pacific | 17 | | | | | | | | ✓ | | | ✓ | | | |
| rro | Waima | Austronesian | Pacific | 9 | | ✓ | | | | | | | | | ✓ | | | |
| sel | Selkup, Tym dialect | Uralic | Eurasia | 18 | | | | | | | | ✓ | ✓ | | ✓ | | | |
| spa | Spanish | Indo-European | Eurasia | 20 | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | | |
| tay | Atayal | Austronesian | Pacific | 19 | | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| wyb | Wangaaybuwan-Ngiyambaa | Australian | Pacific | 15 | | | | | | ✓ | | | | | ✓ | | | |

### 4.3.6 Relationship to plosive voicing: summary

To summarize, Maddieson (2010) claims that voicing contrasts in plosives and fricatives are not independent, a claim that appears to be corroborated for all fricatives other than the velar spirants. However, when the investigation is extended to include voiced fricatives, regardless of whether or not they are paired, then /β, ð, ɣ/ exhibit higher rates of violation than do voiced sibilants. Moreover, the data suggests that the nature of the violations differ between paired and unpaired spirants: paired spirants that incur cross-manner contrast violations tend to occur in inventories that exhibit a voiceless laryngeal contrast, while unpaired spirants that incur cross-manner presence violations tend to occur in inventories that have only a single stop series.

## 4.4 Inventory size

It is a commonly held belief among phonologists that inventory size correlates roughly with segmental complexity: more complex segments are found in larger inventories, a hypothesis explored by Lindblom and Maddieson (1988). Though they do not provide an operational definition or exhaustive ranking of segment complexity, they divide segments into three articulatory classes: *basic*, *elaborated* and *complex*. The basis for the distinction between these classes are "largely intuitively based" (Lindblom and Maddieson, 1988, pg. 68). *Basic* articulations are those that do not deviate from the "default", as for example voiceless obstruents and voiced sonorants. *Elaborated* articulations involve some deviation from the default, such as voicing in fricatives or configurations of articulation that depart from the near-rest position of the articulators (e.g., retroflex). Finally, *complex* articulations combine two or more elaborated articulations, such as the segment transcribed [tɬʰ], which combines lateral release with aspiration. Lindblom and Maddieson (1988) conclude that, "Initially system 'growth' occurs principally in terms of basic consonants. Once these consonant types reach saturation, further growth is then achieved by first adding only elaborated articulations, then by invoking also complex segment types" (pg. 70). In more phonological terms, Lindblom and Maddieson (1988) restate the findings invoking *markedness*: "Small paradigms tend to exhibit 'unmarked' phonetics whereas large systems have 'marked' phonetics" (pg. 70).

According to Lindblom and Maddieson (1988), voiced fricatives are classified as elaborated "since they invoke both non-spontaneous voicing and the control of the additional noise source" (Lindblom and Maddieson, 1988, pg. 67). If the voiced non-sibilant fricatives [β, v, ð, ɣ] are elaborated articulations, then we expect to find them only in inventories that have already exhausted segments characterized as having a basic articulation, that is, we expect to find them primarily in larger inventories. To address this, for each voiced spirant [β, v, ð, ɣ], I assess whether inventory size is a predictor for its presence in an inventory. Finally, I consider whether there is an interaction between within-manner violations (paired vs. unpaired voiced spirants) and inventory size.

Consonant inventory sizes in UPSID are roughly normally distributed around a mean of 23 consonants with a slightly longer right tail. One language was removed from the analysis carried out in this section due to its unusually large consonant inventory, Ju|'hoan (`ktz`), with 82 consonants, which would unduly affect the model. As a basis of comparison, I first consider how voiceless and voiced plosives and sibilants distribute across inventory sizes, followed by voiceless spirants, and then finally voiced spirants. The model used is non-parametric regression analysis, which was chosen because it was not known whether the logit of the regression curve is linear, and in fact the assumption of linearity did not seem to hold when plotted (as will be seen in some of the plots below). The regression analysis was used to estimate the conditional expectation of the binary response variable (whether the inventory contains a voiced spirant) with inventory size as the predictor variable. Specifically, I used the `ksmooth` function in R, which implements the *Nadaraya-Watson estimator*.

The following graphs are interpreted as follows: the circles (located at either 0 or 1), are languages of a given inventory size that do or do not contain the segment (or segment class) in question. Note that what looks like one circle may in fact be multiple languages. The smooth line represents the estimated (non-linear) regression function, or the conditional expectation of having a particular segment, as predicted by inventory size.

### 4.4.1 Plosives and sibilants

We begin with voiceless plosives and sibilants. As seen in Figure 4.5, voiceless plosives are found in nearly all languages, and so the probability of a language having voiceless plosives is essentially 1, and independent of inventory size. Likewise, voiceless sibilants have a fairly high probability of being in an inventory, with a minimum probability of approximately 75%; from this minimum the probability rises dramatically, so that the probability of an inventory having a voiceless sibilant is over 0.9 for all but the smallest inventories.

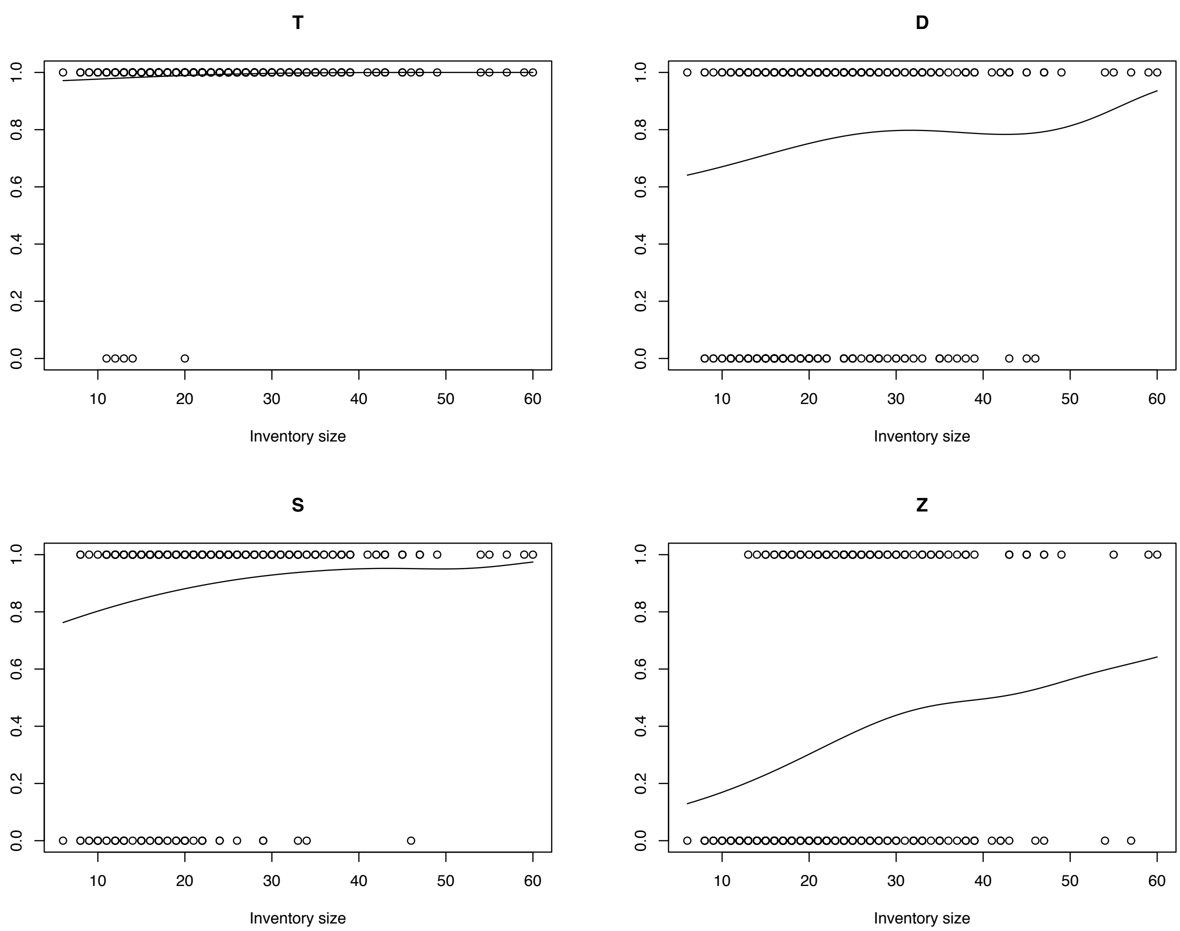



Figure 4.5: Regression estimates of plosives and sibilants by inventory size

We turn now to the voiced stop and sibilant classes. In general, voiced plosives and voiced sibilants are not highly marked segments, but there is nonetheless a difference between the voiceless and voiced classes. The probability of a language containing voiced plosives begins at approximately 0.6 for the smallest inventories, then steadily increases with inventory size until the probability nears 1 for very large inventories. The voiced sibilant shows the closest relationship to inventory size so far: the smallest inventory to have a voiced sibilant has 13 consonants, and in general the probability of having a voiced sibilant in very small inventories, is less than 0.2; as inventory size increases, the probability of having a voiced sibilant increases nearly linearly.

### 4.4.2 Spirants

Each panel of the following two graphs shows the probability of a language having a voiceless spirant (Figure 4.6) or voiced spirant (Figure 4.7) as predicted by inventory size. One of the most obvious features of these graphs is that the general shape of the probability curves is roughly the same across voiceless-voiced pairs, though the voiced spirants are generally rarer than their voiceless counterparts.

As expected, the voiced spirants are rarer than voiced sibilants, with none reaching even 40%. Of all the voiced spirants, the velar voiced spirant [ɣ] shows a linear increase with inventory size, though with a shallower slope, reflecting the fact that it is a relatively rare segment, and the smallest inventories in which [ɣ] appears have 14 consonants. The probability of an inventory containing either of the bilabial or dental spirants does not appear to depend strongly on inventory size; for the dentals [θ, ð] in particular, the slight rise for larger inventories could well be an artefact of there being so few inventories with dental spirants. Moreover, both the bilabial and dental voiced spirants appear in small inventories; in fact, the smallest inventory in the UPSID database, that of Rotokas (`roo`), with 6 consonants, contains both [β] and [ð]. Finally, the density curves for [f] and [v] stand out because of their unusual shape: [f] appears in some very small inventories, and [v] does not, but for both the density curve seems to increase linearly with inventory size at first, and then exhibits an odd shape for large inventories with more than 30 consonants.

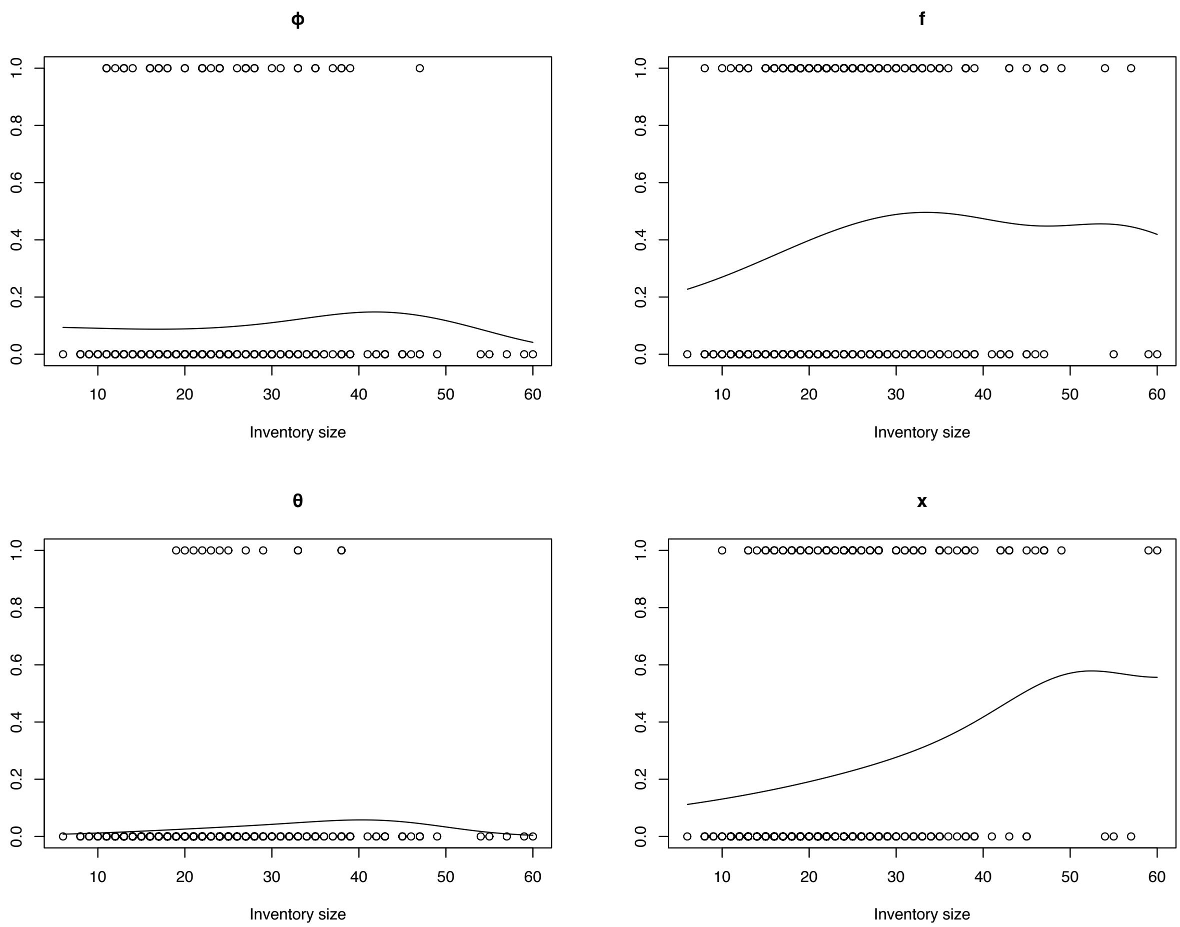



Figure 4.6: Regression estimates of voiceless spirants by inventory size

To summarize, these graphs show that the conditional expectation of an inventory of a given size containing a voiced spirant does not differs from both the stop and sibilant obstruent consonants; in particular, they do not distribute across inventory sizes like voiced sibilants do. As the next section will show, the relationship between voiced spirants and

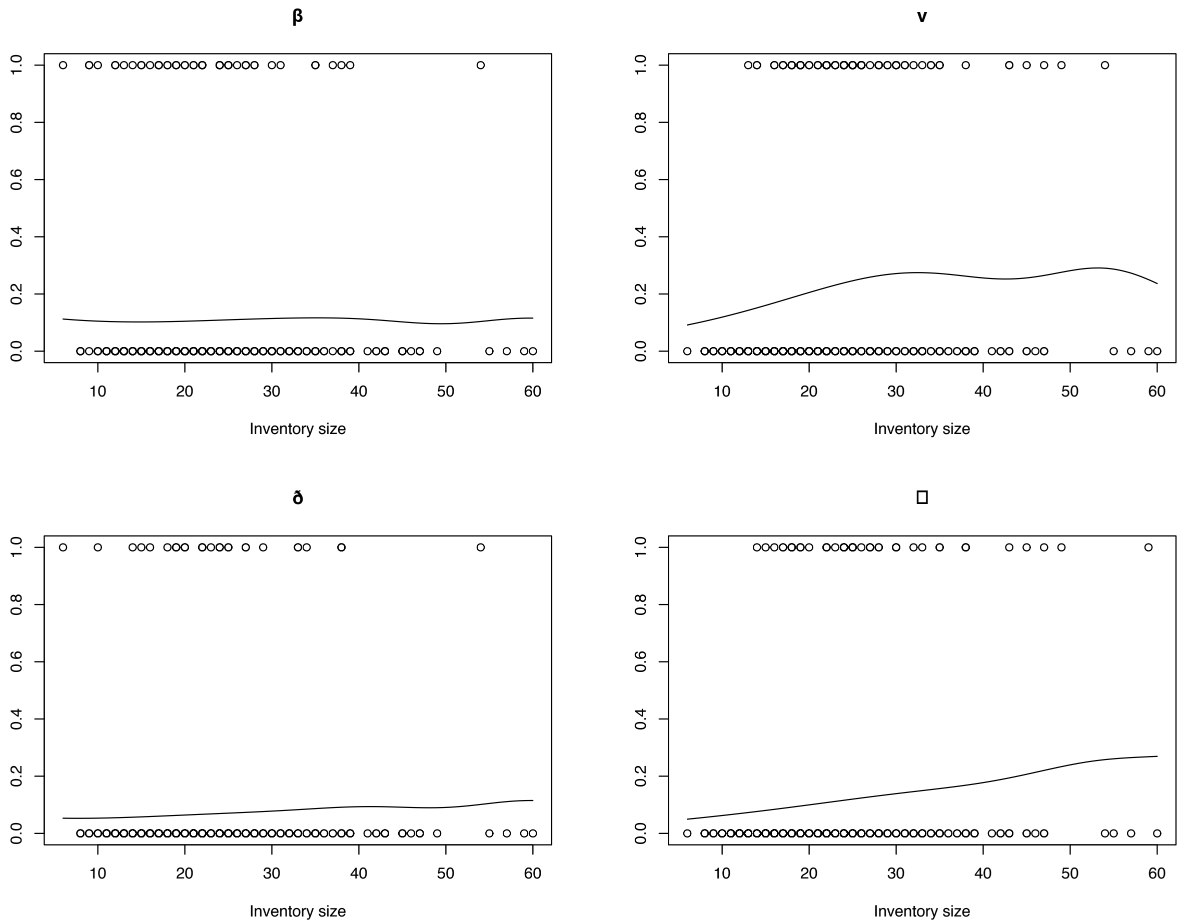



Figure 4.7: Regression estimates of voiced spirants by inventory size

inventory size is rather more subtle than has previously been assumed.

### 4.4.3 Paired vs. unpaired voiced spirants by inventory size

In the following panels we partition languages into those with the voiceless counterpart and those without, and plot the same regression function as above. It is important to note that this is more informative for some pairs than others; in particular, the dentals /θ, ð/ are so rare that the regression plot is not likely to be informative. Consider first the distribution of /β/ across inventory sizes. Figure 4.8 shows three panels plotting the distribution of /β/ across inventory sizes; on the left is the regression estimate for /β/ regardless of whether or not /ɸ/ is in the inventory, repeated from Figure 4.7; the center panel shows how /β/ is distributed across inventory sizes that have /ɸ/ (i.e., when it is *paired*), while the rightmost panel shows the distribution of /β/ when the inventory does not have /ɸ/ (i.e., when it is *unpaired*).

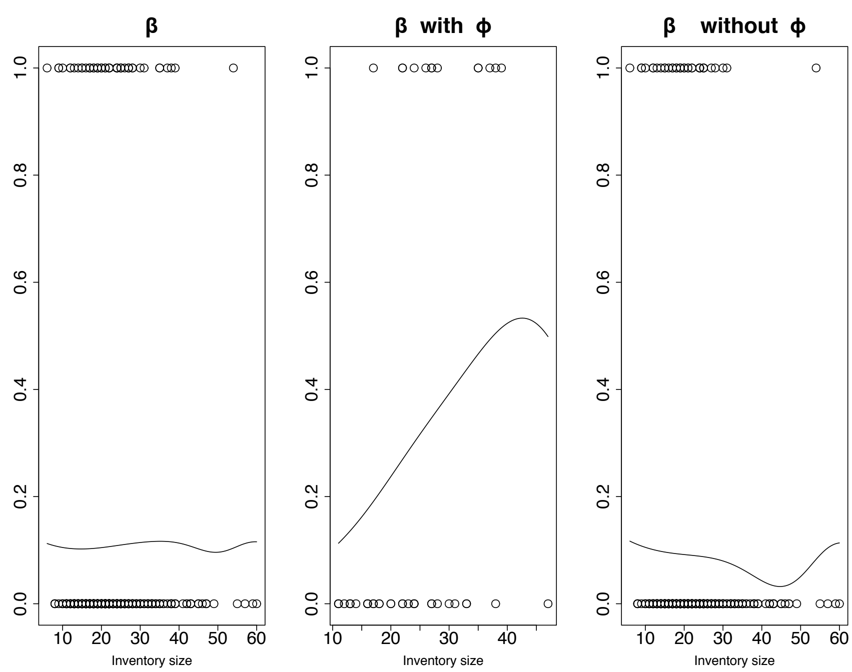



Figure 4.8: Regression estimates of /β/ with (centre) and without (right) /ɸ/ by inventory size

Partitioning the inventories into those with and without the voiceless bilabial spirant /ɸ/

drastically changes the picture. In Section 4.4.2, in which I considered only whether or not an inventory had /β/, it appeared that /β/ was simply rare and insensitive to inventory size, as shown by the relatively flat regression line in the left panel. However, as the two panels to the right show, this is not an accurate representation: /β/ is much more likely to be unpaired in small inventories than in large inventories, and is nearly unattested in large inventories if /ɸ/ is not also there. Turning to the labiodental spirant /v/ we see a similar effect when inventories are partitioned by whether they do or do not have the voiceless counterpart /f/, shown in the center and rightmost panel of Figure 4.9

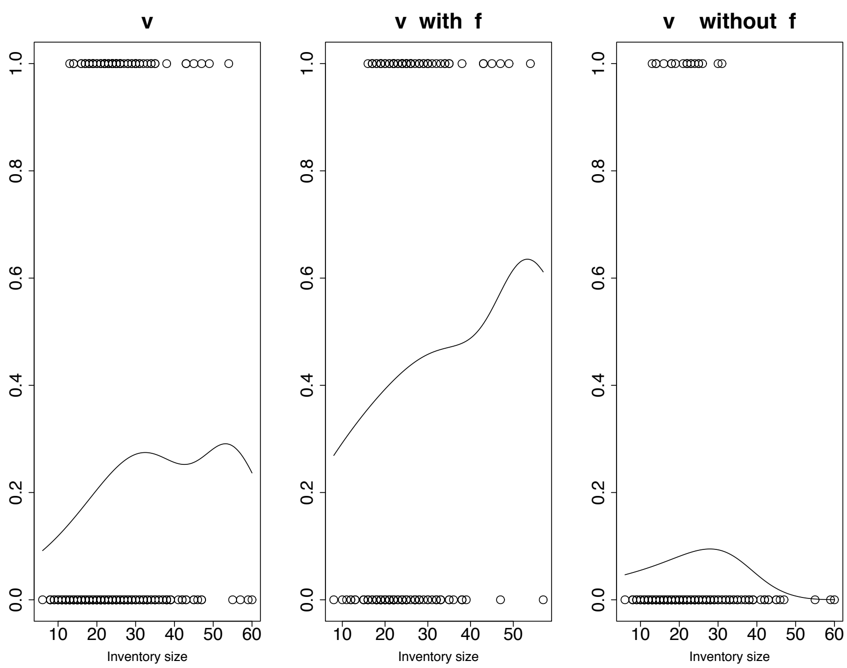



Figure 4.9: Regression estimates of /v/ with (centre) and without (right) /f/ by inventory size

If a language does not have [f], then its probability of having [v] is low (less than 0.2), and is virtually impossible for very large inventories. However, if an inventory does have [f], then the probability of having [v] increases in a roughly linear fashion with inventory size. In other words, inventories tend not to increase in size by adding [v] unless they already have [f], and in fact for sufficiently large inventories, having [f] appears to be a predictor for the presence [v].

The general pattern, namely that unpaired voiced spirants are more likely to be found in smaller inventories is repeated for velar spirants, though to a lesser degree; the dental spirants are included for completeness, but, as previously mentioned, the general rarity of the dental spirants mean that these graphs are perhaps best left uninterpreted. These graphs are shown in Figure 4.10 and Figure 4.11, respectively.

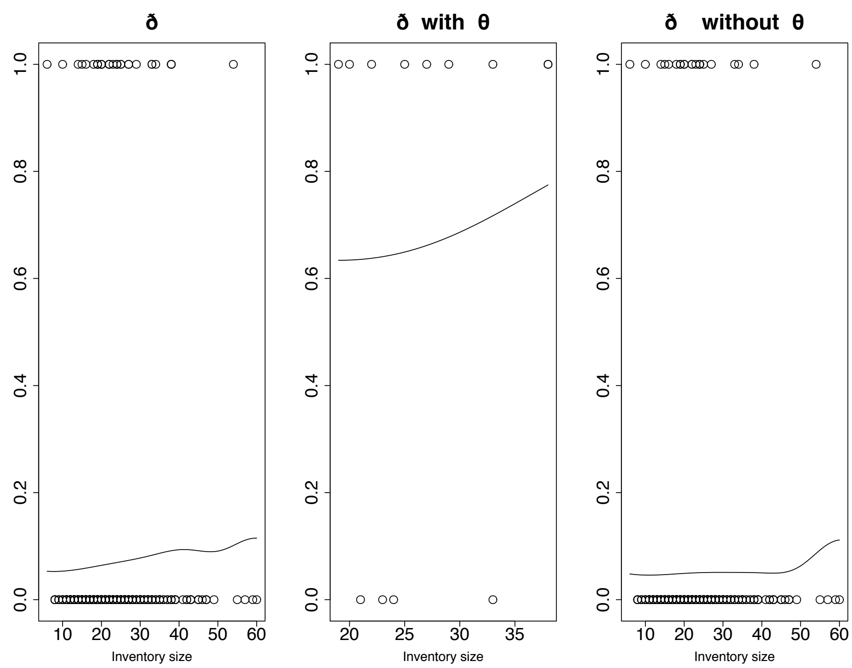



Figure 4.10: Regression estimates of /ð/ with (centre) and without (right) /θ/ by inventory size

The regression estimates above suggest that the relationship between voiced spirants and inventory size is more subtle than previously assumed: in small inventories, voiced spirants are reasonably likely to appear unpaired, but as inventory size increases, the probability of an unpaired voiced spirant decreases quite rapidly to nearly 0, while the probability of a paired voiced spirant increases. This is perhaps surprising: in a large inventory that is already crowded, one might expect that adding two segments is more costly than adding

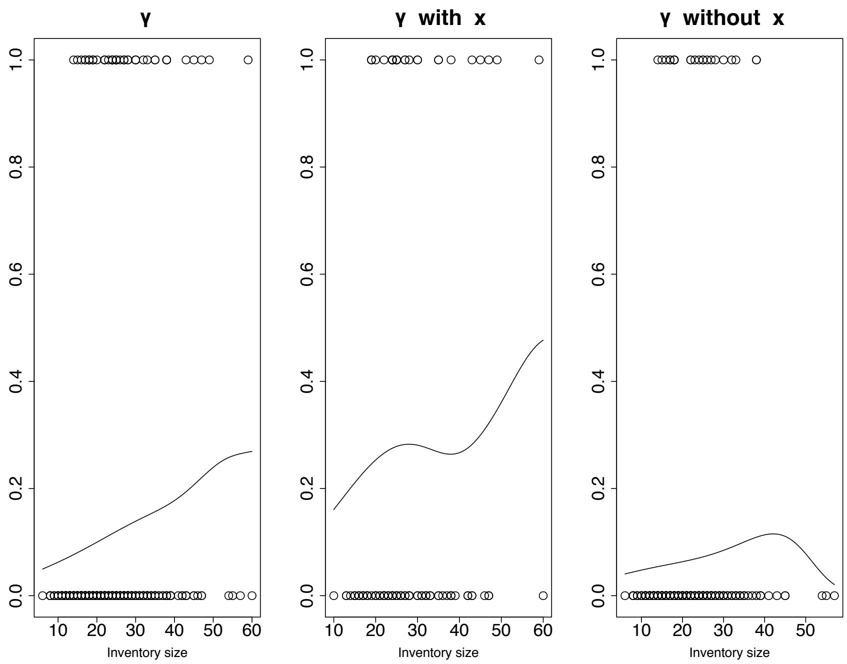



Figure 4.11: Regression estimates of /ɣ/ with (centre) and without (right) /x/ by inventory size

one, but this does not appear to be the case. Rather, it appears that in larger inventories the implicational relations of voicing hold, but that smaller inventories are less constrained by this relation.

## 4.5 The role of place

Thus far, I have shown that the voiced sibilants and spirants have different relationships to inventory size and the presence of voicing contrasts in the plosive system; the implications of these results are considered more fully in Section 4.6. The results of this chapter support the claim that, from a typological perspective, the introduction of the term *voiced spirants* is motivated by the fact that the voiced fricatives do not pattern as a unified class.

The distinction between sibilant and spirant frication type is grounded in the different aerodynamic tensions inherent to the maintenance of voicing and frication that each class instantiates. Nevertheless, place of articulation plays a role in the instantiation of the aerodynamic tensions between voicing and frication.[12] For explosive stop consonants, place of articulation matters because it affects the size of the oral cavity: the further forward the place of constriction, the larger the oral cavity, and thus it takes longer for supraglottal pressure to match subglottal pressure in [b] vs. [g] (Ohala, 1983). It is therefore easier to maintain a fully voiced [b] than [g].[13]

There is reason to believe that place of articulation has an effect on the patterning of individual voiced spirants. Recall that in Section 4.3.5, with respect to the cross-manner presence violations shown in Figure 4.4, rates of violations were low for /v/ and the sibilants, but high for /β, ð, ɣ/, possibly pointing to some typological differences. This may relate to the fact that at the labiodental place of articulation, a contrast between a fricative /v/ and approximant /ʋ/ is considered possible, hence the distinct symbols offered by the IPA; no language contrasts a "fricative" vs. an "approximant" at the bilabial, dental, and velar places of articulation.[14]

Furthermore, the interaction between aerodynamic tensions of voicing and spirant frication is further complicated by the role of place, similar to the role of place in the production of explosive stop consonants. The bilabial spirants /ɸ, β/ are unique in that there is no filter downstream the locus of constriction; this means that only channel turbulence is involved in

[12]The role of place of articulation has been incorporated into discussions of sonority, as in Steriade (1982); the discussion here focusses exclusively on the voiced spirants, and makes no attempt to incorporate them into the sonority hierarchy.

[13]It is well known that if an inventory has a gap in its voiced stop series, it is likely missing [g] Maddieson (1984). It is important to note that not all casees of gaps have phonetic explanations. For example, if an inventory has a gap in its voiceless stop series, it is likely missing [p]. As argued in Maddieson (2013b), so-called "missing-[p]" languages are geographically and geneologically concentrated, suggesting a historical and areal explanation. In contrast, "missing-[g]" languages are geographically and geneologically dispersed, suggesting an aerodynamic explanation.

[14]Though see the discussion in Section 3.5.1 about the manner contrast between /v/ and /ʋ/ in Dutch, and the more general discussion about this contrast in Section 5.4.2.3.

generating frication. In contrast, the labiodental spirants /f, v/ are unique among the spirants in that there is a baffle (the upper teeth) downstream from the locus of constriction.[15] The coronal /ð/ can vary in its articulation, produced either as an interdental or dental; in English, /ð/ (limited to function words) if often so short as to be realized as a stop or tap. Finally, the velar /ɣ/ is produced with the largest surface area, and conditions least amenable for the maintenance of voicing and frication, due to its posterior position which does not allow passive expansion of the vocal tract. Furthermore, because of the relatively large area of constriction, the airstream is more likely to be laminar, and hence /ɣ/ is likely to be realized as an approximant [ɰ].[16] Velars are also particularly prone to palatalization before high and/or front vowels, as in the realization of Greek /x, ɣ/ as palatal [ç, ʝ] before front vowels, and the voiced palatal is often realized as an approximant [j].

The difference between stops and spirants is introduced by what it takes to create the narrow channel for turbulent airflow, which differs based on the articulators involved, since each place of articulation introduces unique articulatory and aerodynamic challenges for the maintenance of voicing and frication. In some sense, then, voiced spirants are inherently more idiosyncratic than the voiced stops. Nevertheless, what the voiced spirants have in common is that, unlike sibilants, they are not produced with a grooved tongue that channels the air stream towards the teeth, and thus distinguishing between sibilant and spirant frication is well grounded. Because of the presence of a baffle, together with its anterior position and large cavity behind the locus of constriction, of the voiced spirants, /v/ is best suited to maintaining both voicing and frication.

The idiosyncracy of voiced spirants is reflected in the fact that they rarely occur in an

[15]As I discussed in Section 1.3, SPE used the feature [+strident] for labiodentals /f, v/ and sibilants /s, z, ʃ, ʒ/ because of the presence of a baffle downstream from the locus of constriction.

[16]An additional complication that the voiced velar /ɣ/ presents is that it is sometimes realized as further back in the mouth, as a voiced uvular fricative /ʁ/, which itself may be somewhat trilled, introducing other aerodynamic considerations.

inventory together, a fact that also extends to the voiceless spirants.[17] Even collapsing the bilabial and labial places of articulation, few languages exploit all three major places of articulation.[18] Figure 4.12 shows the number of languages in UPSID that have one, two or three spirants.[19]

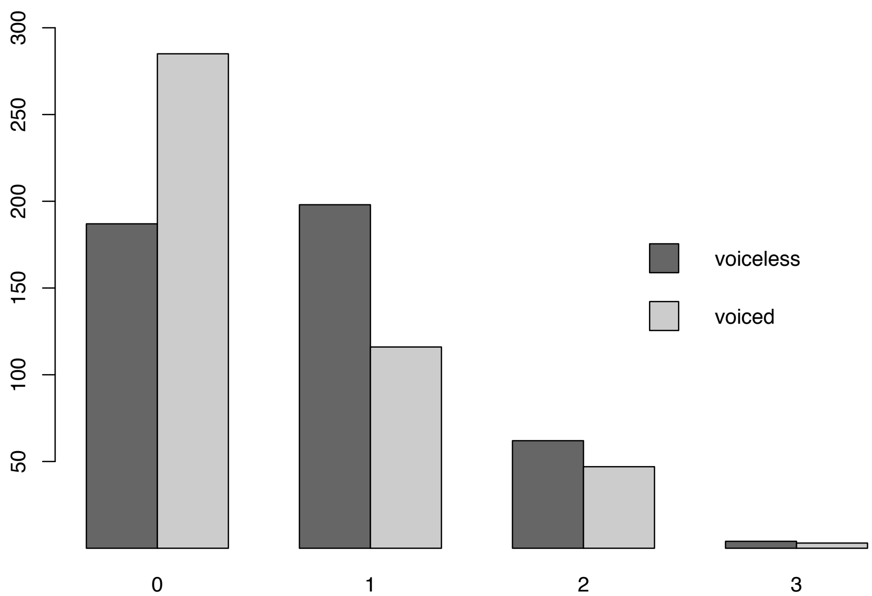


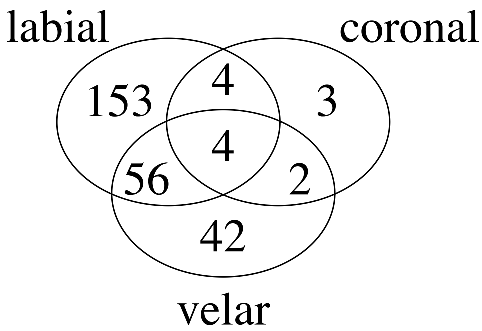

Figure 4.12: Breakdown of languages by number of voiceless and voiced spirants; bilabial and labiodental collapsed into single labial category, and counted once only.

The main difference between voiceless and voiced spirants is whether a language has at least one spirant. For voiceless spirants, slightly more languages have at least one voiceless spirant rather than none at all, while for voiced spirants, the number of languages without a voiced spirant is double the number of languages with one voiced spirant. For more than one spirant, the number of languages that exploit two or three major places of articulation for either voiceless or voiced spirants drops markedly.

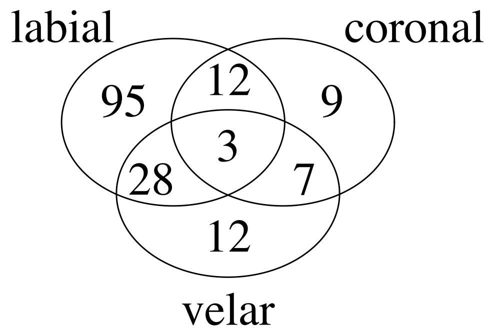
[17]This fact calls into question the notion of *feature economy* (Clements, 2003).

[18]This discussion ignores those languages which are transcribed with /ʋ/ rather than than /β/ or /v/.

[19]Note that no language has four voiced spirants, and only Iaai `(iai)` has four voiceless spirants[20].




(a) Voiceless spirants




(b) Voiced spirants

Figure 4.13: Venn diagrams showing cooccurrence of major place of articulation within voiceless and voiced spirant classes

Restricting attention to just those languages with at least one spirant, the Venn diagrams in Figure 4.13 show how languages are distributed by place of articulation with respect to the voiceless and voiced spirants. Replicating the observation of Maddieson (1984), the most common place of articulation for spirants is labial, and the majority of languages with only one spirant exploit the labial place of articulation; for both voiceless and voiced segments the second most frequent place exploited is velar, followed by coronal. Along the same lines, if a language exploits two of the three major places of articulation, it is most likely to exploit the labial and velar places, though this fact is more robust for voiceless spirants than for voiced.

Only four languages—Bashkir (`bak`), Greek (`ell`), Iaai (`iai`), and Spanish (`spa`))—have a full series of voiceless spirants, and only three languages—Greek (`ell`), Mari (`mhr`) and Spanish (`spa`)—have a full series of voiced spirants, listed in Table 4.19.[21] In Greek

[21]UPSID lists two additional languages as having three voiced spirants at all major places: Kabardian (`kbd`) with /v, ð, ɣ/ and Chiquihuitlan Mazatec (`maq`) with /β, ð, ɣ/. However, none of my sources identified Kabardian as having dental spirants (Kuipers, 1960; Colarusso, 1992; Gordon and Applebaum, 2006; Matasović, 2007), though it does have a rich fricative inventory; for Mazatec, none of the sources referenced by UPSID include /ð, ɣ/ in the description (Pike and Pike, 1947; Jamieson, 1977a,b). As a result, these entries were updated in my database, and the changes are reflected here.

`(ell)`, both voiceless and voiced spirants derive historically from stops. In both Mari and Spanish voiced stops /b, d, ɡ/ alternate with voiced spirants /β, ð, ɣ/, and whether the voiced stops or voiced spirants are taken as basic is a matter of disagreement across sources.[22] The inclusion of Bashkir `(bak)` in this count could be challenged on multiple grounds: /f/ only occurs in loans, /v/ is only articulated as [v] in unadapted Russian loans (elsewhere it is [w]), and the velar /x/ is described by some as uvular (note that in this case, /x/ would be paired with the post-velar voiced fricative, transcribed in some sources as [ɣ], but in others as [ʁ]) (Poppe, 1964).

Table 4.19: Languages with spirants at labial, coronal and velar places of articulation in UPSID

| ISO | Language | Family | Area | Inv. Size | ɸ | β | f | v | θ | ð | x | ɣ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| bak | Bashkir | Turkic | Europe | 25 | | | ✓ | ✓ | ✓ | ✓ | ✓ | |
| ell | Greek | Indo-European | Asia | 19 | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| iai | Iaai | Austronesian | Pacific | 38 | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | |
| mhr | Mari, Meadow (Cheremis) | Uralic | Europe | 25 | | ✓ | ✓ | | | ✓ | ✓ | ✓ |
| spa | Spanish (Eur.) | Indo-European | Asia | 20 | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ |

This brief discussion shows that most languages have only one voiced spirant, and that when a language exploits all three places of articulation, then this is likely a result of lenition and/or alternation with the voiced stop series, as in Greek and Spanish. In fact, even in languages with both labial and velar voiced spirants can be the product of lenition, since as Gurevich (2004) notes, the alternation between voiced stops and spirants is often accompanied by flapping of the voiced coronal stop /d/. Further study is needed to more fully understand how the aerodynamic tensions inherent to voicing, spirant frication, and place of articulation interact in order to yield both the idiosyncratic and class behaviour of the voiced spirants.

[22]Thanks to Johannes Dellert for clarifying the Mari data.

## 4.6 Discussion

This chapter investigated the typological character of the voiced spirants with respect to inventory structure. The basic question underlying these investigations was to determine whether the voiced sibilants and spirants pattern as a homogenous class of voiced fricatives. Two specific claims from the literature were investigated: (1) there exists an implicational relationship between voiced fricatives and stops, whereby if an inventory has a voicing contrast in the fricatives, then it also has a voicing contrast in the stops; (2) voiced fricatives are *elaborated* articulations and hence are predicted to appear only in inventories of a particular size.

