THE ROLE OF MORPHOLOGICAL STRUCTURE IN PHONETIC VARIATION

Ruaridh Keith Purse

A DISSERTATION

in

Linguistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2021

Supervisor of Dissertation

_____

Meredith Tamminga, Associate Professor of Linguistics

Graduate Group Chairperson

_____

Eugene Buckley, Associate Professor of Linguistics

Dissertation Committee:

Jianjing Kuang, Associate Professor of Linguistics

David Embick, Professor of Linguistics

THE ROLE OF MORPHOLOGICAL STRUCTURE IN PHONETIC VARIATION

ACKNOWLEDGMENT

So many things about the PhD process have been a surprise. It seemed to go altogether too fast and yet it almost seems like I've always been part of Penn Linguistics. Spending the latter half of the programme amidst a global pandemic was certainly unexpected, and required a significant adjustment of my initial vision for this dissertation. But even as I feel some regret about that, I'm surprised that I feel a much stronger sense of accomplishment and gratitude. Gratitude because I could not have managed it without the support of those around me, who have surprised me time and again with their generosity. Before I submit this dissertation, I have to take some time to thank some people in particular.

I owe so much to Meredith Tamminga, my supervisor. Equal parts brilliant and kind, you have always known when and how to push me to think deeper—with more clarity—and when to celebrate with me. That you have consistently expressed trust in me and my ideas (even since I had less of filter for them) means more than I can express. I look forward to our continued collaboration and friendship as I am sure I have many more things to learn from you. Among them, I hope I can come to model some small part of your intellectual insight, professional savvy, and empathy for anyone who might look up to me in the same way I have looked up to you.

My committee members, Jianjing Kuang and David Embick have also played key roles in the development of this dissertation and my development as a scholar. Thank you, Jianjing, for continually supporting my forays into unfamiliar methods and believing in my ability to make them work eventually. Thank you for challenging me to reconsider my biases and stop to think about my data in a different way. You have improved my rigour as a scientist. Thank you Dave, for helping me to frame my work in a wider context, and for your consistent encouragement that I can do so impactfully. I would always leave our discussions feeling energised and stimulated and with far too many new insights to remember. I hope we can have many more.

Many other faculty at Penn and further afield have been so generous with their time and their insight. I thank Mark Liberman, Gene Buckley and Rolf Noyer in particular for their

guidance as I continue to poke at phonology and phonetics to see what falls out. I thank Joe Fruehwald, Alice Turk, Lauren Hall-Lew, and Nicole Holliday for their collaboration and mentorship, setting me up to succeed at Penn and beyond. I also thank Danielle Turton, Patrycja Strycharczuk and Sam Kirkham, for engaging with my work and making me feel at home in our corner of linguistics.

A very special thank you has to go to Yosiane White, you absolute beacon of light. I'm so glad we have had each other from the very beginning of our PhDs to the very end. I have relied on you more times than I can count to help me articulate a problem, navigate bureaucracy, or generally vent my feelings. Thank you for your endless patience in letting me procrastinate from my own work by talking about yours. It's bittersweet that we will be graduating together and going our separate ways, but I know that we will be lifelong friends, and before long we will sing our hearts out at karaoke or bare our souls over wine and a campfire again.

Thank you to the rest of my cohort at Penn Linguistics: Nari Rhee, Ryan Budnick, Nikita Bezrukov, and Faruk Akkus. Thank you to those who came before me and paved the way: Betsy Sneller, Sabriya Fisher, Duna Gylfadóttir, Rob Wilder, Amy Goodwin-Davies, Luke Adamson, Kajsa Djärv, Nattanun Chanchaochai, Milena Šerekaitė, Aletheia Cui, Ava Irani, Ava Creemers, Lacey Wade, and Wei Lai. And thank you to those who have come after: Johanna Benz, Lefteris Pararounas, Alexanderos Kalomoiros, Ollie Sayeed, Aini Li, and Gwen Hildebrandt. You have all taught me so much and given me so much joy during my time here. I can't wait to meet you all again in various corners of the world.

My non-linguist friends have played at least as big a role in getting me where I am today. Thank you in particular to Andrew Potter for your honesty, your linear algebra lessons, and for always being the better friend. You consistently remind me that I'm not alone, even when my brain plays tricks on me. Thank you also to Hamish Leiper, Tom Holm, Cameron Maitland, and Martin McVey for support and the best distraction. I can't think of a better group of friends to escape with. Thank you also to Matthew Barnbrook and Liam Brierley, my other ex-Edinburgh friends. You were my early models for what

being a graduate student meant, and how to be a full human being at the same time. Thank you for so many nights dancing or gaming or just talking until we saw the sun.

Thank you to Drew Amacker, for lending your voice to my research, and for making Philadelphia home all these years. Thank you for regularly reminding me of a world outside Linguistics, but always managing to think of a new problem that 'Linguistics should care about' for my next big project. Thank you for all the food and cinema, all the encouragement, and all the love and support through the highs and lows of this process. I couldn't have done it without you.

Lastly, I wish to thank my family. To Heather McNicoll, I was always amazed by your unwavering curiosity about the world; I wish I could have discussed this dissertation with you. To Helen Purse, thank you for always being one of my biggest cheerleaders, even while you were disappointed I kept deciding to move away and study more. To my brother, Calum Purse, thank you for looking out for me as a big brother my whole life and for being my friend today. And to my parents, Robert and Alison Purse, thank you for being there—even from a distance. Your love and patience will always mean the world. Thank you for everything.

ABSTRACT

THE ROLE OF MORPHOLOGICAL STRUCTURE IN PHONETIC VARIATION

Ruaridh Keith Purse

Meredith Tamminga

This dissertation is situated in broad debates about the architecture of the phonological grammar, and the sensitivity of gradient phonetic parameters to morphological structure. It takes, as its primary case study, a linguistic variable that is of prevailing interest to sociolinguists and phonologists alike: English Coronal Stop Deletion (*old*∼*ol'*; CSD). While CSD is robustly sensitive to the morphological class of words in which coronal stops are contained, its alignment with the small class of other morphology–phonetics interactions is not straightforward.

I approach this problem from several angles, incorporating diverse methodologies. In the first place, I provide new articulatory evidence suggesting that CSD does indeed have its primary locus in the gradient phonetics, demonstrating that the magnitude of tongue tip raising to a coronal stop constriction is gradiently conditioned by morphology. Moreover, this variation is typologically distinct from the majority of other examples of phonetic phenomena conditioned by morphology, which primarily concern durational parameters.

In the rest of the dissertation, I problematise CSD's status as exceptional in this way, probing how well explanations for other morphology-sensitive phonetic phenomena (i.e. effects of prosody and word predictability) account for CSD patterns. In two perception experiments, listeners do not show perceptual sensitivity to the covert tongue tip raising observed in articulation, but do reflect an association between morphological complexity and increased duration. Finally, a large-scale corpus study shows only measures of word frequency that are relative to a word's larger morphological paradigm predict CSD patterns accurately. This suggests that morphological structure was a key missing element in predictability accounts of the variable.

Ultimately, surface CSD may amount to the confluence of more than one type of morphologically conditioned phonetic phenomenon. This dissertation sets the stage for continued

progress towards an account integrating these different factors, and generates new puzzles in the asymmetry between production and perception for variable phonology and phonetics.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

While early perspectives on generative grammar strictly separated linguistic variation from the study of the grammar (e.g Chomsky, 1965), a strand of concurrent research in variationist sociolinguistics began using tools from formal grammatical analyses to account for variation (Weinreich et al., 1968; Cedergren and Sankoff, 1974), and started a research tradition that has been exceedingly fruitful. To take a few well-known examples from varieties of English alone, variationist research has uncovered robust systematicity in the vocalisation of /r/ (Labov et al., 2006), progressive suffix choice (*working*∼*workin'*), and negative concord (*I don't like nothing*; Labov 1972a). It is now increasingly clear that systematic variation permeates all levels of linguistic structure. As such, a mutually informative relationship is developing between descriptions of variation and formal models of the grammar. Many contemporary grammatical models are explicitly moulded to account for variation, and evaluated on this basis (e.g. Adger, 2006; Parrott, 2007; Coetzee and Pater, 2011). By the same token, our understanding of linguistic variables is informed by what these models tell us about the architecture of the grammar. A key sense in which this is true concerns how different self-contained levels of linguistic structure relate to one another, and where such interaction is prohibited. In general, adherents to a generative framework understand these levels of structure—or 'modules'—to be arranged in a strict order (syntax, morphology, phonology, phonetics) such that a given level's interactions—or 'interfaces'—are shared only with the levels immediately before and after it. Even more strictly, the relationship between modules is generally theorised to be 'feed-forward', such that a module's only input is the output of the immediately preceding module, and any influence between modules is lim-

ited to what structure a given module passes on by one step. This means, for instance, that the phonetics (the domain of the physical articulation and perception of language) should only be directly influenced by the phonology (the domain of the abstract sound system) and not by morphology (the structure of words) or syntax (the structure of sentences).

The 'modularity' diagnostic has been useful for variationists looking to pinpoint the 'locus' of variation for different variables: which level of structure is implicated. To take an important example for this dissertation, so-called English 'Coronal Stop Deletion' (CSD) occurs at dramatically different rates depending on the morphological status of the word in which a target coronal stop would appear. This observation is commonly taken as evidence that Coronal Stop Deletion is indeed a phonological process rather than a phonetic one (Coetzee and Pater, 2011), since only phonology (and not phonetics) should be influenced by morphological structure. However, as resources for observing the fine-grained detail of phonetic implementation become more accessible, we should look to interrogate this conclusion, and compare our assumptions about the properties of variable phenomena and different levels of structure with what actually occurs in speech production.

The separation of phonetics from morphological structure and cases where it appears to be violated are the focus of this dissertation. In this introductory chapter, I offer insights into the interfaces between morphology, phonology, and phonetics, and an overview of cases where the phonetics appears to be sensitive to morphological structure, *contra* our understanding of the strict feed-forward organisation of these modules. I also provide an overview of the vast literature on Coronal Stop Deletion as a phonological variable whose patterns are commonly used to craft models of the phonological grammar. This is the backdrop for Chapter 2, in which I report the results of an articulatory study providing new perspectives on the implementation of Coronal Stop Deletion. In it, I demonstrate that standard phonological models are insufficient to capture what appears to be widespread phonetic gradience in so-called Coronal Stop Deletion. Instead, apparent Coronal Stop Deletion looks to be implemented using diverse strategies, including what seems like an important example of a morphologically-conditioned phonetic phenomenon that is not necessarily explained by

existing accounts. The remaining chapters of the dissertation constitute forays into new puzzles and alternative accounts for apparent Coronal Stop Deletion as a consequence of other morphologically-conditioned phonemena, which gain new importance in light of Chapter 2's articulatory results. Chapter 3 explores the role of listeners, and the degree to which the details of articulatory gradience are evident in perception. Chapter 4 provides new evidence on existing discussions of potential durational differences that covary with the articulatory ones. Chapter 5 delves into discussions of lexical frequency and processing that have previously fallen short of explaining Coronal Stop Deletion results, and demonstrates that morphological structure was a key missing piece in the very measures used to evaluate it. Finally, Chapter 6 is a discussion of what we can conclude from all these studies, and what avenues of research seem most promising to shed more light on the representation of Coronal Stop Deletion.

1.1 Morphology and phonetics

A major issue at play in this dissertation is the organisation of the morphology, phonology, and phonetics as aspects of linguistic cognition. In generative schools of thought, these modules are often understood to exist in a strict feed-forward relationship (Pierrehumbert, 2002; Bermúdez-Otero, 2007). In other words, the phonology can reflect the morphological structure of words that it takes as its input, and the phonetics can similarly reflect the phonology, but the phonetics should not directly reflect the morphology beyond what is encoded in the intervening phonology. However, instances where morphological structure appears to be reflected in the phonetics are occasionally reported. In this section, I aim to provide an overview of these phenomena—along with typical explanations that preserve a modular architecture—and point out that research in variationist sociolinguistics is a potential source of more morphology-phonetics interactions that appear to be typologically different (in that they do not, at first glance, primarily concern phonetic duration). Before this, however, it is important to expand upon what we understand to be the basic properties of the phonetic and phonological modules, how they differ, and how they are related.

1.1.1 Phonetics and phonology

A great deal of ink has been spilled trying sort out the division of labour in the "sound"[1] systems of language. That is, how linguistic concepts are given form that can be produced, perceived and understood. While I don't purport to comprehensively review all of the issues at play here, this section amounts to a sketch of the general properties of different parts of these systems, and how they are understood to interact. The classic division that is central to this discussion is between the domains of "Phonetics" and "Phonology". Phonology deals with abstract symbols, including how a finite set of them are organised and interact in a given language. Phonetics, on the other hand, is concerned with the physical—kinematic and acoustic—properties of speech. For example, the English word *bat* has a phonological form that can be abstractly represented as a string of three phones—/b/, /a/, and /t/—whose order and identity distinguish this word from other words in English. In order to actually produce a phonetic signal that is recognisable as *bat*, however, a speaker must coordinate the movement of their lungs, larynx, lips, tongue and jaw in real time to generate appropriate vibrations in the air. Those vibrations that are 'appropriate' are those that can be understood to correspond to the intended phones and not other ones. For example, executing a /b/ with a very long voice onset time (VOT) following the release burst would likely veer into the territory of the signal associated with /p/, a contrasting category that would alter the meaning of the intended word. However, small non-contrastive variation within abstract categories is ubiquitous and uncontroversial.

The idea of linguistic form as doubly represented in phonetics and phonology is an old one. Ladd (2011) dates the emergence of the phonemic principle to the late 19th century. More abstractly, this dual representation relates to a classic conceptual move within theoretical linguistics to separate the notion of linguistic knowledge from that of language in practice. Examples of this idea can be found in Saussure's (1916) *Langue* versus *Parole*, and both Chomsky's (1965) Competence versus Performance and (1986) I(nternal)-language

[1] While this dissertation is centrally concerned with spoken language and how characterising representations in spoken languages interacts with the complex relationship between articulation and acoustics, questions around the properties of linguistic form are also logically relevant to signed languages.

versus E(xternal)-language distinctions. Though different in their formulation, all of these theories share a desire to capture abstract linguistic knowledge as an idealised representation that is distinct from the messy realisation of language. But while phonetics and phonology can be conceptually distinguished, they must be intimately related, because phonology is shaped by phonetic limitations and tendencies and phonological representations are necessarily phonetically implemented.

In some approaches, the conceptual separation of phonetics and phonology is paralleled by an organisation of the brain into highly specialised modules (Fodor, 1983). A version of this, which resembles the simple sketch of the production of *bat* that I just outlined, is described as the 'consensus view' of phonetic implementation by Pierrehumbert (2002: 101). "Lexemes (the phonological representations of words) are abstract structures made up of categorical, contrastive elements. The phonetic implementation system relates this abstract, long-term, categorical knowledge to the time course of phonetic parameters in particular acts of speech". In early Generativist descriptions, this is even formalised in terms of universal articulatory correlates of invariant phonological features (e.g. Chomsky and Halle, 1968: 293). In other words, symbols would represent direct instructions for their articulation, and any modulation of that spellout process must come from some extra-grammatical source. Such modulations, presumably, include such details as a speaker's identity or the particular language variety being spoken. In reality, we know that the relationship phonetics and phonology is somewhat more complex than invariant phonological representations of words that directly correspond to articulatory movements, and efforts to properly account for it are ongoing and varied.

Broadly, the facts complicating the basic symbols-and-spellout perspective on phonology and phonetics can be grouped into two classes. In the first place, there is an abundance of evidence for phonetic knowledge that is learned. Equivalent phonological categories—far from having universal phonetic correlates—are implemented in ways that are language-specific (e.g. Cho and Ladefoged, 1999), diachronically changeable (e.g. Zellou and Tamminga, 2014), and socio-stylistically variable (Foulkes et al., 2010). This points to at least

some aspects of gradient phonetics that are arbitrary and learned, as opposed to some consequence of speaker physiology. Speakers must acquire and implement different phonetic realisations as appropriate. Secondly, the characterisation of phonology, pre-phonetic implementation, as wholly invariant is likely to be insufficient. Research in variationist sociolinguistics in particular has uncovered evidence of many examples of apparent phonological variation. That is, there are cases in which language users exhibit variable pronunciations of a single phoneme, with variants that are not attributable to phonetic variation in the implementation of a single phonological target. For example, variable British English t-glottaling in words like *bottom* ([bɒtəm]∼[bɒʔəm]) generally entails a discrete alternation between a raised tongue tip and spread glottis on one hand, and a lowered tongue tip and constricted glottis on the other. Heyward et al. 2014). This kind of variable is most parsimoniously accounted for in terms of a /t/ phoneme that is variably interpreted as [t] or [ʔ] allophones before the stage of gradient phonetic implementation. Correspondingly, plenty of formal phonological machinery has been adapted to account for variation. For example, Cedergren and Sankoff (1974) propose variable versions of standard phonological rules, which Guy (1991b) implements in a Lexical Phonology framework.

Taking these types of evidence (for learned phonetics and variable phonology) into account, several other popular frameworks eschew the strict division of phonological and phonetic representations altogether. Usage-based accounts like Exemplar Theory describe a mental lexicon in which words are represented with clouds of whole-word memory traces for the phonetic signal of each instance in which a word was encountered (Pierrehumbert, 2002; Bybee, 2002). The activation of these traces is then at once a mechanism of lexical, phonological, and phonetic access in speech production and perception. Articulatory Phonology (Browman and Goldstein, 1985), on the other hand, describes stored phonological representations themselves as fundamentally spatiotemporal. Thus, there is no earlier stage of representation in terms of segments and features that needs to be interpreted in phonetic dimensions. The representations already consist of complex and coordinated articulatory movements across time.