A crucial component of the argument was the methodological decision to focus the investigations on violations, rather than on positive correlations. The questions are thus better reframed as attempting to determine the properties of inventories that appear to allow voiced spirants to appear (1) when there is no voicing contrast in the fricatives, and (2) when the inventories are small. By distinguishing between the sibilants and spirants, I showed that neither statement holds true of the spirants, but does hold true of sibilants. The general conclusion to be drawn from this chapter is that voiced sibilants and voiced spirants are not a homogenous class, typologically speaking.

It is widely accepted that voiced fricatives are relatively complex articulations, requiring the simultaneous maintenance of turbulent airflow and voicing which are in aerodynamic tension. And, in general, there does appear to be a pattern that the most complex articulations tend to be found in only the largest inventories. When voiced sibilants and spirants are grouped together as a single category, there does appear to be a fairly linear relationship between inventory size and the presence of voiced fricatives. When voiced sibilants are distinguished from the voiced spirants, we see that the previous linear relationship is driven almost entirely by the voiced sibilants. Specifically, voiced sibilants rarely occur in very small inventories, but voiced spirants do.

Further investigation into the relationship between inventory size and the presence of voiced spirants revealed that whether or not the spirant is paired or not is important. Unpaired voiced spirants, that is, voiced spirants that occur in an inventory without their voiceless counterpart, tend to occur in very small inventories; the likelihood of an inventory having a paired voiced spirant increases relatively linearly as inventory size increases, like the sibilants. Further, unpaired voiced spirants are more likely to appear in inventories that do not have a voicing contrast in the stops, and in particular have only a single stop series. The two factors, inventory size and the presence of more than one stop series, are certainly interrelated.[23] Furthermore, there is a clear areal confound: among the 11 languages that have both small consonant inventories and voiced stops ($n < 12$), only two areas are represented, Pacific and Americas.[24] Thus, while there are some small languages with a voicing contrast in the stops, in most cases, a language with a stop voicing contrast will typically not have a particularly small consonant inventory.

The findings on the relationship between inventory size and the presence of voiced spirants are thus related to the findings on the relationship between voiced spirants and the presence of voiced stops in an inventory. These issues are related not just empirically, but conceptually as well. The claim that there is an implicational relationship between an inventory having a

[23]The reason for this is that stops tend to surface as series with at least three members (labial, coronal, and dorsal), and so the introduction of a laryngeal contrast in the stops typically adds at least three consonants to the inventory. Counterexamples to this generalization exist, but these languages are remarkable in that they tend to lack either nasals or fricatives, both highly marked inventory gaps.

[24]Three of the four Pacific languages are spoken in Papua New Guinea (Rotokas (`roo`), Naasioi (`nas`), Gadsup (`gaj`)); the fourth, Ekari (`ekg`), is spoken in Indonesia. Of the remaining languages spoken in the Americas, three are spoken in Columbia (Cubeo (`cub`), Waimaha (`bao`), both Tucanoan, and Andoque (`ano`), a language isolate); two are spoken in Brazil (Pirahã (`myp`), Maxakali (`mbl`)); one in Peru (Yagua (`yad`)), and one in the United States (Cherokee (`chr`)).

voicing contrast in the fricatives and a voicing contrast in the stops amounts to an idea that more complex articulations are, in some sense, "licensed" by simpler articulations. In other words, the underlying assumptions about the relationship between stop and fricative voicing assume that inventories are structured such that simpler articulations are used first, and that more complex articulations make use of simpler articulations in "building up" towards more complex articulations. Acquisition data appears to support this view, as basic articulations tend to be acquired before more complex articulations.

Framing the relationship between voiced stops and fricatives in terms of licensing reflects the more nuanced approach taken in this chapter. If voiced fricatives are licensed by voiced stops, then what, specifically, is doing the licensing? Is it the fact that producing voiced fricatives is easier if the speaker already produces voiced stops? This assumption translates to the *cross-manner presence* implicational relation: "the presence of one or more voiced fricatives implies that the inventory has one or more voiced stops". Examining violation rates showed that for /v, z/, the implication generally holds. That is, the violation rates for /v, z/ were very low, meaning that /v, z/ tend to appear in inventories that already have voiced stops. The remaining voiced spirants exhibited much higher violation rates, meaning that /β, ð, ɣ/ do tend to appear in inventories where there are no voiced stops. Part of this result has to do with inventory size: we know from the investigation of inventory size that voiced stops do not appear in very small inventories, but that (unpaired) voiced spirants are likely to appear in these inventories.

Maddieson (2010) does not frame the implicational relation between fricatives and stops in terms of the presence of voiced members. Rather, Maddieson claims that for a language to have a voicing *contrast* in the fricatives, it must have one in the stops. By and large, this claim is upheld for sibilants and the voiced spirants /β, v, ð/. If a language has a voicing contrast in the sibilants, it has one in the stops, and likewise, a language that has one or more of the pairs /ɸ, β/, /v, f/, or /θ, ð/ tends to surface in inventories that have both voiceless and voiced stops. The velars /x, ɣ/ have a higher violation rate. Of the violations incurred by sibilants and /β, v, ð/, they fall into two groups: for approximately one-third, the violation is incurred because the inventory only has a single stop series; for the remaining two-thirds, the violation is incurred because there is a laryngeal contrast in the stops, but the contrast is not realized as a voicing contrast. Of these two kinds of violations, the first appears to be a more serious violation. What is the phonological patterning of the voiced fricatives in these languages? In an inventory with a labiodental voicing contrast in the fricatives but with no laryngeal contrast in the stops, do /f, v/ pattern as a phonological pair, and what does the evidence for this look like? Answering these questions in depth extends beyond the scope of the present chapter, but identifies an area of future research that pursues questions that investigate the relationship between typology and phonological patterning.

The second kind of violation appears to be less of a violation, and related to the fact that laryngealizations such as aspiration or glottalization are much more difficult to produce on fricatives than on stops. These violations are thus best seen as a reflex of the way that laryngeal states interacts with manner. Nonetheless, phonological questions remain about the status of the voiced fricative pair in these languages. In some languages, all lenis segments pattern together, as in German, an aspirating language, where laryngeal contrasts in both stops and fricatives are neutralized in coda position. In other languages, like Turkish (also an aspirating language) stops undergo word final laryngeal neutralization, but voiced fricatives /v, z/ do not devoice. Further research into these systems is required in order to understand whether the presence of a laryngeal contrast in both stops and fricatives is explained by phonetic factors, or whether there is a deeper phonological connection to be made.

Regardless, it is important to keep in mind that inventory size and the presence of a voicing contrast in the stops is highly correlated. Because of this, in a average to large consonant inventory, it is difficult to discern whether the presence of a /v, f/ in an inventory is licensed by inventory size or by a voicing contrast in the stops. Disentangling whether asymmetries between voiced spirants and sibilants are due to inventory size or voicing contrasts in stops is beyond the scope of the present work. However, making an issue as to whether or not these factors are independent may be a red herring, as the overall picture still stands. Some inventories exhibit what may be considered the "standard conception" of laryngeal contrasts in stops and fricatives: such inventories have a voicing contrast in both stops and fricatives, and they tend not to be unusually small. Other inventories exhibit a more marked overall configuration: they tend to be small, voiced spirants appear without their voiceless counterparts, and they tend not to have voiced stops. What to make of these two kinds of inventories?

Consider first what I will call, for ease of exposition, the *unmarked inventories*, that seem to support the hypothesis that voiced fricatives tend to appear only if the inventory also has voiceless fricatives and voiced stops. Such inventories fuel tacit assumptions about symmetry in laryngeal contrasts: sonorants are voiced and tend to be unpaired (with a voiceless counterpart), but voicing in obstruents creates "pairs" of segments that differ only in voicing. Tempting though this view might be, I show in Chapter 5 that voiced spirants rarely live up to this promise in terms of their phonological patterning. Time and again, /v/ exhibits ambiguous behaviour with respect to obstruents and sonorants, regardless of the kind of laryngeal contrast instantiated in the stops.

Turning to the relatively marked inventories, we see that voiced spirants, but not voiced sibilants, are likely to show up unpaired, in small inventories. Given the relative markedness of voiced spirants compared to their voiceless counterparts, why is a small inventory more likely to have, say, /β/ rather than /ɸ/? As mentioned above, one possibility is that in such inventories the voiced spirants are in fact approximants, and so comparing them to voiceless spirants is irrelevant. Based on the descriptions available, this seems only partially correct: the voiced spirants in these languages are often described as varying between fricative and approximant realizations, sometimes in apparent free variation, and other times in accordance with a rule (say, determined by position within a word). It might therefore be better to characterize the situation not as one of markedness, but rather as getting the most perceptual salience for a given articulatory configuration, given the overall inventory. Specifically, while both voiced and voiceless fricatives require the same configuration of the oral articulators, in uncrowded inventories the voiced fricatives may in fact be the more salient option, so long as they are not required to maintain voicing and frication in all environments in order to maintain contrast. Voiceless spirants, as opposed to voiceless sibilants, are generally quiet, and exhibit a relatively high degree of confusability; furthermore, they are relatively stable in their acoustic realization. While it may be difficult to produce both voicing and frication on any given token, if all that is required of voiced spirants is to produce a sound at a particular place of articulation that is voiced, continuant, and non-nasal, then either [β] or [β̞] will fulfill that function, as long as there are no approximants at the same place of articulation against which the voiced spirant must contrast. Note that the possibility of varying between a fricative and approximant realization is not available for /z/: the articulatory configuration required for sibilants—namely the grooved tongue that channels the air into a jet stream—does not appear to admit an approximant realization; insufficient airflow will simply generate a voiceless fricative. The hypothesis is therefore that determining the markedness of a segment is not just a function of its articulatory configuration, aerodynamic challenges, or acoustic realization, but also a function of what else is in the surface inventory.

The contribution of this chapter has been to show that, typologically, voiced spirants and sibilants do not form a unified class. The assumption that they do has led to typological generalizations about voiced fricatives that are in fact only true for the voiced sibilants. This chapter has shown that the voiced spirants do not have the same typological identity as voiced sibilants, not just in their rates of violation, known from Maddieson (1984), but also in their relationship to voiced stops and inventory size. The results of this chapter contribute evidence to the thesis that voiced spirants are not simply a subclass of voiced fricatives.

APPENDIX

## 4.A Modelling violation rates

### 4.A.1 A simplified demonstration

A convenient model for violation rates is a Binomial model with a beta prior distribution. Conceptually, a Bayesian framework models the probability of attaining a proportion of violations given the data (together with prior beliefs about the true rate of violation). I begin with a simplified demonstration of the model parameters.

Consider Figure 4.A.1, which models the violation rates of plosives (voiced plosives with no voiceless plosives), bilabials (/β/ without /ɸ/) and labiodentals (/v/ without /f/), which can be read off of the middle column of Table 4.5, and repeated beside Figure 4.A.1 for convenience. The $x$-axis denotes the violation rate: note that the peak (mode) of each curve is centered at the violation rate. The area under each curve sums to 1, and the width is inversely proportional to the amount of data for that violation rate; thus the curve for the plosives is the narrowest because there are 324 languages with voiced plosives in UPSID, while the width of the curve for the bilabials is widest because there are only 51 languages that have /β/. Below the curves are 95% credible intervals for the corresponding beta distributions; I return to credible intervals in Section 4.A.2.

Each curve in Figure 4.A.1 is a posterior beta distribution, which is determined not only by the data, but also by the *prior*. Intuitively, these curves can be interpreted as probability distributions that correspond to our belief about the true rate of violations, given our prior biases (specified explicitly) and the data (here from UPSID). In this work the prior can be

| | Exceptions | Violation rate |
|---|---|---|
| plosives | 5/324 | 1.5% |
| /ɸ, β/ | 38/51 | 74.5% |
| /f, v/ | 19/93 | 20.4% |

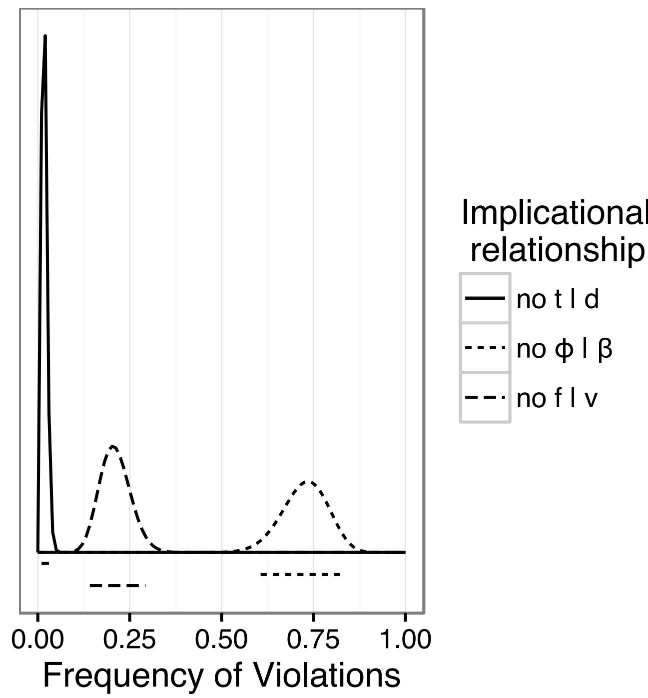



Figure 4.A.1: Beta distributions of violation rates

interpreted as the violation rate already observed in previous experience with languages. In Figure 4.A.1 I selected a prior of 1:1, meaning that I encoded my previous experience as two languages with the voiced member in question (i.e., a voiced stop series, /β/, or /v/); of these two languages, one lacked the voiceless counterpart (i.e., a voiceless stop series, /ɸ/, or /f/). In order to model violation rates in general I adopt a prior of 1:9.

**The effect of different priors** Figure 4.A.2 illustrates the effect of three different priors on the beta distributions on violation rates for the plosives, bilabials, and labiodentals. For each panel, a prior equivalent to having seen 10 languages was used. The leftmost panel uses the uniform prior, like in Figure 4.A.1, but with prior violation rate of 5/10, rather than 1/2; it is therefore a stronger prior, though still rather weak overall. The centre panel uses what I call the *obstruent bias prior*: it assumes that of the 10 previously encountered languages, 1 incurred a violation. The rightmost panel uses what I call *sonorant bias prior*: it assumes that of the 10 previously encountered languages, 9 incurred a violation. As can be seen from the relative position of the beta distributions, the prior does have an effect on the model: relative to the uniform prior, the obstruent bias prior shifts the curves to the left, while the sonorant bias prior shifts the curves to the right. Figure 4.A.2 also shows that the effect of the prior is greater when there is less data; the curve representing the violation rate for bilabial spirants shifts more than for plosives.

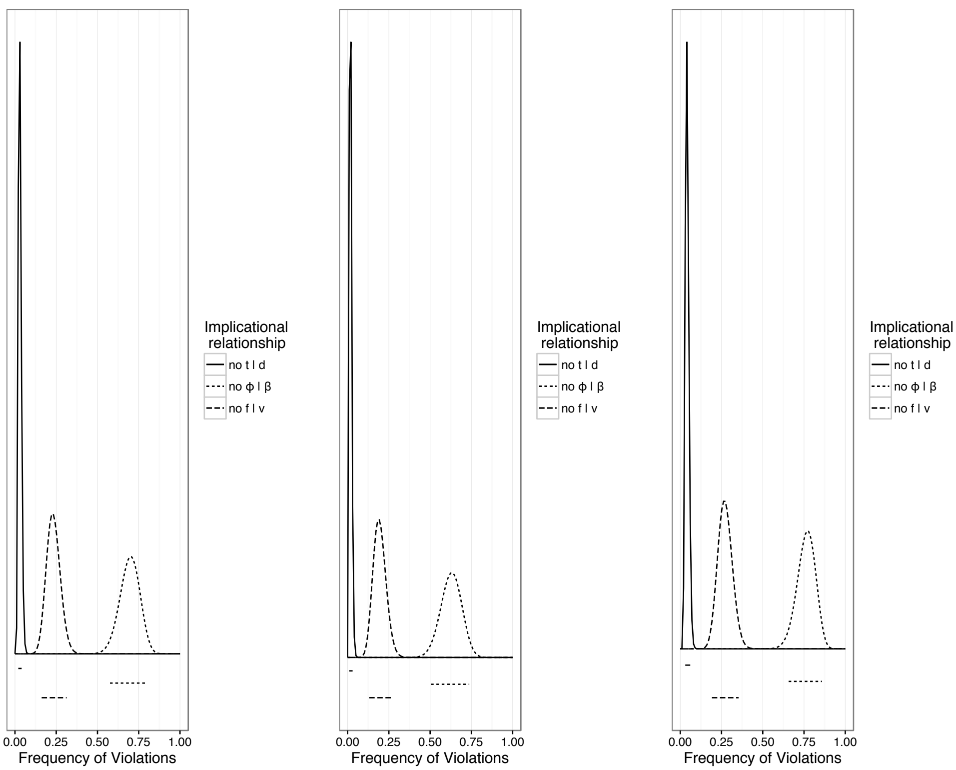



Figure 4.A.2: Effect of different priors in modelling violation rates
Left: uniform prior (5:5); Centre: obstruent bias (1:9); Right: sonorant bias (9:1)

**The strength of the prior** As a final illustration, Figure 4.A.3 shows what happens when the strength of the prior is manipulated. In all three panels a uniform prior is used, meaning that it is assumed that half of previously encountered languages instantiate a violation.

What differs is the strength of the prior, which corresponds to the number of languages previously encountered, with the strength of the prior increasing to the right. The effect of the strength of the prior is stark: when only 10 languages have been previously encountered, the effect of the data on the model is relatively strong. In other words, if you've only seen 10 languages with /β/ (as in the leftmost panel), and half of them lack /ɸ/, then seeing 51 languages with a violation rate of 74.5% means that the data will have more of an effect on your belief about the true violation rate. However, if you've seen 1000 languages with /β/ (as in the rightmost panel), and 500 languages lacked /ɸ/, then a dataset of 51 languages is not likely to change your mind very much about the true violation rate. And indeed, all of the curves in the rightmost panel are clustered around the 50% violation mark, because of the overly strong prior.

In the following work I have elected to model the data with a weak obstruent bias of 1/10 (10% violation rate). This encodes the traditional bias that spirant pairs are obstruents, but does so weakly so as not to overpower the data, since the number of languages containing voiced spirants is small, ranging from 31 languages with /ð/ to 93 languages with /v/.

### 4.A.2 Binomial model with a beta distribution: model details

Proportional data such as that considered here can be considered the result of successive *Bernoulli trials*, where a Bernoulli trial is a random experiment consisting of two possible outcomes, "success" or "failure". If we consider a "success" to be a violation of the implicational relations of voicing, we consider each language with a voiced spirant as an experiment, and whether or not it lacks a voiceless member is coded to be a success or failure, respectively. The beta distribution is a convenient distribution for modelling proportional data since it is

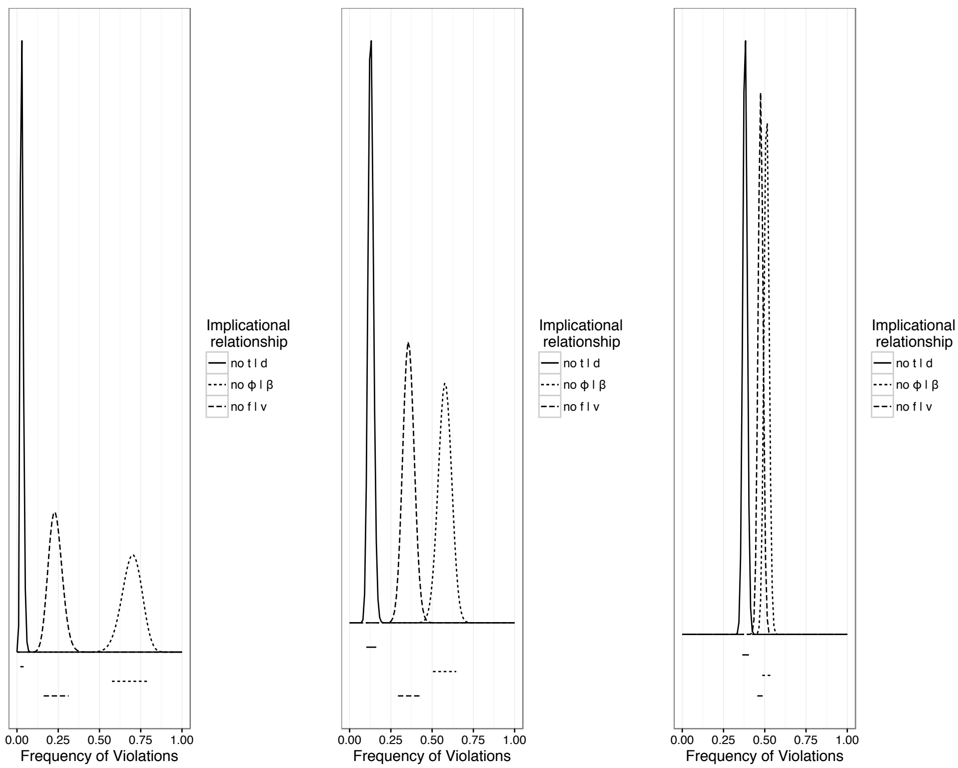



Figure 4.A.3: Effect of the strength of the prior in modelling violation rates; uniform prior Left: 10 languages (5:5); Centre: 100 languages (50:50); Right: 1000 languages (500:500)

the *conjugate prior* of the binomial.[25] The model uses a Beta($\alpha$, $\beta$) prior distribution, where $\alpha$ and $\beta$ determine prior beliefs about the rate of violations, based on previous experience. For example, Beta(1, 1) represents a uniform distribution prior; in this work I use the prior distribution Beta(1, 9), a relatively weak prior with a violation rate at 10%.

[25]For a more in-depth explanation, see Vasishth (????). Briefly: a Bayesian model assigns a *prior* distribution to the probability of attaining a particular proportion, and updates the distribution based on the data; this is called the *posterior* distribution. In general, the prior and posterior distributions need not be related, but if a beta-distribution is used for the binomial, then both the prior and the posterior are beta distributions; thus a conjugate prior refers to the fact that both the prior and posterior are part of the same family of distributions.

When the prior is a Beta($\alpha$, $\beta$) and the data follows a binomial distribution, the posterior distribution follows a Beta distribution with updated hyperparameters: if $n_v$ and $n_n$ are the number of observed violations and non-violations, respectively, then the posterior is a Beta($\alpha + n_v$, $\beta + n_n$). To see how these parameters relate to the more familiar representation of the data, we consider the contingency table for /x, ɣ/ given in Table 4.8. Of the 451 total languages, 23 of these exhibit a violation (i.e., have /ɣ/ without /x/), 27 exhibit a non-violation (i.e., have both /ɣ/ and /x/). Note that the first column, namely languages without /ɣ/, but with or without /x/ is irrelevant. Thus for the pair /x, ɣ/ with prior Beta(1, 9), the shape parameters defining the posterior Beta distribution would be $\alpha = 1 + 23$, $\beta = 9 + 27$.

**Full example: within-manner violation rates** Turning our attention back to within-manner violation rates, consider Figure 4.A.4 which shows the Beta(1, 9) distributions of the proportion of violations given in Table 4.5, together with 95% credible intervals as bars below the curves.

Figure 4.A.4 shows a clear distinction between the plosives and sibilants on the one hand, and the spirants on the other, much as was gleaned from the bare rates of violations in Table 4.5: the curves representing the beta distribution of the proportion of violations for the plosives and sibilants have a mode near 0, while the modes for the remaining curves are much higher. This replicates the results of Table 4.5. However, Figure 4.A.4 provides more information than just the point estimates: it shows that for the plosives and sibilants our estimate of the proportion of violations is very robust, evidenced by the narrowness of the curve, while for the spirant pairs it is considerably less so. This is mostly a result of the relative quantity of data, as the spirants are considerably rarer than the plosives or sibilants.

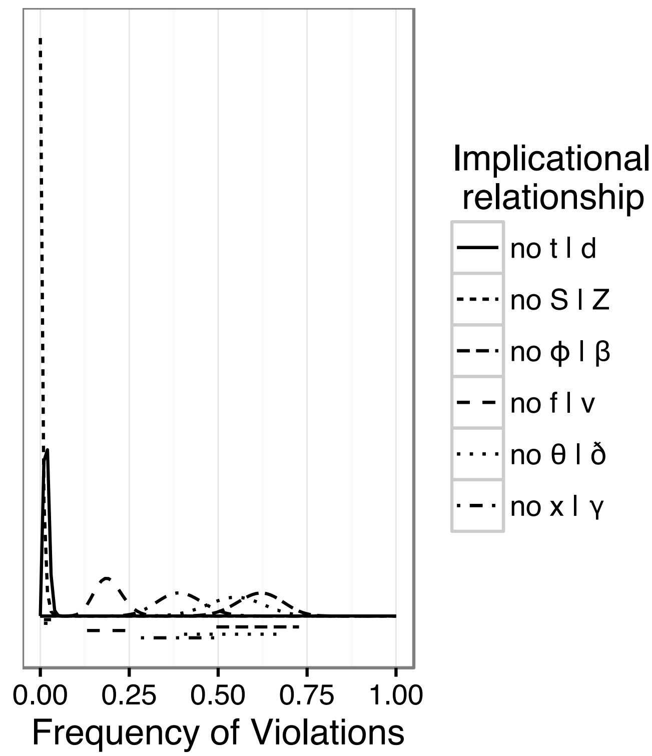



Figure 4.A.4: Proportion of violations of plosives, sibilant and spirant pairs

#### 4.A.2.1 Credible intervals

For ease of exposition, rather than show the beta distributions associated with violation rates, I show only the credible intervals corresponding to the beta distributions. Credible intervals are analogous to confidence intervals but differ in interpretation. In frequentist statistics, a 95% confidence interval means that, if 100 confidence intervals were constructed from a given dataset using the same method, 95 would contain the true parameter; in contrast, the Bayesian credible interval means that the probability that the interval contains the true value is 95%. This interpretation makes sense because in Bayesian statistics the true value is assumed to be a random variable, and thus the credible interval can be interpreted as a probabilistic statement of the parameter itself. Figure 4.A.5 shows the 95% credible intervals of Figure 4.A.4 with thick lines, extended to 97.5% credible intervals with the thin lines; henceforth I will refer to these as 95(/97.5)% credible intervals, where the larger number in brackets refers to the larger credible interval in thin lines.

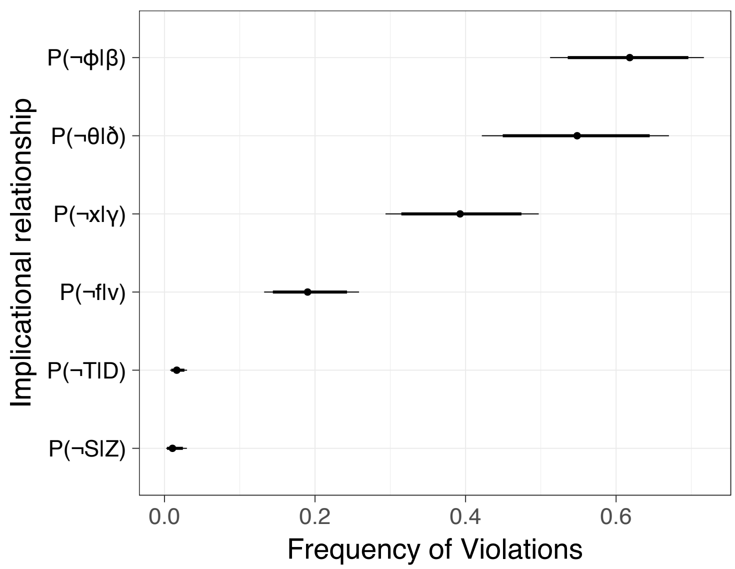



Figure 4.A.5: Proportion of violations of spirant pairs

Like the beta distributions of Figure 4.A.4, position of the mode (marked by a solid circle) is a function of both the prior (set at a 1/10 violation rate) and the rate of violation in the data. The width of the credible interval is determined in part by the amount of data that contributes to the posterior distribution, thus the widest credible interval is for the dental pair /θ, ð/, which has the least amount of data and therefore the most uncertainty, and the credible intervals for the plosives and sibilants are reduced to a single point, since they have the most data and therefore the least uncertainty. The other factor that determines the width of the credible interval is the probability specification, or how conservative one wants to be about the posterior. Figure 4.A.6 shows how the width of credible intervals change based on the probability specification. The leftmost panel uses 95(/97.5)% credible intervals: it is the most conservative, and as a result it has the most difficulty distinguishing between the violation rates of spirant pairs. The rightmost panel uses a 50% credible interval; the violation rates are obviously clearly distinguished, but such intervals are useless because they are no better than chance. The centre panel uses a 90(/95)% credible interval, and is the one used in the remainder of this work, unless otherwise specified.

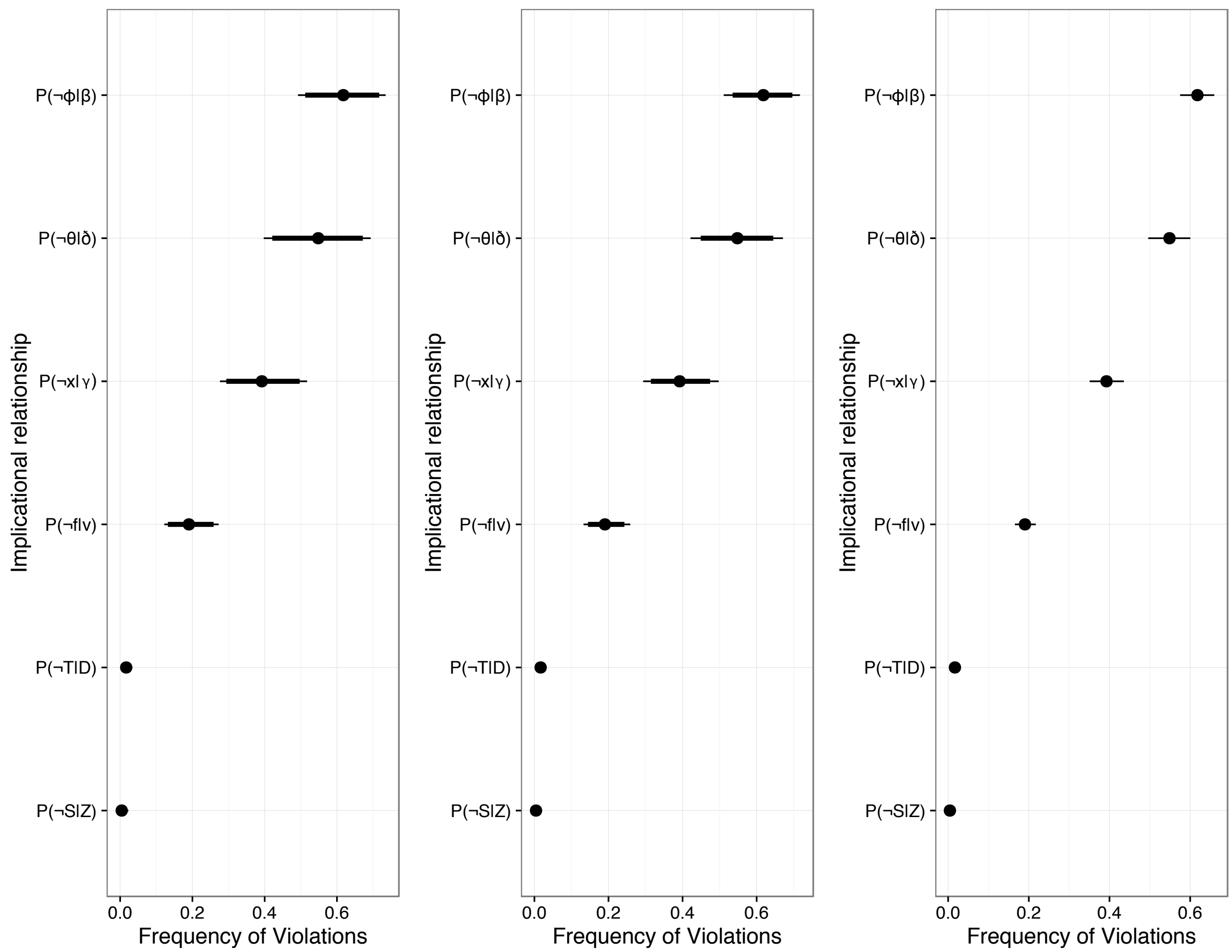



Figure 4.A.6: Effect of probability specification for credible intervals
Left: 95(/97.5)%; Centre: 90(/95)%; Right: 50(/75)%

CHAPTER 5

**FEATURAL SPECIFICATIONS OF** /ʋ/

## 5.1 Introduction

In this chapter I propose that /v/ is both an obstruent and a sonorant, and in many cases is specified as such featurally. That is, in addition to the feature [sonorant], I propose that there exists a feature [obstruent], and that both features are required in order to capture the cross-linguistic patterning of voiced spirants. The feature [obstruent] is not new: Clements and Osu (2002) proposed that implosive stops in Ikwere are specified as [−sonorant] *and* [−obstruent], based on their patterning under nasal harmony. I extend their analysis by arguing that /v/ is generally specified as [+sonorant] and [+obstruent], a feature specification which straightforwardly accounts for Russian /v/'s patterning with respect to voicing assimilation.

Though I begin the chapter with an analysis of Russian /v/, the primary goal of the chapter is to establish the cross-linguistic phonological identity of /v/ and the features [obstruent] and [sonorant]. Together, the features [obstruent] and [sonorant] yield four logical possibilities: [+sonorant, +obstruent], [+sonorant, −obstruent], [−sonorant, +obstruent], [−sonorant, −obstruent]. The feature specification [−sonorant, −obstruent], which is how the non-explosive stops of Ikwere are analysed, does not apply to voiced spirants. Although in some languages /v/ does not show evidence of patterning with either sonorants or obstruents, I argue that this is because [sonorant] and [obstruent] are not active in the phonology of that language, and I propose that the specification of [−sonorant, −obstruent] is not available to /v/ due to the phonetic underpinnings of [obstruent] and [sonorant].

One of the main claims in this chapter is that the standard conception of phonological features as binary partitions is at the root of the "problem of /v/", particularly with respect to [sonorant]. By introducing [obstruent], we can capture some of the granularity that is afforded by the sonority hierarchy, which proves to be relevant for accounting for the voicing assimilation facts of a variety of languages. In spite of the tighter link assumed between voicing assimilation and the sonority hierarchy, the data does not support using the full power of a finely grained sonority hierarchy in order to account for laryngeal phenomena.

The patterning of *ambiguous* /v/ is not unique to Russian, and I propose that all cases of ambiguous /v/ patterning be treated as [+sonorant, +obstruent]. In some languages, /v/ patterns as a voiced obstruent with respect to voicing assimilation, and in these languages I propose that /v/ is specified as [−sonorant, +obstruent]; however, I show that not only is this phonological specification rarer than has previously been assumed, but that it seems to require contrast with a labiodental approximant. In other words, the featural specification [−sonorant, +obstruent] generally arises in inventories in which it can contrast with a labial approximant specified as [+sonorant, −obstruent]. Another class of languages, exemplified by Serbian, exhibit sonorant patterning of /v/; that is, /v/ is outside the obstruent class entirely, as it does not trigger regressive voicing assimilation (RVA), but is also not a target of RVA or final devoicing (FD), and hence these languages suggest a classification of [+sonorant, −obstruent].

The features [obstruent] and [sonorant] are very nearly complements of each other, but not quite. I argue that by distinguishing between segments that are *exclusively* sonorant or obstruent, and those that are both or neither, we achieve a more parsimonious explanation for voicing phenomena across a broad range of languages. The notion of *exclusivity* allows us to introduce a finer-grained distinction of category membership than the all-or-nothing perspective required by traditional features, without having to invoke scalar features or mere phonetic similarity.

This chapter is structured as follows: in Section 5.2, I begin with a brief summary of Clements's analysis of Ikwere nasal harmony and the introduction of [obstruent]. In Section 5.3 I show that, phonetically, /v/ is both [+obstruent] and [+sonorant], and that analysing /v/ as [+sonorant, +obstruent] straightforwardly accounts for its patterning in Russian and other ambiguous /v/ languages. In Section 5.4, I present a typology of /v/ with respect to [obstruent, sonorant] and voicing assimilation; this section is primarily descriptive, but it serves the important purpose of showcasing just how similar the patterning of /v/ is in language after language. Section 5.5 shows how the present typology relates to other typologies of voicing assimilation that have taken the obstruent-sonorant divide for granted, and shows that the "problem of /v/" undermines other attempts to characterize the voicing assimilation of a language. These sections furthermore show that regressive voicing assimilation (RVA) is generally asymmetric with respect to triggers and targets, and that the assumption that the triggers and targets of RVA are generally the same only holds for stops and sibilants. Finally, in Section 5.6, I turn my attention to the question of what it means for a segment to be specified as [+sonorant, +obstruent]. Section 5.7 briefly considers the featural representation of voiceless spirants, while Section 5.8 returns to questions about the nature of voicing and its interaction with sonorancy and obstruency. Section 5.9 concludes the chapter with a discussion.

## 5.2 The feature [obstruent]

The feature [obstruent] is introduced in a series of papers by Clements and Osu in their analysis of Ikwere nasal harmony (Clements and Osu, 2002, 2003, 2005). To briefly summarize the phenomenon in question: Ikwere (`ikw`) contrasts oral and nasal vowels, but has no phonemic nasal consonants. Nasal vowels trigger nasal harmony within the phonological root; targets of nasal harmony include sonorants and the non-explosive stops [ḅ, 'ḅ].

Based on airflow and pressure traces of [ḅ, 'ḅ], Clements and Osu (2003) shows that although these sounds are produced with complete closure, they are not characterized by an increase in supra-glottal pressure that characterizes explosive stops such as [p, b], and therefore are not obstruents. On the other hand, they are also clearly not sonorants, since they lack the periodic and well-defined formant structure characteristic of sonorant sounds (Ladefoged, 1997). Clements and Osu (2003) proposes a binary feature [obstruent], based on the definition of obstruency given by Stevens (1983):

> Another class of consonants, called *obstruent*, is defined in the articulatory domain by the presence of a pressure increase within the vocal tract during the production of the consonant. This pressure increase occurs because a complete closure or a sufficiently narrow constriction is made within the vocal tract to contain the air.
>
> The acoustic consequence of this pressure increase is that turbulence noise is generated in the vicinity of the constriction at some point during the production of the sound. This noise can occur either throughout the constriction interval (as in a fricative consonant) or at the release of a closure (as in a stop consonant) (. . . ). Presumably, a listener is sensitive to the presence or absence of this type of noise in the sound, and this attribute, then, defines the natural class of obstruent consonants. (Stevens, 1983, pg. 254)

Crucially, [obstruent] and [sonorant] are not coextensional, and together yield a three-way classification of stops: explosive stops, non-explosives stops, and sonorant stops (i.e., nasals), shown in (1).

(1) Stop classification (Clements and Osu, 2003, pg. 89)

| | explosive stops | nonexplosive stops | sonorant stops |
|---|---|---|---|
| [obstruent] | + | − | − |
| [sonorant] | − | − | + |

Adopting [obstruent] as a feature readily accounts for the natural class of segments that undergo nasalization in Ikwere: all [−obstruent] segments are targets of nasal harmony.

An important component of Clements & Osu's analysis of Ikwere is the link between the representation of non-explosive stops, and the phenomenon in question, i.e., nasalization. It is well known that obstruents resist nasalization due to the difficulties of lowering the velum while simultaneously increasing the supraglottal pressure required for obstruency in the form of either an oral burst or frication (Ohala and Ohala, 1993). Cross-linguistically, vowels, glides, and liquids are more prone to nasalization than obstruents and exhibit implicational relationships between segment classes (see Cohn (1993) for a discussion of the phonetics and phonology of nasalization, and references therein).

Nasalizability looks rather like sonority (*cf.* Walker (2003)), but Hume and Odden (1996) propose that the nasalizability scale tracks *impedance*: how much airflow is impeded by a given supralaryngeal articulation. As Clements and Osu (2003) explains, the similarity between nasalizability and sonority stems from the fact that both are related to impedance: "high-sonority sounds and those that are most susceptible to nasalization are both characterized by low impedance (low resistance to airflow), while low-sonority sounds and those that are most resistant to nasalization are both characterized by high impedance (high resistance to airflow)" pg. 15. By focussing on impedance, there is a clear place on the nasalizability scale for the non-explosive stops: non-explosive stops are between voiced fricatives and liquids in terms of their nasalizability.

To summarize, Clements & Osu introduce the feature [obstruent] in addition to [sonorant]. They show that there exists a class of segments, the Ikwere non-explosive stops [ḅ, 'ḅ] that, phonetically, are neither obstruents nor sonorants, and as such are featurally specified as [−sonorant, −obstruent]. Their featural specification as [−obstruent] explains their amenability to nasalization, which is recast as an obstruency scale.

In what follows I adopt a similar approach to the featural specification of /v/ in voice assimilation systems. I begin with a case study of Russian, arguing that Russian /v/ is [+sonorant, +obstruent]. Not only does this allow for a straightforward analysis of the Russian facts, but it allows us to make sense of the broader typology of /v/'s patterning, the topic of Section 5.4. To preview the big picture somewhat, I show that there are three kinds of /v/, corresponding to three distinct featural specifications: (1) /v/ that is [−sonorant, +obstruent] patterns strictly with obstruents, and is instantiated by Maltese; (2) /v/ that is [+sonorant, −obstruent] patterns strictly with sonorants, and is instantiated by Serbian; (3) /v/ that is [+sonorant, +obstruent] patterns with both obstruents and sonorants, as exemplified by Russian. The fourth possibility, [−sonorant, −obstruent] is not available for voiced spirants. The patterning of Russian /v/ is therefore a natural part of the typology, and central to understanding the cross-linguistic identity of /v/.

## 5.3 Russian /v/ as [+sonorant, +obstruent]

In this section I argue that Russian /v/ is specified as [+sonorant, +obstruent]. I show that such an analysis straightforwardly accounts for the categorical voicing phenomena in Russian. I then discuss some aspects of the variability of /v/, and propose that this variability is due to an inherent tension between the dual specification of [+sonorant, +obstruent]. Finally I briefly discuss how the present analysis relates to previous analyses of Russian /v/.

### 5.3.1 Featural specification of Russian /v/

Admitting the feature [obstruent] allows us to capture the dual nature of Russian /v/ quite easily. Of the four logical possibilities introduced by having both [sonorant] and [obstruent], two can be ruled out immediately: [+sonorant, −obstruent] and [−sonorant, +obstruent]. The specification [+sonorant, −obstruent] is equivalent to specifying /v/ as [+sonorant], and would not account for its devoicing under RVA; the specification [−sonorant, +obstruent] is equivalent to specifying /v/ as [−sonorant], and would not account for the fact that it does not trigger RVA.

Of the other two logical possibilities, both will technically solve the problem of Russian /v/. Consider first the specification of Russian /v/ as [−sonorant, −obstruent], like the non-explosive stops in Ikwere. Then the class of targets for devoicing are all [−sonorant] sounds, and the class of triggers for RVA are all [+obstruent] sounds. If we instead consider the specification of Russian /v/ as [+sonorant, +obstruent], then the class of targets for devoicing are all [+obstruent], and the class of triggers for RVA are all [−sonorant, +obstruent] sounds. Table 5.1 summarizes the two analyses.

Table 5.1: Comparing two possible feature specifications for /v/ using [obstruent] and [sonorant]; RVA = Regressive Voicing Assimilation, FD = Final Devoicing

| | Analysis 1 | Analysis 2 |
|---|---|---|
| /v/ | [−sonorant, −obstruent] | [+sonorant, +obstruent] |
| /p, t, k, b, d, g, s, z/ | [−sonorant, +obstruent] | [−sonorant, +obstruent] |
| /m, n, l, r/ | [+sonorant, −obstruent] | [+sonorant, −obstruent] |
| Triggers of RVA | [+obstruent] | [−sonorant, +obstruent] |
| Targets of RVA & FD | [−sonorant] | [+obstruent] |

At first glance, Analysis 1 seems simpler, since both the class of triggers and targets are specified by just one feature. In contrast, the class of triggers of RVA in Analysis 2 is the set of segments that are specified as [+obstruent] *and* [−sonorant]. Nevertheless, in spite of the apparent simplicity of Analysis 1, I will argue that Analysis 2 is superior, both because of the interpretation of [sonorant] and [obstruent], and for the range of phonological data that

Analysis 2 can account for.

First, if we consider the definitions of [obstruent] and [sonorant] provided above, it is clear that /v/ is *both* obstruent and sonorant. Articulatorily, /v/ is produced with sufficiently narrow constriction so as to produce at least some turbulence, but acoustically is characterized by periodicity and well-defined formant structure. Clements and Osu (2002) ultimately conclude that [+sonorant, +obstruent] ought to be universally excluded—a point of disagreement with the present analysis—but they nevertheless acknowledge that such sounds are physically possible, citing voiced fricatives:

> Unlike other mutually exclusive feature values such as [spread glottis] and [constricted glottis], such sounds are physically possible: laxly-articulated voiced fricatives such as *z* and *v* often combine the clearly-marked formant structure characteristic of sonorants with the turbulence noise component characteristic of obstruents, and thus qualify as "sonorant obstruents" under the proposed definitions of these features. (Clements and Osu, 2002, 34)

The non-explosive stops in Ikwere and other languages are characterized by complete closure of the articulators without the accompanying increase in supra-glottal pressure and subsequent burst that characterizes explosive stops. In contrast, /v/ exhibits the narrow articulatory configuration of an obstruent, and at least some high frequency noise. In spite of their non-obstruence, the non-explosive stops of Ikwere do not exhibit the acoustic properties of sonorants. In contrast, Russian /v/ does exhibit acoustic properties that are shared by other sonorants: voicing throughout its duration, combined with clear formant structure and high periodicity.

In Chapter 2 we saw that the degree to which [v] in Russian is characterized by frication depends in part on the environment. Recall that in the WIS environment, the spectral centroid on the 1500 Hz high-pass filtered signal of [v] was similar to [f], and was parallel to the relationship between [z] and [s]; in the WMU environment, the spectral centroid of Russian [v] differed from [f] to a much greater degree than [z] did from [s]. In light of the variability in realization of Russian /v/, we might ask whether [+sonorant] is a suitable feature specification for /v/ in WIS position, and whether [+obstruent] is a suitable feature specification for /v/ in WMU position. First, the underlying feature specification for /v/—namely, the feature specification that is required to capture its phonological patterning—must allow for positional variability. Second, the fact that Russian /v/ exhibits a range of realizations that are more obstruent and more sonorant supports the analysis that Russian /v/ is *both* an obstruent and a sonorant. Finally, it is important to recognize that in spite of the variability, the realization of Russian /v/ is not purely sonorant: even in the WMU position, tokens of Russian [v] are still characterized by some frication.

Phonetically, then, Russian /v/ is a good candidate for being specified as [+obstruent] and [+sonorant]. In terms of its phonological patterning, we have already seen that Russian /v/ patterns with both obstruents and sonorants with respect to voicing assimilation: like obstruents, /v/ devoices word-finally and as a result of regressive voicing assimilation (RVA) before voiceless obstruents; like sonorants, /v/ does not trigger RVA, so sequences such as /tv/ are realized as [tv], and not as [dv]. Beyond the voicing assimilation facts, there are further reasons to consider Russian /v/ as both an obstruent and a sonorant. Recall from Section 1.5.3 that in terms of two-consonant clusters, Russian /v/ exhibits the distributional properties of both obstruents and sonorants. Like obstruents, /v/ can appear as the first member of two-consonant clusters, a position generally[1] reserved for obstruents; when in this

[1]Russian is famous for its typologically unusual phonotactics, allowing up to four consonants in the onset ([fskrɨl] 'he opened') and coda ([tʃorstf] 'stale'), sonorant-obstruent sequences in onsets ([rtut$^j$] 'mercury') and obstruent-sonorant sequences in codas ([ʒezl] 'staff') (Proctor, 2009, pg. 126). Such clusters have been been traditionally viewed as counterexamples to the Sonority Sequencing Principle (SSP), but Proctor (2009) argues that sonority violations are only permitted at word-edges, and argues that word-medially, Russian syllable structure generally follows the SSP. Based on a corpus study based on Sharoff (2002) of Russian two-consonant onset clusters, Proctor further concludes that these clusters are dispreferred and that "the modern lexicon of Russian demonstrates an overwhelming preference for syllable structures in which onsets,

position, it can precede stops, fricatives, nasals, liquids, and glides. Like the glides, liquids and nasals, it can also appear in second position, and in this position, consistent with its patterning as a sonorant with respect to the active phonological process of RVA, can disagree in voicing specification with the previous obstruent.

The sonorant patterning of /v/, beyond its non-involvement with RVA, has long been used to justify the sonorant status of Russian /v/. For example, /v/ patterns with sonorants /j, n, m/ in a group of verb stems that Jakobson (1948) calls the *narrowly closed full-stems* class. These verb stems end with consonants /j, v, n, m/, and drop this consonant before a consonantal suffix, as in (2); in verb forms that end with an obstruent other than /v/, the consonant does not delete.

(2) Narrowly-closed full-stems: deletion of sonorants before consonantal suffix

| | 3rd plural | past feminine | |
|---|---|---|---|
| a) | [delaj-u-t] | [dela-l-a] | ‘do’ |
| b) | [stan-u-t] | [sta-l-a] | ‘become’ |
| c) | [ʒiv-u-t] | [ʒi-l-a] | ‘live’ |

Padgett (2002) notes that the sonorant patterning of Russian /v/ is also seen in some of its phonotactic behaviour, citing Zalizniak (1975). Like other sonorants, /v/ requires following sonorants to have a non-syllabic realization, thus [mavr] ‘Moor’ and [ʒanr] ‘genre’; in contrast, a syllabic pronunciation is possible following true fricatives, thus [tsifr] $\sim$ [tsifr̩] ‘figure (gen. pl.)’.

To summarize, based on the phonetic properties of /v/, its distribution and phonotactic

and to a lesser extent also codas, conform to typologically-standard sonority sequencing principles” (Proctor, 2009, pg. 130).

patterning, and its behaviour under RVA and FD, it is clear that Russian /v/ exhibits properties of both obstruents and sonorants, justifying the featural specification of [+sonorant, +obstruent].

Further support for considering Russian /v/ comes from considering the nature and representation of voicing assimilation. It is well-known that obstruents are prone to devoicing due to the aerodynamic tensions inherent in maintaining sufficient airflow for voicing and achieving the articulatory configuration required for obstruency. Devoicing is therefore crucially not simply a property of "non-sonorants", but rather a property of obstruency. I am not aware of any reports that non-explosive stops (often more simply referred to as implosives) are prone to devoicing; in fact, implosion is one strategy of *maintaining* voicing through the duration of a stop consonant, and Clements and Osu (2002) maintain that the Ikwere non-explosive stops are voiced through their closure. Therefore the class of targets for devoicing is better understood as [+obstruent], as in Analysis 2, rather than as [−sonorant], as in Analysis 1. With respect to the class of triggers of RVA, both analyses pick out obstruents as triggers. The difference is that Analysis 2 further picks out those obstruents that are not also [+sonorant]. This amounts to a difference in *exclusivity*: RVA is triggered by segments that are *exclusive* obstruents, a term I elaborate upon in Section 5.6.3.

Under the assumption that Russian /v/ is featurally specified as [+sonorant, +obstruent], the analysis of voicing facts in Russian becomes reasonably straightforward. For simplicity's sake, consider what it would mean to represent Russian voicing phenomena in a rule-based formalism with bundled features and a binary feature [voice]. An appropriate rule for regressive voicing assimilation would be something like (3a), while the appropriate rule for final devoicing would be something like (3b).[2]

[2]The presentation of rules in (3) should not be interpreted as advocating for a particular phonological formalism.

(3) Russian voicing phenomena rules

a. RVA: [+obstruent] → [$\alpha$voice] / ______ [−sonorant, +obstruent, $\alpha$voice]

b. FD: [+obstruent] → [−voice] / ______ #

For the stops, sibilants, and traditional sonorant consonants (nasals, liquids, glides), the rules are equivalent to using [−sonorant] in place of both [+obstruent] and [−sonorant, +obstruent]. Because /v/ is an obstruent, it devoices before voiceless obstruents, and so /v supe/ → [f supe] 'in the soup'; because /v/ is not specified as [−sonorant, +obstruent]—i.e., it is not an "exclusive obstruent", as I will discuss later—it is not a trigger for voicing assimilation, and so /ot-vesti/ → [otvesti], *[odvesti] 'lead away'. Similarly, as an obstruent, /v/ undergoes FD, resulting in the alternation [prav-a] ∼ [praf] 'right (fem./masc.)'. The rules in (3) are ordered with respect to one another, and final devoicing feeds regressive voicing assimilation, as in [pojezd-a] ∼ [pojest] 'train (gen./nom.sg)'.

Allowing Russian /v/ to be specified as both [+obstruent] and [+sonorant] allows other facts about Russian phonology to be straightforwardly accounted for as well, since processes and distributional constraints that involve [+sonorant] segments include /v/, and processes and distributional constraints that involve [+obstruents] also include /v/. Although the analysis presented here depends on the dual specification of Russian /v/ as [+sonorant, +obstruent], representing both [+obstruent] and [+sonorant] on a single segment carries with it an inherent tension, which I argue is seen in some "glitches" specific to /v/.

### 5.3.2 Glitches and variability

The first glitch is that, although /v/ does not trigger voicing assimilation in general, it does if and only if it precedes an obstruent, as seen in (4). As described by Kavitskaya

(1998), this is “a case of positional identity: /v/ is phonologically an obstruent when it is followed by an obstruent or word-finally, and a sonorant when it is followed by a sonorant” (pg. 225).

(4) Russian /v/ as a trigger for RVA

| | | |
|---|---|---|
| a) /pod vsemi/ | [potfsemi] | ‘underneath everyone’ |
| b) /ot vdovɨ/ | [odvdovɨ] | ‘from the window’ |
| c) /k vzdoxam/ | [gvzdoxam] | ‘to the sighs’ |

The second glitch is that, in word-final position, /v/ always devoices to [f], but preceding voiced obstruents variably retain their voicing when they precede word-final /v/; thus [tr$^j$esf] ∼ [tr$^j$ezf] ‘sober (short adj.)’ (Kavitskaya, 1998, pg. 230).

Recent phonetic investigations by Kulikov (2012), however, show that the triggering status of /v/ is more subtle than previously assumed. In cases where /v/ devoices to /f/, preceding voiced obstruents devoice, so that /pod vsemi/ is indeed realized as [potfsemi]. In cases where /v/ precedes a *voiced* obstruent, however, a preceding voiceless obstruent was only voiced some of the time. As described by Kulikov (2012, pg. 126), “Voicing in /tvd/ clusters was observed less often, but it was a regular pattern for speakers 3, 6, and 11 even when reading the list. The other speakers did not assimilate /t/s before /v/ followed by a voiced obstruent in the list condition. Speakers 8 and 13 produced half of underlying /t/s in /tvd/ clusters as voiced and half as voiceless.” This suggests that the two “glitches” are in fact variants on the same underlying issue, and requires some comment as to the relationship between [voice], [obstruent], and [sonorant].

For true obstruents specified as [−sonorant, +obstruent], being specified with either valuation of [voice] is equivalent to traditional representations of [voice] on [−sonorant] segments.

Because /v/ is specified as [+obstruent], it is also specified as [+voice], in spite of its specification for [+sonorant]. I propose that either valuation of [voice] on [+sonorant, +obstruent] is an unstable configuration, leading to different interpretations, and hence variability.

First we consider the word-final case, where under final devoicing, /v/ is specified as [[+ obstruent, + sonorant, − voice]]. Since Russian sonorants are voiced, such a configuration may be simplified to [[+ obstruent, − sonorant, − voice]], thereby representing neutralization with [f] and triggering devoicing. Alternatively, the special status of /v/ may allow it to stay as is, and so word-final devoiced /v/ is in fact a voiceless obstruent sonorant that is better represented as [v̥].[3] This analysis predicts that forms that do not simplify are realized as [trʲezv̥], while forms that are simplified are realized as [trʲesf]. Although Kavitskaya (1998) describes this phenomenon as /v/ devoicing to [f], she does not present acoustic data of these forms, and the perceptual difference between [v̥] and [f] is certainly very small. Future avenues of research would be to test if there is a difference between the word-final devoiced /v/ in these variants.

Explaining the variability in RVA sequences is trickier. Kulikov (2012) shows that greater variation in voicing is found in stop-final prepositions such as /ot/ ‘from’ and /nad/ ‘over’, suggesting that this may confound the results somewhat, as sequences of /dvs/ and /tvd/ arise from preposition + noun forms. Nevertheless, an asymmetry is reported between the devoicing and voicing cases, which calls for at least a sketch of an explanation.

If word-final cases of /zv/ can be realized as [zv̥], then why does this not happen in clusters of the form /dvs/, which are uniformly realized as [tfs]? And why does /tvd/ exhibit variability? I suggest that this has to do with different pressures that the environment places

[3]Transcriptions such as [v̥] highlight the inadequacy of a feature system that distinguishes only between a voiced obstruent fricative [v] and a voiced sonorant approximant [ʋ]. Because I do not take the stance that Russian /v/ is intermediate, I simply choose [v] to represent tokens of /v/ when I have to refer to them in this chapter.

on the tensions inherent to the specification of [voice] on [+sonorant, +obstruent]. It is well known that Russian admits sonority reversals at word-edges, and with word-final devoiced /v/, the boundary does not contribute as much pressure for /v/ to be reanalyzed as [[+ obstruent, − sonorant, − voice]]. In contrast, the preobstruent position of /v/ in sequences such as /dvs/ exerts more pressure to resolve the unstable configuration.

The case of voicing /tvd/ to either [dvd] or [tvd] also exhibits instability. When /d/ spreads its voicing to /v/, one possibility is that /v/ is treated as a standard voiced sonorant, and so blocks voicing assimilation. Some presentations of voicing facts in Russian describe sonorants as transparent to voicing assimilation (Hayes, 1984), meaning that voicing assimilation occurs between two obstruents that are separated by a sonorant, so that /tmz/ is realized as [dmz]. Kulikov (2012) refutes this claim, showing that voicing assimilation does not apply through sonorants. Because /v/ is specified as [+sonorant], it may be interpreted as a blocker for voicing assimilation. The other possibility is that the [+obstruent] interpretation wins out, and therefore /v/ triggers voicing of the preceding voiceless obstruent.

Accounting for the "glitches" of voicing assimilation has been a long-standing sticking point for analyses of Russian /v/, and these glitches are made all the more difficult to account for given their variability. Certainly, my attempts to account for the variability in /v/'s trigger status leave open many questions, but the main one has to do with the phonetic realization of /v/ in these sequences: does the variability in /v/'s trigger status correspond with more or less sonorant realizations? Answering this question is left for future research. Nevertheless, I maintain that by recognizing the dual specification of /v/, as well as the inherent tension of such a specification, we may be able to shed some light on the unique variability of /v/ in voicing phenomena.

### 5.3.3 Relationship to other analyses

As I discussed in Chapter 1, the intermediacy of Russian /v/ between obstruents and sonorants has long been recognized, and all analyses of the Russian voicing facts must capture this intermediacy in some way. In analyses using SPE-style rules, and with only [sonorant] to distinguish obstruents from sonorants, the intermediacy is captured obliquely through rule ordering. Consider, for example, the analysis of Russian voicing assimilation by Hayes (1984)[4], in which [v] is posited to be /w/ underlyingly. Four ordered rules are required to derive the correct surface forms, paraphrased for simplicity in (5). In order for this analysis to go through, sonorants must also be targets of voicing assimilation, devoicing before voiceless consonants, and then revoice *after* /w/ obstruentizes to either [f], if voiceless, or [v], if voiced.

(5) a. *Final Devoicing*

C → [−voice] / ______#

b. *Voicing Assimilation*

In a consonant cluster, assign voicing of the last obstruent to all consonants on its left.

c. *W Strengthening*

[− consonsonantal, +labial] → [−sonorant]

d. *Sonorant Revoicing*

[+sonorant] → [+voice]

The advantage of positing that /v/ is underlyingly a sonorant /w/ is that it is easy to account for the various ways that it patterns as a sonorant. But the analysis is clearly problematic:

[4]Kiparsky (1985) is a similar approach.

not only does /v/ never surface as [w] in Russian, but also there is no evidence that sonorants are prone to devoicing.[5] In other words, in order to account for the patterning of a single segment, voicing assimilation has to apply far more broadly than there is evidence for, the underlying form posited for Russian /v/ is notably abstract, and an extra rule that ensures the obstruent realization of /v/ must be posited.

More recent analyses have tackled the intermediacy of Russian /v/ more directly, either by representation, constraint, or feature. Abstracting away from the details, we see a long history within the literature of having to treat /v/ as intermediate in some way. For example, Kavitskaya (1998) argues that /v/ is underspecified with respect to sonority, and posits a constraint No-released-v that states that, unlike obstruents, /v/ may not be released, and hence has the same sonority specification as the following segment (pg. 237). Reiss (2018) argues instead that /v/ is underspecified with respect to a binary feature [voice]. Working within the theory of contrastive specification (Dresher et al., 1994; Dresher and Zhang, 2005), Hall (2007) argues that /v/ is specified for neither the [Laryngeal] node, which is the source for voicing on obstruents, nor with [SV], which is the source of voicing for sonorants.

It is beyond the scope of this work to delve into issues surrounding underspecification, but I briefly consider the analysis of Reiss (2018) to show why such approaches fail *in principle* for the variability associated with /v/. Reiss (2018) proposes that Russian /v/ is underspecified with respect to [voice]; he denotes underspecified Russian /v/ as /V/. Briefly, Reiss (2018) proposes that /v/ is specified as [−sonorant], but is unspecified with respect to voice; in his framework, all consonants other than /v/, both obstruent and sonorant, are specified for [voice].[6] Because /v/ is not specified for [voice], it cannot trigger voicing assimilation;

[5]Russian sonorants, like sonorants in other languages, can devoice gradiently in certain positions, but do not devoice in pre-obstruent position; see Padgett (2002) for a discussion of sonorant devoicing, as well as for a more thorough critique of Hayes's analysis.

[6]For full details of this framework, the reader is directed to Reiss (2018).

because it is [−sonorant], it can always acquire a voicing specification of a segment to its right. When /v/ has the specification [− sonorant, − voice], it surfaces as [f], and when it has the specification [− sonorant, + voice], it surfaces as [v].[7] The problem arises when we consider final devoicing, in which /v/, like all [−sonorant] segments, acquires [−voice] by rule in word-final position: "So, /V/ neutralizes with the voiceless obstruent /f/ to [f] in final position". Standard rule ordering between FD and RVA accounts for the devoicing of final clusters.

This analysis runs into a problem when we consider the glitches mentioned above. Because the featural specifications of /v/ are filled in with either valuation for [voice] at some point in the derivation, /v/ is either neutralized with [f] or is a fully obstruent [v]. There is no way to account for the unique variability of /v/ in this framework. The above proposal to treat a dual specification of [+sonorant, +obstruent] as inherently in tension at a phonological level is incomplete, but it represents a serious attempt to tackle the unique status of /v/, including its variability, head on.

Perhaps the analysis that, in some way, comes closest to the present one is that of Padgett (2002), because he explicitly represents the dual membership of /v/ as both an obstruent and sonorant in terms of features.[8] In particular, Padgett (2002) posits that Russian /v/ is in fact /ʋ̞/, and is featurally specified as [+sonorant] and [−wide]. The feature [wide] refers to constriction degree: obstruents and /ʋ̞/ are specified as [−wide], while nasals, liquids, and glides are specified as [+wide]. Articulatorily, [wide] refers to the degree of oral aperture required in the articulation of a segment; correspondingly, Padgett calls segments such as [ʋ̞]

[7]For brevity I omit the other features which clearly must apply, such as [labial]. Note also that, for ease of exposition, I do not use the notational conventions used by Reiss (2018) to denote the set-theoretic nature of features and natural classes. The argument I present here does not hinge on this innocent misrepresentation.

[8]It should be noted here that Padgett (2011) supplants the main part of his analysis of Russian voicing assimilation from Padgett (2002). However, Padgett (2011) focusses primarily on issues having to do with the domain of voicing phenomena, and he does not touch upon the problem of /v/ in that analysis. For this reason I discuss only the original analysis, which specifically deals with /v/.

*narrow approximants* and notes that other segments that share this property are segments traditionally transcribed as /β, ð, ɣ/. By analysing Russian /v/ as [+sonorant] and [−wide], Padgett classifies /v/ with both obstruents and sonorants.

Featurally, both analyses manage to single out /v/ as belonging to both the obstruent class, as either [+obstruent] or [−wide], and to the sonorant class, as [+sonorant]. However, in spite of the surface similarities, [obstruent] is not simply a notational variant of [wide]. First, the definition of obstruency refers not to constriction degree, but rather is based on impedance to airflow. Certainly, [−wide] will often pick out the same segments as [+obstruent] (and, for that matter, [−sonorant]): for example, stops and sibilants are [−sonorant], [+obstruent], and [−wide]; liquids and glides are [+sonorant], [−obstruent], and [+wide].[9] This is because constriction degree affects airflow impedance, which correspondingly affects the acoustic properties of resonance. However, while [wide] arguably does the job for Russian /v/, it is not generalizable beyond this problem. As discussed above, there is already precedent for a feature [obstruent] in order to account for the nasalizability of non-explosive stops in Ikwere. It is not clear what additional work [wide] does. A further problem is that [wide] relies on a close mapping of constriction degree and phonological patterning; as we saw in Chapter 3 showed that Russian /v/ does not exhibit the inherent intermediacy across environments that is predicted by Padgett (2002). The feature [obstruent], while still precisely defined, is satisfied by a greater range of articulatory configurations.

The history of Russian /v/ in the phonological literature has been to determine language-specific rules, constraints, and representations. These approaches have either captured its intermediacy—that is, the fact that it is neither a true sonorant nor a true obstruent—or its dual membership as both a sonorant and an obstruent. Analysis 1 posits that /v/ is [−sonorant, −obstruent], and captures the intuition that /v/ is somehow in between

[9]A discrepancy is found with nasals, which are [+sonorant] and [−obstruent], but [−wide].

obstruents and sonorants, but not truly one or the other. Analysis 2 posits that /v/ is [+sonorant, +obstruent], and captures the intuition that /v/ is a member of both categories. In Section 5.6.3, I show that both of these intuitions can be encompassed by the notion of exclusivity, and that identifying /v/ as both obstruent and sonorant, but as neither exclusively, subsumes previous analyses. Until then, I continue to adopt the assumption that Russian /v/ is [+sonorant, +obstruent].

In the next section I survey the cross-linguistic identity of /v/ with respect to laryngeal phenomena. The patterning of /v/ in Russian and other related Slavic languages has attracted attention because of it appears to be an anomaly; it is generally assumed that obstruents and sonorants occupy different worlds that are neatly partitioned into [−sonorant] segments and [+sonorant] segments, particularly with respect to voicing assimilation phenomena. The main claim of the present chapter is that the standard partition of consonant inventories as *either* sonorants *or* obstruents is incorrect, and that, in general, /v/ sits at the intersection of the set of obstruent consonants and sonorant consonants.

## 5.4 Typology of /v/ patterning

We have seen that by specifying Russian /v/ as [+sonorant, +obstruent], the analysis of voicing facts in Russian is straightforward, and accounts not only for the categorical voicing phenomena, but provides an avenue to consider the unique variability of /v/. In this section I situate the Russian patterning in a cross-linguistic setting. The combination of [obstruent] and [sonorant] together allow for four logical possibilities, and circumscribe a possible typology: [+sonorant, +obstruent], [−sonorant, +obstruent], [+sonorant, −obstruent], and [−sonorant, −obstruent], as shown in Table 5.1.

Table 5.1: Language candidates for the typology of [sonorant] and [obstruent]

| | [+sonorant] | [−sonorant] |
|---|---|---|
| [+obstruent] | Russian | Maltese |
| [−obstruent] | Serbian | N/A |

### 5.4.1 [+sonorant, +obstruent]

Russian is not the only language in which /v/ is specified as [+sonorant, +obstruent], a fact that is often cited in the Russian /v/ literature. The ambiguous patterning of /v/ with respect to voicing assimilation is attested in other Slavic languages such as Bulgarian (`bul`), Macedonian (`mkd`), and Czech (`ces`), and in non-Slavic languages such as Hebrew (`heb`) and Hungarian (`hun`). In all of these languages, /v/ patterns with both obstruents and sonorants in voicing assimilation, exhibiting the same asymmetry as in Russian. Like obstruents, /v/ undergoes devoicing before voiceless obstruents and word-finally.[10] Like sonorants, /v/ does not trigger voicing assimilation, and so sequences such as /tv/ are licit, and contrast with sequences such as /dv/.

Although the non-uniqueness of the Russian /v/ patterning is well known, the ubiquity of /v/'s intermediacy is under-appreciated. Further, there are patterns of variability in the realization of /v/ that surface again and again across languages, a fact that is rarely discussed at length in the phonological literature. In this section I discuss the patterning, distribution, and variability of /v/'s realization in several languages. This, together with the following section that considers candidates for [−sonorant, +obstruent] specification, shows that /v/ nearly always patterns as intermediate between obstruents and sonorants.

I begin with Czech (`ces`), another Slavic language in which /v/ patterns ambiguously.

[10]Neither Hebrew nor Hungarian have final devoicing.

#### 5.4.1.1 Czech

According to Hall (2003b), citing Hála (1962), a variant of the Russian patterning is reported in some Czech dialects, in which /v/ devoices following voiceless obstruents. In general, the voicing facts of Czech are like those of Russian: the voicing of the rightmost obstruent in a consonant cluster triggers voicing assimilation on all preceding consonants in the cluster, so that, for example, /z xɪbɪ/ → [s xɪbɪ] 'from a mistake'. Czech also has final devoicing of obstruents, and final devoicing feeds regressive voicing assimilation, yielding alternations such as [ɦvost] ∼ [ɦvozdem] 'forest (nom.sg.) ∼ (inst.sg.)'.[11] Like in Russian /v/ devoices before voiceless obstruents, so that /v pole/ → [f pole] 'in a field', but /v/ does not trigger voicing of preceding voiceless obstruents. In some dialects, /v/ patterns exactly like in Russian, so that forms such as [tvor̝it se] 'to take shape' are attested; in dialects with PD, this form is realized as [tfor̝it se].

Progressive devoicing is found in another Czech consonant, the trill-fricative /r̝/, represented orthographically as <ř>. Like /v/ and other sonorants, /r̝/ does not trigger voicing assimilation, but like /v/ and other obstruents, it undergoes devoicing word-finally and when it precedes a voiceless obstruent. Further, like /v/ in dialects with PD, and unlike all other obstruents and sonorants, /r̝/ becomes devoiced when it follows voiceless obstruents as well. Also like /v/, /r̝/ has a sonorant past, and is a reflex of a palatalized rhotic /rʲ/; synchronically, it is considered to be the soft counterpart to the Czech trill /r/ under palatalization. At first glance, one might think that /r̝/ is also specified as [+sonorant, +obstruent], but I argue that, in fact, /r̝/ is specified as [−sonorant, +obstruent]; that is, it is a true obstruent.

The phonetic character of /r̝/ has been the subject of much research. Šimáčková et al. (2012) describe /r̝/ as a period of sibilant-like frication preceded by one or more trill-like

[11] For a fuller exposition of the Czech data, see Short (1993); Dvořák (2010); Hall (2003b,a).

contacts. (Pavlík, 2014) shows, based on a naturalistic news-reading corpus, that the most common variant is a single contact, contra what they describe as the most common description of /r̝/ being realized with two or more contacts. An electromagnetic articulography study by Howson et al. (2015) shows that articulatorily, [r̝] is more similar to [ʒ] in terms of the grooving of the tongue. Komova and Howson (2014) argues that /r̝/ is produced with breathy voice, arguing that such a laryngeal configuration allows for greater airflow to maintain turbulence given the rhoticization. Acoustically, Šimáčková et al. (2012) describe /r̝/ as realized with significant sibilant-like frication. Phonetically, /r̝/ is an obstruent.

Phonologically, it is insufficient to consider only the parallelism between /v/ and /r̝/ under progressive devoicing. Across word boundaries, /r̝/ triggers voicing assimilation: /jak r̝ɛknɛʃ/ → [jag r̝ɛknɛʃ] 'when you say'. In contrast, /v/ does not trigger voicing assimilation across word boundaries, except in the dialect of Moravia, which is distinguished from other dialects by the fact that all voiced segments, sonorants and obstruents, trigger voicing assimilation across word boundaries.

Returning to Czech /v/, Hall (2003b) is the only source I have encountered that describes progressive devoicing as a feature of particular dialects.[12] Šimáčková et al. (2012) describe variability with respect to the behaviour of /v/ in voicing assimilation across the Bohemian and Moravian dialects of Czech, but do not mention progressive devoicing of /v/.

That said, variable progressive devoicing of /v/ after voiceless obstruents is reported in Bulgarian, Macedonian, Polish, and Hungarian; nasals, glides, and liquids are not reported to devoice in this position. I have not found any mention in the English literature that Russian /v/ exhibits variable progressive devoicing, though it is apparently discussed in the Russian

[12] "According to Hála, however, some dialects do devoice /v/ after voiceless obstruents, resulting in [f] or in a voiceless fricative that retains some of the stoplike character of /v/ mentioned earlier." (Hall, 2003b, pg. 101)

literature on /v/. According to Vladimir Kulikov, /v/ exhibited partial devoicing following voiceless obstruents in his data (Kulikov, 2012), but devoicing was variable and gradient, and did not result in [tf]; as such it is not distinguished from the variable and gradient devoicing of sonorants, as in /tl, tr/.[13] Without further confirmation of the Czech data, I will treat Czech /v/ similarly: /v/ is subject to variable progressive devoicing following voiceless obstruents. In light of this, I hypothesize that variable progressive devoicing is attributable to issues of timing between the laryngeal and supralaryngeal gestures, and leave this hypothesis to be explored in future research. In sum, although it has been reported that in some dialects of Czech both /v/ and /r̝/ pattern similarly with respect to voicing assimilation, I do not analyse them as having the same featural specification due to both phonetic and phonological reasons: /v/ is specified as [+sonorant, +obstruent], while /r̝/ is specified as [−sonorant, +obstruent].

Another aspect of Czech patterning deserves mention here. According to Šimáčková et al. (2012), in the Moravian dialect, sonorants also trigger voicing assimilation across word-boundaries, and /v/ is therefore also a trigger in this environment; in the Bohemian dialects, /v/ variably triggers voicing assimilation across word-boundaries, "about half of the time" (pg. 228). Hungarian is also reported to have dialects that exhibit obstruent patterning of /v/ Kiss and Bárkányi (2006), to which we briefly turn.

#### 5.4.1.2 Slovak

Slovak, like Czech and Russian, has regressive voicing assimilation, hence /od suːɟɪc/ → [ɔtsuːɟɪc] 'to condemn', and final devoicing, hence /xlɑd/ → [xlɑt] 'cold (n.)'. Where Czech has /r̝/, Slovak cognates have /r/, and Hall (2003b) points out that since /r/ is a sonorant, it is not subject to progressive devoicing. Across certain morpheme boundaries, voicing

[13] Thanks to Vladimir Kulikov for discussion on this matter (personal communication, March 2017).

assimilation can be triggered by sonorants as well, thus /vlak mɛʃkaː/ → [vlag mɛʃkaː] 'the train is late'.[14]

Like other ambiguous /v/ languages, Slovak /v/ exhibits asymmetry with respect to RVA, in that it undergoes devoicing but does not trigger voicing of preceding obstruents. However, unlike other languages discussed so far, Slovak /v/ does not undergo final devoicing with the remaining obstruents. Instead, coda /v/ is realized as [w], as in (6).

(6) Slovak coda /v/ (Hall, 2003b)

| | | | |
|---|---|---|---|
| a) /stav/ | → | [staw] | 'position nom.sg.m' |
| b) /stavba/ | → | [stawba] | 'building nom.sg.m' |

#### 5.4.1.3 Hungarian

Hungarian (hun) /v/ patterns ambiguously with respect to voicing assimilation (7), exhibiting the same patterning as in Russian.

(7) Hungarian /v/ (Kiss and Bárkányi, 2006)

| | | | | |
|---|---|---|---|---|
| a) /haːt-ba/ | → | [haːdba] | 'back illat' | |
| b) /laːb-toːl/ | → | [laːptoːl] | 'foot abl' | |
| c) /sav-toːl/ | → | [saftoːl] | 'acid abl' | |
| d) /keːt vaːr/ | → | [keːtvaːr] | 'two castles' | *[keːdvaːr] |

Kiss and Bárkányi (2006) note that, unlike sonorants, but like other obstruents (specifically, voiced fricatives such as /z, ʒ/), /v/ can appear as the second member of a word-final

[14] The details of this process will not be discussed here. For further information, see the sources in Hall (2003a).

consonant cluster. Thus words such as [elv] ‘principle’ and [terv] ‘plan’ occur, as do [borz] ‘badger’. Kiss and Bárkányi (2006) note (pg. 181, footnote 6) that /j/ can occur word-finally after sonorants, in which case it is realized as a noisy fricative [ʝ], as in [feːrʝ] ‘husband’.

The patterning of Hungarian /v/ appears to be subject to the same dialectal variation as seen in the Slavic languages. In the WTH dialect, we find [borodva] ‘razor’, corresponding to Educated Colloquial Hungarian (ECS) [borotva] (Kiss and Bárkányi, 2006). However, in these dialects, sonorants also trigger voicing assimilation (8a); there is also the widespread variability (at times varying even in the speech of a single speaker) the progressive devoicing of /v/ (8c).

(8) Western Transdanubian dialect of Hungarian (Kiss and Bárkányi, 2006)

a. /haːt nem/ → [haːdnɛm] ‘well not’

b. [borodva] ‘razor’

c. [huːʃfɛːt] ‘Easter’

Even this region, however, exhibits more variability than this description suggests. Figure 5.1 shows a linguistic map of southwestern Hungary from Király (2005), a language atlas of the Somogy and Zala counties, containing data from 99 villages in this region. The /tv/ sequence in the target word, <*kötve*> ‘tied’, is realized as [dv, tv, tf].[15] In standard Hungarian, forms such as /kœt+ve/ would be realized with the cluster [tv]. In this region, however, there is a fair degree of variability: while in some villages /v/ triggers voicing assimilation ([dv]), in others /v/ devoices ([tf]).[16] According to Péter Rebrus, Péter Szigetvári, and Miklós

[15]Thanks to Peter Rebrus who provided me with scanned copies of the relevant maps, as well as an explanation of the linguistic atlas and its data. Special thanks also to Mate Szabo, who translated the target sentence for me. Note that some forms are transcribed with uppercase V and D; it is not clear what these denote, but these forms only increase the number of variable realizations.

[16]This fact is discussed for both Hungarian and Hebrew in Barkai and Horvath (1978).