While many contemporary models of the phonological grammar are, on the whole, moving towards less strict divisions between phonetics and phonology, there are elements of this kind of ordering of representations that are worth salvaging. In particular, there are a number of sources of evidence for a phonemic level of representation that is not always captured in more phonetically rich frameworks. McQueen et al. (2006) show that listeners are able to learn new acoustic properties of a phoneme, depending on the phonological categories in their inventories (Lisker and Abramson, 1970; Kazanina et al., 2006), and generalise it to instances of that phoneme in different contexts. A number of studies on speech errors in production show that they are commonly structured in terms of reorganising a phonemic representation, and errors that unambiguously involve individual features or syllabic units are comparatively rare (Shattuck-Hufnagel and Klatt, 1979; Shattuck-Hufnagel, 1992). Another particularly compelling source of evidence for phonemic representation is the phonological equivalence of discrete variants or alternants that seem phonetically distinct. Phenomena like the previously-mentioned British English t-glottaling are difficult to account for without abstract phonemic representations with multiple realisations (Heyward et al., 2014). But similar evidence comes from alternants that speakers use consistently in certain contexts; Scobbie et al. (2009) show that some Dutch speakers realise /r/ with two discretely different places of articulation in onsets (uvular) and codas (alveolar approximant), but in a shadowing task participants do not exhibit a delay even when their own /r/ productions don't match what they hear—suggesting phonological equivalence (Mitterer and Ernestus, 2008). This suggests we should not do away with some abstract level of representation like the phoneme.

Another property of models that model phonology and phonetics as more strictly partitioned and ordered that is important concerns differential sensitivity to morphosyntactic structure. This is the 'feed-forward' aspect of modular feed-forward frameworks, such that each component of the grammar communicates only with the component immediately following it. As such, categorical phonological alternations frequently interact with morphological categories, but interaction between the morphology and the gradient phonetics ought

to be prohibited. These gradient phonetic properties would not reflect morphological structure because there is an intervening level of categorical phonological structure, which is the only level with which the morphology should interact under strict feed-forward modularity. Bermúdez-Otero (2010) spells out the typological argument against a direct relationship between morphosyntactic structure and phonetics, proposing impossible behaviours such as 'mark dual number by adding /-no/ and increasing gestural overlap by 20%' and 'form eventive nominalizations by adding /-ti-/ and lengthening VOT by 30%'. Of course, this dissertation is chiefly concerned with cases where the morphology *does* seem to influence the phonetics, but these form rare exceptions to a general rule and (as I will discuss) seem tightly constrained in ways that suggest the separation of morphology and phonetics holds.

It is clear that modern descriptions of phonetics and phonology need to incorporate the flexibility to account for phonetic grammar and phonological variation, while capturing some important aspects of classic feed-forward modular grammars. One approach to this problem is to further stratify the phonology and phonetics beyond just a level of abstract phonological representation and its implementation in the phonetics. For example, Coetzee and Pater (2011) describe 'early' and 'late' components of a phonological grammar, with different properties. In their description, early phonological processes can interact with the lexicon, and therefore may be sensitive to morphological structure and/or have particular lexical exceptions. Early phonology should also be categorical and insensitive to global phonetic parameters like speech rate. On the other hand, processes in the late phonology are described in terms that closely resemble classic descriptions on phonetic phenomena. In this description, they should not interact with the lexicon, and therefore should be insensitive to morphological structure and affect all words, but Coetzee and Pater also allow for non-categorical variation and sensitivity to speech rate within this domain. The primary reasoning here is that early phonology must produce categorical outputs that are interpretable for further manipulation in later stages of the phonology, while late phonology is not beholden to any subsequent modules. Other, fleshed out, accounts of phonology and phonetics describe a similar continuum of representation like categoricity vs. gradience and

sensitivity to morphosyntax (Cohn, 2007), or partially ordered but overlapping components with these properties (Scobbie, 2007). In this dissertation, the relevant observation is that morphological sensitivity ought to be relegated to early stages of these continua, while variation along gradient parameters ought to be situated in the late phonology or phonetics. I describe phenomena of the latter type as 'phonetic' throughout this dissertation, because its specific characterisation is not the primary focus. Rather, I focus on places with these gradient 'phonetic' phenomena appear to be sensitive to morphological structure, in violation of how these components are understood to be ordered.

1.1.2 Morphology-sensitive phonetics

We have seen that while extreme versions of feedforward modularity are insufficient to capture all the facts, there is some relationship between important properties that have traditionally defined the phonetics–phonology interface. The properties of categoricity versus gradience, invariance versus variability, arbitrariness versus naturalness, and sensitivity versus insensitivity to higher order linguistic structure appear to form simultaneous continua upon which we can characterise phenomena in phonology and phonetics. For example, early phonological processes are generally observed to be categorical and invariant, and are much more likely to be phonetically unnatural and sensitive to morphosyntax, than processes in the late phonology and phonetics. Given this observation, findings of cases where properties like these ones come apart are still of interest and continue to generate discussion (Strycharczuk, 2019), even if we eschew the strictest form of feed-forward modularity in the grammatical architecture. This section, and this dissertation more broadly, is primarily concerned with the potential for phonetic (i.e. gradient, variable) phenomena to be sensitive to morphological structure, which represents a significant 'coming apart' of these representational continua.

Before examining specific cases of phonetic sensitivity to morphological structure, or morphological influence on phonetic implementation, the first order of business is to expand upon what is meant by 'sensitivity to' or the 'influence of' morphology when it comes

to phonetics. Bermúdez-Otero (2010) gives hypothetical examples of modulating global phonetic parameters, like degree of gestural overlap, as primary exponents of specific morphosyntactic constructions, like dual number. These examples are deliberately extreme to demonstrate the requirement for some restrictions on the relationship between morphosyntax and phonetics like those provided by the representational continua I have just alluded to. His second hypothetical example, concerning longer (but not contrastively so) VOTs in eventive nominalisations is perhaps slightly more realistic. Fine-grained language-specific differences in the VOT of equivalent phonological categories are well-documented (Cho and Ladefoged, 1999), and must be learned. However, once again this is framed in terms of a phonetic parameter that might be manipulated as a primary exponent of a specific morphosyntactic construction. As I will explain, this does not adequately describe the phenomena that are actually observed and discussed in terms of a morphology–phonetics interaction.

By and large, the morphology–phonetics interactions described in the literature are limited to a few key domains: phonetic lengthening at morphological boundaries, incomplete neutralisation of morphophonological contrasts, more general effects of the quantity and frequency of morphologically related words, and sociolinguistic variables whose rate of application is modulated by the morphological class of the word they target. Considering these phenomena, it seems apparent that morphologically-sensitive phonetic effects are constrained both in terms of their morphological triggers and the phonetic parameters that are affected. First, their triggers them seem to be best characterised as general properties of morphological structure, rather than specific morphosyntactic constructions. Despite some inconsistencies across studies, phonetic lengthening effects in English have previously been reported at boundaries preceding *-s*, *-ed*, and *-ing* suffixes (Sugahara and Turk, 2009), rather than targeting one in particular. The apparent ubiquity of the effect across suffixes, and the locality of lengthening to around the juncture between stem and suffix, suggests that the morphological boundary itself (however that is represented) is what induces lengthening. Similarly, the multitude of effects reported concerning how word pronunciations are

influenced by the form and frequency of morphologically related words implicate properties of the mental lexicon on the whole, especially how morphology is relevant to its organisation and mechanisms of lexical access. Secondly, in terms of the phonetic parameters involved, they are primarily durational in nature. This is obviously true for phonetic lengthening, but even the literature on incomplete neutralisation is dominated almost exclusively by work on processes of obstruent devoicing that primarily retain residual contrasts in the length of the preceding vowel. This limitation in the observed effects seems meaningful, since voiced and voiceless obstruents are distinguished by several other phonetic cues, the incomplete neutralisation of which is rarely or never reported, not to mention other relevant morphophonological contrasts in the same languages. For example, there are no reports of incomplete neutralisation of umlaut between German plural and singular nouns, which are an key site of evidence for incomplete obstruent devoicing. And in cases that seem exceptional to the modulation of duration according to where morphological structure, like /l/-final rhymes (Sproat and Fujimura, 1993; Strycharczuk and Scobbie, 2016, 2017), there are still effects of morphological structure on parameters that are strongly associated with duration, like /l/ darkness (Sproat and Fujimura, 1993; Lee-Kim et al., 2013; Strycharczuk and Scobbie, 2016, 2017; Turton, 2017). Given these generalisations, it is clear that the interaction between morphology and phonetics is constrained. Speakers do not, as Bermúdez-Otero (2010) points out, employ arbitrary differences in phonetic implementation as primary exponents of specific morphosyntactic constructions. Rather, general properties of morphological structure or paradigmatically related words are implicated in word production.