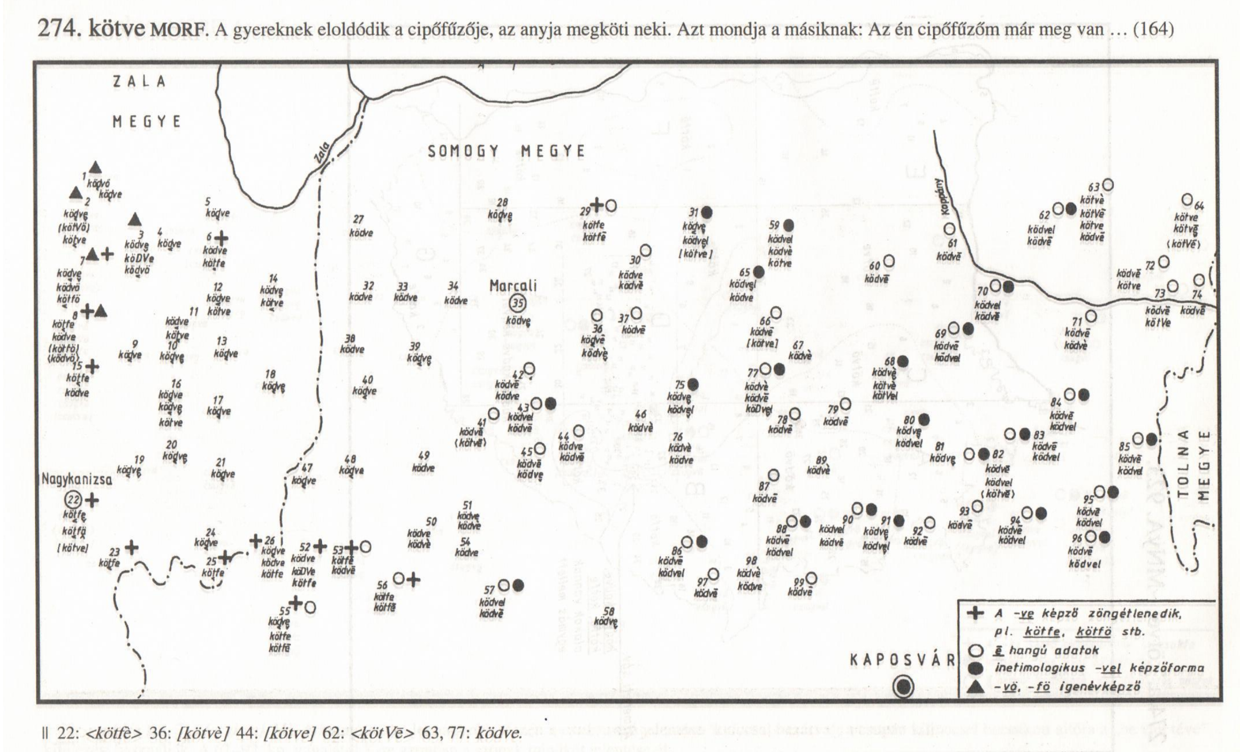



Figure 5.1: Realization of verbal stem <*köt*> + adverbial participle suffix <*va*>; target sentence translates to: *The shoelace of the kid becomes untied, and the mother ties it for them. The [kid] says to the other one: my shoelaces are now…*

Törkenczy, at least some of these dialects exhibit obstruent patterning of /v/, contrary to the description in Kiss and Bárkányi (2006) (p.c).

The standard description of Hungarian is that /v/ patterns ambiguously. Nevertheless, in the WTH dialect region, there is significant variability in the realization of /tv/ clusters, where /v/ can pattern either with obstruents in triggering voicing assimilation, with sonorants in not triggering voicing assimilation, on its own by devoicing after voiceless obstruents, or with all voiced segments in the patterning described by Kiss and Bárkányi (2006) as the WTH dialect.

The variability of /v/ patterning, which we will see more of presently, is important. True obstruents and true sonorants are not reported to undergo such variability. I return to this issue when I introduce *exclusivity*, but I submit that representation of segments reflect the difference in variability that is available to obstruents, sonorants, and /v/.

Table 5.2 lists the languages that qualify as "ambiguous /v/" languages, that undergo devoicing before voiceless obstruents (/tv/ → [tv]), but do not trigger voicing of preceding voiceless obstruents (/vt/ → [ft]). In all languages that I have come across, /v/ undergoes some degree of variable progressive devoicing when it follows voiceless obstruents.

Table 5.2: Ambiguous /v/ languages. Variable and/or gradient devoicing of /v/ after voiceless obstruents is attested in all cases.

| | | /vt/ | /tv/ | |
|---|---|---|---|---|
| `(rus)` | Russian | [ft] | [tv] / [tv̥] | Padgett (2002) |
| `(bul)` | Bulgarian | [ft] | [tv] / [tf] | Scatton (1993) |
| `(mkd)` | Macedonian | [ft] | [tv] / [tf] | Friedman (1993) |
| `(ces)` | Czech | [ft] | [tv] / [tf] | Hall (2003b) |
| `(slk)` | Slovak | [ft] | [tv] | Hall (2003b) |
| `(hun)` | Hungarian | [ft] | [tv] / [tf] | Kiss and Bárkányi (2006) |
| `(heb)` | Hebrew | [ft] | [tv] / [tf] | Barkai and Horvath (1978) |

The languages considered thus far all have regressive voicing assimilation, the clearest evidence for ambiguous /v/ patterning. I now turn to languages that do not have regressive voicing assimilation, but nevertheless exhibit the same kind of intermediacy of /v/.

#### 5.4.1.4 German

German contrasts /v, f/, and obstruents neutralize at the end of the prosodic word (Wagner, 2002); sonorants do not devoice.[17] German /v/, like other obstruents, neutralizes in final position and is realized as [f], as in (9).

[17]German is typically described as having syllable-final neutralization; Wagner (2002) argues that neutralization is not conditioned by the syllable, but by the prosodic word. Nothing much hinges on this analysis either way.

(9) German final devoicing of /v, z/ (Beckman et al., 2009)

a. [aktiveː] ‘active, fem. nom. sg.’

b. [aktif] ‘active’

German does not have RVA, and consonant clusters with mixed voicing are permitted across syllable boundaries. However, like sonorants, German /v/ can occur after a voiceless obstruent, yielding word-initial clusters such as /ʃv, tsv, kv/, as in k <*Quietschen*> ‘squeaks’ and /ʃvitʃən/ <*Schwitzen*> ‘to sweat’.[18]

Phonetically, Hamann and Sennema (2005) compares German /f, v/ with the labiodental inventory in Dutch, /f, v, ʋ/. Motivated by perceptual studies that showed that German listeners consistently misperceived Dutch [ʋ] tokens as German [v], Hamann and Sennema (2005) assessed the acoustic realization of these segments. Based on measures of spectral centroid on the 500 Hz high pass filter, harmonicity, and duration, Hamann and Sennema (2005) shows that German [v] tokens are produced with measures that are either intermediate (duration), or like the Dutch [ʋ] tokens. Hamann and Sennema (2005) concludes that acoustically, German [v] is an approximant.

The laryngeal neutralization facts in German are well known, and the devoicing of /v, z/ have been studied extensively; sonorants in German do not devoice in final position. Either specification of German /v/ for [sonorant] has its challenges: if assigned to [+sonorant], then the phonotactic distribution and acoustics is accounted for, but not final devoicing; if assigned [−sonorant], as is traditional, then the devoicing is accounted for, but not the

[18]Thanks to Carrie Ankerstein for these examples (personal communication, February 2017); these clusters are also discussed in Lulich (2004). Although I have not done a systematic acoustic investigation of /v/ in German, my impression of German speakers uttering these words is that there is some variability in the realization of /v/ in these sequences: some speakers clearly produce a glide [w], others a fricative [v], and still others a devoiced segment, either [f] or [ʍ]. It is not clear how much of this variability is dialectal, and whether the same speaker might produce all of these forms.

phonotactic distribution or the acoustic realization.

#### 5.4.1.5 Icelandic

In Icelandic, the voiceless and voiced labiodentals contrast in word-initial position, evidenced by the minimal pair [fauː] 'obtain' vs. [vau] 'calamity', but Árnason (2011) notes that the degree of frication of Icelandic /v/ is often so weak as to be approximant-like, even in word-initial position; nevertheless, Árnason regularly transcribes /v/ as [v] in detailed phonetic transcriptions, while simultaneously transcribing the corresponding labiodental voiced continuant in Faroese (`fao`) as [ʋ].

Regardless, if Icelandic /v/ is in fact a sonorant, then it is not surprising that it occurs after voiceless obstruents in tautosyllabic clusters, as in (10), a position otherwise reserved for /r, l, j/.

(10) Icelandic clusters with /v/ as second member (Árnason, 2011, pg. 164-165)

a. [θvoɔː] 'to wash'
b. [tʰveiːr̥] 'two'; [tvɛrkʏr̥] 'dwarf'
c. [kʰveɛːða] 'to speak'; [kvɛntʏr̥] 'Gvendur, a man's name'
d. [skvaːpi] 'fat, obesity'

In general, sonorants do not devoice when they follow voiceless obstruents, but /j, v/ are reported to undergo variable devoicing following /s/ (11), reminiscent of the variable devoicing of /v/ discussed for canonical examples of ambiguous /v/ languages.

(11) Icelandic /sv, sj/ clusters

a. [sjauː]/[s̥jauː] ‘to see’

b. [svar̥tʏr̥]/[sfar̥tʊr̥] ‘black’

Beyond onset clusters, there is evidence to treat Icelandic /v/ as a sonorant. Word-internally, clusters of the form /p, t, k, s/ + /v, j, r/ are syllabified as onsets of the following syllable, evidenced by the presence of a long vowel in the preceding syllable, hence [tʰvɪː.svar̥] ‘the evening-gen’. If [s] were the coda of the previous syllable, the /ɪ/ would be realized as short. The same phenomenon occurs when these clusters are word-final, thus /skroːkv/ ‘to lie, truncated’, though due to word-final devoicing of all word-final voiced consonants, sonorants included, the realization is [skroːkv̥], or [skroːkf].

Nevertheless, in spite of its patterning as a sonorant, there is also evidence that /v/ is less sonorous than other sonorants, and in fact exhibits some voicing alternations with /f/. In terms of clustering ability, /v/, unlike other sonorants, can appear as the first member of a consonant cluster, as in [vjɛːl̥] ‘machine’; the reverse cluster [jv] is not permitted and suggests that /v/ is less sonorous than /j/. Word finally, /v/ can occur before other sonorants, as in [tʰœvr] ‘to do magic’.[19] Finally, alternations such as [haːva] ‘to have’ and [haft] ‘had’ illustrate devoicing of /v/ before voiceless obstruents.

#### 5.4.1.6 Swedish

Swedish (`swe`) has a two-way laryngeal contrast in which lenis stops are realized as prevoiced, and fortis stops are realized as aspirated (preaspirated in initial position; either pre- or post-aspirated in medial position). Although monomorphemic voiced clusters exist (e.g., [magda] ‘Magda’), all underlying voicing discrepancies in Swedish are always resolved in favour of the voiceless obstruent, regardless of where it is in the cluster (Ringen and

[19] Thanks to Linda Ösp Heimisdóttir for these examples and a discussion in Icelandic phonology (personal communication, February 2017).

Helgason, 2004).[20]

(12) Swedish voicing assimilation; data from Ringen and Helgason (2004)

| | | | | | |
|---|---|---|---|---|---|
| a) | /ʃʷœːpa/ | [ʃʷœːpa] | ∼ | [ʃʷœːʰpa] | ‘to buy’ |
| | /ʃʷœːp+de/ | [ʃʷœːpte] | ∼ | [ʃʷœːʰpte] | (past) |
| | /ʃʷœːp+t/ | [ʃʷœːpt] | ∼ | [ʃʷœːʰpt] | ‘bought supine’ |
| b) | /vɛːga/ | [vɛːga] | | | ‘weigh’ |
| | /vɛːg+de/ | [vɛːgde] | | | (past) |
| | /vɛːg+t/ | [vɛːkt] | ∼ | [vɛːʰkt] | ‘weighed supine’ |

The laryngeal system of Swedish does not obviously bear on questions of /v/’s categorization with respect to voicing assimilation, since no voiced segments trigger voicing assimilation. The non-triggering status of /v/ could therefore either be due to the non-triggering status of voiced obstruents, or the non-triggering status of voiced sonorants. Nevertheless, the distribution of /v/ exhibits both obstruent and sonorant characteristics. As discussed by Riad (2014), /v/ can appear before liquids in onsets, as in [vrɑːk] ‘wreck’, and after liquids in codas, as in [torv] ‘turf’, patterning with other obstruents, and in contrast to other approximants in the language. Like sonorants, Swedish /v/ can appear as the rightmost member of a cluster, a position reserved for /v/ and the approximants /ʝ, l, r/, hence onset clusters such as [spl̥ɪtːra] ‘to splinter’, and [skʋ̥ɛ̞tːa] ‘to splash’ (pg. 57). Note that the devoicing of /v/ in this position cannot be taken as indicative of obstruent status: first, Riad (2014) notes that the devoicing is not sufficient to be neutralized with [f], and second, the other sonorants also devoice in this position.[21]

[20] The forms in (12) have been transcribed in full from Ringen and Helgason (2004), who provide forms in the orthography, and only transcribe the relevant consonants; the orthographic forms given are <*köpa*> and <*väga*>.

[21] Riad (2014) describes the segment /ʝ/ as either a palatal approximant or fricative, and both occur in onset position with some idiolectal variation: “Factors that seem to have a positive influence on frication

#### 5.4.1.7 Georgian

Butskhrikidze and van de Weijer (2001) describes Georgian `(kat)` consonants as divided into three groups based on differences in sonority: "stops and affricates, fricatives, and the class of sonorants" (pg. 91); she identifies /s, z, ʃ, ʒ, x, ɣ/ as the fricatives, and includes /v/ with the class of sonorants, /l, r, m, n/. The allophones of /v/ are [v, ɸ, w], detailed in (13).

(13) Allophones of Georgian /v/ Butskhrikidze and van de Weijer (2001)

a. [v]: when word-initial position and intervocalically
[viri] 'donkey'; [tavi] 'head'

b. [ɸ]: when followed by voiceless (either aspirated or glottalized) obstruents
[ɸpikrob] 'I think'; [ɸc'er] 'I write'

c. [w]: when preceded by an obstruent, especially dorsal
[q'wavili] 'flower'; [kwa] 'stone'

The distribution of allophones is reminiscent of the general patterning of /v/ in ambiguous /v/ languages, particularly because it devoices when it precedes voiceless obstruents; devoicing is not reported for Georgian sonorants. In post-obstruent position, /v/ retains its voicing and makes no demands on the laryngeal status of preceding obstruents, like other sonorants. Phonotactically, /v/ stands in the same intermediate position as in other languages. It would take us too far afield to fully delve into Georgian phonotactics, but /v/ is less sonorous than other sonorants. Georgian admits long sequences of consonants, but the order of consonants

are stress, initial position, and vowel height. When non-initial, it also seems to matter if the /j/ is preceded by a vowel or a consonant, a consonant typically inhibiting frication somewhat." (pg. 59) Although /j/ also presents some classification difficulties, it does not exhibit quite the same two-facedness: it can only appear as the rightmost consonant in onset clusters, like sonorants, but it can also appear word finally after a consonant.

is strictly determined by the ordering shown in (14).

(14) Order of consonants in Georgian clusters (Butskhrikidze, 2002, pg. 108)

| | |
|---|---|
| I | /b p p' m/ |
| II | /r/ |
| III | /d t t' j c c' ǯ č č' z s ž š/ |
| IV | /g k k' ɣ x χ'/ |
| V | /v/ |
| VI | /r l m n/ |

Butskhrikidze (2002) explains: "One consonant from each set can combine in the strict order given and form maximally a six-member cluster, e.g. /brdɣvna/ 'to fight', /prckvna/ 'to peel'. Any set can be skipped, but the order between the sets should be respected. Consonants in the sequence must have the same laryngeal specification and must be regressive. For instance, the five-member sequences are: bdɣvr-, p'rč'χv-, etc." (pg. 108). In long sequences with /r/ as the second member, /r/ in this position is syllabic, and so /v/ follows less sonorous segments, but precedes more sonorous segments, a distribution which patterns the distributional flexibility of /v/ in Russian. As mentioned previously, in post-obstruent position /v/ is realized as labialization on the preceding consonant, particularly if said obstruent is dorsal, but when /v/ follows coronal, Butskhrikidze transcribes these forms with [v], as in [tr̩tvili] 'hoar-frost'.

Butskhrikidze and van de Weijer (2001) argues that /v/ is a sonorant, and is specified only for labiality.[22] Supporting a sonorant classification, /v/ breaks up vowel hiatus, thus /me-rje-e/ → [merjeve] 'milkman', and the allophone /w/ is realized as labialization on a

[22]For example, /v/ alternates with /m/ and /o/, is lost if followed by round vowels /o, u/, and the process of *v-metathesis* is blocked if /v/ would be followed by /m/.

preceding obstruent. Nevertheless, /v/ does not easily inhabit the sonorant set either. For example, sonorants /r l m n/ can all be syllabic, as in (15), but /v/ is not reported to be syllabic.

(15) Syllabic sonorants in Georgian (Butskhrikidze, 2002, pg. 88)

| | Nom | Dat | | |
|---|---|---|---|---|
| a. | naɣm-i | naɣm-s | [naɣm̩s] | ‘mine’ |
| b. | ipn-i | ipn-s | [ipn̩s] | ‘ash tree’ |
| c. | saxl-i | saxl-s | [saxl̩s] | ‘house’ |
| d. | tetr-i | tetr-s | [tetr̩s] | ‘white’ |

To summarize, although Georgian /v/ is classified as a sonorant in Georgian, it exhibits similar intermediacy as /v/ other ambiguous languages.

#### 5.4.1.8 Summary of ambiguous /v/ patterning

Russian /v/ patterns with both sonorants and obstruents with respect to voicing assimilation, and the specific patterning of /v/ with respect to RVA is attested in other languages, both Slavic and not. In this section I have shown that the intermediacy of /v/ extends beyond the standard description of /v/ as a target and non-trigger of RVA. In particular, /v/ exhibits variable progressive devoicing in many languages, such as Czech, Polish, and Hungarian. Diagnosing the status of /v/ is complicated by the fact that in many of the “ambiguous /v/” languages, certain dialects exhibit pre-sonorant voicing across certain word boundaries.

Of particular note is that, in addition to languages like Russian and Hungarian, /v/ exhibits intermediacy in other languages as well, in terms of active phonological processes, phonotactic distribution, and allophonic realizations. In Icelandic, Swedish, Georgian, and German, /v/ exhibits devoicing either word-finally or beside voiceless obstruents (generally unlike other sonorants), and yet is frequently described as having a fairly sonorous articulation, and sonorant allophones.

In all of these languages, /v/ is a candidate for a [+sonorant, +obstruent] specification. Certainly, the brief descriptive survey provided here does not even begin to provide an in-depth phonological analyses of the voicing phenomena and phonotactics of each language individually. Nevertheless, a clear pattern emerges: in various languages, with or without regressive voicing assimilation and/or final devoicing, /v/ is intermediate between obstruents and sonorants in strikingly similar ways: like sonorants, it does not spread its voicing, but like obstruents, it readily devoices.

The fact that /v/ is so frequently intermediate ought to spur us to reconsider its default assignment as [−sonorant]. Moreover, the cross-linguistic similarities in allophone distribution should call into question another default: when a language has a segment that is transcribed as /v/ but is produced with "weak frication", a tacit assumption is that /v/ in these cases has been mistranscribed, and really ought to be considered a sonorant, likely /ʋ/.[23] The ubiquity of intermediate patterning suggests that Russian /v/ may not be a problem about Russian, but rather is a problem about /v/ itself.

In the next section I consider languages in which /v/ is standardly considered to pattern with obstruents, and are candidates for a [−sonorant, +obstruent] classification.

[23]For this latter default, I can only present anecdotal evidence: when I have discussed the patterning of *v*-like objects in different languages with various linguists, a first response to sonorant-like patterning is to make some remark of the form, "Well, maybe that's really just a /ʋ/ then".

### 5.4.2 [−sonorant, +obstruent]

The standard conception of /v/ is that it is classified as a fricative obstruent. In this section I consider cases where /v/ is reasonably classified as [−sonorant, +obstruent]. I begin with Maltese, which has regressive voicing assimilation and final devoicing. Unlike Russian, Maltese /v/ patterns as both a target and trigger for RVA.

#### 5.4.2.1 Maltese

The Maltese consonant inventory is given in Table 5.3. Borg (1997) describes the laryngeal contrast in Maltese as one of voicing, with voiceless stops being realized with little to no aspiration, but Galea (2016) describes voiceless stops as realized with aspiration (unless preceded by [s]), including in consonant clusters (thus /ptɐlɐ/ → [$\text{p}^\text{h}\text{t}^\text{h}$ɐːlɐ] 'holiday'); Borg (1997) describes voiced stops as realized with full voicing.

Maltese obstruent clusters must agree in voicing (/niktbu/ → [nígdbu] 'we write'), and voiced obstruents devoice in final position (/kiteb/ → [kítep] 'he wrote') (Borg, 1997, pg. 250). Borg (1997) notes further that "Some speakers display optional secondary voicing of historically voiceless stops in resonant environment in certain Romance loans", but does not provide any examples (pg. 250).

Table 5.3: Maltese consonant inventory

| | Labial | | Alveolar | | Palatal | | Velar | | Glottal |
|---|---|---|---|---|---|---|---|---|---|
| Stop | p | b | t | d | | | k | g | ʔ |
| Affricate | | | ts | dz | tʃ | dʒ | | | |
| Fricative | f | v | s | z | ʃ | ʒ | | | h |
| Nasal | | m | | n | | | | | |
| Lateral | | | | l | | | | | |
| Rhotic | | | | r | | | | | |
| Approximant | | w | | | | j | | | |

Various segments have been incorporated via loan words, primarily via Italian; these include /p, v, c, dz, tʃ/. Contrast between /v/, /f/, and /w/ is evidenced by the near-minimal triple [fɐːr] 'it overflowed', [vɐːrɐ] 'statue', and [wɐpt] 'during' (Galea, 2016, pg. 24). Borg (1997) notes that the high-frequency form /iːvɐ/ 'yes' has a nonstandard variant [iːwɐ] (pg. 253). In spite of its loan status, /v/ triggers voicing assimilation, as in (16).[24]

(16) Patterning of Maltese /v/

a) /ʃ + vɐnn [ʒvɐnn] 'what van'

b) /ʃ + vɛrs/ [ʒvɛrs] 'what a verse'

The patterning of /v/ in Maltese represents the "expected" patterning, based on standard assumptions that /v/ is an obstruent fricative. Nevertheless, it is worth pointing out that even Maltese /v/ may well exhibit some aspects of sonorant patterning, beyond the non-standard variant [iːwɐ] for [iːvɐ]. Borg (1997) describes a process of morphophonemic ima:la in Romance quadrilateral plurals, where the quality of the vowel in the plural depends on the quality of the stressed vowel in the singular; the quality of the vowel is irrelevant for the present purpose. Rather, consider the forms in (17), which are reproduced faithfully from Borg (1997); the first column is the singular, the second column is the plural. The forms are appear to be rendered in some mix of orthography and transcription (presumably with the exception of [r] in the form 'cap'); for example, the cluster <*bs*> in 'walking stick' is realized as [ps].

(17) Maltese Romance quadrilateral plurals

[24]Thanks to Luke Galea for these examples and confirmation of the Borg (1997) data (personal communication, June 2017).

| | | |
|---|---|---|
| a) ber[r]ítta | brí:ret | ‘cap’ |
| b) furkétta | frí:ket | ‘fork’ |
| c) čavétta | čwí:vet | ‘key’ |
| d) kappéll | kpí:pel | ‘hat’ |
| e) bastún | bsa:ten | ‘walking stick’ |

In general, the consonantal pattern is that the first CVC sequence in the singular becomes CCV in the plural, thus [furk] → [frík] in ‘fork’. The form for [cap] shows that if the second consonant is not in a cluster, it copies, thus [berí] → [brí]. Suprisingly, the plural of ‘key’ does not quite follow this pattern: instead of <*čv*>, which would be realized as [dʒv], the onset cluster is [tʃw]. Further investigation into Maltese morphophonology is required to determine whether this is a general pattern, or whether this is a singular form.

The patterning of /v/ under RVA shows that /v/ patterns unambiguously with obstruents, in spite of some sonorant-type behaviour in the ima:la data and the sonorant variant [iwɐ] for ‘yes’. Maltese /v/ is therefore specified as [−sonorant, +obstruent].

Perhaps surprisingly, Maltese is the only language that I was able to find that exhibited unambiguous obstruent patterning of /v/ under RVA. Nevertheless, other kinds of evidence bear on the obstruent classification of /v/.

#### 5.4.2.2 Dutch

Take, for example, Dutch `(nld)`, which is one of the most famous languages in the laryngeal phonology literature. Dutch contrasts voiceless and voiced stop obstruents /p, b, t, d, k, g/, as well as voiced and voiceless fricatives /f, v, s, z, ʃ, ʒ, χ, /; voiced stops are fully voiced. In addition to the labiodental fricatives /f, v/, Dutch also has voiced labiodental approximant /ʋ/; the minimal triplet /feː/ ‘fairy’ ∼ /veː/ ‘cattle’ ∼ /ʋeː/ ‘ache’ shows that all three labiodentals are contrastive (Hamann and Sennema, 2005). Nonetheless, there is significant geographical variation in the production of voiced fricatives as partially or fully devoiced, even in onset position; for details, see Pinget (2015) and references therein. There is also variation in the realization of /ʋ/; for example, in the Maastricht dialect, it is produced as /β̞/ (Gussenhoven and Aarts, 1999).

Both voiced and voiceless stops trigger RVA (18), but voiced fricatives do not. If a fricative (either voiced or voiceless) is the rightmost member of the obstruent cluster, then the entire cluster is realized as voiceless (19).

(18) Dutch voicing assimilation: stop rightmost consonant (Grijzenhout and Krämer, 2000)

a) /ʋɑs + baːr/ → [vɑz.baːr] 'washable'

b) /stoːv/ + /peːr/ → [stoːf peːr] 'stewing pear'

(19) Dutch voicing assimilation: fricative rightmost consonant (Grijzenhout and Krämer, 2000)

a) /ʋɛrk + zaːm/ → [ʋerk.saːm] 'active; effective'

b) /slaːp/ + /zɑk/ → [slaːp sak] 'sleeping bag'

Although Dutch /v/ is not a trigger for voicing assimilation, it patterns with /z/ in that both undergo progressive devoicing. The obstruent classification of Dutch /v/ is evidenced by its contrast with /ʋ/, and the difference between these two segments in their participation in voicing phenomena. The labiodental approximant /ʋ/ does not devoice word-finally, nor does it devoice when adjacent to a voiceless obstruent, while /v/ devoices word-finally and patterns with /z/ in undergoing progressive devoicing. There are also distributional restrictions: /v/ appears as the first member of a consonant cluster, while /ʋ/ only appears as the second member of clusters.

#### 5.4.2.3 A note about the /v/ ∼ /ʋ/ contrast

The contrast between /v/ and /ʋ/ is typologically very rare. Aside from Dutch, few other languages are known to have such a contrast, all Niger-Congo. One group is a set of closely related Edoid languages spoken in Nigeria: Auchi (Yekhee),[25] Edo (bin), Eruwa (erh), Ghotuo (aaa), Isoko (Uzere dialect) (iso), Okpe (oke), Oloma (olm), Uneme (une), Urhobo (urh) and Uvbie (evh). Additionally, the Bantu languages Kalanga (kck) (spoken in Zimbabwe and Botswana), and Shona (sna) (spoken in Zimbabwe) employ the /v/ ∼ /ʋ/ contrast. A notable feature of these languages is that they all have very rich labial systems: in addition to the three-way contrast /f, v, ʋ/, many of these languages also have /w/, and some even have additional labial segments. For example, Oloma also contrasts breathy voiced /v̤/.

Although none of these Niger-Congo languages have voicing assimilation, the phonological classification of /v/ as an obstruent is supported by its patterning under nasal harmony. For example, Edo, Eruwa, Okpe, Urhobo, Uvbie all exhibit nasalization of sonorants and /ʋ/, but not /v/ and other obstruents.[26] Ikwere also contains /v/, which contrasts in voicing with /f/; it also contrasts with /w/. Clements and Osu (2003) classifies Ikwere /v, z/ with obstruents that do not undergo nasalization before nasal vowels, further supporting the [+obstruent] classification of /v/.

The phonological systems of the languages presented in this section are very different from each other, but all exhibit obstruent patterning of /v/. Additionally, all of these languages also contain a contrast between /v/ and a labial approximant, whether /ʋ/ or /w/. In this

[25] Also known as a dialect of Etsako (ets).

[26] Patterning information taken from the online PBase at http://pbase.phon.chass.ncsu.edu/.

vein, the only Slavic language that is reported to have regressive voicing assimilation in which /v/ patterns with obstruents is Polish, which, unique among Slavic languages, also has /w/. Nevertheless, there are several subtleties to the Polish patterning that require comment, to which we now turn.

#### 5.4.2.4 Polish

Discussions of ambiguous /v/ languages often refer to the patterning of Polish `(pol)` as an example of /v/ patterning with other obstruents (Padgett, 2002; Kiss and Bárkányi, 2006). Kiss and Bárkányi (2006) remark that "in Polish, voiceless obstruents assimilate to /v/" (pg. 221, footnote 42), but the data are in fact more complicated than this. There are significant differences in the voicing phenomena between the two main dialects, Warsaw Polish (WP) and Cracow Polish (CP), and the patterning of /v/ differs as well; moreover, the patterning of /v/ within the prosodic word does not mirror its patterning across word boundaries.

Table 5.4: Consonant inventory of Polish; adapted from Swan (2002)

| | Labial | | Dental | | Postalveolar | | Palatal | | Velar | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stop | p | b | t | d | | | | | k | g |
| Affricates | | | ts | dz | tʃ | dʒ | tɕ | dʑ | | |
| Fricative | f | v | s | z | ʃ | ʒ | ɕ | ʑ | x | |
| Nasal | | m | | n | | | ɲ | | | (ŋ) |
| Lateral | | | | l | | | | | | |
| Rhotic | | | | r | | | | | | |
| Glide | | w | | | | | | j | | |

The inventory of Polish is provided in Table 5.4. Stops, sibilants, and affricates trigger voicing assimilation word-internally in both dialects, as seen in morphological paradigms (20), and this is mirrored in the static distributional requirements, as in (21). Gussmann notes that the voice-uniformity constraint is applied to loanwords as well, as in (22).

(20) Polish RVA

a) /mwod-ɨ/ [mwodɨ] ‘young (nom.sg.m.)’

b) /mwod-ʃ-ɨ/ [mwotʃɨ] ‘younger (nom.sg.m.)’

c) /koɕ-b-a/ [koʑ-b-a] ‘the process of mowing (nom.sg.)’

d) /koɕ-i-tɕ/ [koɕitɕ] ‘mow (infin.)’

(21) Polish obstruent clusters (Gussmann, 2007, pg. 291)

a) [ptak] ‘bird’

b) [bʒɛk] ‘coast’

c) [matka] ‘mother’

d) [pʲɛgʒa] ‘whitethroat’

(22) Polish voicing requirement applying to loans (Gussmann, 2007, pg. 291)

a) [fudbɔl] ‘football’

b) [raktajm] ‘ragtime’

When /v/ precedes a voiceless obstruent, it devoices to [f], as in (23).

(23) Polish /v/ target of RVA (Hall, 2003b, pg. 109)

a) /v-gdɨɲi/ [v-gdɨɲi] ‘in Gdynia (loc.sg.)’

b) /f-polst͡se/ [f-polst͡se] ‘in Poland (loc.sg.)’

Hall (2003b) reports that in rightmost position, /v/ undergoes progressive devoicing, as in (24), but Gussmann (2007), notes that there is variability in the realization of /v/ following voiceless obstruents, present both in morphological alternations, as in (25) and in static distributions, as in (26). Gussmann notes that “The variants which violate voice uniformity cannot be connected today with any regional variety of the language but rather they appear to be individual or idiolectal" (pg. 308).

(24) Polish /v/ target of PVA (Hall, 2003b, pg. 109)

a) [dva] 'two'

b) [tfuj] 'your (nom. sg. m.)'

(25) Variable PVA of /v/ in Polish (alternations) (Gussmann, 2007, pg. 308)

a) [ʃɛvɛk] 'seam, dim.' [ʃfɨ] ∼ [ʃvɨ] 'seam, nom. sg.'

b) [tsɛrcɛvnɨ] 'Orthodox, nom. sg.' [tsɛrkfʲi] ∼ [tsɛrkvʲi] 'Orthodox church, gen. sg.'

(26) Variable PVA of /v/ in Polish (distribution) (Gussmann, 2007, pg. 308)

a) [dva] 'two' [tfuj] ∼ [tvuj] 'your (nom. sg. m.)'

b) [dʐvʲik] 'crane' [tʃfartɛk] ∼ [tʃvartɛk] 'Thursday'

Sonorants do not devoice in this environment, and so the treatment of Polish /v/ as the voiced counterpart to /f/ in voicing assimilation is not entirely accurate. Moreover, in realizations where /v/ does not devoice, it patterns exactly as Russian /v/ with respect to voicing assimilation, devoicing before voiceless obstruents and failing to trigger voicing in presonorant position.

The characterization of Polish /v/ as an obstruent comes from its patterning in word-initial position in the Warsaw dialect where, across word boundaries, it triggers voicing assimilation, and does not undergo variable devoicing (Gussmann, 2007).[27]

(27) RVA triggered by /v/ in Warsaw Polish across boundaries (Gussmann, 2007, pg. 309)

[27] In discussing loanwords, Gussmann (2007) also provides the example of 'Nashville', pronounced [nɛʒvʲil], and does not identify that there is any progressive devoicing of /v/ in this form.

| | | | |
|---|---|---|---|
| a) | <*gotów pisać*> | [gɔtuf p$^{j}$isaɕ] | ‘ready to write’ |
| | <*gotów drukować*> | [gɔtuv drukɔva] | ‘ready to print’ |
| b) | <*smak wina*> | [smag v$^{j}$ina] | ‘taste of wine’ |
| | <*los wygrany*> | [lɔz vɨgranɨ] | ‘winning number’ |

A complication arises in Cracow Polish, where all voiced consonants, obstruents, sonorants and vowels alike, trigger voicing assimilation in this context (Bethin, 1984; Gussmann, 2007; Cyran, 2012). The examples in (28) show how Warsaw Polish (WP) and Cracow Polish (CP) differ with respect to the triggers of voicing assimilation at word boundaries. In CP, therefore, the obstruent status of /v/ cannot be readily determined, since triggering voicing assimilation is a property of all voiced segments.

(28) Voicing across word boundaries: Warsaw vs. Cracow Polish (Cyran, 2012, pg. 154)

| | | WP | CP | |
|---|---|---|---|---|
| a) <*bra**k** **o**ceny*> | ‘lack of mark’ | [k ɔ] | [g ɔ] | ___ V |
| b) <*bra**k** **j**asności*> | ‘lack of clarity’ | [k j] | [g j] | ___ S |
| c) <*bra**k** **w**ody*> | ‘lack of water’ | [g v] | [g v] | ___ $C^{+v}$ |
| d) <*bra**k** **p**ieczątki*> | ‘lack of stamp’ | [k p] | [k p] | ___ $C^{-v}$ |
| | | | | |
| e) <*obra**z** **a**nioła*> | ‘picture of angel’ | [s a] | [z a] | ___ V |
| f) <*obra**z** **m**istrza*> | ‘picture of master’ | [s m] | [z m] | ___ S |
| g) <*obra**z** **b**urzy*> | ‘picture of storm’ | [z b] | [z b] | ___ $C^{+v}$ |
| h) <*obra**z** **cz**łowieka*> | ‘picture of man’ | [s t͡ʃ] | [s t͡ʃ] | ___ $C^{-v}$ |

To summarize: Polish /v/ devoices word-finally and before voiceless obstruents in all dialects. In word-internal clusters of a voiceless obstruent followed by /v/, /v/ is variably realized either as [v] (thereby replicating the asymmetry in Russian) or as [f] (undergoing progressive devoicing). Across word-boundaries, all voiced segments trigger voicing assimilation in Cracow Polish, but only voiced obstruents, including /v/, trigger voicing assimilation in Warsaw Polish. Therefore, only across word-boundaries in Warsaw Polish does /v/ pattern unambiguously as a phonological voiced obstruent with respect to voicing assimilation.

What do these data tell us about the classification of /v/ in Polish? The patterning of /v/ in Cracow Polish is the same patterning seen in Russian, Czech, and other Slavic languages with ambiguous /v/, with the exception of pre-sonorant voicing. Hence, /v/ in Cracow Polish does not appear to be a true obstruent, and is a candidate for being specified as [+sonorant, +obstruent]. A full analysis of Cracow Polish voicing phenomena, particularly pre-sonorant voicing, is beyond the scope of this work.

Warsaw Polish presents an interesting challenge with respect to the patterning of /v/, since it patterns "ambiguously" word-medially, but patterns with obstruents in triggering voicing assimilation across word-boundaries. Based on its patterning within words, I propose that /v/ is generally specified as [+sonorant, +obstruent]. In order to account for the patterning of /v/ as a trigger in word-initial position I hypothesize that word-initial /v/ is produced with sufficient fortition that it is interpreted as purely [−sonorant, +obstruent], and thus triggers voicing assimilation. This analysis is plausible given the results of Chapter 2: in WIS position, /v/ generally exhibited more frication than in the WMU environment, a result that held across all languages. The interpretation of /v/ as a true obstruent in word initial position is likely facilitated by its contrast with /w/; nevertheless, to verify this hypothesis, a detailed phonetic study of Polish /v/ and voicing assimilation is required, and left to future research.

Of the cases looked at so far, the only languages that exhibit unambiguous patterning that is consistent with a [−sonorant, +obstruent] classification of /v/ is Maltese and Dutch based on voicing assimilation, and Ikwere and other Niger-Congo languages based on nasal harmony. All of these languages further exhibit a contrast between /v/ and a labial approximant. To what extent is the presence of a labial approximant a necessary and/or sufficient criterion for an obstruent classification of /v/? The patterning of /v/ in Polish shows that the inclusion of /w/ in an inventory is not a sufficient criterion for obstruent patterning, since Warsaw Polish /v/ patterns ambiguously within words. As I show later, the patterning of /v/ in Greek shows that the presence of /w/ is also not necessary. Before addressing the Greek case, however, I wish to highlight a case of /v/ patterning that is typically included among the ambiguous /v/ languages, Hebrew.

#### 5.4.2.5 Hebrew

Hebrew `(heb)` /v/ also patterns ambiguously with respect to voicing assimilation, seen in (29). Hebrew /v/ derives historically from two sources in Biblical Hebrew: merger with /w/, and alternation (due to spirantization) with /b/ (Bolozky, 1978).[28]

(29) Hebrew /v/ (Barkai and Horvath, 1978)

| | | | | |
|---|---|---|---|---|
| a) [sagar] | ‘he closed’ | [lizgor] | ‘to close’ | (/s/ → [z]) |
| b) [hitnageʃ] | ‘he collided’ | [hitnakʃut] | ‘collision’ | (/g/ → [k]) |
| c) [ʃevet] | ‘tribe’ | [ʃifti] | ‘tribal’ | (/v/ → [f]) |
| d) [kvuca] | *[gvuca] | ‘group’ | | |

Bolozky (1978) describes regressive voicing assimilation in Modern Hebrew as optional, but obligatory in fast speech. Like other ambiguous-/v/ languages, Bolozky notes that

[28] The connection to between /v/ and /b/ is considered important in light of analyses of Russian /v/ that posit that [v] is underlyingly sonorant /w/, mirroring the historic development of /v/ from *w*, as in **?**.

progressive voicing assimilation of /v/ may occur, yielding variants such as [kfar] alongside the standard [kvar] 'already'. In fast speech, /v/ can trigger voicing assimilation, yielding [gvar], an observation that he repeats in Bolozky (1997).

Bolozky (1997) attributes the patterning of Hebrew /v/ to an influence from Slavic, noting the parallel in optional devoicing of /v/ in Polish and Hebrew:

> It appears that *v* not causing voicing assimilation is due to substratum influence of Slavic languages such as Russian and Polish, where (owing to *v* originating from a historical glide *w*) the same phenomenon is observed (and furthermore, where at least intra-morphemically in Warsaw Polish, **progressive** assimilation occurs instead—which may also surface in the speech of some Polish-born Israelis). (pg. 292; emphasis in the original)

In later papers, Bolozky notes that while voicing assimilation is still optional, dependent somewhat on the casualness of speech, it is "especially productive before stops, and when a voiced obstruent precedes a voiceless one in an initial cluster…or when the articulatory distance between points of articulation is maximal" (Bolozky, 2013). Furthermore, the previously exceptional patterning of /v/ has regularized and triggers voicing assimilation (Bolozky, 2006, 2013).[29] Possibly relevant is that Hebrew has borrowed many words with English /w/. Although /w/ has not spread beyond proper names, only older speakers pronounce English words such as *Washington* with initial [v], and the same demographic pronounces /kvar/ as either [kvar] or [kfar]; younger speakers tend to pronounce /kvar/ as [gvar], and *Washington* with initial [w].[30] The hypothesis that the Hebrew phonological system is undergoing at least a partial reorganization is interesting in light of the fact that, of all the Slavic languages, only Polish (which also has [w]) exhibits obstruent patterning of /v/, as with Maltese; of course, without in-depth sociophonetic study, such a hypothesis can only be speculative.

[29]Similarly, the exceptional patterning of /x/ as not undergoing voicing assimilation has also regularized.
[30]Thanks to Shmuel Bolozky for discussion on the Hebrew facts (personal communication, March 2017).