The leading explanations for these morphologically-sensitive phonetic effects are correspondingly general. A common refrain in this literature involves invoking paradigm uniformity constraints, which are classically used to explain phonological effects whereby the same morpheme is realised with the same form in different contexts, even when it violates other phonological constraints (Benua, 1995; Burzio, 1994; Kenstowicz, 1995; McCarthy and Prince, 1995). The application of paradigm uniformity to phenomena like incomplete

neutralisation requires an extension of the theory from its normal role in enforcing categorical adherence to a single phonological form, to stipulating that paradigmatically related phonological forms also exert a gradient influence in the phonetics (Ernestus and Baayen, 2006; Frazier, 2006; Roettger et al., 2014). A similar explanation is sometimes put forth for lengthening effects at morphological boundaries, where paradigm uniformity applies the same timing properties from when the end of a stem co-occurs with the end of a word (e.g. *lap#*), as when that same stem appears with a suffix (e.g. *lap-s*). The requirement for extending paradigm uniformity into gradient effects does not necessarily apply here; we can simply stipulate that word-final prosodic lengthening (e.g. a π-gesture) is categorically applied to every instance of a stem. In fact, prosodic explanations need not rely on paradigm uniformity at all. Some approaches represent sublexical constituents, circumscribed by morphological boundaries, as being directly encoded in the prosody in the same way as larger syntactic constituents, without reference to the representation and timing of paradigmatically related words. This focus on prosody has the advantage of capturing the generalisation that morphology–phonetics interactions are primarily durational in nature, but fails to properly account for phenomena where the shape of the larger morphological paradigm seems to generate phonetic differences between isomorphic words (i.e. incomplete neutralisation). On the other hand, we can take seriously the notion that morphology–phonetic interactions are induced by relationships between paradigmatically related forms in the lexicon, and ask about other aspects of the episodic representation of words, such as their quantity and frequency. Several studies do just that, and deduce that paradigm uniformity effects should vary in magnitude according to the strength of representation of related forms, along with more general phonetic effects corresponding to a word's frequency. In the next subsections, I review research in different types of morphology–phonetics interactions, and point to which of these explanations is dominant in the literature. I end this section with an area of research that is not commonly linked to the morphology–phonetics literature and which seems at first glance to disrupt many of the generalisations laid out here: morphologically-conditioned sociolinguistic variables, including English Coronal Stop

Deletion which is the primary case study for this dissertation. As this dissertation will lay out in much greater detail, I argue that Coronal Stop Deletion in not exceptional and may covertly exhibit many of the same properties, including sensitivity to duration and paradigmatic effects.

1.1.2.1 Durational correlates of morphological structure

One place where morphological structure seems to straightforwardly affect phonetic implementation involves segment duration. In a number of studies, suffixed English words are shown to be pronounced with slightly longer segment durations than their morphologically simplex homophones. In general, stem-final rhymes before Level II English suffixes (e.g. [ak] in *tacks* and *packed*) are 4-6% longer than identical sequences in monomorphemes (Sugahara and Turk, 2009; Seyfarth et al., 2018). Similarly, Schwarzlose and Bradlow (2001) focus on stem-final consonants *preceding* a suffix (e.g. [k] in *tacks*, *tucks*, and *macs*), and found they were 3 to 5 ms longer than the penultimate segment in monomorphemic homophones (*tax*, *tux*, *max*). In terms of the suffix itself, laboratory studies find that a suffixal [s] marking either 3SG or plural (e.g. *wrecks*, *laps*, *hearts*), is produced about 9ms longer on average than an [s] at the end of a monomorphemic homophone (*Rex*, *lapse*, *Hartz*) (Walsh and Parker, 1983; Seyfarth et al., 2018), and even a coronal stop for an *-ed* suffix (pronounced [t,d]) may be slightly longer than the same stop in a monomorphemic homophone (Lociewicz 1992, cf. Mousikou et al. 2015; Seyfarth et al. 2018). However, the results for suffix duration are mixed in corpus data that measures average duration across words without directly comparing homophone pairs (Song et al., 2013; Plag et al., 2017). Similarly, some studies do not report the pre-boundary stem lengthening effect, but all such cases seem to involve /l/-final stem rhymes, specifically /iːl/ (Sproat and Fujimura, 1993), /uːl/ (Strycharczuk and Scobbie, 2016) and /ʊːl/ (Strycharczuk and Scobbie, 2017).

A leading explanation for the general occurrence of lengthening effects appeals to the prosody. Derivationally, it may be that the morphological boundary between a stem and a Level II English suffix is aligned with a prosodic boundary. This has been specifically

formulated as nested Prosodic Word constituents (Sugahara and Turk, 2009). Alternatively, we could conceptualise the same effect in terms of Paradigm Uniformity. In other words, the timing for *free-s* is influenced by the word-final prosodic lengthening at the end of *free*, but *freeze* is unaffected (Seyfarth et al., 2018). In this way, these durational effects may only be indirectly related to morphology, through prosody. This the essence of the Indirect Reference Hypothesis (e.g. Inkelas, 1990), in which elements of morphosyntax may have apparent phonetic correlates without breaking modularity. The phenomenon at hand motivates looking for this prosodic structure even at a sublexical level. I will explore this matter further in the domain of perception in Chapter 4.

1.1.2.2 Incomplete neutralisation

While durational effects corresponding to morphological structure may find a compelling prosodic explanation, some similar effects are not so straightforwardly reconciled. One such effect is popularly known as incomplete neutralisation. Neutralisation describes a classic phonological process whereby a contrast that is made in one environment is lost in another (Jakobson et al., 1951). While neutralisation may affect a number of features, phonological literature has focused in particular on the loss voice contrasts. Table 1.1 illustrates an especially famous case: the process of syllable-final obstruent devoicing in German.

Rad	[ʁaːt]	'wheel'	∼	*Räder*	[ʁæːdɐ]	'wheels'
Rat	[ʁaːt]	'council'/'advice'	∼	*Räte*	[ʁæːtə]	'councils'

Table 1.1: Neutralisation of syllable-final obstruent voicing in German

In the left column of this table is a pair of singular German nouns that are traditionally described as homophones, whose stems are nonetheless distinct in their respective plural forms. The canonical explanation for the neutralisation pattern in German is that only voiceless obstruents are allowed in syllable-final position. This means that underlyingly voiced obstruents come to lose their voicing feature when they appear syllable-finally and should be indistinguishable from underlyingly voiceless obstruents in the same position. In Table 1.1, the plural form *Räder* features a voiced obstruent [d] in onset position, but when

the plural suffix is removed and this obstruent becomes word-final in *Rad*, it is devoiced and transcribed with an identical phonological form to *Rat*. This is the type of morphophonological process that was standardly thought to be the epitome of classic categorical phonology. However, several investigations into the phonetic realisation of obstruent voicing neutralisation have discovered that homophones-by-derivation like German *Rad* and *Rat* are not completely identical. Rather, there exist fine-grained subphonemic differences in the production that point to residual 'voicing' in words with underlyingly voiced obstruents. In other words, the neutralising process of obstruent devoicing is 'incomplete'. As well as in German (Mitleb, 1981; Port et al., 1981; Port and O'Dell, 1985; Roettger et al., 2014), small but systematic differences between word- and syllable-final voiceless obstruents with different underlying specifications are reported for Dutch (Ernestus and Baayen, 2006; Warner et al., 2004), Russian (Dmitrieva et al., 2010; Kharlamov, 2014), Polish (Slowiaczek and Dinnsen, 1985), and Catalan (Dinnsen and Charles-Luce, 1984; Charles-Luce, 1993).

Ordinary phonemic voicing contrasts have a number of phonetic correlates, in addition to glottal activity during the obstruent itself. In the transition into and out of voiced obstruents, the F0 and F1 of adjacent vowels are slightly lower compared to voiceless obstruents (Hombert et al., 1979; Kingston and Diehl, 1994); constriction durations for voiceless obstruents are significantly longer on average than for voiced obstruents (Kluender et al., 1988); and vowels preceding voiceless obstruents are significantly shorter on average than those preceding voiced obstruents (Chen, 1970). The latter phonetic correlate for obstruent voicing mentioned—longer preceding vowels—is the one most commonly reported to not be completely neutralised in obstruent devoicing processes. In a meta-study on the most thoroughly investigated German case, Nicenboim et al. (2018) estimate a residual contrast of 10ms in vowels preceding underlyingly voiced and voiceless obstruents, based on reports of differences ranging from 6ms and 16ms. Cross-linguistically, most studies report a range of 10–15ms differences between vowels preceding underlyingly voiced obstruents and vowels preceding underlyingly voiceless obstruents (where both obstruents are canonically assumed to be voiceless on the surface), but some studies report much smaller differences. In Dutch,

Warner et al. (2004) observe vowels preceding underlyingly voiceless obstruents are just 3.5ms shorter than vowels preceding obstruents that are voiceless-by-derivation. As well as preceding vowel duration, some other phonetic parameters have occasionally been found to show small incomplete neutralisation effects. In Russian (Kharlamov, 2012) and in German (Port and O'Dell, 1985), there are reports of both slightly longer constrictions and shorter glottal pulsing for underlyingly voiceless obstruents, compared to underlyingly voiced obstruents that have been devoiced. These latter findings notwithstanding, it is striking that cues to incomplete neutralisation are overwhelmingly found in terms of preceding vowel duration, despite the fact that preceding vowel duration is just one of many cues to voicing contrasts more generally.

Another well-known neutralisation process is American English flapping, in which underlying /t/ and /d/ both become [ɾ] in certain prosodic contexts (Kahn, 1980). Unlike final obstruent devoicing, the neutralisation of American English flapping is symmetrical: both /t/ and /d/ are transformed into something that is featurally distinct from both, rather than one being changed to resemble the featural specifications of the other. However, just like final obstruent devoicing, several studies report very similar incomplete neutralisation effects in flapping. The relevant phonetic parameters include longer preceding vowel length for flaps with underlying /d/ than underlying /t/; shorter closure durations for flaps with underlying /d/ than underlying /t/; and, to a lesser extent, a smaller reduction in spectral intensity for flaps with underlying /d/ than underlying /t/ (Fisher and Hirsh, 1976; Fox and Terbeek, 1977; Patterson and Connine, 2001; Herd et al., 2010). These effects are similarly subtle, and some studies report data from particular speakers (Joos, 1942) and particular phonological environments (Huff, 1980) that show no such residual contrasts. Once again, and in a process where the outcome of the neutralisation process is voiced rather than voiceless, durations appear to be the key residual cues to underlying voicing.

While most cases of incomplete neutralisation focus on feature-level and segment-level contrasts, another example of incomplete neutralisation concerns vowel length. A striking example of this comes from Japanese, which contrasts long and short vowels (e.g. [obasan]

'aunt' vs. [obaasan] 'grandmother'). However, there is a minimum weight requirement of 2 morae for any licit prosodic word (Itô, 1990; McCarthy and Prince, 1993). This bimoraicity requirement means that some underlyingly monomoraic words are lengthened when there are no other elements affixed to them. For example, in the reciting of telephone numbers the word for the number 2, [ni], is realised as [nii]. Additionally, monomoraic nouns (e.g. [ki] 'tree') are lengthened when they are not followed by a case marking particle ([kii]). Braver and Kawahara (2016) performed an experiment to elicit underlyingly monomoraic nouns with and without case-marking particles, as well as underlyingly bimoraic nouns (with long vowels). They report that in cases where the long vowel is derived due to the absence of a case-marking particle, it is more than 30ms shorter on average than cases where a long vowel is underlying. Braver (2019) notes that this large difference is remarkable, because unlike word-final obstruent devoicing, there is evidence to suggest that derived long vowels are independently treated as 'long' by separate phonological processes. Similar incomplete neutralisation of vowel duration has been suggested to take place in Swedish (Bruce, 1984; Hayes, 1995), St Lawrence Island Yupik (Krauss, 1975; Hayes, 1995), Tongan, and Wargamay (Hayes, 1995). However, it is difficult to rule out possible confounding effects of stress-based lengthening in these languages.

Incomplete neutralisation effects are, on the whole, very small and show a high level of variance in terms of their magnitude. Indeed, while some perceptual studies have demonstrated that many listeners can distinguish between voiceless and 'devoiced' obstruents with above-chance accuracy (Port and O'Dell, 1985; Roettger et al., 2014), performances on these tasks generally show fairly high error rates. In fact, Fourakis and Iverson (1984) argue that German incomplete neutralisation is the result of a hypercorrect 'spelling pronunciation' in which the orthographic representations presented to participants in a laboratory setting influences pronunciation. This production–perception asymmetry is reminiscent of the case of 'near-merger' (Labov, 1972b) in which speakers maintain a small difference between classes of words in production but report that there is no such difference. However, near-mergers may differ from incomplete neutralisation in that the former is typically seen as the last

stage in a merger-in-progress, before any contrast in obliterated. The small contrasts observed in incomplete neutralisation phenomena are often the reflexes of historic contrasts that should have disappeared before the living memory of a language's speakers. As such, incomplete neutralisation poses an extra problem: how can speakers acquire and reproduce a contrast that is too small for them to reliably be aware of it, presumably across multiple generations?

Unlike the cases of prosodic lengthening reviewed in §1.1.2.1, the incompletely neutralised contrasts reviewed in this section do not directly mark differences in morphological structure. German *Rad* and *Rat* are isomorphic: monomorphemic singular nouns. However, they still imply a relationship between morphology, as it is represented in the lexicon, and phonetics. A common explanation for incomplete neutralisation makes reference to paradigm uniformity, as well as exemplar theoretic accounts in which phonetic forms stored in memory for a given word and morphologically related words are coactivated during production (Ernestus and Baayen, 2006; Bybee, 2001; Pierrehumbert, 2003). In other words, *Rad* is treated as if the word-final obstruent were somewhat more voiced than that in *Rat*, because *Rad* is morphologically related to—and coactivated with—words like *Räder*, in which the corresponding obstruent is fully voiced. However, this perspective on incomplete neutralisation is challenged by two pieces of evidence. Firstly, Braver (2014) reports that American English Flapping in nonce words, with presumably no corresponding exemplars or morphological paradigm stored in speakers' memory, also exhibits incomplete neutralisation effects. Secondly, Kaplan (2017) demonstrates that incomplete neutralisation in Afrikaans is asymmetric: while small preceding vowel duration differences exist between underlyingly voiceless and devoiced obstruents (e.g. *Rat* and *Rad*), no such difference is observed between voiced obstruents with and without devoiced obstruents (e.g. *Räder* versus *Leder* 'leather'). A standard exemplar theoretic framework should have *Räder* coactivate with *Rad*, regardless of which one is the target word. In this way, the pronunciation of *Räder* should show signs of slight devoicing, compared to a word like *Leder* with no morphological relatives in which the corresponding obstruent is devoiced.

1.1.2.3 Paradigmatic effects

Despite some complications, the leading explanations for incomplete neutralisation phenomena make reference to the influence of morphologically related words on phonetic form. But as Roettger et al. (2014) point out, we might predict words to be differently influenced by related words depending on those related words' frequency, recency, or quantity. That is, a given word should be most strongly influenced by related words that are strongly coactivated (e.g. because they are highly frequent in general). Investigations into the effects of these different properties has yielded mixed results. It is fairly well documented that frequent words are themselves realised with shorter (Wright, 1979; Kawamoto, 1999; Aylett and Turk, 2004; Gahl, 2008) and more reduced (Munson and Solomon, 2004; Gahl, 2008; Lin et al., 2014) pronunciations than infrequent words. However, some recent investigations have found words that are highly frequent compared to the frequency of their morphological relatives are enhanced and pronounced with *less* phonetic reduction. Some early examples of this result are focused on 'pockets of uncertainty' between functionally equivalent forms that directly compete for use in the same position. For example Kuperman et al. (2007) investigate Dutch compound linking morphemes (*-e(n)-* and *-s-*) and find that the linking morpheme itself is longer in duration when it is the most likely choice for a given compound, and Cohen (2015) finds a similar effect in terms of unreduced vowel quality in the most relatively frequent in a pair of two contextually licensed (and therefore competing) Russian suffixes. The notion of a 'pocket of uncertainty' aligns fairly closely with mainstream variationist conceptions of the linguistic variable, and enhancement type results with those found in some explorations of variant frequency. For example, Bürki et al. (2011) find that variable word-medial schwa in French (e.g. *fenêtre* [f(ə)nɛtʁ] 'window') is longer in words that appear relatively more frequently with schwa compared to without it. Kuperman et al. (2007) conceptualise this mechanism, dubbed the Paradigmatic Signal Enhancement Hypothesis, in terms of speaker confidence. Speakers feel confident in selecting the most relatively frequent form among competing forms and do not 'hold back' in production.

This literature has been extended to investigations into different forms that appear in the same position but do not represent direct competition because they are not functionally equivalent. Tomaschek et al. (2019) find greater durations for word-final /s/ in American English when it represents a common ending for a given stem. A further extension sees a number of studies measuring variables that are represented consistently within their paradigms. Tucker et al. (2019) and Tomaschek et al. (2021) report enhancement of American English stem vowels (duration and quality) when they appear with their most frequent inflectional endings. Bell et al. (2021) investigate American English compound words, and find that final segment duration in the first word of a compound is longer when the second word is a highly likely ending (i.e. the first word frequently appears in a compound with the second). And, relatedly, Lõo et al. (2018) show Estonian whole words are produced with longer durations when they have fewer related words in their paradigm. Rather than focusing on a 'pocket of uncertainty' between forms in direct competition (e.g. Kuperman et al., 2007; Bürki et al., 2011; Cohen, 2015), or between forms that appear in a similar position even though they are not functionally equivalent, these studies seem to suggest that the paradigmatic enhancement effect that comes with selecting a favourable (frequent) member of a paradigm permeates the whole word, even reinforcing the phonetic form of the stem itself. However, these effects are not at all consistent in the wider literature. While Cohen (2015) finds enhancement of words that are frequent compared to a competing word, she reports reduction for words that are frequent relative to their whole inflectional paradigm. Plus, several studies on English prefixes find they are *shorter* when they appear in a word that is frequent relative to its base, or with a more opaque relationship to its base. This has been replicated for *dis-* (Smith et al., 2012; Plag and Ben Hedia, 2017), *mis-* (Smith et al., 2012), *un-* and *in-* (Ben Hedia and Plag, 2017; Plag and Ben Hedia, 2017) prefixes. These results are commonly framed in terms of the morphological 'segmentability' of complex words, such that a word whose base is relatively infrequent or opaque (e.g. *distort*) is less likely to be broken up into morphological pieces and is produced quickly as a whole word.