#### 5.4.2.6 Greek

Recall that in Section 3, Greek was selected as a language in which /v/ patterns with obstruents, even though it did not have /w/ or any other labial approximant against which /v/ could contrast. However, the determination of Greek /v/'s obstruent status was based primarily on phonotactic considerations. In Greek, /v/ is a reflex of */b/, and /f/ is a reflex of */$p^h$/, meaning that distributional facts about /v, f/ reflect their history as obstruents. Finding evidence for the obstruent classification of /v/ in terms of its synchronic active phonology is not easy, since Greek syllables are generally open, and the only segments that can end a word in the native lexicon are /n, s/. Furthermore, all voiced segments can voice a preceding /s/ across word boundaries, meaning that there is no process of regressive voicing assimilation triggered only by obstruents that we can use as a diagnostic for /v/'s classification. Baltazani (2007) shows that Greek "s-voicing" is gradient rather than categorical, and depends both on place and manner of articulation: "there is more assimilation before homorganic segments and also there is more assimilation before voiced obstruents than before sonorants" (pg. 10). Unfortunately, /v/ and /z/ were not included in the study,[31] and so it is not known if /v/ triggers s-voicing like obstruents, like sonorants, or as intermediate. This is a question left for future research.

In Chapter 2, /v/ exhibited the same relationship to /f/ as /z/ did to /s/, showing that type of frication did not have an effect on the relationship between voiced and voiceless members of a pair, but again this does not directly address the question at hand. In Chapter 3, Greek /v/ consistently had the highest relative spectral centroid, and was significantly different from Serbian /v/ in both environments.[32] To summarize, classifying Greek /v/

[31]The possible triggers for s-voicing that were included in the study were the voiced stops /b, d, g/ and the sonorants /l, m/.

[32]Russian /v/ was not significantly different from Greek in the WIS environment, but not significantly different from Serbian in the WMU environment.

with respect to [obstruent] and [sonorant] is not clear. Phonetically, it appears that /v/ is a true voiced fricative. The phonotactics of Greek and the distribution of /v/ facilitate a more obstruent realization, as Greek is characterized by open syllables and /v/ only occupies the first position of a consonant cluster. Phonologically, the evidence is somewhat weak. Based on the phonetic realization and phonotactic evidence, Greek /v/ appears to be a candidate for a [−sonorant, +obstruent] classification.

If the classification of Greek /v/ as [−sonorant, +obstruent] is correct, then the presence of a labial approximant against which /v/ contrasts is not a necessary condition for the obstruent classification of /v/. The patterning of word-internal /v/ in Polish, which contrasts /v/ with /w/ shows that it is also not a necessary condition. The obstruent patterning of /v/ in Greek and Polish therefore suggest that the presence of surface contrast with /w/ is neither a necessary nor sufficient condition to ensure obstruent patterning of /v/.

Nevertheless, the presence of a labial approximant appears to make the unambiguous obstruent classification of /v/ much more likely, both phonetically and phonologically, suggesting that perhaps a weaker hypothesis is plausible. Based on my own observations perusing grammars and segment databases, there appears to be a strong correlation between the description of /v/ and whether or not a consonant inventory also contains a labial approximant. In languages without a labial approximant, /v/ is typically described as a weak fricative with sonorant-like realizations (where variants are described as free or as conditioned by position); in languages with a labial approximant, /v/ is described as a fricative with no suggestion that /v/ is anything other than an obstruent. This suggests that contrast plays at least some role in determining the realization and phonological classification of /v/. Disentangling the precise role that both surface and underlying contrast has on the realization of /v/ and other voiced spirants is left to future research.

#### 5.4.2.7 Summary of obstruent /v/ patterning

To summarize, it is possible for /v/ to exhibit an obstruent classification, either due to its phonetic realization or due to its phonological patterning, but the situation obtains much more rarely than previously assumed. Of all the languages that I have examined with regressive voicing assimilation, only Maltese exhibits unambiguous obstruent patterning of /v/, meaning that /v/ participates as a trigger for RVA, together with other voiced obstruents, and unlike sonorants. In all other RVA systems that I have come across, the exclusive classification of /v/ as an obstruent is undermined either by its variability or a lack of synchronic phonological evidence that distinguishes obstruents from sonorants. In systems that do not have RVA, /v/ tends to pattern as an obstruent in blocking nasal harmony, while labial approximants undergo nasalization. Furthermore, the classification of /v/ as an obstruent is heavily favoured in systems that contrast it with a labial approximant. The important take-away from this section is that although /v/ is in many cases realized as and patterns as an obstruent, the unambiguous classification of /v/ as [−sonorant, +obstruent] is much rarer than previously assumed.

I now turn to the last of the logical possibilities for /v/.

### 5.4.3 [+sonorant, −obstruent]

In other Slavic languages, like Serbian described in Section 1.5.2, /v/ is reported to pattern with sonorants. Languages in this category exhibit symmetric voicing assimilation of the stops and sibilants, but /v/, like sonorants, is neither a trigger nor a target for regressive voicing assimilation. In spite of the non-participation of /v/ in RVA, the patterning of /v/ in many of these languages is subject to significant positional variability.

#### 5.4.3.1 Slovene

Consider, by way of example, the patterning of /v/ in Slovene `(slv)`, which is rarely described as having a fricatival realization, but exhibits a similar allophony pattern as /v/ in Russian. Herrity (2000) describes four other surface forms of /v/, all of which are sonorant. When /v/ is in the rightmost position of the onset (i.e, immediately preceding a vowel or vocalic /r̩/), it is realized as a fricative. In coda position, it is realized as the offglide [u̯], as seen in (30); this realization is also seen when word-initial /v/ follows a vowel-final word, an apparent case of resyllabification.

(30) /v/ → [u̯] in Slovene

| | | |
|---|---|---|
| a) /ʒiːv/ | [ʒiːu̯] | ‘alive’ |
| b) /ɔːvca/ | [ɔːu̯tsa] | ‘sheep’ |
| c) /bi vstaːli/ | [biu̯staːli] | ‘(they) would have got up’ |

Word-initial /v/ followed by a voiced consonant is realized as [w], as is word-final /v/ when it follows a sonorant consonant, like vocalic /r̩/, as in (31). This realization of /v// is also found when /v/ is between two consonants, when either both of them are sonorants, or neither are. Herrity notes that there is variability between [w] and [u].

(31) /v/ → [w]

| | | |
|---|---|---|
| a) /vzeːti/ | [wzeːti] | ‘to take’ |
| b) /tʃr̩v/ | [tʃr̩w] | ‘worm’ |
| c) /skrvjoː/ | [skrwjoː] | ‘with blood’ |
| d) /odvzeːti/ | [odwzeːti] | ‘to take away’ |

When /v/ is followed by a voiceless consonant, it is realised as [ʍ] (32). This realization is interesting because although Slovene /v/ is not considered a counterpart to /f/ in the obstruent system, it is devoiced in precisely the same environment as ambiguous /v/.

(32) /v/ → [ʍ]

| | | |
|---|---|---|
| a) /vsaːk/ | [ʍsaːk] | 'every' |
| b) /predvsɛm/ | [pretʍsɛm] | 'the day before yesterday' |

Again, Herrity notes that there is variability between [ʍ] and [u], possibly due to morphological considerations: e.g., when /v/ is a preposition, it tends to be realized as [u], as in /v piːvu/ > [u piːvu] 'in the beer', though again there is variability between [u, w, ʍ].

While all realizations of underlying /v/ are sonorant, when /f/ is realized as [v] due to voicing assimilation, it is always realized as a voiced fricative [v], and never as one of the aforementioned sonorant realizations; thus, /ʃeːf govoriː/, 'the boss says', is [ʃeːv govoriː], not *[ʃeːu̯ govoriː].

The Slovene facts are interesting because the allophones of Slovene /v/ are more sonorant versions of the realizations of Slovak /v/. In both Slovak and Slovene, /v/ is realized as a sonorant in coda position and as devoiced before voiceless obstruents. After obstruents, Slovak /v/ is realized as [v], and Slovene /v/ is realized as [w]. In both languages, /v/ is realized as [v] prevocalically in simple onsets; in Slovak sequences like /tv/ also are realized with [v].

I propose that, although Slovene /v/ is part of the obstruent system, it is also not wholly part of the sonorant system, since other sonorants do not have such a wide range of realizations, nor are any of them reported to devoice before voiceless obstruents. The patterning of Slovene /v/ is repeated in Latvian and Lithuanian as well, to which I now turn.

#### 5.4.3.2 Latvian and Lithuanian

The Baltic languages of Latvian (`lav`) and Lithuanian (`lit`) exhibit the Slovak and Slovenian pattern. Both languages have voicing assimilation of obstruent clusters, and both have /v/, though the precise realization appears to differ between the two. Mathiassen (1997) notes for Latvian that "[v] is regarded as a fricative, not an approximant…It differs from [f] only with respect to the feature [+voiced]" pg. 23; in his Lithuanian grammar, Mathiassen (1996) notes that the Lithuanian /v/ and /vʲ/ are "approximants (and consequently not the voiced counterparts of the fricatives [f], [fʲ])" (pg. 23). In spite of these different descriptions, by the same author, the patterning of /v/ and its palatalized counterpart is described identically in the two grammars: in both languages the rightmost voiced obstruent triggers voicing assimilation of the entire cluster, except /v/: thus we find [svaiks] 'fresh', not *[zvaiks], in Latvian (Mathiassen, 1997, pg. 26), and [ʃʲvʲentas] 'sacred', not *[ʒʲvʲentas], in Lithuanian. Mathiassen does not give examples of /v/ before voiceless obstruents, but Kariņš (1996) describes different realizations of /v/ depending on its position in the syllable: when /v/ is in the onset, it is realized as a fricative, as in words such as [viːrs] 'man', but when /v/ is in the coda, it is vocalized as an offglide [w], as shown in (33). When coda /v/ precedes a vowel, it resyllabifies as the onset of the following syllable and is pronounced as [v].

(33) Realization of Latvian /v/; repeated from Kariņš (1996, pg. 13)

| base form | pronunciation | gloss |
|---|---|---|
| tev-i | [tevi] | ‘you acc sg’ |
| tev-is | [tevis] | ‘you gen sg’ |
| tev | [tew] | ‘you dat sg’ |
| nav | [naw] | ‘3rd does not have’ |
| tev nav | [tew nau] | ‘you sg. do not have’ |
| tev ir | [te.vir] | ‘you sg. have’ |

Ambrazas et al. (2006) notes that in Lithuanian both /v, j/ are vocalized as [w, i] word-finally and before a consonant (pg. 48), and that [w] is a possible realization of /v/ intervocalically.

Although I am not aware of any detailed phonetic studies on the Baltic languages, at least not in the English literature, the data presented here are interesting in light of the fact that the same author, writing grammars of each language, provides different descriptions of their phonetic realization, with Latvian having [v], while Lithuanian has [ʋ]. In spite of these differences, /v/ patterns identically, and like the sonorantizing Slavic /v/ languages.

Finally I turn to Ukrainian, which is also reported to have sonorant /v/, but the system of voicing assimilation differs from the remaining Slavic languages in that only voiced obstruents trigger voicing assimilation.

#### 5.4.3.3 Ukrainian

Ukrainian (`ukr`) is unique among Slavic languages in that only voiced segments trigger voicing assimilation (Butska, 1998). In (34), we see that /b/ triggers voicing of preceding /s/ to [z], but /k/ does not trigger devoicing of /b/ to [k].

(34) Ukrainian voicing assimilation; data from Butska (1998)

| | | | | |
|---|---|---|---|---|
| a) | [prosɪ-tɪ] | 'request, inf' | [proz$^j$ba] | 'request, nom. fem.' |
| b) | [korob-ok] | 'box, gen. pl.' | [korob-ka] | 'box, dim. nom. fem.' |

An important exception to this patterning is when single-consonant, word-initial prefixes are added, such as /z/ 'with': in this case, the prefix is devoiced, as in /z toboju/ [stoboju] 'with you'; Bethin (1984) argues that in this case, the prefix is syllabified with the following voiceless obstruent.

Ukrainian /v/ does not pattern with other voiced obstruents in triggering voicing assimilation. Like Russian, clusters of the form voiceless stop + [v] are attested, as in [morkva] 'carrot'. Because voiced obstruents do not generally undergo devoicing, there are no alternations between [v] and [f]. In syllable-final position, /v/ is realized as a glide [w]. Bethin (1984) takes these facts to suggest that /v/ in Ukrainian is specified as [+sonorant].

#### 5.4.3.4 Summary of sonorant /v/ patterning

The languages presented here show that although /v/ does not pattern with obstruents, its realization under the conditions for voicing assimilation differ from those of other sonorants as well. In particular, while liquids, nasals, and glides have relatively stable realizations across environments, the realization of /v/ varies significantly. Consider, for example, Table 5.5, which summarizes the realizations reported for /v/ in the environments relevant for voicing assimilation phenomena: word-final position, before (voiceless) obstruents, and after (voiceless) obstruents; the voicing value of adjacent obstruents is not, generally relevant, as the realization of /v/ is described in terms of syllable position. As can be seen in Table 5.5, /v/ is reported to have quite sonorant realizations in these positions, ranging from an offglide [u̯], to labiovelar [w], to labiodental approximant [ʋ].

Table 5.5: Sonorant /v/ realizations.

| | | /v/# | /vt/ | /tv/ | |
|---|---|---|---|---|---|
| (hbs) | Serbo-Croatian | N/A ([v]) | [vt] | [tv] | Browne (1993) |
| (ukr) | Ukrainian | N/A ([w]) | [wt] | [tʋ] | Shevelov (1993) |
| (bel) | Belarusian | [w] | [wt] | [tv] | Mayo (1993) |
| (slv) | Slovene | [u̯] | [ʍt] | [tw] | Herrity (2000) |
| (lav) | Latvian | [w] | [wt] | [tv] | Kariņš (1996) |
| (lit) | Lithuanian | [w] | [wt] | [tʋ] | Mathiassen (1996) |

Table 5.5 exhibits across-the-board sonorant realizations; although Serbian is transcribed above as /v/ based on Browne (1993), as we saw in Chapter 2, Serbian /v/ is systematically realized with a shorter duration and weak frication, suggesting that Serbian /v/ ought to be transcribed as [ʋ]. This naturally raises the question as to whether any of these languages have a [v] at all.

Descriptions acknowledging the sonorant-like realizations of /v/ are common to the literature on all these languages; also common to them all is the treatment of this segment as /v/. Some grammars explicitly acknowledge this inconsistency, but generally justify the choice to represent the segment as a fricative because in at least some contexts, particularly word-initial, it exhibits enough frication for the authors to count the segment as [v] sometimes, and hence as /v/.

Consider, for example, the findings of Chapter 2 with respect to Serbian; the relationship between frication and voicing showed that Serbian [v] was consistently "outside" the system of fricatives, as the interaction between frication and voicing was highly significant independent of environment, and the spectral centroid of [v] compared to [f] did not parallel that of [z] to [s]. Nevertheless, Serbian did exhibit some frication, and more in the WIS environment than in the WMU environment. In the WIS environment, the median spectral centroid value for Serbian [v] was approximately 4500 Hz, and in the WMU environment around 4000 Hz;

these are not the spectral centroid values of a pure sonorant with just formant structure and no frication.

Returning to the phonological identity of /v/ in these languages, it is clear that, regardless of whether /v/ is realized with some frication in some environments, /v/ patterns with sonorants with respect to voicing assimilation phenomena, and suggests that /v/ in these languages be specified as [+sonorant, −obstruent].

### 5.4.4 [−sonorant, −obstruent]

Specifically discussing the phonetics, phonotactics, and phonology of Russian /v/, I rejected the classification [−sonorant, −obstruent] in Section 5.3. The question is, is [−sonorant, −obstruent] impossible in principle? Based on the phonetic definitions of [obstruent] and [sonorant], yes. Any segment that could reasonably be considered 'v'-like—whether in fact [v, ʋ, w, β, β̞, f]—will exhibit either the acoustic traits of sonorants, the increase in supraglottal pressure of obstruents, or both of these simultaneously.

Phonologically, however, there are languages in which it is reported that /v/ is neither an obstruent, nor a sonorant. Botma and van't Veer (2014) investigate the class of voiced spirants /β, v, ð, ɣ/, and argue that in many cases, voiced spirants are more appropriately viewed as sonorants rather than as fricatives, based on their typological identity, phonetic properties, and phonological patterning. They investigate 70 languages selected from the PBase (Mielke, 2008) that have unpaired voiced spirants; that is, inventories that have any of /β, v, ð, ɣ/, without the corresponding voiceless spirant /ɸ, f, θ, x/. For each language, they identify cases where the voiced spirant patterns with both obstruents and sonorants, just sonorants, just obstruents, or neither obstruents nor sonorants. I focus just on the 29 languages with unpaired /v/; of these, /v/ patterns with both obstruents and sonorants in 7 languages, with just obstruents in 1 language, with just sonorants in 5 languages, and with neither obstruents nor sonorants in 16 languages. The PBase, from which these languages are selected, is not geographically or geneologically balanced, and so I do not try to interpret the meaning behind the number of languages in each category.[33] The list of these languages is given in (35), organized by patterning.[34]

(35) Reported patterning of /v/ in Botma and van't Veer (2014)

a. **/v/ patterns with both obstruents and sonorants**

Altaic: Evenki
Austronesian: Nalik
Dravidian: Kuvi; Malayalam
Eskimo-Aleut: Inupiaq
Indo-European: Marathi
Mixe-Zoque: Mixe

b. **/v/ patterns with obstruents**

Austronesian: Sie

c. **/v/ patterns with sonorants**

Dravidian: Kolami; Koraga, Mudu
Mixe-Zoque: Popoluca
Sino-Tibetan: Mikir
Trans-New Guinea: Daga

d. **/v/ patterns with neither obstruents nor sonorants**

Austronesian: Banoni; Kedang; Kiribati; Rapanui; Wolio
Dravidian: Gondi, Adilabad; Gondi, Koya; Irula; Kui
Hokan: Mojave; Yavapai

[33] For example, Adilabad and Koya are two closely related dialects of Gondi, a Dravidian language, so likely should not count as two distinct data points.

[34] Two of the languages listed additional unpaired voiced spirants: Nalik (Austronesian) also has /β/; Sie (Austronesian) also has /ɣ/.

Mayan: Tzotzil
South Caucasian: Georgian
Uto-Aztecan: Chemehuevi; Pima Bajo; Tepehuan

What is the status of /v/ in the last category? Botma and van't Veer (2013) state, "Languages in which offending fricatives pattern neither with obstruents nor with sonorants provide no evidence for their phonological status" pg. 51. That there is no evidence for the classification of /v/ because there is no available diagnostic phonological phenomenon is very different from assigning a specification of [−sonorant, −obstruent].

To illustrate, consider Rapanui (`rap`), an Austronesian language spoken on Easter Island. The following description is based on Du Feu (1996). There are only 10 consonants in Rapanui: voiceless stops /p, t, k/; nasals /m, n, ŋ/; "single flapped apico-alveolar [r]; a glottal stop [ʔ]; a "glottal-approximant fricative" /h/, which is voiced intervocalically, and a "labio-labial voiced fricative" [v], which is "strongly labialized" (Du Feu, 1996, pg. 182-183). Ignoring /v/, the distinction between sonorants and non-sonorants aligns perfectly with the distinction between voiced and voiceless consonants. There are no word-final consonants, nor are there any consonant clusters; syllable structure is strongly (C)V. As far as I can tell, there are no processes that appear to rely on the distinction between sonorant-type voicing and obstruent-type voicing, or even any kind of voicing at all. Given these facts, obstruent vs. sonorant is not relevant to Rapanui phonology, and hence assigning [−sonorant, −obstruent] to Rapanui /v/ is inappropriate given the phonology of the language.

To take a second example, there is no phonological evidence to distinguish obstruents from sonorants in Banoni (`bcm`), another Austronesian language, spoken on Papua New Guinea. Unlike Rapanui, Banoni does contrast voiceless and voiced stops /p, t, k, b, d, g/; voiced stops are completely oral and not prenasalized, a feature which is unusual for languages of the

Solomon Islands (Lincoln, 1976b). The segment transcribed as /v/ represents [β], [w] [β͡ɣ], but in some varieties also represents [b] or [m] (Lincoln, 1976a).[35] Nevertheless, although Banoni contains a voicing contrast in the obstruents and contains voiced sonorants, there is again no process that I was able to uncover that distinguished obstruents from sonorants, and so /v/ also not categorized as [−sonorant, −obstruent].

Based on the phonetic definitions of [sonorant] and [obstruent] I propose that [−sonorant, −obstruent] is universally unavailable for continuants, unlike what is possible for non-continuants. This reflects a major difference in the way that sonority, voicing, and continuancy interact: non-continuants instantiate three of the four logical possibilities—[+sonorant, −obstruent] (nasal stops), [−sonorant, +obstruent] (explosive stops), and [−sonorant, −obstruent] (non-explosive stops)—but [−sonorant, −obstruent] is not available for continuants.

### 5.4.5 Summary of the typology of featural specifications of /v/

Adding the feature [obstruent] together with [sonorant] yields four logical possibilities. Clements and Osu (2002) showed that stops instantiate three of the four possibilities, omitting [+sonorant, +obstruent]. In this section I have shown that /v/ patterns in a way that is consistent with three of the four possibilities, omitting [−sonorant, −obstruent]. Based on the phonetic definitions of [sonorant] and [obstruent], I argued that supralaryngeal voiced continuants cannot be specified as [−sonorant, −obstruent].

Cross-linguistically, /v/ appears to readily occupy a place that is less sonorous than other sonorants and more sonorant than other obstruents. The intermediacy of /v/ is instantiated most obviously in languages like Russian, Hungarian, and other "ambiguous /v/" languages,

[35] According to Lincoln (1976b), /β/ and /ɣ/ merge to labiovelar fricatives before back rounded vowels.

but is also found elsewhere, in languages without RVA. The intermediacy of /v/ is instantiated distributionally, but also exhibits allophony that recalls the distribution of allophones in ambiguous /v/ languages.

The only language that I am aware of to exhibit exclusive obstruent patterning of /v/ with respect to RVA is Maltese. Nevertheless, there are languages in which obstruent classification of /v/ can be argued for based on other factors; for example, either distributional facts, such as Greek, or its patterning under nasal harmony, as in Ikwere. In interesting observation worth pursuing is that many languages that exhibit unambiguous obstruent classification of /v/ also contain either /w/ or /ʋ/, suggesting that contrast plays an important role in more unambiguous classifications of /v/.

The final possible configuration is sonorant /v/ languages, in which /v/ does not pattern either as a trigger or a target to RVA. Grammars for these languages typically report that /v/ is realized as [v] in prosodically strong positions, but have sonorant allophones either in coda position or after consonants. The non-participation of /v/ in languages such as Serbian suggests that a [+sonorant, −obstruent] specification is appropriate. Nevertheless, /v/ does not appear to rest wholly within the sonorant system either, as other sonorants do not exhibit the kind of allophony exhibited by /v/. Furthermore, in some languages, such as Slovene, the sonorant allophones appear to also have devoiced variants that occur in precisely the same environments that /v/ devoices to [f] in Russian. There are two possible explanations for this. One solution is that /v/ is indeed specified as [+sonorant, −obstruent] in all these languages, and any patterning that is reminiscent of ambiguous /v/ is chalked up to default phonetic implementations. Another possibility is that in languages like Slovene, /v/ is in fact ambiguous and featurally specified as [+sonorant, +obstruent], but due to language specific implementation, all allophones are simply more sonorant versions than those in a canonical ambiguous /v/ language. Resolving this issue is beyond the scope of the present work, and

I leave this to future research.

In sum, the present section showed that three of the four logical possibilities of [obstruent] and [sonorant] are instantiated for /v/ with respect to voicing assimilation and its phonotactic distribution. There is, to my knowledge, no language that instantiates a three-way contrast within any given language. Languages that contrast two of the the three options, specifically [−sonorant, +obstruent] and [+sonorant, −obstruent] are rare and, if Dutch is any indication, the contrast is not stably maintained as one of manner or sonority, but also enhanced with devoicing. Clements and Osu (2002) uses this fact as an argument against allowing [+sonorant, +obstruent], but collapses "laxly-articulated voiced fricatives such as *z* and *v*", given the assumption that sibilants and spirants pattern as a homogenous group. However, as seen in the previous three chapters, /v/ and /z/ do not always pattern as a unified class. In fact, we have seen, phonetically, typologically, and now in terms of phonological patterning, that in some languages /v/ patterns with /z/ as an obstruent, in others it patterns distinctly from /z/ as a sonorant, and in still others it patterns intermediately.

Rather than maintain the universal ban on [+sonorant, +obstruent] suggested by Clements and Osu (2002), I propose that the different manifestations of [obstruent] and [sonorant] across continuants and stop reflect the interaction between sonority/obstruency and manner of articulation. The absence of a three-way contrast between the three kinds of /v/ likely reflects the articulatory and perceptual difficulties in maintaining a three-way contrast between a sonorant /v/, obstruent /v/, and sonorant obstruent /v/; as discussed previously, the contrast between /v/ and /ʋ/ is rare, and potentially requires more extensive cueing, given the devoicing of /v/ in Dutch.

The typology presented in this section focussed on /v/. How does this typology relate to the general understanding of voicing assimilation in a cross-linguistic perspective? In the following section I show that, by not focussing on the problem of /v/, existing typologies of voicing assimilation are necessarily incomplete.

## 5.5 Typologies of voicing assimilation

The two primary phenomena of interest have been Final Neutralization (FN) and Regressive Voicing Assimilation (RVA). The four-way typology established by the presence or absence of these two phenomena is attested, as summarized in Table 5.1 with an example of each.[36]

Table 5.1: Typology of presence of FN and RVA

| | | RVA | |
|---|---|---|---|
| | | Yes | No |
| FN | Yes | Russian | German |
| | No | Serbian | Berber |

Presenting the typology in this way encodes a crucial assumption about RVA, namely that within the obstruent system, both voiced and voiceless segments act as both triggers and targets. Table 5.2 also presents a typology of FN and RVA, but this time distinguishes between the triggers and targets of RVA, restricting attention to the stops and sibilants. Only language types 1–4 are attested, where stops and sibilants are symmetric with respect to their patterning under RVA. Languages where stops and sibilants are triggers but not targets, or targets but not triggers, are not attested, regardless of whether or not they have FN.

In general then, robust examples of types 5–8 are unattested. That is, RVA is generally a symmetric process, whereby voiced stops and sibilants are both targets and triggers, and voiceless stops and sibilants are both targets and triggers; the symmetric nature of RVA is

[36] For the time being, I am not distinguishing here between word-internal FN and word-final FN (see Wetzels and Mascaró (2001) for discussion on this).

Table 5.2: Patterning of voiced stops (/b, d, g/) and sibilants (/z, ʒ/) as targets and triggers of FN and RVA.

| | | RVA | | |
|---|---|---|---|---|
| | FN | Target | Trigger | Voiced stops & sibilants |
| 1 | ✓ | ✗ | ✗ | German |
| 2 | ✗ | ✗ | ✗ | Berber, English (Level 2) |
| 3 | ✓ | ✓ | ✓ | Latvian, Maltese, Warsaw Polish, Russian, Slovak |
| 4 | ✗ | ✓ | ✓ | Hungarian, Hebrew, Serbian |
| 5 | ✓ | ✓ | ✗ | *not attested* |
| 6 | ✗ | ✓ | ✗ | *not attested* |
| 7 | ✓ | ✗ | ✓ | *not attested* |
| 8 | ✗ | ✗ | ✓ | *not attested* |

independent of whether or not a language has FN. Not all languages exhibit uniformity within the stops and sibilants. For example, Dutch (`nld`) exhibits symmetric RVA within the stop system, but clusters with a fricative as the rightmost member (regardless of its underlying voicing specification) surface as voiceless, thus /slaːp + zɑk/ → [slaːp sak] 'sleeping bag' (Grijzenhout and Krämer, 2000). Another example comes from Turkish (`tur`), in which stops are targets of FN, thus /kab/ → [kap] 'container (nom. sg.)', but voiced /z, v/ are not targets, thus [kaz] 'goose' and [av] 'hunting'. The closest language that comes to being a type 8 language would be Ukrainian, since voiced obstruents trigger voicing assimilation but do not devoice because voiceless obstruents are not triggers.

Table 5.3 expands the typology given in Table 5.2 by adding a separate column for the patterning of /v/. The table should be read as follows: under the column headed "Voiced stops & sibilants", I list languages for each type where the voiced stops and sibilants conform to the patterning indicated; under the column headed "/v/", I list languages where /v/ conforms to the patterning indicated. For example, German stops and sibilants are targets of FN, and German /v/ is also a target of FN; these facts are indicated by listing German under both columns, in the row denoting type 1.

Table 5.3: Patterning of voiced stops (/b, d, g/) and sibilants (/z, ʒ/) vs. the spirant /v/ as targets and triggers of Final Neutralization (FN) and Regressive Voicing Assimilation (RVA).

| | | RVA | | | |
|---|---|---|---|---|---|
| | FN | Target | Trigger | Voiced/lenis stops & sibilants | /v/ |
| 1 | ✓ | ✗ | ✗ | German | German |
| 2 | ✗ | ✗ | ✗ | Berber | Latvian |
| 3 | ✓ | ✓ | ✓ | Maltese, Warsaw Polish, Russian, Slovak | Maltese, Warsaw Polish *(word boundaries only)* |
| 4 | ✗ | ✓ | ✓ | Latvian, Hungarian, Hebrew, Serbian | Hebrew (younger speakers) |
| 5 | ✓ | ✓ | ✗ | *not attested* | Russian |
| 6 | ✗ | ✓ | ✗ | *not attested* | Slovak, Hungarian |
| 7 | ✓ | ✗ | ✓ | *not attested* | *not attested* |
| 8 | ✗ | ✗ | ✓ | *not attested* | *not attested* |

Few languages in Table 5.3 can be classified as a single type for stops, sibilants, and /v/, which, based on the descriptive survey in Section 5.4, is not surprising. Of those that do (German, Hebrew, Warsaw Polish), the typology above does not allow for the subtlety in patterning that pertains to /v/. The typology presented in Table 5.3, or even that in Table 5.2, has not been proposed in the literature, but it illustrates the difficulty that any typology of voicing assimilation will have.

Many typologies of laryngeal phenomena posit are abstract, but all of them begin with some kind of descriptive typology that characterizes a language's type based on what phenomena are present or absent. Nevertheless, the majority of these typologies do not address the ambiguity of /v/, or if they do, only address it peripherally. For example, Jansen (2004) includes a phonetic study of Hungarian voicing assimilation, but does not include /v/ as a consonant of study, and merely mentions its anomalous patterning in passing, noting that it is sonorant like. Wetzels and Mascaró (2001) does not address the patterning of /v/ in their typology of voicing assimilation, focussing instead on whether or not a language has final neutralization (and whether it occurs word-medially or word-finally) and voicing assimilation.

Cho (1990b) identifies two parameters of voicing phenomena: (1) Devoicing, which includes three types (coda devoicing, cluster devoicing, no-devoicing), and (2) Spreading, which includes two types, languages that spread voicing, and languages that do not. Finally, she includes a rule of *Universal Devoicing* that is imposed in all languages. The two parameters yield six different language types, and Universal Devoicing encompasses another language type, exemplified by English and Swedish, as in (36).

(36) Typology of voicing assimilation (Cho, 1990b, pg. 143)

| | | Spreading | | | |
|---|---|---|---|---|---|
| | | | + | | − |
| Devoicing: | coda-devoicing | (1) | Dutch, Catalan | (2) | German |
| | cluster-devoicing | (3) | Serbo-Croatian | (4) | Kirghiz |
| | no-devoicing | (5) | Ukrainian, Dakota | (6) | Kannada |

Several of the languages listed in (36) have been addressed. Kirghiz and Dakota do not have /v/. The standard dialect of Catalan does not have /v/, though it is present in some dialects, such as the Majorcan dialect in the Camp de Tarragona (Carbonell and Llisterri, 1992); Catalan has pre-sonorant voicing, in which all voiced segments trigger voicing of preceding voiceless obstruents (Strycharczuk, 2012). Kannada, which has no devoicing nor spreading of voicing assimilation (and so is not relevant to the issues at hand), has /ʋ/, which is realized as labio-dental [v] before front vowels; before /ɑ/ and consonants it is realized as [β], and before back vowels it is [w]; Kannada /v/ exhibits a sonorant classification, which positional variability in line with sonorant-/v/ languages, though there is no devoicing in Kannada.[37] The typology in Table (36) does not address /v/'s special status across languages. In her dissertation work, Cho addresses Russian /v/, proposing that /v/ is a sonorant but is then specified for [voice] at an intermediate stage of the derivation so that it can devoice. Cho (1990a) discusses the variability in patterning of /v/, but explains this with language-specific encodings for the featural valuation of /v/, including the point during the derivation at which /v/ is specified for [voice].

Lombardi (1999) presents a typology of voicing assimilation that focusses on the interplay between voicing assimilation and final neutralization, framed in terms of positional faithfulness constraints, such as those in (37).

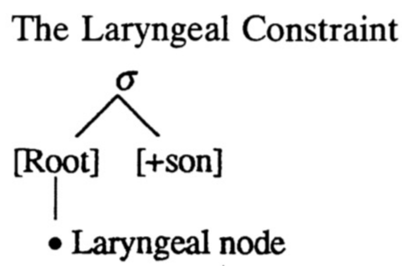
(37) Positional faithfulness constraints in Lombardi (1999)

a. IDentOnset(Laryngeal) (abbreviated IDOnsLar):
   Consonants in the position stated in the Laryngeal Constraint. . . should be faithful to underlying laryngeal specification
b. IDent(Laryngeal) (IDLar)
   Consonants should be faithful to underlying laryngeal specifications

The *Laryngeal Constraint* stipulates that "laryngeal neutralization is the result of a licensing constraint active in some languages which allows laryngeal features only in [(38)]", meaning that only syllable onsets license laryngeal contrasts. Ignoring the details of the analysis, note that the constraint encodes the distinction between sonorants and obstruents by representing the Laryngeal node as a dependent on the Root node, distinct from [sonorant].

[37] Kannada information obtained at `http://ccat.sas.upenn.edu/plc/kannada/grammar/KaGram1.pdf`.

(38) Laryngeal Constraint (Lombardi, 1999, pg. 267)




To summarize, previous typologies of laryngeal phenomena have treated the patterning of Russian /v/ as idiosyncratic and of limited relevance. As a result of this, the widespread ambiguity of /v/ has flown under the radar, cropping up periodically as a nuisance that needs to be dealt with. In contrast, I maintain that the patterning of /v/ shows that the binary distinction between obstruents and sonorants is flawed, and that there does not exist a binary partition of all consonants into either [−sonorant] or [+sonorant] (or whatever representational variant of this system is used).

## 5.6 What is [+sonorant, +obstruent]?

Throughout this dissertation I have referred to /v/ in languages like Russian as “ambiguous”, following its treatment in the literature. However, analysing /v/ as [+sonorant, +obstruent] shows that it is not, in fact, ambiguous: it readily devoices because it is [+obstruent]; it fails to trigger voicing because it is [+sonorant]. There is nothing ambiguous about Russian /v/.

What does it mean for /v/ to be specified as [+sonorant, +obstruent]? Features are generally interpreted as specifying set membership. Segments specified as ‘+’ are part of the same set, and all segments that are specified as '−' are not part of this set. Sometimes, both '+' and '−' are argued to have equal phonological status, as has been argued for [voice] (e.g., Wetzels and Mascaró (2001)); other times only the positive valuation has status, as has been suggested for [voice] by Cho (1990b) and many others. No matter the assumptions regarding feature valence, a segment cannot be specified as both [+voice] and [−voice], or as being specified for [nasal] and also not specified as such. Certainly, these are impossibilities, both physiological and logical.

The specification [+sonorant, +obstruent] is different. The definitions of [obstruent] and [sonorant] are not co-extensional, and therefore the dual specification is neither a physiological nor logical impossibility. Nevertheless, such specifications are frowned upon in phonology, as exemplified by the following discussion in Reiss (2018):

> The copious literature on the inconsistent behavior of Russian /v/ tends to focus on its putative status as a segment between an obstruent and a sonorant. Jakobson (1978), for example, says the segment "occupies an obviously intermediate position between the obstruents and the sonorants". Of all people, Jakobson should not have made such a statement. As one of the architects of distinctive feature theory, he knew that features give us the tools to transcend the vague traditional phonetic categories. Given a binary feature system, there are no "intermediate positions" (other than 'unspecified'). There are just points in a discrete multidimensional space, each defined by sets of valued features. A set of valued features can be specific enough to define a natural class containing a single segment, or it can be incomplete with respect to some features and thereby define larger sets of segments. That is the whole point of the combinatoric feature system – each combination is different from the others.

In this section I propose that the combinatoric theory of features that assumes that features cleanly partition an inventory without intermediate positions is flawed. However, such a system need not be "vague": it is constrained, principled, and can be formalized. Furthermore, I argue that such a system provides a more transparent account of voicing assimilation, a clearer relationship to the phonetic substrate on which features must be induced, and has a natural place for variation.

First, however, I must address what [+sonorant, +obstruent] is not: it is not mere intermediacy on the sonority scale (Section 5.6.1), and not a third category (Section 5.6.2). The specification [+sonorant, +obstruent] is a lack of *exclusivity*, as I explain in Section 5.6.3.

### 5.6.1 [+sonorant, +obstruent] is not mere intermediacy

With three categories [−sonorant, +obstruent], [+sonorant, +obstruent], [+sonorant, −obstruent], it is tempting to view [+sonorant, +obstruent] as merely intermediate to the other two. In other words, we might interpret [+sonorant, +obstruent] as a notational variant of, say, [3 sonorant], reminiscent of scalar features, such as those proposed for vowel height (Clements, 1991) or sonority (Selkirk, 1984). In such a system, segments that fall along the sonority hierarchy are assigned a scalar value that matches their ranking on the sonority scale. For example, Selkirk (1984) attempts to quantify the sonority scale, though she acknowledges that the absolute numbers are mostly used for exposition; in her system, voiceless stops receive a value of 0.5, while vowels receive values greater than or equal to 8. Other scalar approaches differ based on how fine-grained the scale. Phonological process are then framed in terms of these scalar values.[38]

In fact, such a solution has been offered to account for the patterning of ambiguous /v/. Barkai and Horvath (1978) argue, on the basis of arguments similar to those proposed here, that /v/ is of intermediate sonority between obstruents and sonorants. They further suggest that in languages such as Russian, Hungarian, and Hebrew, the intermediate sonority of /v/

[38]For further discussion on approaches to quantifying the sonority scale, see Parker (2002) and references therein.

is encoded as a scalar value in the sonority hierarchy, shown in (39).[39]

(39) Placement of ambiguous /v/ on the sonority hierarchy (Barkai and Horvath, 1978, pg. 83)

| stops | fricatives | v | nasals | j | r | l |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

In this framework, voicing assimilation is formulated with respect to scalar values (40):

(40) Voicing assimilation rule (Barkai and Horvath, 1978)

m.sonorant $\rightarrow$ [$\alpha$ voice] / ______ [n.sonorant, $\alpha$ voice], where $m \leq 3$, $n \leq 2$

Such a rule explicitly encodes the fact that only consonants less sonorous than /v/ may trigger RVA, but any consonant of sonority equal to or less than /v/ can be a target for it. In formulating the RVA rule with specific reference to values on the sonority scale, Barkai and Horvath (1978) capture the intermediacy of /v/ but the account lacks explanatory power.[40] Why do languages have this rule, and not another rule that reverses the values for $m$ and $n$? More generally, the problem with this approach is that there is no principled way to constrain *which* scalar values the phonology can make reference to. This analysis raises more questions than it answers, and problematically, the questions seem to be unanswerable: What does it mean, phonetically, for a segment to be specified with [5 sonorant]? How fine-grained can the scale be? Are the scalar units meaningful, or is the difference in sonority between values 2 and 3 incomparable to the difference in sonority between values 5 and 6?

[39]Barkai and Horvath (1978) use ‘y’ to refer to the palatal glide, but I have changed this to reflect the IPA transcription used throughout this dissertation.

[40]Arguments against this analysis are also put forth by Padgett (2002).