An early segmentability finding of this type focused on the relationship between English /t/-final stems and suffixes, and on non-durational effects of phonetic reduction. Specifically, Hay (2004) argues that word-medial /t/ was more likely to be deleted in a word like *swiftly* than in *softly* because *swiftly* is highly frequent relative to its base (*swift*), while *softly* is infrequent relative to its base (*soft*). This means that *swiftly* is most efficiently produced as a whole-word unit, while *softly* should be produced with cues to decomposition, including the /t/ at its morphological boundary. This finding is particularly relevant to the present dissertation, which focuses on the production of coronal stops. Other work looking at frequency dynamics within a paradigm and the effect on coronal stop production have been focused on Dutch. Schuppler et al. (2012) find an enhancement effect where morphological /t/ marking 3SG is more likely to be deleted if this form is infrequent relative to the form without 3SG marking. Hanique and Ernestus (2011) find the opposite effect for /t/ marking past tense in Dutch irregulars, such that it is both less likely to be deleted and longer when this form is infrequent relative to other reflexes of the same lemma. We will explore this idea further in Chapter 5.

1.1.2.4 Morphologically-conditioned variable alternations

The previously reviewed examples of morphologically conditioned phonetics have primarily concerned durational parameters, but the effect of morphology is not (it seems) limited to this. Morphological structure is an important factor in the patterning of a number of variables in the variationist literature. These are traditionally framed in terms of discrete phonological alternants. In English, for example, flapping is sensitive to paradigm uniformity, and is licensed in *capitalistic* (because it is also licensed in *capital*) but not in *militaristic* (because it is also unlicensed in *military*) (Steriade, 2000; Herd et al., 2010). Similarly, word-final *-ing* is far more commonly realised with an alveolar nasal (as opposed to a velar nasal) when it represents a progressive suffix (e.g. *working*) than in a word like *awning* where it is tautomorphemic with the rest of the stem (Tagliamonte, 2004).

However, a number of variables that are sensitive to morphological structure have been

reframed in more gradient phonetic terms. Hayes (2000) notes that /l/ is more likely to be produced as dark when it is followed by a morphological boundary, but Sproat and Fujimura (1993) argue that /l/ darkening is a gradient phenomenon best described in terms of the magnitude and timing of the dorsal gesture. Several subsequent studies find that the magnitude of the dorsal gesture does indeed vary according to the presence of a subsequent morphological boundary (Sproat and Fujimura, 1993; Lee-Kim et al., 2013; Strycharczuk and Scobbie, 2016, 2017; Turton, 2017). This is noteworthy because these same /l/s do not exhibit the commonly-observed lengthening effects at morphological boundaries (Sproat and Fujimura, 1993; Strycharczuk and Scobbie, 2016, 2017). Instead, the presence morphological structure has phonetic correlates in the magnitude of lingual gestures.

Another example of a variable that is robustly conditioned by morphology is English Coronal Stop Deletion. Since I take Coronal Stop Deletion as my primary case study in this dissertation, the following section is devoted to reviewing previous research on the variable, and setting up the subsequent chapters in which its architectural properties are probed. Ultimately, it appears that Coronal Stop Deletion may not be exceptional, but represent a confluence of more than one other type of morphologically-sensitive phonetic phenomenon.

1.2 Coronal Stop Deletion

English Coronal Stop Deletion (CSD, sometimes called '/t,d/ Deletion') is the variable surface absence of an underlying word-final coronal stop following a consonant. This means that a word like *act* is sometimes pronounced with a coronal stop (e.g. [ækt]), and sometimes without (e.g. [æk]). The phenomenon is sketched as a phonological rule in (1.1)[2].

$$C^{\text{COR}} \rightarrow \emptyset / (C) C \_\_ \# \tag{1.1}$$

Since its first description as a process of cluster simplification (Labov et al., 1968),

[2]Some theories posit that CSD applies on multiple levels of a stratified phonology, including the stem level. In these cases the pound sign used in (1.1) can be taken to denote a domain-final position, if that domain is smaller than a word.

CSD has become one of the most thoroughly investigated phenomena in English variationist sociolinguistics. In fact, CSD has been found to be a feature of varieties of English around the world, including Southern British English (Baranowski and Turton 2020, cf. Tagliamonte and Temple 2005), General American English (Guy, 1980), African American English (Fasold, 1972; Labov, 1972b), Chicano English (Santa Ana, 1991), Appalachian English (Wolfram and Christian, 1976; Hazen, 2011), Canadian English (Walker, 2012), New Zealand English (Holmes and Bell, 1994), Singapore English (Lim and Guy, 2005), Hong Kong English (Hansen Edwards, 2016), Nigerian English (Gut, 2007), and English-lexified Jamaican Creole (Patrick, 1991). Among these studies, there are some striking consistencies in CSD's sensitivity to aspects of both phonological and morphological context.

One basic effect of phonological context on CSD is that CSD is far more common when a target coronal stop is followed by a consonant (e.g. *west square*) than a when it is followed by a vowel (e.g. *west avenue*). Several accounts attribute this effect to a variable resyllabification of the coronal stop to be the onset to a following vowel (Guy, 1991a; Kiparsky, 1993; Reynolds, 1994). This resyllabification will necessarily bleed CSD, which only targets coda stops in stem-final or word-final position. The following segment effect has additionally been found to be systematically modulated by the strength of the intervening syntactic boundary (Tamminga, 2018)[3], argued to be the result of an extra-grammatical effect of production planning. The preceding environment of an eligible coronal stop is also found to play a role in the likelihood that CSD will apply, but this effect is less obviously linked to sonority due to mismatches between the hierarchy of CSD-favouring environments and the sonority hierarchy. For example, CSD occurs more often following sibilants than stops (Guy, 1980, 1991a).

Of particular interest to many researchers is the observation that CSD is also sensitive to morphological context Guy (1991b). The most robustly observed effect of morphological context on CSD is that deletion occurs more frequently in monomorphemes (e.g. *pact*) than in words where the coronal stop constitutes a past tense *-ed* suffix (e.g. *packed*).

[3]It is likely that syntactic boundary strength will coincide with the strength of prosodic boundaries .

Undeniably, this effect has a 'functional' flavour: speakers are more likely to retain stops that are more informative, i.e. when they mark past tense. However, CSD rates for past forms are not significantly affected by an accompanying auxiliary verb (Guy, 1980). In such cases (e.g. *have walked*), tense is redundantly marked by both the auxiliary verb and the *-ed* suffix, so we might expect a higher rate of CSD if the morphological effect on CSD were truly driven by functional pressures.

1.2.1 Categoricity and gradience

For many variable phenomena, there is ongoing debate around the fundamental nature of their representation. More specifically, it can be unclear whether a variable phenomenon is the outcome of a non-obligatory phonological process, or is caused by phonetic or contextual variation in the implementation of an otherwise invariant phonological form. To probe this issue, we can make use of our understanding of how a variable that is represented as a phonological or phonetic process should be implemented. That is, where phonology and phonetics are held to be separate components of the grammar (but see Browman and Goldstein 1985 et seq.), they are typically characterised as differing with regard to categoricity[4] and gradience (Scobbie, 2007). Some phonetic variation is inevitable due to the fact that humans are inconsistent in how they physically produce language, but this kind of variation manifests as a distribution across a continuum—or several continua—representing dimensions of articulation, acoustics, and timing. Phonology, on the other hand, concerns computations over strings of discrete symbols (cf. work on gradience in phonological representation and derivation, e.g. Hayes, 2000; Smolensky and Goldrick, 2016), so variation is this domain should result in alternations between these discrete categories. Therefore, a growing body of work makes use of a variety of techniques to explore the implementation of variable phenomena, especially in varieties of English. For example, Zsiga (1995) presents electropalatography evidence that palatalised segments within words (e.g. *press*

[4]We use 'categorical' to refer to representations and implementations that comprise a set of discrete categories, rather than falling across a continuum. 'Categorical' is never be used to mean 'invariant' in this document.

$\sim$ *pressure*) are not equivalent to those that occur variably at word boundaries (e.g. *press you* [pɹɛʃju]), suggesting that the former type may be the result of a phonological process and the latter the result of coarticulation. Conversely, Ellis and Hardcastle (2002) find that individuals speakers differ in whether word-final nasals consistently fully assimilate to the following segment's place of articulation (e.g. *ban cuts* [bæŋkʌts]) or produce residual lingual contact at the alveolar ridge. Finally, Turton (2017) presents ultrasound evidence that light and dark coda /l/ differ in their sensitivity to rime duration, which she takes to support a categorical distinction between them.

The matter of categoricity and gradience with regard to CSD has yet to be thoroughly investigated in the same way, despite the large body of work devoted to describing the phenomenon. Using examples in her own acoustic data from speakers of York English, Temple (2014) highlights the need for this work by pointing to a number of previously overlooked ambiguities in acoustic-impressionistic analyses of CSD. For example, there are cases in which multiple different processes might affect a coronal stop or its environment. For example, in York English a /kt/ cluster (e.g. in *act*) may be replaced with a single glottal stop, such that it is not clear what combination of deletion and glottal replacement processes (targeting /k/ or /t/ or both) has given rise to the output. And even in cases of apparent CSD that are not ambiguous in this way, there is a general potential for lenition and coarticulation to give the illusion of phonological deletion. This much is evident from Browman and Goldstein's (1990) report of inaudible tongue tip raising for word-final coronal stops. The data for this observation comes from X-Ray Microbeam recordings of two instances of the sequence *perfect memory*, and Browman and Goldstein (1990) concluded that the stop in *perfect* was rendered inaudible not from a deletion rule, but rather how the articulators were coordinated across time. Either the coronal closure and release were masked by the overlap of dorsal or labial closures[5], or as a phonetic process in which the speaker undershot their target of the alveolar ridge and never made a complete closure in

[5]While articulatory overlap is not the same as deletion, Bermúdez-Otero (2010) notes that it is fairly straightforward to account for it as a phonological process linking adjacent segments to the same timing unit (e.g. /kt/ $\rightarrow$ [$\widehat{\text{kt}}$]).

the first place.

If, as has been shown for at least some isolated cases, apparent CSD is a gradient phonetic phenomenon more generally, this must be reconciled with the fact that it is sensitive to morphological structure. Indeed, the presence of morphological conditioning is itself sometimes taken as evidence for a phonological representation of CSD. This is because for theorists who assume some kind of feed-forward modularity in speech production (e.g Pierrehumbert, 2002; Bermúdez-Otero, 2007), the morphology and phonetics do not share an interface: phonology must intervene. A crucial argument in favour of phonological intervention is typological. That is, if morphosyntactic structure could directly inform phonetic implementation, we would predict widespread and varied phonetic effects for different types of morphosyntactic structure. A common example where morphosyntactic structure does seem to have phonetic correlates is when prosody intervenes. When morphosyntactic and prosodic boundaries are aligned, corresponding phonetic effects can be found, primarily concerning durational parameters (e.g. Selkirk, 2011).

Crucially, evidence of inaudible stops due to articulatory overlap or undershoot does not preclude the existence of cases of categorical deletion. Indeed, Lichtman (2010) conducted an investigation of the Wisconsin Microbeam Database with a follow-up EMA study and found relatively frequent categorical coda /t/ deletion, alongside evidence of covert tongue tip raising. Every speaker in this study exhibited at least some such tokens with no residual linguoalveolar gesture. It is important to note, however, that the majority of tokens included in this study were singleton stops, and not part of clusters as is typically held to be the environment where CSD is possible. As such, some of these instances may be the result of different processes such as /t/-glottaling. Indeed, Heyward et al. (2014) provide evidence that instances of intervocalic phrase-medial /t/-glottaling in the ESPF DoubleTalk Corpus typically feature no evidence of residual tongue tip raising. Instead, this phenomenon resembles true categorical allophony in that /t/ loses its tongue tip articulation and is realised via glottal constriction only. Restricting the sample to just word-final coronal stops in clusters, an earlier EMA study of mine (Purse and Turk, 2016) looked at a small sample

from twelve speakers of British Englishes across a number of tasks and found that inaudible tokens with no accompanying tongue tip raising were relatively rare. However, there was a small effect of morphological class on the phonetics among the speakers of Southern British English, such that regular past *-ed* suffixes were produced with higher normalised tongue tip raising than coronal stops at the end of monomorphemes. Whether or not structure preserving segmental deletion is a possible CSD variant, several studies have found that the majority of cases of inaudible coronal stops do not fit this description. This raises questions about whether CSD should be represented as a categorical segmental process at all, and how we can account for the phenomenon's sensitivity to morphological context otherwise.

1.2.2 Formal accounts of CSD

CSD has many properties—variability; sensitivity to phonological and morphological context—that make it an interesting puzzle for phonological and phonetic theory. Several potential solutions to this puzzle have been proposed, which vary in their capacity to account for implementations other than categorical deletion. Some such proposals approach the problem using classic segmental phonology. In one well-known example, Guy's (1991b) Exponential Hypothesis attributes differences in deletion rates to the stages of word-formation. In this account, monomorphemes undergo CSD at a higher rate because a deletion rule is, variably, applied at multiple levels of a stratified morphophonology (e.g. Kiparsky, 1982). However, *-ed* suffixes are attached at a late stage of this process, and are only eligible for one level's CSD process. Table 1.2 is a sketch, according to the Exponential Hypothesis, of the morphophonological formation of example words that undergo CSD at some frequency. Cells are shaded where CSD cannot apply, and the process is effectively counterfed, because the relevant coronal stop constitutes a suffix that is separated from the stem and not attached until a later level of the morphophonology.

An abstract segmental approach lends itself to a lot of well-established theoretical machinery, and various rule-based and constraint–based accounts of CSD take advantage of this (Coetzee and Pater, 2011). However, mainstream segmental phonology has very lit-

	Monomorphemes	Semiweak Past	Regular Past
	mist	*kept*	*missed*
Level 1	mɪst	kɛp + t	mɪs # d
Level 2	mɪst	kɛpt	mɪs + d
Postlexical	mɪst	kɛpt	mɪst

Table 1.2: Relevant input to each level of phonology, from words in morphological classes with different CSD rates. # and + indicate morphological boundaries separating suffixes from stems according to the Exponential Hypothesis.

tle capacity to account for gradience in the implementation of CSD, and thus is difficult to reconcile with results of frequent tongue tip raising for inaudible stops (Browman and Goldstein, 1990; Lichtman, 2010; Purse, 2019). This is especially true for a stratal account like Guy's (1991b) Exponential Hypothesis, which posits CSD rules even at early levels of phonology. Since the output of early phonology forms the input to subsequent levels of phonology, it is important that early phonological rules be at least categorical if not structure-preserving Myers (1995). Otherwise, we would need to define the featural content of the outputs of a gradient process (i.e. everything on a continuum between /t,d/ and Ø).

Another well-known approach to explaining patterns of CSD is found in usage-based models like Exemplar Theory (Bybee, 2002). In this model, each time a word is produced, the phonetic signal is based on a selected exemplar from all the speaker's memory traces of encountering that word, and similar words. This means that if speakers perceive CSD as categorical, they will select from exemplars of a word with and without a coronal stop, and vary categorically in production. However, in other Exemplar Theoretic models, production targets are more explicitly informed by a region of multiple exemplars in phonetic space (Pierrehumbert, 2002). In the case of CSD, if this region is wide enough to activate exemplars with both deleted and retained coronal stops, speakers might compute a coronal stop target that is intermediate between full alveolar closure and complete absence. Thus, there is potential for gradience in this approach. On the other hand, in strong versions of Exemplar Theory language users do not explicitly store morphological structure in memory

or factor it into speech production calculations, so the effect of morphology on the rate of CSD is not directly explained. Bybee (2002) suggests that the morphological effect is actually emergent from coincidental frequency effects that regular *-ed* suffixed words simply occur less frequently compared to monomorphemes, and typically in phonetic contexts that favour stop retention. As such, the effect of grammatical class on rate of CSD is a side-effect of the fact that many of the words in these classes accumulate different proportions of deletion-favouring and -disfavouring representations. However, several CSD studies have observed morphological effects that remain even when measures of lexical frequency are controlled for.

A third option for explaining both morphological sensitivity and the potential for gradience in the implementation of CSD is revealed when we begin to move away from a segmental representation of the effect, at least in terms of its primary locus. Instead, it is possible that what has been described as deletion has the same source as other morphologically sensitive phonetic phenomena reviewed in Section 1.1.2, which are primarily durational in nature. This alternative account is a common refrain throughout this dissertation, especially as I begin to demonstrate the gradient phonetic implementation of apparent CSD that other accounts struggle to capture. Generally speaking, we would expect for phonetic lengthening to be accompanied by less overall lenition, including more fully realised stops. On the other hand, when coronal stops are articulated in shorter intervals they are more likely to be reduced or seemingly omitted as the same target is more difficult to reach (Parrell and Narayanan, 2018), and as other constrictions in the same cluster overlap with the coronal stop constriction (Browman and Goldstein, 1990). With these potential effects in mind, the findings of lengthening effects at morphological boundaries align neatly with robustly attested morphological patterns in rates of CSD application (Guy, 1980). Monomorphemes, with no morphological boundary to induce lengthening effects, are more likely to have word-final coronal stops that are deleted (or otherwise rendered inaudible. On the other hand, *-ed* suffixes, which are immediately preceded by a morphological boundary, may induce some degree of lengthening (Lociewicz, 1992; Sugahara and Turk, 2009) that could lead to

more fully articulated instances of their coronal stop exponent due to the mechanisms just outlined. These lengthening effects would allow for grammatical conditioning of phonetic parameters that does not necessarily violate modularity insofar as they are framed as a matter of sublexical prosody. This the essence of the Indirect Reference Hypothesis (e.g. Inkelas, 1990), in which elements of morphosyntax may have apparent phonetic correlates because they are encoded in an intermediate level of phonology that deals with timing and phrasal prominence. A corollary of this hypothesis is, of course, that the phonetic correlates found should be those that are relevant for prosody more broadly, like duration.