In spite of the surface similarity between specifying Russian /v/ as [3 sonorant] and [+sonorant, +obstruent], they are not equivalent, and introducing [obstruent] does not suffer the same problems. First of all, the definitions of [obstruent] and [sonorant] clearly identify the articulatory and acoustic correlates of the + and − valuations of each feature. Secondly, I argue that the trigger and target classes are principled: all [+obstruent] segments devoice as a result of RVA and FD, but only [−sonorant, +obstruent] segments can trigger it.[41]

A deeper problem with specifying /v/ as [3 sonorant] is that /v/, regardless of its patterning, is always intermediate with respect to the inventory of which it is a part. Recall from Chapter 2 that /v/ had the highest harmonicity of all the fricatives; based on the definition of harmonicity, it is clear that, *within a given inventory*, a fully voiced /v/ will be less sonorous than stops, voiceless fricatives, and voiced sibilants, but more sonorous than nasals, liquids, and glides. From the perspective of the learner then, what distinguishes the intermediacy of /v/ in Greek from that in Russian from that in Serbian? In contrast, the overlapping features [obstruent] and [sonorant] are both available for any "intermediate" /v/, and specific valuations are assigned by the learner based on how the patterning of /v/ compares to the patterning of exclusive obstruents and sonorants, a discussion I return to in Chapter 7.

Although I maintain that [+sonorant, +obstruent] is not equivalent to specifying Russian /v/ as [3 sonorant], the voiced triggers of RVA appear to obey an implicational hierarchy reminiscent of the sonority hierarchy: if /v/ triggers RVA, then so do voiced sibilants; if voiced sibilants trigger RVA, then so do voiced stops. The maximal system is exemplified by Maltese, where voiced stops, sibilants and /v/ are all triggers; Russian exemplifies a

[41]There is another problem with Barkai and Horvath's analysis. Simply identifying /v/ with a specific value on the sonority scale cannot derive the positional identity of /v/ identified earlier. Recall that /v/ *does* trigger voicing assimilation when it precedes an obstruent, so that /ot vdovɨ/ → [odvdovɨ]. The rule formulated in (40) does not encode such positional identity, and hence cannot account for the triggering behaviour of Russian /v/ in pre-obstruent position.

system where voiced sibilants and stops are triggers; in Dutch only voiced stops are triggers of RVA. I am not aware of any system in which, say, /v/ (and/or other voiced spirants) and voiced stops trigger RVA but voiced sibilants do not, or where voiced fricatives, spirants and sibilants together, trigger voicing assimilation but voiced stops do not. These facts suggest that a greater recognition of the role that sonority/obstruency plays in RVA is important in order to adequately capture the facts. In some ways, then, the introduction of [obstruent] is a minimal retreat from the binary division enforced by [sonorant], and an acknowledgment that there is something profoundly right about linking voicing assimilation with the sonority hierarchy. The approach here incorporates this intuition without sacrificing a feature-based analysis of voicing phenomena.

### 5.6.2 [+sonorant, +obstruent] is not a third category

Similar to the preceding discussion on intermediacy, it is important to stress that [+sonorant, +obstruent] is not a third category, represented by, perhaps, a ternary feature. This perspective is motivated by the earlier phonetic results of this dissertation. Recall that Chapter 3 showed that tokens of Russian /v/ were not produced with acoustic properties that situated them intermediate to those of Greek and Serbian. In the WIS position, Russian /v/ was realized like Greek, but in the WMU position, Russian /v/ was realized like Serbian. Only two environments were tested in Chapter 3, and so it is not clear if all tokens of Russian /v/ are realized as either obstruent-like [v]s or sonorant-like [ʋ]s, and it is possible that Russian /v/ exhibits a wider range of realizations than either Greek or Serbian. Regardless, this does not admit a ternary feature with Russian occupying a third position.

### 5.6.3 [+sonorant, +obstruent] is a lack of *exclusivity*

The notion of *exclusivity* is an intuitive one. For example, when stops and sibilants are referred to as "true" obstruents, what is conveyed is that their classification as obstruents, and not as sonorants, is not in question, either phonetically or phonologically; the same goes for the "true" sonorants. When the status of a segment, like Russian /v/ is in doubt, its patterning is compared to "true" obstruents and "true" sonorants. Instead of "true", I propose the term *exclusive*: stops, sibilants, and voiceless spirants are *exclusive obstruents*; liquids, nasals, glides, and vowels are *exclusive sonorants*. Russian /v/ is both an obstruent and a sonorant, and in virtue of this, it is neither an exclusive obstruent nor an exclusive sonorant.

Exclusive obstruents and exclusive sonorants give the impression of binarity. Discussions of obstruents and sonorants—and analyses that attempt to capture the "dual" or "intermediate" status of Russian /v/—are in fact discussions of those ends of the sonority/obstruency hierarchy that correspond to the binary distinction that is encoded with a single feature [sonorant]. By introducing the feature [obstruent], we can reintroduce a finer representation of the sonority/obstruency scale in featural terms without resorting to a scalar feature [$n$ sonorant].

Segments that are specified as [−sonorant, −obstruent], as the nonexplosive stops, and segments that are specified as [+sonorant, +obstruent], as /v/ in Russian, share non-exclusivity in being neither exclusive members of the obstruent class, nor exclusive members of the sonorant class. Interestingly, there is reason to believe that, in at least some languages, non-exclusivity may itself be a natural class. Consider, for example, the Tsou language, and Austronesian language spoken in Southern Taiwan. Tsou syllables are maximally [CCV:]; in such clusters, $C_1$ is released (Wright, 1996). The complete list of consonant clusters is listed below, taken from Kehrein and Golston (2004, pg. 344).[42]

| $C_1/C_2$ | Labial | Coronal | Dorsal |
|---|---|---|---|
| Labial | * | pt pt$^s$ pɗ ps ɓs ft ft$^s$ fs fz vt$^s$ vz (pz ɓz) | pk fk (vk) |
| Coronal | tp tɓ tf tv t$^s$p t$^s$f t$^s$v sp sɓ sf sv zv (t$^s$ɓ) | t$^s$z | tk t$^s$k sk |
| Dorsal | kv (kp) | kɗ ks (kt$^s$) | * |

Tsou is remarkable because it contains consonant clusters that differ in laryngeal specification: voiceless obstruents combined with voiced implosives /ɓ, ɗ/ and voiced fricatives /v, z/. Kehrein and Golston (2004) argue that laryngeal contrasts are licensed only on subsyllabic constituents such as onset, nucleus, and coda, and are not properties of segments themselves. They show that voicing, aspiration, and glottalization generally occur only once per onset, nucleus, and coda. Their framework predicts that within a given subsyllabic constituent, clusters such as [pɗ] or [sv] are prohibited. Kehrein and Golston (2004) discuss Tsou as an apparent counterexample.

First, with respect to the voiced fricatives /v, z/, Kehrein and Golston (2004) argue that they are realized with little to no frication, based on spectrograms published in Wright (1996). Tsou /v, z/ were historically semi-vowels **w*, **j*, and alternate with vowels, so that [i] alternates with [z] and [o, u] alternate with [v]. Kehrein and Golston (2004) conclude based on this that /v, z/ are in fact sonorants. Second, they argue that clusters of the form [ɓs, ɓz, sɓ] are banned because Tsou does not have implosive fricatives, so that the absence of implosion on [s, z] is an artefact of this general ban, and not a true counterexample. Finally, with respect to clusters of the form voiceless stop + implosive, they argue that this represents late phasing of implosion on the stop cluster, noting that this interpretation is supported by the absence of clusters implosive + voiceless stop.

[42] The chart is taken from Kehrein and Golston (2004), who cite Wright (1996), due to their more compact representation of the data.

Another interpretation, not entirely out of line with the explanations proposed by Kehrein and Golston (2004) is that in Tsou /v, z/ are specified as [+sonorant, +obstruent], and /ɓ, ɗ/ are specified as [−sonorant, −obstruent].[43] What these two groups have in common is that they are neither exclusive obstruents, nor exclusive sonorants.[44] Nevertheless, it does not appear appropriate to suggest that /v, z/ and /ɓ, ɗ/ share a featural specification. Rather, non-exclusive obstruents appear to licit different laryngeal-oral timing relationships within the subsyllabic constituent. Pursuing this hypothesis is beyond the scope of the present work, and is left as an avenue of further research.

## 5.7 The representation of voiceless fricatives

The proposal advanced in this chapter argues that the voiced sibilants and spirants are generally not part of the same class, and in many cases are featurally distinct. This brings up the question of the voiceless fricatives. If voiced sibilants and spirants are not a coherent class, then what of the voiceless sibilants and spirants? With respect to their classification for [obstruent] and [sonorant], the voiceless fricatives are clearly [−sonorant, +obstruent], in terms of their phonetic definitions. This specification predicts that voiceless fricatives will uniformly trigger regressive voicing assimilation, and as [+obstruent] segments will be targets for RVA. The patterning of voiceless /h/ in Hungarian challenges this assumption, which exhibits the mirror image patterning of /v/ (Szigetvári, 1998): when a voiced ob-

[43]It is difficult to assess, based only on the spectrograms in Wright (1996) whether [v] resembles the [v] in Greek or Serbian, and Wright does not quantify the degree of frication of [v] in various environments to allow this issue to be adjudicated. It is clear, however, that the spectrograms for Tsou [z] do not exhibit the kind of high frequency noise typically seen in [z]. I continue to use the transcription [z] for consistency with the literature on Tsou, but it does not appear that the transcription is appropriate.

[44]It is worth pointing out that non-exclusivity is easily captured in set-theoretic terms using the symmetric difference operation. Given sets $A$ and $B$, the *exclusively* $A$ points are picked out by taking the set difference $A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}$. Similarly the *exclusively* $B$ points are picked out by $B - A$. The *symmetric difference* of $A$ and $B$, denoted $A \triangle B$, is the union of $A \setminus B$ and $B - A$, in other words, all the points that are either exclusively one or exclusively the other. Non-exclusivity is thus the complement of the symmetric difference.

struent precedes /h/, it is realized as devoiced, thus /ad+hat/ → [athat] 'may give', but in preconsonantal position (where /h/ is strengthened to its voiceless velar allophone [x]), it does not become voiced, thus /potroh+ba/ → [potroxba], *[potroɣba] 'into the abdomen'.[45] Similarly, the voiceless pharyngeal fricative /ħ/ in Cairene Arabic is a trigger for RVA, hence /mazħuum/ → [masħuum] 'crowded', but is not a target for RVA, hence /jiħzar/ → [jiħzar] 'to be cautious' (Kabrah, 2008, pg. 31). The voiced pharyngeal fricative /ʕ/ patterns symmetrically with respect to RVA as an obstruent, as do the uvulars /χ, ʁ/; laryngeals pattern as sonorants in not participating either as targets or triggers.[46]

The asymmetric patterning of voiceless back fricatives suggests that in addition to distinguishing between the triggers and targets of RVA, it may be appropriate to incorporate a distinction based on voiced vs. voiceless triggers and targets. For example, if Hungarian /h/ and Cairene Arabic /ħ/ are specified as [+sonorant, +obstruent], then we would have to say that for voiceless segments, the triggers and targets of RVA are mirror images of each other, as in Table 5.1.

Table 5.1: Distinguishing triggers and targets of RVA by voicing?

| | Triggers | Targets |
|---|---|---|
| Voiced | [−sonorant, +obstruent] | [+obstruent] |
| Voiceless | [+obstruent] | [−sonorant, +obstruent] |

The problem with this approach is that it does not sit well with the definitions of [obstruent] and [sonorant]. Clements and Osu (2002) speculate on how [obstruent] and [sonorant] interact with respect to laryngeals:

> Do we ever find a three-way distinction within the class of continuants among obstruents, sonorants and a third type of sound which is neither of these? Such sounds would theoretically be produced with no buildup in oral air pressure and

[45] it is important to note that the realization of /h/ in preconsonantal position is as the velar /x/, and so the non-participation of underlying /h/ cannot be due to its laryngeal articulation in that position.

[46] Kabrah (2008) uses /ɣ/ to denote the voiced uvular fricative, rather than the IPA symbol /ʁ/.

> without voicing, or without a clearly-marked formant structure. The obvious candidates to fill this slot are the so-called voiceless sonorants, including the laryngeals [h] and [ʔ]. These sounds have been notoriously difficult to classify in feature terms in the past, due to their behavior in some respects as obstruents, in others as sonorants. An analysis of these sounds as [−obstruent, −sonorant] sounds might go some way toward explaining this ambiguity" (pg. 34−35).

Based on the patterning of back consonants in Cairene Arabic, however, this does not seem to be quite right. Laryngeals pattern with sonorants with respect to RVA, in that they are neither triggers nor targets. Furthermore, having to specify different triggers and targets of RVA for voiced and voiceless continuants does not seem particularly parsimonious. Furthermore, although the [−obstruent] classification of laryngeals is well grounded, since they do not involve any build up of oral air pressure, the [−sonorant] classification seems less well motivated. Although [h] is not highly resonant, its formant structure is often clear enough that it has been classified with vowels, as in Chomsky and Halle (1968).[47]

Another approach is to assign segments such as Cairene Arabic [ħ] a [−sonorant, −obstruent] classification, as seen in Table 5.2, so that it can be accounted for in the same framework as Russian /v/.[48]

Table 5.2: Possible segment classifications

| | [+sonorant] | [−sonorant] |
|---|---|---|
| [+obstruent] | v | p, b, s, z |
| [−obstruent] | m, l, h | ħ |

If these featural specifications are correct, then this calls for a revision of the rules for RVA that I argued for in Chapter 5. In particular, while the targets of RVA are [+obstruent] segments, the triggers for RVA are simply all [−sonorant] segments. Nevertheless, this

[47] For some discussion of the featural specification of laryngeals, see Vaux and Miller (2011) for their relationship to fricatives, Botma (2011) for their relationship to sonorants, and Hall and Żygis (2010) for their relationship to obstruents.

[48] Cairene Arabic does not have /v/, so it is not known how it would pattern within the Cairene Arabic system of RVA.

analysis raises more questions than it solves: First, why does Cairene Arabic /h/ get specified as [+sonorant, −obstruent], but Hungarian /h/ (which patterns as Cairene Arabic /ħ/) gets specified as [−sonorant, −obstruent]?[49] Second, if the triggers of RVA are now the [−sonorant] segments, then we can no longer make sense of the Tsou patterning, where both voiced continuants /v, z/ as [+sonorant, +obstruent] and non-explosive stops /ɓ, ɗ/ pattern together as non-triggers. Third, since the voiced pharyngeal /ʕ/ patterns with obstruents, what might account for the fact that voiced /ʕ/ is specified as [−sonorant, +obstruent], while voiceless /ħ/ is specified as [−sonorant, −obstruent]?

It is beyond the scope of this work to solve the problems that back voiceless continuants pose for the obstruent-sonorant divide. Further research into the articulatory, aerodynamic, and acoustic properties of these consonants is required in order to understand how they might be defined with respect to [obstruent] and [sonorant].

## 5.8 Sonorancy, obstruency, and voicing

In this chapter I have argued that Russian /v/ is featurally specified as [+sonorant, +obstruent]. The advantage to this approach is that it incorporates a greater degree of granularity in the feature system than what is provided by a single binary feature [±sonorant], without invoking a scalar [sonorant] feature that mirrors the Sonority Hierarchy. This analysis raises several questions, particularly with respect to the relationship between voicing and sonority.

As I discussed in Chapter 1, the distinction between obstruents and sonorants is well-entrenched both phonologically and phonetically. In particular, many phonological frame-

[49]Szigetvári (1998) takes pains to clarify that in pre-obstruent position, where /h/ is not voiced, it is a velar [x], and not a laryngeal; it is not clear how this fact helps in this case.

works reify the obstruent/sonorant divide with representations that are meant to capture the distinct nature of voicing on obstruents and sonorants. For example, Cho (1990b) and Wetzels and Mascaró (2001) argue that [voice] is only specified on obstruents; for Cho (1990b) [voice] is privative, while for Wetzels and Mascaró (2001), [voice] is binary, but they nevertheless share the assumption that sonorants are not specified for voicing. Rice (1993) proposes that sonorants are specified underlyingly with a Sonorant Voice (SV) node, and obstruents are represented with [voice] as a dependent on the Laryngeal node, thereby distinguishing obstruents and sonorants representationally.

The proposal to include [obstruent], and to treat /v/ in languages such as Russian as [+sonorant, +obstruent] is difficult to reconcile with these perspectives: does a segment specified as [+sonorant, +obstruent] receive a [voice] specification because it is [+obstruent], or does it not receive a [voice] specification because it is [+sonorant]?

The general argument against specifying [voice] on segments specified as [+sonorant] or [SV] is that to do so is redundant (for specific arguments against specifying [voice] on sonorants in theories that admit underspecification, see Rice (1993) and references therein). The absence of a [voice] specification also accounts for the fact that sonorants do not, and arguably cannot, contrast strictly in terms of voicing.[50] Nevertheless, in spite of the importance that contrast plays in phonological theory, non-contrastiveness usually finds a way to creep in.

Consider, for example, the analysis of Slavic /v/ patterning in Hall (2007), which relies on the *Contrastivist Hypothesis* framework and *Successive Division Algorithm* (SDA) of Dresher (2009), together with the system of representing voicing contrasts by Avery (1996), shown

[50]See Botma (2011, section 2.3) for an overview of the relationship between voicelessness and sonorancy, as well as Lombardi (1991). As Botma (2011) points out, voiceless sonorants typically exhibit patterning associated with aspirated segments, rather than voiceless segments.

in Figure 5.1.

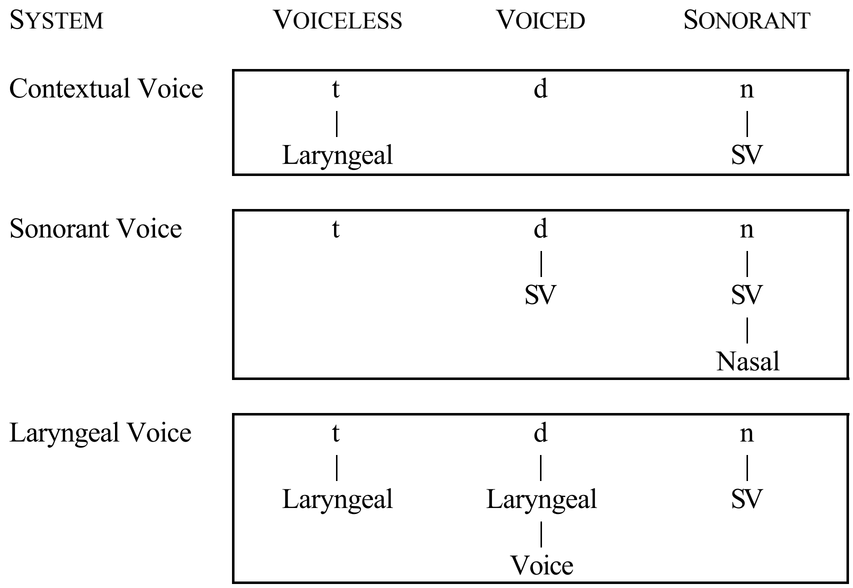



Figure 5.1: Voicing systems under Avery (1996)

Following Avery (1996), Hall argues that the Slavic data are best accounted for by analysing them as a mix of Laryngeal Voice and Contextual Voice systems. In a Laryngeal Voice (LV) system, obstruents bear a Laryngeal node, while sonorants bear an SV ("Spontaneous Voicing") node (Rice and Avery, 1989); obstruents are further distinguished with a Voice node for voiced obstruents. In Contextual Voice systems, sonorants still bear an SV node, but only voiceless obstruents are specified as having a Laryngeal node; voiced obstruents are left entirely unspecified for voicing. Hall (2003a) argues that systems like Russian, Czech, and Slovak ought to be regarded as an instantiation of a mixed system: sonorants are specified for SV, all non-/v/ obstruents specified with a Laryngeal node and Voice specified on voiced obstruents (as in an LV language), but /v/ is entirely unspecified for voicing features. Hall (2007) offers the following representational configurations for Russian consonants, in Figure 5.2. Under the assumption that assimilation is feature spreading, the patterning of /v/ follows straightforwardly: since /v/ is not specified for any voicing features, it cannot be a trigger for voicing processes; also because it is left unspecified for voicing features, it can act as a landing site for spreading of the Laryngeal node.

<table>
<tr><th>VOICED OBSTRUENTS</th><th>VOICELESS OBSTRUENTS</th><th>/v/</th><th>SONORANTS</th></tr>
<tr><td>/d/<br>|<br>Laryngeal<br>|<br>Voice</td><td>/t/<br>|<br>Laryngeal</td><td>/v/</td><td>/n/<br>|<br>SV</td></tr>
</table>

Figure 5.2: Representation of Russian voicing contrasts Hall (2007)

The reliance on the SDA and the kind of underspecification that it entails means that /v/ is also not specified for other features either, and hence is a target not just for voicing assimilation, but also for total assimilation. As Hall (2007, pg. 86) explains, "In Russian, Slovak, and Polish, /v/ has a phonemic voiceless counterpart /f/, but no phonemic counterpart specified for [Voice]. . . either /v/ will be impossible to identify once it has received the features [Laryngeal] and [Voice] or the rule of regressive assimilation must be revised". Hall's solution is to posit that additional, non-contrastive features—which he dubs *prophylactic features*—are present and visible to the phonetic component. For Russian, /v/ is endowed with prophylactic features Labial and/or Continuant, as seen in Figure 5.3.[51] "What prophylactic features prevent is the conflation of different phonetic implementation rules for distinct segments from which the same feature is non-contrastively absent."

Prophylactic features are non-contrastive and not involved in phonological computation, but are nevertheless present and visible to the phonetic module in order to prevent total assimilation. It is hard not to see this as a technical solution to a theory-internal problem, and various questions remain, but I leave these issues aside for now.[52] Rather, what I wish

[51] I follow Hall's convention of distinguishing prophylactic features by using a different typeface.

[52] For example, how does the learner distinguish between contrastive and non-contrastive specifications

| VOICED OBSTRUENTS | VOICELESS OBSTRUENTS | /v/ | SONORANTS |
|---|---|---|---|
| /d/<br>Laryngeal<br>Voice | /t/<br>Laryngeal | Labial<br>and/or<br>Continuant<br>/v/ | /n/<br>SV |

Figure 5.3: Representation of Russian voicing contrasts, with prophylactic features Hall (2007)

to emphasize is that the reliance on a strict dichotomy between obstruents and sonorants is problematic, even in a theory that allows fairly enriched representations that can distinguish multiple kinds of participation and non-participation in voicing phenomena.

One might argue that the introduction of prophylactic features has more to do with the SDA and the Contrastivist Hypothesis, and that the intuition that sonorants are specified as SV ought to be maintained. Indeed, as I discussed in Chapter 1, the division between obstruent and sonorant voicing is well established, grounded in the aerodynamic constraints that dictate voicing. Nevertheless, it is arguable that the phonetic grounding for representing [voice] differently on obstruents and sonorants is most robust for stop consonants produced with an egressive airflow, such as the difference between, say, [m] and [b], than it is for spirants. For stops, the compensatory largyngeal configurations that allow voicing to occur throughout the closure of [b] are known to be quite different than those required for voic-

in a given language? Are there universal constraints on prophylaxis, or is it always available to "save" the derivation? It is beyond the scope of this work to delve into the details of the Contrastivist Hypothesis and attendant issues, but the introduction of prophylactic features is symptomatic of a more general problem that has been widely noted: reference to non-contrastive features is often necessary. These issues go beyond the scope of the present work, but for a discussion of these issues with respect to the Contrastivist Hypothesis, see Nevins (2015), and references therein.

ing of [m], which is reasonably considered "spontaneous" (Chomsky and Halle, 1968). For spirants, however, an appeal to the laryngeal configuration is insufficient, as further aerodynamic considerations play a significant role in whether the resulting sound is characterized by turbulent or laminar flow. In other words, for a given articulatory configuration (laryngeal and supralaryngeal), the amount of airflow determines whether a voiced, labiodental continuant will be produced with frication, as [v], or as an approximant, as [ʋ].[53]

Clements and Osu (2003) consider the SV hypothesis with respect to the definitions of [sonorant] and [obstruent]. Recall that SV is predicated on there being spontaneous voicing, which as Clements and Osu (2003) point out, "is a mechanical consequence of the absence of intraoral air pressure buildup in the production of nonobstruent sounds, provided that other necessary conditions for voicing, such as loose adduction of the vocal folds, are also present" (pg. 12), and argue that SV simply refers to [−obstruent]. Importantly, the use of both [obstruent] and [sonorant] tease apart the articulatory/aerodynamic and acoustic components of a binary distinction afforded by [±sonorant] or SV vs. non-SV consonant.

A different criticism of SV is levied by Reiss (2018), who argues that the special voicing status of SV "already follows from a combinatoric feature system that +Voiced combined with +Sonorant might behave differently from +Voiced combined with −Sonorant" (pg. 37). The alternative that Reiss (2018) proposes is a framework in which segments are sets of valued features, and that the features are universal; all of the work of accounting for phonological phenomena rests in the rules, rather than the representations; the only representational tool

[53]Ohala (1983, pg. 202) cites further evidence that stop and fricative voicing differs with respect to laryngeal activity: "There is evidence that the state of the glottis may not be the same during voiced fricatives and voiced stops. Hirose and Ushijima (1978) report electromyographic (EMG) data for laryngeal muscles which suggest that, at least for medial consonants, [z] has a less constricted glottis than [b, d, g]. If so, this may be necessitated by the fact that voiced fricatives, but not voiced stops, require greater airflow for the sake of maintaining the trans-oral constriction pressure differential." It is not clear from this discussion if voiced spirants exhibit a laryngeal configuration that is the same as voiced sibilants or voiced stops, or something else entirely.

that Reiss (2018) permits himself is underspecification. In particular, he analyses Russian /v/ as being unspecified for Voice, which allows him to account for the Russian /v/ data through ordered rules. Abstracting away from the details of his framework, under Reiss's analysis, because /v/ is underlyingly not specified for Voiced, it can be filled in as −Voiced to undergo final devoicing, but it cannot trigger voicing assimilation because, at the stage of the derivation that RVA applies, /v/ is not specified as +Voiced. Like other analyses of Russian /v/, then, Reiss (2018) must rely on a language-specific representation of /v/, as well as a specific rule-ordering that is required in order to fill in the right feature values for /v/ at the right time. What this analysis fails to do is provide a general theory of /v/, or a theory about why regressive voicing assimilation appears to be inherently asymmetric.[54]

The analysis that Reiss (2018) proposes is motivated by a more general philosophy with respect to features. In particular, Reiss (2018) assumes that there exists a universal feature set, and that cross-linguistic variability is accounted for by ordered rules. In the following chapter, I sketch out a model that addresses questions of universality and innateness with respect to phonetic and phonological features.

Returning to the question of how [voice] should be represented, I propose that [voice] is specified on all segments, regardless of their [obstruent]/[sonorant] status. First, the analysis of RVA presented in Chapter 5 distinguishes between triggers and targets of voicing assimilation by reference to whether they are exclusive obstruents (i.e., [−sonorant, +obstruent]) or non-exclusive obstruents (all segments specified as [+obstruent]), and so reference to the voicing specification is not relevant. Second, as I discuss in the following section, the articulatory, aerodynamic, and acoustic bases for distinguishing between spontaneous and

[54]Empirically, an additional shortcoming in Reiss's analysis is that he cannot account for the "glitches" discussed in Section 5.3.2—namely the variability that appears to be inherent to /v/—since his rules treat the trigger status of /v/ under RVA as symmetric, regardless of whether /v/ appears before a voiced or voiceless obstruent.

non-spontaneous voicing conflict precisely at /v/, making it difficult to maintain a robust distinction. For simplicity, I adopt the view that [voice] is binary, but a fuller development of the representation of [voice] within this framework is an issue left to future research.

## 5.9 Discussion

In this chapter I have proposed that the segment typically transcribed as <*v*> corresponds to three different phonological specifications across languages. Building on work by Clements and Osu (2002), I argue that there is evidence for a feature [obstruent], and that voiced spirants in so-called "ambiguous /v/" languages are featurally specified as [+sonorant, +obstruent]. A wide range of languages contain a sonorant obstruent /v/, including languages that do not otherwise have regressive voicing assimilation. When viewed within a framework that includes [obstruent] and [sonorant], the patterning of Russian /v/ forms a natural part of the typology circumscribed by these two features.

In the preceding discussion I characterized [+sonorant, +obstruent] as a minimal retreat from the binary division that the traditional [sonorant] feature posits. In other ways, assuming [obstruent] is not minimal at all, and is in fact somewhat radical, given the assumption that features are binary partitions, as articulated above in the quote by Reiss (2018). I reject Reiss's claim that acknowledging intermediacy somehow fundamentally undermines the combinatoric and precise nature of distinctive feature theory. The disagreement stems from a difference in what features represent. Reiss (2018) takes the view that segments are sets of valued features, and assumes an innate universal feature set. In contrast, I take the view that features and segments represent not simply primitives of the computational system, but are themselves the output of the cognitive system, induced by the learner over the course of acquisition (*c.f.* Mielke (2008), Blevins (2007)); I return to the emergence of [obstruent] and [sonorant] in Chapter 6. Nonetheless, the generalizations *are* categorical, and manipulated computationally by the phonology. In this dissertation I have focussed exclusively on the patterning of /v/, thereby focussing just on its status as a cognitive category that already exists in the phonological system; I delve into the question of feature emergence in Section 7.

The problem of Russian /v/ has been a thorn in the side of linguistic theory for several decades. While some of the controversies it has engendered have been solved, such as the issue of sonorant transparency in voicing assimilation, the fundamental problem has never gone away. The analysis provided here attempts to provide a simple analysis of the patterning of Russian /v/, eschewing some of the more elaborate formalisms available. The representation of /v/ as both [+obstruent] and [+sonorant] is, in some ways, simplistic, requiring only that we allow at least some features to overlap. Such an approach is grounded in the phonetic properties of both obstruency and sonority, without trying to establish a one-to-one relationship between phonetics and phonology so that we "read the phonology off the phonetics". I proposed that the variability of /v/ is due to the fact that its dual specification presents an inherent phonological tension. The inclusion of a four-way contrast afforded by [obstruent] and [sonorant] allows for the connection between voicing assimilation and sonority/obstruency to be made more explicit, without resorting to a fine-grained scalar feature that tracks the sonority hierarchy. Finally, I showed that in spite of being featurally "opposite", segments that are non-exclusively obstruents and non-exclusively sonorants may pattern together in licensing laryngeal contrasts within the subsyllabic constituent.

CHAPTER 6

# THE QUANTAL NATURE OF [SONORANT] AND [OBSTRUENT]

## 6.1 Introduction

In Chapter 5 I argued that the inclusion of the feature [obstruent] accounts for the cross-linguistic patterning of /v/, both in terms of its cross-linguistic patterning and its ambiguous patterning in Russian and other such languages. The feature [obstruent] has precedent in work by G.N. Clements and S. Osu, who argue in a series of papers that the domain of nasalizability in Ikwere are the non-obstruent consonants (Clements and Osu, 2002, 2003, 2005). Clements and Osu (2003) define [sonorant] and [obstruent] such that they are not simply mirror images of each other, as they are defined in different domains: [sonorant] is acoustically defined, while [obstruent] is defined in articulatory/aerodynamic terms. Sounds that are [+sonorant] have periodic, well-defined formant structure; sounds that are [+obstruent] are those that are produced in such a way that there is sufficient air pressure buildup in the oral cavity.

Clements and Osu (2002) refer to quantal theory in motivating a binary feature [obstruent] over a gradient notion of larynx-lowering[1]:

> Under the view that articulatory variation between distinct phonemic categories is marked by rapid shifts in spectral properties, while articulatory variation within any single phonemic category is not (see Stevens' quantal theory of speech, 1989), one would expect to find a categorical property distinguishing implosives and other non-explosive stops from explosive stops. (pg. 10)

In order to justify the categoricity of [obstruent], Clements and Osu (2002) show that implosives and other non-explosive stops (such as glottalized stops) can be grouped together

[1]Namely in terms of both degree of lowering and rapidity with which the larynx lowers.

in virtue of the fact that they are not characterized by an increase in supraglottal pressure. Moreover, they show that appealing to other factors, such as larynx lowering or airstream mechanism (ingressive vs. egressive), does not serve to group all non-explosive stops together. Rather, the non-explosives are characterized by an absence of an increase in supraglottal pressure. Furthermore, they show that [−obstruent] is not the same as [+sonorant], since non-explosive stops are not characterized by periodic, well-defined formant structure. On this basis, Clements and Osu conclude: "It seems, then, that the feature [−obstruent] provides an adequate quantal basis for distinguishing implosive (and other nonexplosive) stops from explosive stops" (pg. 15).

This is not quite right, however. Although their argument supports a categorical interpretation of obstruence, *Quantal Theory* (QT), as articulated in the work of Kenneth Stevens and Jay Keyser, is about much more than categoricity. Specifically, QT claims that there are regions of acoustic space that are relatively invariant to small perturbations in the articulatory domain; between regions of stability, small articulatory changes lead to drastic acoustic changes.[2] It is not immediately clear how to reconcile the articulatory/aerodynamic definition of [obstruent] within this framework. Building on work by (Stevens, 1989; Stevens and Keyser, 1989; Keyser and Stevens, 2006; Stevens and Keyser, 2010), in this chapter I consider what it would mean for both [sonorant] and [obstruent] to be defined in terms that are consistent with QT, as well as some implications of these definitions for feature theory. Finally I turn to broader issues within the and the question of innateness and emergence.

[2]For simplicity, I put aside the acoustic-auditory domain.

## 6.2 Quantal Theory

QT posits a particular relationship between the articulatory and acoustic domains, such that, all else being equal, gradual changes in the articulatory domain can have rapid changes in the acoustic domain; a similar relationship is posited to hold between the acoustic and auditory domains. Figure 6.1 schematizes the basics of a quantal relationship between two domains, say the articulatory domain, represented on the $x$-axis, and the acoustic domain, represented on the $y$-axis.

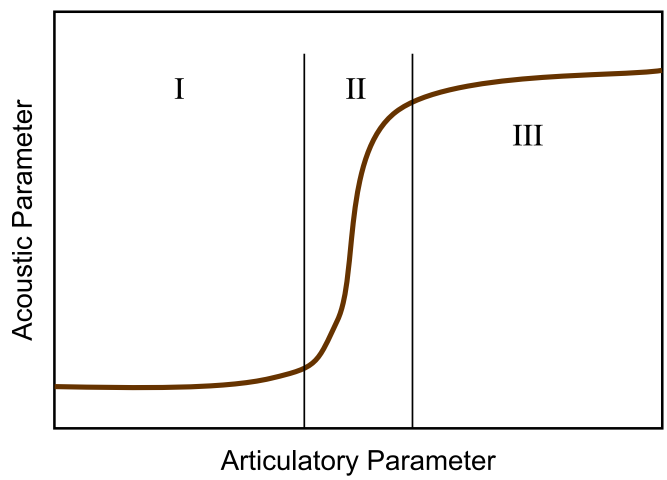



Figure 6.1: Schematization of relation between articulatory parameter and acoustic parameter, from Stevens and Keyser (2010)

The graph is interpreted as follows: for some continuous change in articulatory configuration, the sigmoidal curve represents the acoustic consequence along a particular parameter, all else being equal. If the $x$-axis represents degree of constriction at a particular place of articulation, the $y$-axis might represent acoustic resonance or harmonicity. Regions I and III are considered to be "regions of insensitivity", in which the acoustic consequence of small perturbations in articulatory configuration are negligible. Region II exhibits the rapid acoustic consequence that arises from changes in articulatory configuration over a particular range of values. Regions I and III are acoustically quite distinct, and Stevens (1989) argues that, "the difference in the acoustic pattern between regions I and III should not be regarded as simply a matter of identifying two points on a scale of some acoustic parameter. Rather, the acoustic attribute often undergoes a qualitative change as the articulatory parameter moves through region II" (pg. 4). Furthermore, QT hypothesizes that regions of stability correspond to distinctive features: "The articulatory and acoustic attributes that occur within the plateau-like regions are, in effect, the correlates of the distinctive features" (Stevens, 1989, pg. 5).

QT is attractive because it seemingly provides solid phonetic footing to the definition of distinctive features. Nevertheless, finding empirical support for some of its claims has proved elusive, and serious theoretical criticisms have been levied against it. On the latter point, for example, Studdert-Kennedy (1989) argues that the interpretation of Figure 6.1 is not always made explicit, and the notion of stability is not, itself, stable within QT (pun intended):

> The word "stable" has two possible meanings in the present context: (1) unchanging over some period of time during the execution of a particular utterance, and therefore static; (2) unchanging over different utterance, and therefore invariant. The two meanings are not incompatible: a static pattern might be invariant, and an invariant pattern might be static. Nor are they mutually entailed: a static pattern might vary from one utterance to another, and an invariant property might be a recurrent pattern of change. (Studdert-Kennedy, 1989, pg. 135)

Studdert-Kennedy (1989) goes on to show that in the original presentation of Figure 6.1, Stevens (1989) uses a *temporal* interpretation: "In the acoustic signal, therefore, there will be an alternation between *temporal* regions where the acoustic parameters remain relatively steady, and narrow regions marked by acoustic events where there are rapid changes" (Stevens, 1989, pg. 136; emphasis added). Under a temporal interpretation, the regions of stability Figure 6.1 are stable in the sense that they are static. Later on, however, Stevens (1989) presents Figure 6.1 as an idealized relationship between an articulatory and acoustic space, and in this interpretation stability means invariant. The interpretation I gave above abstracts over the temporal dimension, since the $x$-axis is interpreted as constriction degree; although changing articulatory constriction degree can unfold over time, I do not interpret the quantal relation expressed in Figure 6.1 as temporal. In what follows, I reject the temporal interpretation of QT, and focus only on the invariant interpretation of regions of stability.[3]

Even restricting attention to the articulatory-acoustic domain, thus abstracting away from the temporal interpretation, empirical support for some of the claims of QT has been weak. For example, Stevens (1989) claims that acoustic regions of stability may be characterized by less articulatory precision: "...the precision with which the articulatory state is achieved may be rather lax" (pg. 5). The robustness of the acoustic signal in light of the imprecision of articulatory configuration is thought to account for the fact that certain sounds are cross-linguistically more common than others, and because Stevens (1989) adopts the view that regions of stability are themselves features, the purported acoustic stability plays a special role:

> Features whose correlates are based on acoustic-articulatory or auditory-acoustic relations of this kind have the desirable attributes that (1) only limited articulatory precision is required to obtain a desired auditory response when a feature is implemented, and (2) there is a large difference in the acoustic or auditory patterns for articulatory configurations or states corresponding to a particular feature opposition. (Stevens, 1989, pg. 40)

Tabain (2001) tests this prediction comparing articulatory (obtained via elecropalatog-

[3] At face value, the notion that there are invariant cues in the acoustic signal is itself problematic. However, as I make clear in Section 6.6, the invariance assumed is with respect to a certain segment-feature pairing. In other words, [s] is invariant with respect to its cues for obstruency, but a segment such as [v] may not be.

raphy) and acoustic (calculated using a 1000 Hz high-pass filtered spectral centroid) data of (Australian) English fricatives [θ, ð, s, z, ʃ, ʒ] in a variety of vocalic contexts. Specifically, Tabain tested the hypothesis that because the sibilants [s, z, ʃ, ʒ] are the most typologically frequent, there should be a greater degree of articulatory variability than spirants [θ, ð]. Results showed that "sibilant consonants show less variability overall than nonsibilants" (pg. 84), and "variability in production and variability in the speech signal seem to be well correlated" (pg. 85). As Tabain (2001) notes, this is problematic for QT, since variability in the articulatory domain does not lead to acoustic robustness (as measured by spectral centroid), and that sibilants in general require a precise articulatory configuration of the tongue in order to direct the airstream to the teeth. The results of Tabain (2001) are in line with a general flaw of QT that the theory predicts greater acoustic similarity across languages than is found, and that languages do not, in general seem to systematically exploit articulatory sloppiness, as for example, in the articulation of [s] (*c.f.* Pisoni (1981)).

In spite of these criticisms, QT has some traction within the phonological literature (*c.f.* Clements and Ridouane (2006)) perhaps because it embraces distinctive features. Furthermore, even without reference to QT, the featural distinctions most relevant to this dissertation appear to have something of a quantal nature, namely the distinction between obstruent and sonorant voicing (*c.f.* Halle and Stevens (1971)), and the distinction between fricative and non-fricative continuants. I believe that QT suffers from two flaws: (1) its exclusive reliance on the articulatory-acoustic domain in identifying regions of stability, and (2) its equivocation between regions of stability and features. In some ways, these are the two fundamental claims of QT, as formulated by Stevens (1989) and in subsequent work with J. Keyser, and doing away with these assumptions strips QT bare of its main substantive claims. Nevertheless, as I sketch out in the remainder of this section, there is a conceptual core to the notion of a quantal relation that can be exploited in conjunction with *Emer-*

*gent Feature Theory* (EFT) (Mielke, 2008). By marrying QT with EFT, we can account for both the cross-linguistic robustness of certain features, while also making room for feature ambivalence and ambiguity.

The remainder of this section is structured as follows. First, I consider the quantal nature of [sonorant], as explained by Stevens (1989), and consider this in light of the proposal by Clements and Osu (2002) that there exists an articulatory feature [obstruent] which also displays quantal properties in Section 6.3. I then argue in Section 6.4 that quantal relations exist between more domains than just the articulatory-acoustic domain, and that regions of stability are not features in Section 6.5. Finally, in Section 6.6 I propose that emergence must be viewed as a relationship between segments and features—what I call *segment-feature pairs*, or *sf-pairs* for short—and that emergence looks different across different sf-pairs.

## 6.3 The quantal nature of [sonorant]

Stevens (1989) discusses the quantal nature of [sonorant] in the extended quotation below. There are two components to the argument: first, showing that within sonorant and non-sonorant regions, there are regions of stability (regions I and III in Figure 6.1, above); second, showing that at the "edges" of these regions, there is a threshold effect (region II).

> When a consonant is produced in such a way that the airway between the glottis and the output of the vocal tract has no narrow constriction, air can flow through this airway without appreciable build up of pressure above the glottis. In particular the vocal folds can continue to vibrate in a normal fashion under these conditions, and there is no modification of the vocal-fold vibration pattern relative to that in an adjacent vowel. A consonant produced in this manner is classified as a *sonorant* consonant. Since the vocal-tract transfer function from input to output volume velocity is close to unity (or 0 dB) or slightly greater at low frequencies (below the frequency of the first resonance of the system), then the amplitude of the first and possibly the second harmonic **shows very little change as the vocal tract executes a maneuver between a vowel and a**

> **sonorant consonant**.
>
> Non-sonorant consonants, on the other hand, are produced with a sufficiently narrow constriction in the supraglottal airways that the airflow from the glottis causes a pressure drop across the constriction and a consequent increase in supraglottal pressure. The resulting decrease in transglottal pressure causes a decrease in amplitude of the glottal pulses, and if the constriction is sufficiently narrow, the transglottal pressure decreases to the point where glottal vibration ceases. When a non-sonorant consonant is produced preceding or following a vowel, then, there is a decrease in amplitude of the glottal pulses, which is reflected in a reduced spectrum amplitude in the vicinity of the first harmonic…
>
> These considerations suggest that there is continuity in the low-frequency amplitude between a vowel and a sonorant consonant such as a nasal consonant, whereas there is a decrease in this amplitude as the closure in the vocal tract is formed for a voiced stop consonant… **Evidently the boundary between the sonorant and the non-sonorant responses is about where the low-frequency amplitude in the consonant region is equal to that in the adjacent vowel.**
>
> Apparently, then, listeners make use of the simple property of low-frequency amplitude decreasing or amplitude not decreasing as a way of distinguishing non-sonorant from sonorant consonants, at least in the absence of additional properties (such as strong nasalization in the early part of the vowel in a syllable consisting of a nasal consonant followed by a vowel) that would provide cues that conflict with the low-frequency property. Again we have an example of a threshold effect in an auditory-acoustic relation: the threshold as exceeded when the low-frequency amplitude no longer shows an increase at the time the consonant is released into the adjacent vowel. As indicated above, a decrease in low-frequency amplitude is expected in the sound for a non-sonorant consonant, i.e., for a consonant that is produced with an increase in supraglottal pressure. (pg. 35–36; bold-faced added)

For example, the relationship between the articulatory configurations required for the voiced bilabial egressives [m] and [b]—namely, velum position—exhibits a quantal relationship with respect to the acoustic consequence of resonance, or harmonicity, all else being equal (place, airstream, etc.). For the bilabial egressive nasal [m], the articulatory configuration of a lowered velum allows air to flow freely through the nasal cavity, meaning that there is no appreciable build up of pressure above the glottis, and so [m] is [+sonorant]. Provided that the position of the larynx stays the same and the airflow remains egressive, the raising of the velum (an action which is, of course, continuous through time and space) results in an abrupt change in acoustic realization. Instead of the periodic, well-defined formant structure characteristic of [m], the signal is characterized by an extremely low-amplitude sound with no formant structure (assuming, of course, a fully voiced [b]), together with a rise in supraglottal pressure; upon release, this is realized as a transient aperiodic energy. Thus [b] is [−sonorant], while [m] is [+sonorant].

The quantal nature of [sonorant] discussed above does not account for those cases described by Clements and Osu (2002), namely non-explosive stop consonants. In these cases, the articulatory constriction is sufficient so that, if modal voicing were to be produced, there would be an increase in supraglottal pressure, but the supraglottal pressure does not increase due to other articulatory strategies and airstream mechanisms employed.[4] For example, an increase in supraglottal pressure occurs with an egressive, but not ingressive, airstream. Sounds designated as [−sonorant], as discussed above, and sounds designated as [+obstruent] are coextensional under egressive airstream conditions, but not for ingressive airstream mechanisms, showing that the quantal nature of sonorant depends on the airstream mechanism.

For example, consider the relationship between, say, a voiced, bilabial, explosive stop [b] and a voiced, bilabial, non-explosive stop [ɓ]. According to Clements and Osu (2002), the difference between [b] and [ɓ] is that the first is characterized by an increase in supraglottal pressure, while the second is not, but this is a property of the articulatory parameter—namely larynx lowering—and not an acoustic region of stability. To show that there exists a quantal relationship between [b] and [ɓ], one must show that the continuously varying

[4]The articulatory adjustments that allow for full closure without an increase in supraglottal pressure related to non-explosive stops is discussed in depth in Clements and Osu (2002).

articulatory parameter, say larynx lowering, gives rise to a quantal response in the acoustic domain. Cross-linguistically, some of the acoustic correlates that implosives exhibit compared to explosives are: increased duration, increased amplitude, a lack of transient burst noise upon release, and an increasing amplitude over the duration of the closure, rather than a decreasing amplitude (Ladefoged, 2003). This establishes that there are regions I and III, but falls short of explicit evidence in favour of a threshold effect, i.e., a region II. Clements and Osu (2002) do not establish this stronger claim, rather relying on the fact that there are clear aerodynamic implications of the articulatory configuration. Future research into the relationship between the articulatory and acoustic domains is important in order to assess whether or not [obstruent] exhibits a quantal relationship as hypothesized by Clements and Osu (2002).[5]

For the voiced fricatives, the quantal relationship between the articulatory and acoustic domains appears to be more straightforward, but is in fact subject to some of the same issues. As explained in Section 1.4.1, the threshold between laminar and turbulent flow is quantified by the *Reynold's number*: low values for the Reynold's number is associated with laminar flow, while high values are associated with turbulent flow. It might therefore seem as though the Reynold's number is exactly the threshold required to show that the relationship between constriction degree and resonance (or, conversely, turbulence) is quantal. As I discussed in Section 1.4.1, however, turbulent airflow depends on both constriction degree and airflow, and so the acoustic output of turbulence vs. resonance depends not just on constriction degree, but also on airflow.

[5] One potential challenge that the explosive and non-explosive stops present is that the acoustic consequence of larynx lowering appears to unfold over time, for example in the decrease or increase of amplitude over the course of the closure (Ladefoged, 2003), suggesting that quantal representations may need to abstract over the time domain. Another possibility for stop consonants is that the relevant acoustic response occurs at the moment of release, such that the presence or absence of a transient noise burst is relevant. Disentangling these possibilities is beyond the scope of the present work.

As emphasized in Stevens and Keyser (2010), Figure 6.1 "is the result of a hypothetical experiment in which just one articulatory parameter is manipulated, with all other articulatory parameters remaining constant. For example, if the articulatory parameter represents the degree of vocal-tract constriction formed by the tongue blade, it is assumed that all other parameters, such as glottal opening, vocal-fold stiffness, stiffness of vocal-tract walls, sub-glottal pressure, etc. remain constant" (pg. 11). In the above example, a quantal relation holds between the articulatory parameter varied (the position of the velum) and the acoustic consequence (resonance), but all other aspects of articulation, such as the airstream mechanism and the position of the larynx, stay the same. This fact, that quantal relations hold "all else being equal", is crucial in understanding how different quantal relations interact with one another, and I argue that the additional complexity introduced by aerodynamic considerations is precisely the tool that we need to disentangle [obstruent] and [sonorant] with respect to their quantal nature.

To summarize so far, QT hypothesizes that it is the *relationship* between the articulatory and acoustic domains that is important: regions of stability arise where small perturbations in the articulatory domain (say, additional lowering of the velum in an already lowered velum) have an equally small effect on the acoustic output (perhaps greater amplitude); a threshold effect arises when small articulatory perturbations give rise to discontinuities, namely large changes in the acoustic response. On this reading, then, defining [sonorant] as an acoustic feature, and [obstruent] as an articulatory feature, does not easily square with quantal theory. The quantal nature of [sonorant], as discussed by Stevens (1989), crucially depends on an egressive airflow in an articulatory/acoustic domain, thereby leaving no room for [obstruent]. Furthermore, the hypothesized quantal relationships hold "all else being equal", which makes it difficult to see how such regions of stability could arise across different segment classes. In the following section I sketch an elaboration of the relevant domains in QT in order to explicitly account for the distinction between [obstruent] and [sonorant].

## 6.4 Expanding the set of domains in QT

For a given constriction degree, adjusting the rate of egressive airflow has distinct acoustic consequences, specifically with respect to the presence or absence of turbulent flow. Moreover, there is a threshold (quantified by the Reynold's number), such that small perturbations in airflow lead to rapid changes in the acoustic domain. Beyond the threshold required to generate turbulent airflow, additional airflow leads to a greater intensity of frication, in a relatively continuous manner. Thus, for a given constriction degree and a pulmonic egressive airstream, there exists a quantal relationship between the *aerodynamic* and *acoustic* domains.

Likewise, for a given place of articulation, and voicing specification (or vocal fold configuration), adjusting the height of the larynx has distinct aerodynamic consequences: a raised larynx causes an increase in oral pressure, while a lowered larynx decreases oral pressure. Although I am not aware of any studies that specifically identify a threshold effect, such that small perturbations in larynx height lead to rapid changes in the aerodynamic domain in combination with a relative stability on either side of the threshold, the existence of stable regions suggests that such an effect is likely to be found. Thus, for a given place of articulation and laryngeal configuration, we may hypothesize that there exists a quantal relationship between the *articulatory* and *aerodynamic* domains.

The preceding paragraphs show quantal relations exist between other domains. Specifically, the clause that quantal relations hold "all else being equal" obscures the fact that quantal relations may exist between multiple domains: articulatory-acoustic, articulatory-aerodynamic, aerodynamic-acoustic, and potentially many others. Although QT is formulated over the articulatory-acoustic (and the acoustic-auditory) domains, it is not, in principle, limited to these. By recognizing that quantal relationships may exist across any number of domains, there is space for a greater number of features grounded in regions of stability.

In showing that [sonorant] is quantal, Stevens (1989) refers to constriction degree, supraglottal pressure, and the acoustic resonance/turbulence that results. The introduction of [obstruent] teases apart these various components, and Clements and Osu (2002) argue that [sonorant] is an acoustic feature, while [obstruent] is an articulatory feature. One possible way to connect QT with the hypothesis put forth by Clements and Osu (2002) is to propose that [sonorant] and [obstruent] reflect quantal relationships across different domains. For example, [sonorant], which is standardly thought of as an acoustic feature, may be best thought of as residing in the acoustic-perceptual or acoustic-auditory domain; [obstruent] therefore likely resides in the articulatory-acoustic or aerodynamic-articulatory domain.

It is beyond the scope of this work to assess the precise quantal nature of [obstruent] and [sonorant]: more detailed phonetic models regarding the relationships between different domains are required, and further empirical work is required to see if the relationships across various domains can be teased apart in the lab. Nevertheless, this discussion has shown that it is conceptually possible to expand the set of domain relationships within a QT framework, and that doing so allows for both [obstruent] and [sonorant] to be defined in quantal terms.

The problem with this approach is, of course, the explosion of new features that this approach allows. QT hypothesizes that regions of stability correspond to universal distinctive features:

> We hypothesize that a quantal acoustic/articulatory relation underlies each distinctive feature, and consequently each feature can be said to be based on a defining articulatory range and a defining acoustic attribute. This acoustic attribute may depend upon the associated articulator-free feature of the segment.

> These defining attributes are properties of the human speech production system and are expected to be universal in language. It is hypothesized that the human speech production system is structured in such a way that the sounds that it can generate and the articulatory attributes that produce these sounds define a set of quantal states. (Stevens and Keyser, 2010, pg. 15)

Expanding the quantal relations admitted to the theory greatly increases the number of distinctive features, weakening its predictive power in a seemingly unconstrained fashion. In the next section I propose that regions of stability are not themselves features, but are only potential features, and that features are induced by the learner through acquisition.

## 6.5 Regions of stability are not distinctive features

Section 6.4 showed that, in principle, quantal relations hold across many different domains, not just the articulatory-acoustic and acoustic-auditory. This is particularly salient for the quantal relationships that define turbulent airflow: either articulation or airflow can be held constant, with the other as the varying parameter on the $x$-axis. QT posits that the articulatory-acoustic domain is the grounding for distinctive features, but *a priori*, the aerodynamic-acoustic domain can also function in this respect.[6] Conceptually, quantal relations can hold between many other pairs of domains, which is sufficient to show that the articulatory-acoustic grounding for QT is not necessary.

QT hypothesizes that regions of stability are features, and that these features are universal because they are "properties of the human speech production system". Indeed, to the extent that quantal relations are determined by physics and biology, regions of stability

[6]There is a long tradition within phonology to prioritize articulatorily defined features, and on balance, articulatory parameters appear to be more readily exploited than, say, acoustic cues (for discussion on this point, see Halle (1983)). Nonetheless, the point made here is orthogonal to this question: the articulatory-acoustic domain may well be most robust for many segment-feature pairs, but, as I address in the following section, this fact can be accounted for by an emergentist view of features.

must be, in some sense, universal. Nevertheless, I reject the claim that regions of stability are themselves features. Identifying regions of stability with features confuses the behaviour of features with their phonetic grounding (as, for example, argued by Reiss (2018)).[7]

The main argument against identifying regions of stability with features, however, is that what might be stable for one acoustic parameter may not be stable on another acoustic parameter. Recall that the results of Chapter 2 and Chapter 3 showed that while /v/ exhibited the highest harmonicity across languages, the distribution of energy in the high-frequency range differed considerably across languages, and in the case of Russian, by environment. Which acoustic parameter defines [v] as opposed to [z]: its harmonicity, spectral centroid, or some other measure not considered here? One could imagine that, given the general robustness of harmonicity across environments and languages, it could to be a good candidate on which to base a quantal relation, but doing so would completely obscure the important differences within and across languages that were revealed by the spectral centroid.

If, however, we adopt the view that quantal relations can exist between many different domains, beyond those specified by the original formulation of QT, then we do not have to specify the acoustic measure. And, if we allow features to be emergent, then the importance of regions of stability is that they are the basis on which features are defined. In order to make this more precise, we first have to recognize that feature emergence looks different depending on which segment-feature pairs are under consideration. This is the subject of the next section.

[7]See Cohn (2011) for an overview of phonological primitives and the different assumptions at play in emergentist vs. universalist perspectives.

## 6.6 Emergence of universality

In this section I propose that features are *emergent*, building on a growing consensus within the phonological literature that features are learned or induced over the course of acquisition (Mielke, 2005, 2008; Blevins, 2007; Dresher, 2014). I propose to model this as follows: regions of stability are fixed points on which features can be grounded, but are not themselves features; nevertheless, their stability with respect to certain segment-feature pairs gives the impression of universality, and constrains ambivalent and ambiguous feature specifications. The existence of multiple, possibly conflicting, quantal relationships means that the specific dividing line that makes a segment either [+F] or [−F] must be induced by the learner.

An important component of the model is that emergence looks different depending on the feature and segment in question. To make this explicit, I refer not to the emergence of features alone, but to the emergence of *segment-feature pairs*, where what emerges is the specification of a particular feature for a particular segment. To illustrate, consider the voiceless, unaspirated, bilabial stop [p]. For any reasonable realization of [p] (i.e., ignoring serious mispronunciations), certain features seem uncontroversial and, in universalist approaches to distinctive features, would be candidates for universal features, as in [labial], [−continuant], or [−sonorant]. On the other hand, whether a voiceless, unaspirated [p] is featurally specified to group with the *lenis* or *fortis* series of consonants does not appear to be universal, and indeed has been the subject of much research regarding the appropriate representation of voicing in true voicing vs. aspirating languages (*c.f.* Iverson and Salmons (2011)). In Greek, a true voicing language, the segment [p] is part of the fortis series, namely the set of [−voice] consonants, grouped together with clearly voiceless [f, s]; in English, an aspirating language, the segment [p] is part of the lenis series, namely the set of [+voice] consonants (at least in word-initial position), grouped together with clearly voiced [v, z].[8]

With respect to the specification of [p] as, say, [−sonorant], there is a clear phonetic grounding, and small perturbations in the articulatory or acoustic domain would not change the sonority of [p]. In this sense, the *sf-pair* {[p], [sonorant]} resides in a region of stability. Similarly, the *sf-pair* {[v], [labial]} is also quantal with respect to place of articulation, in that small perturbations in the articulatory domain would not change the labiality of [v]. The stability, or robustness, of certain sf-pairs, gives the appearance of universality, but more importantly, it gives features something to stick their proverbial teeth into. Quantal relations identify a particular parameter along which there is a non-linearity in the response variable, such that a binary division can be inferred.

When multiple quantal relations converge over several sf-pairs, the appearance of universality increases. For many consonants, the regions of stability associated with [±sonorant] and [±obstruent] will pick out the same sets of segments, so that consonants such as [p, b, s, z] are both [+obstruent] and [−sonorant], and consonants such as [m, l, w] are both [−obstruent] and [+sonorant]. Figure 6.1 schematizes the relationship between regions of stability across multiple domains, and their "promotion" to features. Under this view, the notion of exclusivity might better be thought of as "robustness". Multiple quantal relations group the same sets of segments together, meaning that multiple sources of evidence converge on the same grouping and classification.

Certain sf-pairs exhibit what Mielke (2008) calls *ambivalence*, meaning that for a given feature [X] the same segment can be specified as [+X] in one language, and as [−X] in another language. For example, Finnish laterals pattern with continuants (Mielke, 2005, pg. 180), but laterals often pattern with non-continuants as well, as in Basque (Mielke, 2005,

[8]For expository purposes I gloss over other possible laryngeal features.

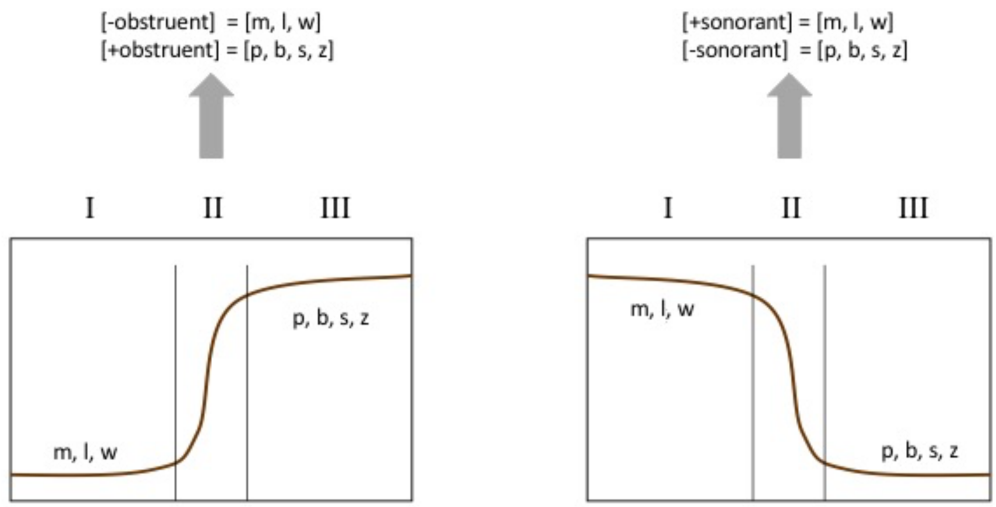



Figure 6.1: Schematization of feature emergence: multiple quantal relationships agree

pg. 179). The ambivalence of classification for /l/ with respect to [continuant] is reflected in different feature theories: for example, Chomsky and Halle (1968) group laterals with [+continuant] sounds, while Halle and Clements (1983) classifies laterals as [−continuant]. The difficulty of determining the continuant status of laterals was noted in Chomsky and Halle (1968), in a discussion of the [continuant] classification of liquids:

> The characterization of the liquid [l] in terms of the continuant-noncontinuant scale is even more complicated. If the defining characteristic of the stop is taken as total blockage of air flow, then [l] must be viewed as a continuant and must be distinguished from [r] by the feature of 'laterality'. If, on the other hand, the defining characteristic of stops is taken to be the blockage of air flow *past the primary stricture*, then [l] must be included among the stops. (pg. 318)

In this case we are faced with two different ways to define [continuant]. The two definitions agree on the edge cases: segments such as [p, t, k] are [−continuant], and segments such as [a, i, j, w, s, f] are [+continuant], regardless of which definition is used. In terms of quantal theory, these two definitions may reflect two different quantal relations, or they may reflect different interpretations on the articulatory domain; this is schematized in Figure 6.2.[9]

[9]It is beyond the scope of this work to determine what the quantal relations for [continuant] might be.

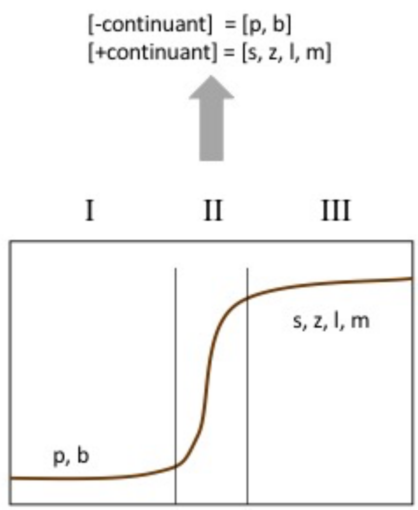



Figure 6.2: Schematization of feature emergence: ambivalence

Equating regions of stability with features is misleading precisely because it prioritizes certain quantal relations over others. In other words, for any feature [F], there are segments that are clearly [+F], and other segments that are clearly [−F], suggesting that a quantal relationship underlies this division. By focussing on the grey areas, as Mielke (2008) does with [continuant] and as I do in this dissertation with the relationship between /v/ and [sonorant], we see that for many features (if not all of them), there are segments that are not so readily identified as either + or −. Rejecting the claim that regions of stability are themselves features, and recognizing that features, as phonological objects, are categories induced by the learner by analogy to the robust cases, shows that quantal relations and emergent feature theory are not in conflict.

Returning to [sonorant] and [obstruent], for most segments, the regions of stability pick out the same sets of segments, but for non-explosive stops and voiced spirants, the quantal

One, albeit speculative, possibility is as follows: If continuance is defined based on the total blockage of airflow, then this may reside in the articulatory-acoustic domain; if continuance is defined based on just the obstruction of airflow in the oral cavity, then the response domain might be something like a somatosensory domain, where the experience of total oral constriction may be paramount. Regardless, what is important is that the two definitions agree at the end points, thus providing phonetic grounding for [continuant].

relations picked out by [sonorant] and [obstruent] do not agree. This is schematized in Figure 6.3, with the voiced labiodental continuant [v] and the voiced bilabial implosive [ɓ].[10]

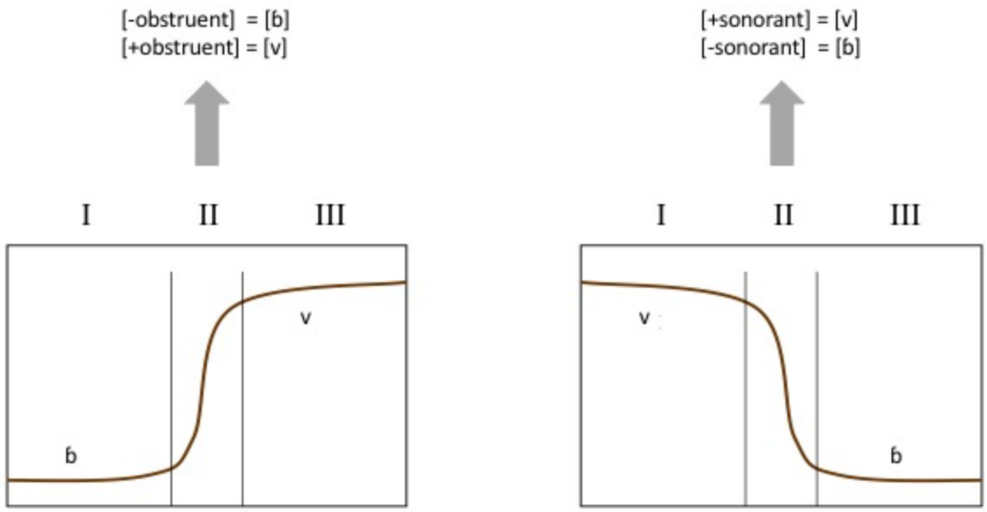



Figure 6.3: Schematization of feature emergence when multiple quantal relationships disagree

Not accounted for in the present model is the fact that quantal theory only accounts for a certain kind of "low level" emergence—namely, the bias that regions of stability introduce to ultimate feature specifications. As shown in Chapter 2, [v] generally exhibits some degree of phonetic intermediacy, regardless of the language in question. Tokens of [v] in all four languages had significantly higher harmonicity than [f, s, z], and, moreover, the harmonicity of [v] tokens across Greek, Serbian, and Russian did not differ.[11] With respect to harmonicity, [v] is clearly the most sonorant-like of [f, v, s, z], and it is not hard to see that including stops, voiceless fricatives and voiced sibilants would not change the "outlier status" of [v] with respect to the obstruent inventory. In other words, for most inventories, [v] will be the most harmonic (and most sonorous) segment, unless it is accompanied by other voiced

[10]Notice that a consequence of this view is that if an inventory lacks segments such as [v, ɓ], then whether or not [sonorant] or [obstruent] is the operative feature does not arise, highlighting the importance of research on less common segments.

[11]In English, in which /v/ and /w/ contrast and [v] tokens were produced with more devoicing than [v] tokens in Greek, Russian, and Serbian, there was quite a bit of overlap between the harmonicity values of [v] and [w].

spirants in the inventory. On the other hand, across all four languages, [v] was also produced with some degree of high frequency energy: even in Serbian, the spectral centroid on the 1500 Hz high pass filtered signal hovered near 4000 Hz, suggesting that at least some turbulence (and hence some increase in supraglottal pressure). Again, it is not hard to see that if one were to measure the spectral centroid of other sonorants such as nasals, liquids (barring some kinds of rhotics) and glides, [v] would be the least sonorant-like.[12] To summarize, the fact that [v] tokens always exhibit high harmonicity compared to other fricatives, and also exhibit some evidence of frication, means that [v] tokens are *candidates* for being specified as [+sonorant, +obstruent] regardless of language.

If [v] tokens are candidates for intermediacy regardless of language, then how do we account for the cross-linguistic differences between Greek and Serbian? The results of Chapter 2 suggest a solution. In languages such as Greek and Serbian, the phonetic cues across environments are stable. Perhaps more importantly, their *relationship* to other phonetic cues is stable. In Greek and English, [v] is reasonably considered the voiced counterpart of [f], because the relationship between [v] and [f] parallels the relationship between [z] and [s] for duration and spectral centroid on the high-pass filtered signal, across both environments tested. In Serbian, [v] cannot be said to be the voiced counterpart of [f] based on any measure, across both environments tested. These results suggest that, insofar as features are expressions of relationships between segments, the stability of the relationship between segments promotes one featural analysis over another. On the other hand, the acoustic parameters for [v] in Russian vary significantly due to environment, and this variability in the relationship, or perhaps the lack of stability, may contribute to the specification of Russian

[12] In general, measuring the spectral centroid on resonant sounds is not interpretable due to the multiple formant peaks. Nevertheless, such a measure can be taken, as can be seen for the nasal [m] in Bjorndahl (2012). Based on a measure of relativized spectral centroid, the spectral centroid calculated on the 1500 Hz high pass filtered signal for [m] is predictably very low, since most of the energy for [m] is located below 1500 Hz.

/v/ as [+sonorant, +obstruent].

This solution elevates the relationship between acoustic cues and the stability of these cues across environments to be the determining factors in assigning featural specifications, and leaves little room for the role of phonological patterning. This cannot be the whole story however, since the valuation of features [sonorant] and [obstruent] extend, of course, to segments beyond /v/. Indeed, the main advantages of features is that they allow classes to be defined in such a way as to abstract away from differences that are not "relevant" to the parameter in question. In spite of the profound differences in how voicing cues are realized in, say, stops and fricatives, features allow us to articulate statements such as /b, v, z/ are the voiced counterparts of /p, f, s/, expressed as [+voice] and [−voice], respectively. Future research is needed to assess whether the relationship between, say, voiced and voiceless stops can be quantified as in Chapter 2 for sibilants and spirants. I propose that the differences between Greek and Serbian are a combination of phonetic and phonological factors. Phonetically, the spectral centroid of Greek [v] exhibits the same relationship to [f] as seen between [z] and [s]; phonologically, Greek /v/ distributes as the voiced counterpart to /f/. Similarly, in Serbian the classification of /v/ as [+sonorant, −obstruent] is likely a combination of the fact that, both phonetically and phonologically, [v] is not the voiced counterpart to [f].

## 6.7 Discussion

In *The Sound Pattern of English*, Chomsky and Halle distinguish between *phonetic distinctive features* and *phonological distinctive features*, where the former are the "universal physical scales that determine the rows of the phonetic matrices", while the latter are "the categories that label the rows of the phonological matrices" (pg. 169). The two types of distinctive features are distinguishable precisely because of the specific grammatical architecture proposed, but as Cohn (2011) notes, and as had been noted by Chomsky and Halle themselves, this distinction is often not observed. In addition to the two roles that distinctive features play between phonology and phonetics, they are often considered "linking objects", say between multiple phonetic domains. For example, Halle (1983) proposes "that the abstract distinctive features constitute the link between specific articulatory and acoustic properties of speech sounds" (pg. 94). In this case, distinctive features do not form the link between the phonetic and phonological realms, but rather act as a linking mechanism between the articulatory and acoustic domains. The fact that distinctive features are linking objects means that they carry an enormous explanatory burden, as Cohn (2011, pg. 20) articulates:

> There is a critical shift from *PSA*[13] to SPE, in which the universally defined set of features accounted for more than the contrasts of phonology. Distinctive features characterize the contrastive elements of phonology as well as natural classes in their binary classificatory function. They are also the basis of phonetic implementation when translated into language-specific scalar values at the output of the phonology and then implemented by an automatic universal phonetic component.

The general linking character of distinctive features is also seen in work that defines features substantively, but then employs said features abstractly. For example, Rice (1993) defines the Sonorant Voice node as an articulatory configuration that allows for spontaneous voicing to occur, thereby defining SV substantively. However, a crucial component of the SV hypothesis is that obstruents are specified for SV when they pattern with sonorants.[14]

[13] *Preliminaries to Speech Analysis* (Jakobson et al., 1952).

[14] An example that is related to the concerns of this dissertation comes from the analysis of /v/ in Hall (2007). Hall argues that in some dialects of Czech, /v/ is specified as SV, but in others, it is not specified as SV (it receives no specification either for SV or Laryngeal Voice in these dialects). Such representational differences are based on the patterning of /v/; crucially, they are not based on whether /v/ in either of these dialects is produced with the articulatory configuration that corresponds to spontaneous voicing, nor whether /v/ tokens in these dialects are produced differently from each other.

The multiple roles that features play lies at the heart of the phonetics-phonology interface: how are abstract, mental categories and concepts *grounded* in the physical domain to which we have (more) direct access? In *SPE*, the grounding of distinctive features was accomplished by universal phonetics, but as Cohn (2011) summarizes, the literature on language-specific phonetics has disproved universal phonetics in a strict sense: "The SPE view of phonetic distinctive features does not provide a sufficient interface with phonetics, as there is not a universal set of phonetic elements (phones) equivalent across languages defined by the set of phonological distinctive features" (pg. 23). As I discussed in Section 6.2, similar criticisms have been levied against QT, which also attempts to formulate distinctive features universally by appeal to regions of stability. In this chapter I have argued that problems of universal phonetics can be circumvented by distinguishing between regions of stability and phonological features if we assume that the relationship between the two is one of emergence. Moreover, I argued that the difference between natural and unnatural classes need not be hard-coded in the grammar, but rather follows from how regions of stability of different quantal relations converge in the former but not the latter.

The goal of the present chapter has been to probe the relationship between the hypothesized quantal nature of [sonorant] and [obstruent], and consider these issues in light an emergentist view of features. My reading of EFT is that it is an attempt to resurrect the distinction between the phonetic and phonological identities of distinctive features. In this chapter I have tried to elaborate on this distinction in one specific way: how might the quantal nature of features [sonorant] and [obstruent] contribute both to robust cases of feature assignment, as well as cross-linguistically ambivalent or ambiguous segment patterning. To the extent that the proposal sketched here is on the right track, several future avenues of research are suggested, particularly with respect to what it means to learn phonetic vs. phonological categories. There has been much recent work on phonetic category learning, particularly within Bayesian modelling approaches and Exemplar Theory (Pierrehumbert, 2001; Shi et al., 2010; Richardson et al., 2015; Antetomaso et al., 2017), but only recently have such approaches begun to seriously venture into phonological categories as abstractions (*c.f.* Pierrehumbert (2016)). Further research into the learning and acquisition of [sonorant] and [obstruent]—both as phonetic categories that exhibit quantal relations, and as phonological categories that account for the patterning of segments under voicing phenomena—is likely to contribute to our understanding of the phonetics/phonology interface.

CHAPTER 7

# CONCLUSION

The motivation for this dissertation came from the anomalous patterning of a single segment in a single language, the patterning of Russian /v/ with respect to voicing assimilation. Although its dual status as an obstruent and sonorant has long been recognized in the phonological literature, its patterning has always been treated as, essentially, idiosyncratic: even with the recognition that /v/ exhibits similar asymmetries in other languages like Hungarian and Hebrew, the existence of such patterning has not been treated as a central concern of laryngeal phonology. I have taken a different approach, placing /v/ at the centre of the phonetic, typological, and phonological investigations in Chapters 2–5. Rather than ask how phonological theory might account for the patterning of /v/, I have instead been motivated by the question, "What can /v/ tell us about phonology and its interface with phonetics?"

To set the stage for these investigations, in Chapter 1 I distinguish between sibilant and spirant frication, and the different ways that voicing interacts with each. Importantly, the introduction of the term *voiced spirants* for the voiced, non-sibilant fricatives /β, v, ð, ɣ/ conveys the fact that the voiced spirants and voiced sibilants do not function as a unified class, phonetically, phonologically, or typologically.

The phonetic investigations of Chapters 2 and 3 focus on the behaviour of [v] as potentially representative of the entire class of voiced spirants. Each of these chapters serves to situate [v] in a broader landscape, acoustically speaking. Four languages are investigated: English, Greek, Serbian, and Russian. Greek, Serbian, and Russian represent a typology of patterning of /v/, where in Greek /v/ patterns with obstruents, in Serbian with sonorants, and in Russian with both; importantly, none of these languages has a labial approximant such as /w/ against which /v/ might contrast. In English, /v/ is considered an obstruent, and the inventory also has /w/. Two environments are used in order to assess the effect of word-internal prosody on sibilant and spirant realization. All words were bisyllabic, and of the form ˈ$C_1VC_2V(C)$; segments in $C_1$ position are in the prosodically strong word-initial stressed (WIS) environment; segments in $C_2$ position are in the prosodically weak word-medial unstressed (WMU) environment.

In Chapter 2, the goal is to understand the relationship between voicing (voiceless vs. voiced) and type of frication (spirant vs. sibilant) using the four segments /f, v, s, z/. Although it is well known that voicing modulates frication, it is less well understood how the relationship between voicing and type of frication is manifested, and how this relationship might differ from language to language. Phonologically, /f, v/ and /s, z/ are considered to be obstruent pairs; featurally, /v, z/ are the [+voice] counterparts of /f, s/. Chapter 2 probes whether the relationship between /f, v/ parallels that of /s, z/, as well as whether this relationship changes as a result of word-internal prosody, tested using the WIS and WMU environments.

Broadly speaking, the results of Chapter 2 show that the relationship between voicing and frication type in English and Greek is independent (though the particular form of the relationship differs) in both environments; in other words, the relationship between [v] and [f] parallels that between [z] and [s]. The relationship between voicing and frication type in Serbian is not independent, as the relationship between [v] and [f] does not parallel that between [z] and [s] in either environment. In Russian, both independence and non-independence are observed: the relationship between [v] and [f] parallels the relationship between [z] and [s] in the strong WIS environment, but not in the weak WMU environment. The results of Chapter 2 show that the phonological relationship of an obstruent pair can indeed be reflected in the phonetics. The fact that voicing and frication type are independent in English and Greek are interpreted to show that the voiced fricatives [v, z] are the voiced counterparts, in a phonetic sense, of [f, s]; this reflects the phonological identity of /v/ as the voiced counterpart to /f/. In Serbian, [v] is not the voiced counterpart of [f] for any measure, for any environment, and reflects the fact that phonologically /v/ is not the voiced counterpart of /f/. The correlation between phonetic and phonological relationships breaks down for Russian /v/, which has a greater range of realizations determined by prosodic factors. Though [v] appears to be the voiced counterpart of [f] in the WIS environment, but not in the WMU environment, these environments do not directly correspond to the phonological ambiguity of Russian /v/. Future research is needed to determine how variability in realization is related to phonological identity, but the methodology and results of Chapter 2 argue for a new approach to the quest for phonetic correlates of phonological categories and relationships. The results further show that focussing on the relationship between segments, in this case [v] and [f], is more revealing of phonetic identity than focussing on the realization of the segment alone. For example, in Serbian, the average spectral centroid of [v] (or rather, [ʋ]) tokens hovers just below 4000 Hz, while for English it hovers just above 4000 Hz. Nevertheless, the difference in the relationship between [f, v] in English vs. Serbian is quite stark. In other words, both English and Serbian [v] tokens exhibit some degree of high frequency energy, but the relationship that [v] tokens exhibit to the overall phonetic system reveals a clear distinction.

Chapter 3 turns the attention to a cross-linguistic comparison of [v] tokens in the three languages whose inventories lack [w]: Greek, Russian, and Serbian. Specifically, Chapter 3 tests the hypothesis that the phonetic identity of /v/ correlates with its phonological identity. The results support a partial correlation, with the relativized centroid values for Greek and Serbian /v/ distinguished in both the WIS and WMU environments. Russian and Greek do not differ significantly in the WIS environment, while Russian and Serbian do not differ significantly in the WMU environment. The results therefore do not support the hypothesis articulated by Padgett (2002), that Russian [v] tokens are inherently intermediate between (Greek) obstruent [v] tokens and (Serbian) sonorant [ʋ] tokens. These results strongly suggest that, although Russian [v] tokens are not produced as inherently intermediate between a fricative [v] and an approximant [ʋ], Russian /v/ does exhibit a particular kind of intermediacy because of its greater variability spanning more fricative-like and more approximant-like realizations. These results challenge us to consider how phonological identity and phonetic identity match up. The partial correlation found for Greek and Serbian support the thesis that phonological representations have phonetic correlates, but the results for Russian reveal that the relationship can be quite subtle. For example, Russian [v] in the WIS environment patterns phonologically as a sonorant in failing to trigger regressive voicing assimilation, but is nonetheless realized as more fricative-like, akin to Greek. Further research is needed to determine how the range of realizations might support (or be supported by) the phonological representation and identity of a sound.

Beyond the results of the phonetic investigations, there is a methodological moral that is theoretically important because of claims in the literature that Russian /v/ is "intermediate". Intermediacy can be interpreted in (at least) two ways. First, /v/ might be intermediate (to sonorants and obstruents) *within* a given language; second, /v/ might be intermediate (to other voiced labiodental continuants) *across* languages. Both of these issues are addressed in the design of the phonetic investigations presented in Chapters 2 and 3. What emerges from these results is that /v/ in particular, and by implication, voiced spirants as a class, are inherently intermediate, but how their intermediacy is resolved on a language-specific basis will depend on a host of other factors.

Chapter 4 investigates the class of voiced spirants from a typological perspective, focusing on their relative markedness with respect to other types of segments. Voiced fricatives are generally considered *marked*, due to the aerodynamic tensions involved in generating both voicing and frication. The markedness of voiced fricatives is reflected in their distribution in consonant inventories, where it has been claimed that voiced fricatives do not tend to surface in small inventories, and that a voicing contrast in the fricatives implies a voicing contrast in the stop system. Previous literature investigating these relationships has taken for granted that voiced fricatives pattern as a homogenous class. However, the results of Chapter 4 show that the voiced spirants have a different typological identity than voiced sibilants such as /z, ʒ/. Voiced spirants frequently surface in small inventories, but only if unpaired, while the correlation between the presence of voiced sibilants (which only occur with a voiceless counterpart) and inventory size is linear. Voiced spirants also tend to appear in inventories without a voicing contrast in the stops, unlike voiced sibilants. These results show that claims about voiced fricatives are in fact claims about the voiced sibilants. The two investigations conducted, on inventory size and the relationship to stop voicing, are not independent: languages without a voicing contrast in the stops are likely to have small inventories. As a result, it is not possible to determine if the presence of unpaired voiced spirants is due to inventory size or the lack of contrast in the stops, but the general picture that emerges is that unpaired voiced spirants surface in inventories with a lot of "space". I leave it as a topic of future research to understand the principles of inventory structure that explain why voiced spirants appear unpaired in some inventories, but paired in others. In sum, this chapter shows that, with respect to the distribution of voiced fricatives in consonant inventories, the voiced spirants and voiced sibilants do not form a unified class.

In Chapter 5, I address how voiced spirants, and /v/ in particular, should be represented in terms of features. Russian /v/ is a problem for phonological theory because phonological theory has insisted on a strict binary division between segments that are [−sonorant] and those that are [+sonorant]. The overarching conclusion to be drawn from these investigations is that the traditional conception of the obstruent-sonorant divide is overly simplistic, and misleading because it prioritizes the clear cases of sonority, rather than the grey areas. The binary division captured by the feature [±sonorant] is not sufficient to account for the phonetic, phonological, and typological identity of segments that inhabit the grey area between canonical obstruents and canonical sonorants, namely, the voiced spirants.

In Chapter 5 I argue for a feature [obstruent], building on work by Clements and Osu (2002), and I propose that Russian /v/ is featurally specified as [+sonorant, +obstruent]. The Russian data are neatly accounted for under such a specification: all segments that are [+obstruent] are targets of devoicing under regressive voicing assimilation and final devoicing, but only segments that are *exclusively obstruents* trigger RVA. The existence of [obstruent] is independently motivated by the patterning of non-explosive stops in Ikwere, but I reject the proposal made by Clements and Osu (2002) that [+sonorant, +obstruent] is universally banned, arguing instead that the lack of a three-way contrast between [−sonorant, +obstruent], [+sonorant, +obstruent] and [+sonorant, −obstruent] is due to perceptual factors. In addition to the analysis of Russian /v/, I proposed that both [−sonorant, +obstruent] and [+sonorant, −obstruent] featural configurations of /v/ are attested. In languages like Greek or Maltese, where /v/ patterns with obstruents, it is featurally specified as [−sonorant, +obstruent]; in languages like Serbian, where /v/ patterns with sonorants, it is featurally specified as [+sonorant, −obstruent]. The configuration [−sonorant, −obstruent] is impossible for voiced spirants. Finally, I introduce the notion of *exclusivity*, and I argue that [+sonorant, +obstruent] is neither intermediacy nor a third category, but rather reflects a lack of exclusivity in categorization as an obstruent or a sonorant.

The proposal to include [obstruent] is further justified by the burden placed on [sonorant], which has traditionally been defined as comprising a bundle of articulatory, aerodynamic, and acoustic properties. In Chapter 6, I argue that [obstruent] and [sonorant] are phonologizations of different quantal relations that overlap for many cases, but differ precisely in their classification of voiced spirants and non-explosive stops. The fact that certain quantal relations identify the same segment classifications leads to the notion of *exclusivity*, where segments such as [m, l, j] that are specified as [−sonorant, +obstruent] are *exclusive sonorants*, and those such as [p, s, ʒ] are *exclusive obstruents*. Voiced spirants, which are generally [+sonorant, +obstruent] and non-explosive stops, which are [−sonorant, −obstruent] are neither exclusive sonorants nor exclusive obstruents. While there is much empirical work to be done to determine the precise quantal relations, the proposed model provides a rigorous basis for including both [obstruent] and [sonorant] in an emergentist framework. One interesting avenue for future research is to determine whether there is a relationship between the robustness of the feature classification and the acquisition of phonological features. In the framework outlined in Chapter 6, the inherent ambiguity of voiced spirants comes from the fact that their underlying quantal relationships suggest conflicting classifications that must be resolved by the learner, and we might therefore ask how segment classification unfolds over the course of acquisition. If the model outlined is on the right track, then we predict that the feature assignments of segments where the quantal relationships suggest non-conflicting classifications should be acquired earlier than those with conflicting classifications. The patterning of exclusive obstruents and sonorants should be more stable than the patterning of non-exclusive segments such as /v/, where stability might be measured in terms of variability or the ability to generalize to new lexical items. Future research is needed to investigate this hypothesis.

The problem of Russian /v/ touches upon a wide range of issues in phonology and phonetics. The scope of these issues is reflected in the literature on Russian /v/, which encompasses a range of theoretical topics and approaches (Hayes, 1984; Kiparsky, 1985; Padgett, 2002; Kiss and Bárkányi, 2006; Lulich, 2004; Hall, 2007). In other words, the problem of Russian /v/ has attracted so much interest because it has served as a battleground to test ideas about phonological architecture, the relationship of phonological representation to phonetic realization, and the explanatory roles of phonology and phonetics in accounting for segmental patterning. In many ways, because these approaches stay within certain theoretical assumptions, namely, that there is a binary division between obstruents and sonorants, they can only partially account for the patterning of Russian /v/, and cannot address the cross-linguistic identity of /v/ as part of the larger class of voiced spirants.

Giving equal consideration to the patterning of segments with robust categorizations as well as to those in the grey areas carries with it other implications for phonological theory. In particular, the nature of voicing assimilation must be revisited. If trigger and target classes are inherently different, as I have proposed by classifying targets as [+obstruent] and triggers as [−sonorant, +obstruent], then voicing assimilation is itself inherently asymmetric. In this case, the asymmetries seen in voicing assimilation likely reflect the phonologization of coarticulatory asymmetries, of the kind proposed by Lulich (2004). Understanding what factors influence the phonologization of such asymmetries (as in Russian), as well as what factors override such asymmetries (as in either Serbian or Maltese) is left as a topic for future research.

The subtitle of Padgett's 2002 article on Russian /v/ is "The mouse that squeaked", and the article begins with the sentence, "Like the mouse that roared, the Russian consonant [v] has a status in phonology out of proportion to its size". But, while /v/ is small, the obstruent-sonorant divide has loomed large in phonology. Even a mouse's tiny squeak released in the chasm of the obstruent-sonorant divide might find itself echoing to the level or a roar. Future research into the voiced spirants and other segments that inhabit the obstruent-sonorant divide promises to further our understanding of the emergence of categoricity.

## BIBLIOGRAPHY


Alexiou, Margaret. 1982. Diglossia in Greece. In *Standard languages: spoken and written*, ed. William Haas. Manchester University Press.

Ambrazas, Vytautas, Emma Geniušienė, Aleksas Girdenis, Nijolė Sližienė, Dalija Tekorienė, Adelė Valeckienė, and Elena Valiulytė. 2006. *Lithuanian grammar*. Baltu lanku leidyba.

Andersen, Henning. 1969. The phonological status of the Russian 'labial fricatives'. *Journal of Linguistics* 5:121–127.

Antetomaso, Stephanie, Kouki Miyazawa, Naomi Feldman, Micha Elsner, Kasia Hitczenko, and Reiko Mazuka. 2017. Modeling phonetic category learning from natural acoustic data. In *Proceedings of the 41st Boston University Conference on Language Development*.

Árnason, Kristján. 2011. *The phonology of Icelandic and Faroese*. Oxford University Press.

Arvaniti, Amalia. 2007. Greek phonetics: The state of the art. *Journal of Greek Linguistics* 8:97–208.

Avery, J. Peter. 1996. The representation of voicing contrasts. Doctoral Dissertation, University of Toronto.

Baayen, R. H. 2008. *Analyzing linguistic data: A practical introduction to statistics*. Cambridge University Press.

Baayen, R.H., D.J. Davidson, and D.M Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59:390–412.

Baltazani, Mary. 2007. On s-voicing in Greek. In *Proceedings for the 7th International Conference on Greek Linguistics*.

Barkai, Malachi, and Julia Horvath. 1978. Voicing assimilation and the sonority hierarchy: evidence from Russian, Hebrew and Hungarian. *Linguistics* 212:77–88.

Barr, Dale J, Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68:255–278.

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. 2015a. Parsimonious mixed models. *arXiv preprint arXiv:1506.04967* .

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015b. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48.

Bates, Douglas M. 2010. *lme4: Mixed-effects modeling with R*. Springer.

Beckman, Jill, Michael Jessen, and Catherine Ringen. 2006. Phonetic variation and phonological theory: German fricative devoicing. In *Proceedings of the 25th West Coast Conference on Formal Linguistics*, ed. Donald Baumer, David Montero, and Michael Scanlon, 76–86. MA: Cascadilla Proceedings Project.

Beckman, Jill, Michael Jessen, and Catherine Ringen. 2009. German fricatives: coda devoicing or positional faithfulness? *Phonology* 26:231–268.

Bethin, Christina. 1984. Voicing assimilation in Polish. *International Journal of Slavic Linguistics and Poetics* 17–32.

Bjorndahl, Christina. 2012. The phonology and phonetics of ambiguity: A case study of /v/. Poster presented at *The 13th Conference on Laboratory Phonology (LabPhon 13)*, July 2012.

Blevins, Juliette. 2007. *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Blevins, Juliette. 2018. Evolutionary phonology and the life cycle of voiceless sonorants. In *Typological hierarchies in synchrony and diachrony*, ed. Sonia Cristofaro and Fernandoñiga Zúñiga, 29–60. Benjamins.

Boersma, Paul. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences*, volume 17, 97–110.

Boersma, Paul, and David Weenink. 2011. Praat: doing phonetics by computer [Computer program]. `http://www.praat.org/`. Version 5.2.22.

Bolozky, Schmuel. 1978. Some aspects of Modern Hebrew phonology. In *Modern Hebrew structure*, ed. R.Ãronson Berman, chapter 2, 11–67. Tel Aviv: Universities Publishing Projects.

Bolozky, Schmuel. 1997. Israeli Hebrew phonology. In *Phonologies of asia and africa*, ed. Alan S. Kaye, volume 1, 287–311. Winona Lake: Eisenbrauns.

Bolozky, Schmuel. 2006. A note on initial consonant clusters in Israeli Hebrew. *Hebrew Studies* 47:227–235.

Bolozky, Schmuel. 2013. Phonology: Israeli Hebrew. In *Encyclopedia of Hebrew language and linguistics*, ed. Geoffrey Khan, Shmuel Bolozky, Steven E. Fassberg, Gary A. Rendsburg, Aaron D. Rubin, Ora R. Schwarzwald, and Tamar Zewi, volume 3: P – Z, chapter 113-122. Koninklijke Brill.

Borg, Alexander. 1997. Maltese phonology. In *Phonologies of Asia and Africa (including the Caucasus)*, ed. Alan S. Kaye, 245–285. Winona Lake, IN: Eisenbrauns.

Botma, Bert. 2011. Sonorants. In *The Blackwell Companion to Phonology*, ed. Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, chapter 08. Blackwell.

Botma, E.D., and B.M. van't Veer. 2014. Voiced fricatives as a phonological borderline disorder. *Phonological Studies* 111–114.

Botma, E.D., and M. van't Veer. 2013. A fraction too much friction: The phonological status of voiced fricatives. *Linguistics in the Netherlands* 30:46–60.

Browne, Wayles. 1993. Serbo-Croat. In *The Slavonic languages*, ed. Bernard Comrie and Greville G. Corbett, chapter 7, 306–387. London and New York: Routledge.

Burton, Martha W., and Karen E. Robblee. 1997. A phonetic analysis of voicing assimilation in Russian. *Journal of Phonetics* 25:97–114.

Butska, Luba. 1998. Faithfulness to [voice] in Ukrainian: An analysis of voicing alternations within Optimality Theory. In *Working Papers from Rutgers University*, volume 1, 59–73.

Butskhrikidze, Marika. 2002. *The consonant phonotactics of Georgian*. LOT.

Butskhrikidze, Marika, and Jeroen van de Weijer. 2001. On v-metathesis in Modern Georgian. In *Surface syllable structure and segment sequencing*, ed. Elizabeth Hume, Norval Smith, and Jeroen van de Weijer, HIL Occasional Papers 4, 91–101. Leiden University: Holland Institute of Generative Linguistics.

Carbonell, Joan F., and Joaquim Llisterri. 1992. Illustrations of the IPA: Catalan. *Journal of the International Phonetic Association* 22:53–56.

Cho, Young-mee Yu. 1990a. Parameters of consonantal assimilation. Doctoral Dissertation, Stanford University.

Cho, Young-mee Yu. 1990b. A typology of voicing assimilation. In *The proceedings of the Ninth West Coast Conference on Formal Linguistics*, 141–156.

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clements, G. Nick. 1991. Vowel height assimilation in Bantu languages. In *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society: Special Session on African Language Structures*, 25–64.

Clements, G. Nick, and Sylvester Osu. 2002. Explosives, implosives, and nonexplosives: the linguistic function of air pressure differences in stops. In *Laboratory phonology 7*, ed. Carlos Gussenhoven and Natasha Warner, 299–350. Berlin: Mouton de Gruyter.

Clements, G. Nick, and Sylvester Osu. 2003. Ikwere nasal harmony in typological perspective. In *Typologie des langues d'Afrique et universaux de la grammaire*, ed. Patrick Sauzet and Anne Zribi-Hertz, volume II, chapter Ikwere Nasal Harmony in Typological Perspective, 70–95. L'Harmattan.

Clements, G. Nick, and Sylvester Osu. 2005. Nasal harmony in Ikwere, a language with no phonemic nasal consonants. *Journal of African Languages and Linguistics* 26:165–200.

Clements, G. Nick, and Rachid Ridouane. 2006. Quantal phonetics and distinctive features: a review. In *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics*. Athens, Greece.

Clements, G.N. 2003. Feature economy as a phonological universal. In *Proceedings of the 15th International Congress of Phonetic Sciences*, ed. M.J. Solé, D. Recasens, and J. Romero.

Coats, Herbert S., and Alex P. Harshenin. 1971. On the phonological properties of Russian U. *The Slavic and East European Journal* 15:466–478.

Cohn, Abigail C. 1993. A survey of the phonology of the feature [±nasal]. *Working Papers of the Cornell Phonetics Laboratory* 8:141–203.

Cohn, Abigail C. 2011. Features, segments, and the sources of phonological primitives. In *Where do phonological features come from*, ed. G. Nick Clements and Rachid Ridouane, 13–42. John Benjamins Publishing Company.

Colarusso, John. 1992. *A grammar of the Kabardian language*. University of Calgary Press.

Crystal, David. 2008. *A dictionary of linguistics and phonetics*. Blackwell Publishing, 6th edition.

Cyran, Eugeniusz. 2012. Cracow sandhi voicing is neither phonological nor phonetic. it is both phonological and phonetic. In *Sound, structure and sense. studies in memory of Edmund Gussmann*, ed. E. Cyran, H. Kardela, and B. Szymanek. Wydawnictwo KUL.

Daniels, W. J. 1972. Assimilation in Russian consonant clusters: I. *Papers in* 5:366–380.

Docherty, Gerard J. 1992. *The timing of voicing in British English obstruents*. Number 9 in Netherlands Phonetics Archives. Berlin: Foris.

Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge University Press.

Dresher, B. Elan. 2014. The arch not the stones: Universal feature theory without universal features. *Nordlyd, special issue on Features* 41:165–181.

Dresher, B. Elan, G. L. Piggott, and Keren D. Rice. 1994. Contrast in phonology: Overview. *Toronto Working Papers in Linguistics* 13:iii–xvii.

Dresher, B. Elan, and Xi Zhang. 2005. Contrast and phonological activity in Manchu vowel systems. *Canadian Journal of Linguistics* 50:45–82.

Du Feu, Veronica. 1996. *Rapanui*. New York: Routledge.

Dvořák, Věra. 2010. Voicing assimilation in Czech. *Rutgers Working Papers in Linguistics* 115–144.

Eager, Christopher D. 2015. Automated voicing analysis in Praat: statistically equivalent to manual segmentation. In *Proceedings of the 18th Internationcal Congress of Phonetic Sciences*.

Ferguson, C. 1959. Diglossia. *Word* 15:325–340.

Forrest, K., G. Weismer, P. Milenkovic, and R.N. Dougall. 1988. Statistical analysis of word-initial voiceless obstruents: preliminary data. *Journal of the Acoustical Society of America* 84:115–123.

Friedman, Victor A. 1993. Macedonian. In Scatton (1993), chapter 6, 249–305.

Galea, Luke. 2016. Syllable structure and gemination in Maltese. Doctoral Dissertation, Universität zu Köln, Malta.

Gordon, Matthew, and Ayla Applebaum. 2006. Phonetic structures of Turkish Kabardian. *Journal of the International Phonetic Association* 36:159–186.

Greenberg, Joseph. 1966. *Language universals, with special reference to feature hierarchies*, volume 59 of *Janua Linguarum, Series Minor*. The Hague: Mouton.

Grijzenhout, Janet, and Martin Krämer. 2000. Final devoicing and voicing assimilation in Dutch derivation and cliticization. In *Lexicon in focus*, ed. Barbara Stiebels and Dieter Wunderlich, number 45 in Studia Grammatica, 55–82. Akademie Verlag.

Gurevich, Naomi. 2004. *Lenition and contrast: The functional consequences of certain phonetically conditioned sound changes*. Routledge.

Gussenhoven, Carlos, and Flor Aarts. 1999. The dialect of Maastricht. *Journal of the International Phonetic Association* 29:155–166.

Gussenhoven, Carlos, and R. H. Bremmer Jr. 1983. Voiced fricatives in Dutch: Sources and present-day usage. *Nowele* 2.

Gussmann, Edmund. 2007. *The phonology of Polish*. Oxford University Press.

Haggard, Mark. 1978. The devoicing of voiced fricatives. *Journal of Phonetics* 6:95–102.

Hála, Bohuslav. 1962. *Uvedení do fonetiky češtiny na obecně fonetickém základě*. Nakladatelství československé akademie věd.

Hall, Daniel Currie. 2003a. A formal approach to /v/: Evidence from Czech and Slovak. In *Formal Approaches to Slavic Linguistics: The Ottawa Meeting*.

Hall, Daniel Currie. 2003b. Laryngeal feature specifications in West Slavic languages. *Toronto Working Papers in Linguistics* 20:93–114.

Hall, Daniel Currie. 2007. The role and representation of contrast in phonological theory. Doctoral Dissertation, University of Toronto.

Hall, T. Alan, and Marzena Żygis. 2010. An overview of the phonology of obstruents. In *Turbulent sounds: An interdisciplinary guide*, ed. Susanne Fuchs, Martine Toda, and Marzena Żygis. Berlin: De Gruyter de Mouton.

Halle, Morris. 1959. *The sound pattern of russian*. Mouton & Co.'s-Grevenhage.

Halle, Morris. 1983. On distinctive features and their articulatory implementation. *Natural Language and Linguistic Theory* 1:91–105.

Halle, Morris, and G. Nick Clements. 1983. *Problem book in phonology*. MIT Press.

Halle, Morris, and Kenneth N. Stevens. 1971. A note on laryngeal features. Technical Report 101, MIT Research Laboratory of Electronics Quarterly Progress Report. Pp. 198-213.

Hamann, Silke, and Anke Sennema. 2005. Acoustic differences between German and Dutch labiodentals. *ZAS Papers in Linguistics* 42:33–41.

Hansson, Gunnar Ólafur. 2001. Remains of a submerged continent: Preaspiration in the languages of Northwest Europe. In *Historical linguistics 1999*, ed. L. Brinton, 157–173. John Benjamins Publishing Company.

Hayes, Bruce. 1984. The phonetics and phonology of Russian voicing assimilation. In *Language sound structure*. Cambridge, Massachusetts: The MIT Press.

Herrity, Peter. 2000. *Slovene: A comprehensive grammar*. London & New York: Routledge.

Hirose, H., and T. Ushijima. 1978. Laryngeal control for voicing distinction in Japanese consonant production. *Phonetica* 35:1–10.

Holton, David, Peter Mackridge, and Irene Philippaki-Warburton. 2004. *Greek: An essential grammar of the modern language*. Routledge.

Hombert, Jean-Marie, John J. Ohala, and William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55:37–58.

Hothorn, Torsten, Frank Bretz, and Peter Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50:346–363.

Howson, Phil, Alexei Kochetov, and Pascal van Lieshout. 2015. Examination of the grooving patterns of the Czech trill-fricative. *Journal of Phonetics* 49:117–129.

Hume, Elizabeth, and David Odden. 1996. Reconsidering [consonantal]. *Phonology* 13:345–376.

Iverson, Gregory K., and Joseph C. Salmons. 2011. Final devoicing and final laryngeal neutralization. In *The Blackwell Companion to Phonology*, ed. Marc van Oostendorp, Colin J. Ewen, and Keren Rice, chapter 69. Blackwell Publishing.

Jakobson, Roman. 1948. Russian conjugation. *Word* 4:155–167.

Jakobson, Roman. 1978. Mutual assimilation of Russian voiced and voiceless consonants. *Studia Linguistica* 32:107–110.

Jakobson, Roman, Gunar Fant, and Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge: MIT Press.

Jamieson, A. R. 1977a. Chiquihuitlán Mazatec phonology. In *Studies in Otomanguean phonology*, ed. W. R. Merrifield, 54, 93–105. Arlington, Texas: SIL-University of Texas.

Jamieson, A. R. 1977b. Chiquihuitlán Mazatec tone. In *Studies in Otomanguean phonology*, ed. W. R. Merrifield, 54, 107–136. Arlington, Texas: SIL-University of Texas.

Jansen, Wouter. 2004. Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch. Doctoral Dissertation, Rijksuniversiteit Groningen.

Jessen, Michael, and Catherine Ringen. 2002. Laryngeal features in German. *Phonology* 19:189–218.

Jesus, Luis M. T., and Christine H. Shadle. 2002. A parametric study of the spectral characteristics of European Portugese fricatives. *Journal of Phonetics* 30:437–464.

Johnson, Keith. 2012. *Acoustic and auditory phonetics*. Malden, MA: Blackwell, third edition edition.

Jongman, A., M. Fourakis, and J. A. Sereno. 1989. The acoustic vowel space of Modern Greek and German. *Language and Speech* 32:221–248.

Jongman, A., R. Wayland, and S. Wong. 2000. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* 108:1252–1263.

Joseph, Brian, and Irene Philippaki-Warburton. 1987. *Modern Greek*. Croom Helm.

Kabrah, Rawiah S. 2008. Regressive voicing assimilation in Cairene Arabic. In *Perspectives on Arabic linguistics: papers from the annual Symposia on Arabic Linguistics*, ed. Ellen Broselow and Hamid Ouali, volume XXII-XXIII, 21–34.

Kariņš, A. Krišjānis. 1996. The prosodic structure of Latvian. Doctoral Dissertation, University of Pennsylvania.

Kavitskaya, Darya. 1998. Voicing assimilation and the schizophrenic behaviour of /v/ in Russian. In *Formal Approaches to Slavic Linguistics: The Seattle Meeting, 1998*, ed. Katarzyna Dziwirek, Herbert Coats, and Cynthia M. Vakareliyska, 225–244. Ann Arbor: Michigan Slavic Publications.

Kehrein, Wolfgang, and Chris Golston. 2004. A prosodic theory of laryngeal contrasts. *Phonology* 21:325–357.

Keyser, Samuel Jay, and Kenneth N Stevens. 2006. Enhancement and overlap in the speech chain. *Language* 82:33–63.

Kingston, John. 1990. Articulatory binding. In *Papers in laboratory phonology I: Between the grammar and physics of speech*, 406–434. Cambridge University Press.

Kiparsky, Paul. 1985. Some consequences of lexical phonology. *Phonology* 2:85–138.

Király, Lajos. 2005. *Somogy–zalai nyelvatlasz [Language atlas of Somogy and Zala]*. Budapest: Magyar Nyelvtudományi Társaság [Hungarian Linguistic Society].

Kiss, Zoltán, and Zsuzsanna Bárkányi. 2006. A phonetically-based approach to the phonology of [v] in Hungarian. *Acta Linguistica Hungarica* 53:175–226.

Kochetov, Alexei. 2017. Acoustics of Russian voiceless sibilant fricatives. *Journal of the International Phonetic Association* 1–28.

Komova, Ekaterina, and Phil Howson. 2014. An airflow examination of the Czech trills. *The Journal of the Acoustical Society of America* 136.

Kreitman, Rina. 2008. The phonetics and phonology of onset clusters: The case of Modern Hebrew. Doctoral Dissertation, Cornell University.

Kuipers. 1960. *Phoneme and morpheme in Kabardian (Eastern Adyghe)*. Mouton & Co.'s-Gravenhage.

Kulikov, Vladimir. 2012. Voicing and voice assimilation in Russian stops. Doctoral Dissertation, University of Iowa.

Kuznetsova, Alexandra, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. 2016. *lmertest: Tests in linear mixed effects models*. URL `https://CRAN.R-project.org/package=lmerTest`, r package version 2.0-33.

Ladefoged, Peter. 1997. Linguistic phonetic descriptions. In *The handbook of phonetic sciences*, ed. William J. Hardcastle and John Laver, 589 – 618. Blackwell.

Ladefoged, Peter. 2003. *Phonetic data analysis : an introduction to fieldwork and instrumental techniques*. Malden: Blackwell.

Ladefoged, Peter, and Ian Maddieson. 1996. *The sounds of the world's languages*. Cambridge, Massachusetts: Blackwell.

Lenth, Russell V. 2016. Least-squares means: The R package lsmeans. *Journal of Statistical Software* 69:1–33.

Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig, ed. 2013. *Ethnologue: Languages of the world, seventeenth edition*. Dallas, Texas: SIL International. URL `http://www.ethnologue.com`.

Lightner, Theodore. 1965. Segmental phonology of Modern Standard Russian. Doctoral Dissertation, MIT.

Lincoln, Peter C. 1976a. Banoni, Piva and Papuanization. *Papers in New Guinea Linguistics* 45:77–105.

Lincoln, Peter C. 1976b. Describing Banoni: An Austronesian language of Southwest Bougainville. Doctoral Dissertation, University of Hawaii.

Lindblom, Björn, and Ian Maddieson. 1988. Phonetic universals in consonant systems. In *Language, speech and mind: Studies in honour of Victoria A. Fromkin*, ed. Larry M. Hyman and Charles N. Li, 62–78. London and New York: Routledge.

Lombardi, Linda. 1991. Laryngeal features and laryngeal neutralization. Doctoral Dissertation, University of Massachusetts Amherst.

Lombardi, Linda. 1996. Restrictions on direction of voicing assimilation: an OT account. *University of Maryland Working Papers in Linguistics* 4:84–102.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in optimality theory. *Natural Language and Linguistic Theory* 17:267–302.

Lulich, Steven. 2004. Russian [v]: An acoustic study. *Folia Linguistica* 38:63–85.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge Studies in Speech Science and Communication. Cambridge: Cambridge University Press.

Maddieson, Ian. 2010. Typology of phonological systems. In *The Oxford Handbook of Linguistic Typology*, ed. Jae Jung Song. Oxford University Press.

Maddieson, Ian. 2013a. Glottalized consonants. In *The world atlas of language structures online*, ed. Matthew S. Dryer and Martin Haspelmath. Leipzig.

Maddieson, Ian. 2013b. The world atlas of language structures online. chapter Voicing and Gaps in Plosive Systems. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL `http://wals.info/chapter/5`.

Maddieson, Ian, Sébastien Flavier, Egidio Marsico, Christophe Coupé, and François Pellegrino. 2013. Lapsyd: Lyon-Albuquerque Phonological Systems Database. In *Interspeech 2013*.

Maddieson, Ian, Sébastien Flavier, Egidio Marsico, Christophe Coupé, and François Pellegrino. 2015. LAPSyD: Lyon-Albuquerque Phonological Systems Databases, Version 1.0. URL `http://www.lapsyd.ddl.ish-lyon.cnrs.fr/`.

Maddieson, Ian, and Kristin Precoda. 1990. Updating UPSID. In *UCLA working papers in phonetics*, volume 74, 104–111. Department of Linguistics, UCLA.

Maniwa, Kazumi, Allard Jongman, and Travis Wade. 2009. Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America* 125:3962–3973.

Matasović, Ranko. 2007. A short grammar of East Circassian (Karbadian).

Mathiassen, Terje. 1996. *A short grammar of Lithuanian*. Slavica Publishers, Inc.

Mathiassen, Terje. 1997. *A short grammar of Latvian*. Slavica Publishers, Inc.

Mayo, Peter. 1993. Belarusan. In *The Slavonic languages*, ed. Bernard Comrie and Greville Corbett, chapter 16, 887–946. London / New York: Routledge.

Mielke, Jeff. 2005. Ambivalence and ambiguity in laterals and nasals. *Phonology* 22:169–203.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford University Press.

Morelli, Frida. 1999. The phonotactics and phonology of obstruent clusters in optimality theory. Doctoral Dissertation, University of Maryland at College Park.

Morén, Bruce. 2003. Affricates, palatals and iotization in Serbian: Representational solutions to longstanding puzzles. *Poljarnyj Vestnik* 6:46–70.

Morén, Bruce. 2006. Consonant-vowel interactions in Serbian: Features, representations and constraint interactions. *Lingua* 116:1198–1244.

Nevins, Andrew. 2015. Triumphs and limits of the contrastivity-only hypothesis. *Linguistic Variation* 15:41–68.

Newton, Brian. 1972. *The generative interpretation of dialect. a study of Modern Greek phonology*. Cambridge University Press.

Nicolaidis, Katerina. 2002. Durational variability in vowel-consonant-vowel sequences in Greek: The influence of phonetic identity, context and speaker. In *Selected Papers on Theoretical and Applied Linguistics from the 14th International Symposium*, ed. Marianthi Makri-Tsilipakou, 280–294. Thessaloniki.

Nirgianaki, Elina. 2014. Acoustic characteristics of Greek fricatives. *Journal of the Acoustical Society of America* 135:2964 – 2976.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In *The production of speech*, ed. P.F. MacNeilage, 189–216. New York: Springer-Verlag.

Ohala, John J., and Manjari Ohala. 1993. The phonetics of nasal phonology: Theorems and data. In *Nasals, nasalization, and the velum*, ed. Marie K. Huffman and Rena A. Krakow, volume 5 of *Phonetics and Phonology*, chapter The Phonetics of Nasal Phonology: Theorems and Data, 225–249. Academic Press.

Ohala, John J., and Maria-Josep Solé. 2008. Turbulence and phonology. Technical report, UC Berkeley.

Ozhegov, Sergei Ivanovich. 1984. *Slovar' russkogo iazyka.*

Padgett, Jaye. 2002. Russian voicing assimilation, final devoicing, and the problem of [v] (*or*, The mouse that squeaked). Unpublished paper.

Padgett, Jaye. 2011. The role of prosody in Russian voicing. In *Prosody matters: Essays in honor of Elisabeth Selkirk*, ed. Toni Borowsky, Shigeto Kawahara, Takahito Shinya, and Mariko Sugahara. Equinox Publishing.

Parker, Stephen G. 2002. Quantifying the sonority hierarchy. Doctoral Dissertation, University of Massachusetts Amherst.

Pavlík, R. 2014. On some quantitative properties of Czech ř in newsreading. *Phonetica* 71:22–49.

Petrova, Olga, and Szilárd Szentgyörgyi. 2004. /v/ and voice assimilation in Hungarian and Russian. *Folia Linguistica* 38:87–116.

Philippaki-Warburton, Irene. 1980. Greek diglossia and some aspects of the phonology of common modern Greek. *Journal of Linguistics* 16:45–54.

Pierrehumbert, Janet, and Rami Nair. 1996. Implications of Hindi prosodic structure. In *Current trends in phonology: Models and methods*, 549 – 584. CNRS, Paris-X and University of Salford: University of Salford Press.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In *Frequency effects and the emergence of linguistic structure*, ed. Joan L. Bybee and Paul J. Hopper. John Benjamins Publishing Company.

Pierrehumbert, Janet B. 2016. Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics* 2:33–52.

Pike, K. L, and E. V. Pike. 1947. Immediate constituents of Mazateco syllables. *International Journal of American Linguistics* 13:78–91.

Pinget, Anne-France. 2015. *The actuation of sound change*. LOT.

Pisoni, David B. 1981. Variability of vowel formant frequencies and the Quantal Theory of speech: A first report. *Phonetica* 37:285–305.

Poppe, Nicholas. 1964. *Bashkir manual*, volume 36 of *Indiana University Publications, Uralic and Altaic Series*. Bloomington: Indiana University.

Proctor, Michael. 2009. Gestural characterization of a phonological class: the liquids. Doctoral Dissertation, Yale University.

R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

Reiss, Charles. 2018. Contrast is irrelevant in phonology: A simple account of Russian /v/ as /v/. In *Beyond markedness in formal phonology*, ed. Bridget D. Samuels, 23–46. John Benjamins Publishing Company.

Riad, Tomas. 2014. *The phonology of Swedish*. Oxford University Press.

Rice, Keren D. 1993. A reexamination of the feature [sonorant]: The status of 'sonorant obstruents'. *Language* 69:308–344.

Rice, Keren D., and Peter Avery. 1989. On the interaction between sonorancy and voicing. In *Toronto working papers in linguistics*, volume 10, 65–82. University of Toronto.

Richardson, Rachael R., Naomi H. Feldman, and William Idsardi. 2015. What defines a category? Evidence that listeners' perception is governed by generalizations. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*.

Ringen, C., and V. Kulikov. 2012. Voicing in Russian stops: Cross-linguistic implications. *Journal of Slavic Linguistics* 20:269–286.

Ringen, Catherine, and Pétur Helgason. 2004. Distinctive [voice] does not imply regressive assimilation: Evidence from Swedish. *International Journal of English Studies* 4:53–71.

Scatton, Ernest A. 1993. Bulgarian. In *The Slavonic languages*, ed. Bernard Comrie and Greville Corbett, chapter 5, 188–248. London / New York: Routledge.

Selkirk, Elisabeth. 1984. On the major class features and syllable theory. In *Lanuage sound structure: Studies in phonology*, ed. Aronoff and Oehrle, 107–136. Cambridge: MIT Press.

Shadle, C. H., S. J Mair, and J. N Carter. 1996. Quantifying spectral characteristics of fricatives. In *Proceedings of the International Conference of Spoken Language Processing (ICSLP 96)*, 1517–1520. Philadelphia, USA.

Shadle, Christine H. 2010. The aerodynamics of speech. In *The handbook of phonetic sciences*, ed. William J. Hardcastle, John Laver, and Fiona E. Gibbon, chapter Chapter 2, 39–80. Blackwell Publishing, second edition edition.

Sharoff, Serge. 2002. The frequency dictionary for Russian. URL `http://www.comp.leeds.ac.uk/ssharoff/frqlist/frqlist-en.html`, (accessed January 19, 2009).

Shevelov, Y., George. 1993. Ukrainian. In *The Slavonic languages*, ed. Bernard Comrie and Greville G. Corbett. Routledge.

Shi, Lei, Thomas L. Griffiths, Naomi H. Feldman, and Adam N. Sanborn. 2010. Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review* 17:443–464.

Short, David. 1993. Czech. In *The Slavonic languages*, ed. Bernard Comrie and Greville G. Corbett. Routledge.

Smith, Caroline. 1997. The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics* 25:471–500.

Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. Doctoral Dissertation, MIT. Published in 1990. New York: Garland Press.

Stevens, Kenneth N. 1983. Design features of speech sound systems. In *The production of speech*, ed. Peter F. MacNeilage, 247–261. New York: Springer-Verlag.

Stevens, Kenneth N. 1987. Interaction between acoustic sources and vocal-tract configurations for consonants. In *Proceedings of the 11th International Congress of Phonetic Sciences (ICPhS 87)*, 385–389.

Stevens, Kenneth N. 1988. *Acoustic phonetics*. MIT Press.

Stevens, Kenneth N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17:3–45.

Stevens, Kenneth N. 1991. Acoustic phonetics. Ms., MIT, Cambridge, Mass.

Stevens, Kenneth N., and Samuel Jay Keyser. 1989. Primary features and their enhancement in consonants. *Language* 65:81–106.

Stevens, Kenneth N., and Samuel Jay Keyser. 2010. Quantal theory, enhancement and overlap. *Journal of Phonetics* 38:10–19.

Stevens, Mary, and John Hajek. 2004. How pervasive is preaspiration? Investigating sonorant devoicing in sienese italian. In *Proceedings of the 10th Australian International Conference on Speech Science & Technology. Canberra: S*, 334–339.

Strycharczuk, Patrycja. 2012. Phonetics-phonology interactions in pre-sonorant voicing. Doctoral Dissertation, University of Manchester.

Studdert-Kennedy, Michael. 1989. Feature fitting: A comment on K. N. Stevens' "On the quantal nature of speech". *Journal of Phonetics* 17:135–143.

Swan, Oscar E. 2002. *A grammar of contemporary Polish*. Bloomington, Indiana: Slavica Publishers, Inc.

Szigetvári, Péter. 1998. Why [h] is not voiced. *Structure and Interpretation. Studies in Phonology. PASE Studies & Monographs* 4:287–301.

Tabain, Marija. 2001. Variability in fricative production and spectra: Implications for the Hyper- and Hypo- and Quantal Theories of speech production. *Language and Speech* 44:57–94.

Vago, Robert. 1980. *The sound pattern of Hungarian*. Washington, D.C.: Georgetown University Press.

Vasishth, Shravan. ???? Bayesian data analysis for psycholinguistics. URL `http://www.ling.uni-potsdam.de/~vasishth/bdalecturenotes.pdf`, lecture notes.

Vaux, Bert. 1998. The laryngeal specifications of fricatives. *Linguistic Inquiry* 29:497–511.

Vaux, Bert, and Brent Miller. 2011. The representation of fricatives. In *The Blackwell Companion to Phonology*, 669–693. Blackwell.

Šimáčková, Šárka, Václav Jonáš Podlipský, and Kateřina Chládková. 2012. Czech spoken in Bohemia and Moravia. *Journal of the International Phonetic Association* 42:225–232.

Wagner, Michael. 2002. The role of prosody in laryngeal neutralization. *MIT Working Papers in Linguistics* 42:373–392.

Walker, Rachel. 2000. *Nasalization, neutral segments, and opacity effects*. New York: Garland.

Walker, Rachel. 2003. Reinterpreting transparency in nasal harmony. In *The phonological spectrum: Volume 1: Segmental structure*, 37–72. John Benjamins Publishing Company.

Wetzels, W.L̃eo, and Joan Mascaró. 2001. The typology of voicing and devoicing. *Language* 77:207–244.

Wiese, Richard. 2000. *The phonology of German*. The Phonology of the World's Languages. Oxford University Press.

Wright, Richard. 1996. Consonant clusters and cue preservation in Tsou. Doctoral Dissertation, University of California Los Angeles.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Yu, Alan C. L. 2011. Mergers and neutralization. In Iverson and Salmons (2011), chapter 80.

Zalizniak, A. A. 1975. Razmyshleniia po povodu 'iazv' A. A. Reformatskogo. *Problemnaia gruppa po eksperimental'noi i prikladnoi lingvistike Instituta russkogo iazyka AN SSSR: predvaritel'nye publikatskii* 71:13–23.

Zec, Draga. 1995. Sonority constraints on syllable structure. *Phonology* 12:85–129.