Chapter 2

Investigating the articulation of Coronal Stop Deletion

In chapter 1, I argued that variable Coronal Stop Deletion (CSD) in English is a site of both robust morphological conditioning and ambiguity at the phonetics–phonology interface. This chapter chiefly comprises an investigation into this ambiguity, focusing on the issue of categoricity and gradience in the implementation of CSD variants. The results, from a comparison of acoustic-impressionistic and Electromagnetic Articulography (EMA) data, demonstrate that complete CSD without residual tongue tip raising is indeed rare and exhibits no obvious patterns in its distribution. Moreover, in exploring the magnitude of tongue tip raising in different contexts, I show effects of morphological class on this gradient phonetic measure. These findings cement CSD as an apparent morphologically-conditioned phonetic phenomenon.

2.1 Articulatory sociophonetics

Articulation and acoustics are associated non-linearly: this much is well understood. Stevens (1972, 1989) describes the relation between articulation and acoustics as 'quantal' such that some large articulatory movements have little acoustic consequence, while some small adjustments are used to produce qualitative (even phonemic) differences. The case of stop consonants is a particularly strong example of this. Only a complete closure at the alveolar ridge will produce the characteristic stop and release burst of a [t] or [d]. However, there is a wide range of tongue tip positions along a superior-inferior dimension, each one of which

(as long as no closure is made) produces an acoustic signal that is no more or less stop-like than the next. Consequently, a coronal gesture towards a stop closure that is undershot by even a very small distance along this axis could be acoustically indistinguishable from a case of true categorical deletion with no residual articulation of a coronal stop. Thus, controversy surrounding issues of categoricity and gradience—phonology and phonetics—in CSD necessitates the exploration of articulatory data. This will allow us to observe the range of articulations underlying the categorical perception of stop deletion or retention that are inextricable from the acoustic signal.

The research that has been conducted on this narrow topic has yielded some mixed results. In a classic work from Browman and Goldstein (1990), X-Ray Microbeam data revealed evidence of inaudible tongue tip raising for word-final coronal stops. For the two instances of the sequence *perfect memory* where this was observed, Browman and Goldstein (1990) conclude that the stop was not rendered inaudible from a deletion rule, but rather the difficult coordination of articulators across time. Either the coronal closure and release were either masked by the temporal overlap of adjacent labial and dorsal closures, or the speaker undershot their target of the alveolar ridge and never made a complete closure in the first place. Similarly, Purse and Turk (2016) observe that apparent CSD without tongue tip raising is rare in the ESPF DoubleTalk Corpus. However, for 9 tokens (25% of all inaudible coronal stops) speakers appeared to produce no tongue tip raising, and some of these featured distinct downward tongue tip movement. It is still not clear, however, whether these 9 tokens constitute categorical deletion or just one extreme of a continuum of lenition. On the other hand, Lichtman (2010) conducted an investigation of the Wisconsin Microbeam Database and a follow-up EMA study, and reports widespread articulatory categoricity in coda /t/ deletion such that all speakers exhibit some tokens with no residual linguoalveolar gesture. It is important to note, however, that the majority of tokens included in this study are singleton stops, not part of consonant clusters as is typically held to be the environment that is eligible for CSD. As such, some of these instances may be the result of different processes such as /t/-glottaling. Indeed, Heyward et al. (2014)

provide evidence that instances of /t/-glottaling in the ESPF DoubleTalk Corpus typically feature no evidence of residual tongue tip raising. Instead, this phenomenon resembles true categorical allophony in that /t/ loses its anterior place of articulation and is realised on another tier entirely.

The paucity of research making use of articulatory data is a problem for sociophonetics in general, which lags behind the fields of traditional phonetics and speech pathology in this respect. Indeed, in Thomas's (2008) review of instrumental phonetic techniques for sociolinguistics, there is no mention of methodologies for measuring articulation. One major obstacle to the incorporation of these methodologies in sociolinguistics is the normalisation of measures across multiple speakers with different anatomies. Studies in articulatory sociophonetics must explore creative solutions for making meaningful comparisons between observations from various individuals. Another potential obstacle lies in the elicitation of the naturalistic speech that is of primary interest to sociolinguists. Where Labov's (1972b) style-shifting paradigm classifies all laboratory speech as a context in which attention-paid-to-speech is bound to be high, any setup for physically observing articulators is likely only to exacerbate this problem. However, Boyd et al. (2015) report that speech from laboratory tasks is largely comparable with speech produced in a sociolinguistic interview. While they do not include a comparison with a context in which articulatory data is gathered, this finding is encouraging in its suggestion that the effects of researcher observation are fairly consistent. It seems likely that even in the present study the effects of conspicuous articulatory methodology should not be so egregious as to level out all the potential variation.

Some areas of sociophonetics have already benefitted from articulatory studies to explore the reality of speakers' production strategies. For example, it is understood that an English approximant /r/ can be produced with various covertly allophonous tongue shapes, which can be broadly categorised in terms of a bunched versus retroflex taxonomy (Delattre and Freeman, 1968), but which are perceptually indistinguishable (Twist et al., 2007). Using Ultrasound Tongue Imaging (UTI), it has been demonstrated that speakers adapt

their /r/ articulation to produce the least effortful allophone given the phonetic context (Stavness et al., 2012) and perturbations of their articulators (Tiede et al., 2011), and that children explore different articulations during acquisition (Magloughlin, 2016). Further, /r/ articulation has been shown to be class stratified in Scottish English (Lawson et al., 2011), and to play a key role in the actuation of s-retraction in North American English (Mielke et al., 2010; Baker et al., 2011; Mielke et al., 2016).

As well as revealing covert ranges of articulatory movement (e.g. incomplete tongue tip raising for coronal stops), and covert allophony (e.g. different lingual configurations for English approximant /r/), articulatory data is particularly informative regarding the matter of timing in speech production. Timing in speech is relative, such that we can conceive of it as the covariation of multiple objects in time, or as variation along a continuum of time between gestures. Relative timing of lingual gestures has been revealed to be a key locus of variation for /l/ darkening (Sproat and Fujimura, 1993; Bermúdez-Otero and Trousdale, 2012; Turton, 2017) and Scottish English /r/ pharyngealisation (Lawson et al., 2018). In these cases, a delayed anterior gesture—sometimes until after voicing ends—leads to a darker /l/ and a more pharyngealised /r/ respectively.

Timing of gestures is also the main locus of variation available in an Articulatory Phonology framework (Browman and Goldstein, 1990). Here, temporal overlap of gestures associated with adjacent speech sounds gives rise to coarticulatory outcomes or the masking or omission of a gesture. Further, Davidson and Stone (2004) explore a timing analysis for English speakers' production of excrescent vocoids that variably appear in phonotactically illegal clusters. They conclude that this is indeed a case of gestural mistiming based the absence of evident vowel targets in their UTI data. This result implies, as Browman and Goldstein (1990) predict, a continuum of overlap between any two given gestures. At one end of a continuum of this kind of overlap, where gestures are maximally separated in time, there is the potential for apparent vowel epenthesis phenomena as in Davidson and Stone (2004). The other, extreme, end of this kind of continuum of gestural overlap is presumably metathesis, in which the underlying order of speech sounds is reversed on the

surface. In cases of both variable vowel epenthesis and elision, and variable metathesis, a clear research goal should be to observe how variants are distributed across a continuum of gestural overlap. A unimodal distribution in the degree of gestural overlap would suggest that the variable perception of vowels or reordering of segments is a product of a gradient timing relationship between gestures. On the other hand, a bimodal distribution of these results would suggest the existence of discrete categories. In the latter case we could surmise that speakers have separate targets for multiple potential surface forms. Articulatory data is crucially poised to provide this kind of evidence. With it, a researcher can disentangle simultaneously produced gestures and is not limited to data on these gestures' completion but also their onset or 'Maximum Acceleration Event' (Perkell and Matthies, 1992) as may be most informative concerning the planning and execution of articulatory timing.

This chapter offers a contribution to this growing body of work on articulatory sociophonetics. The findings presented for word-final coronal stops touch on many of the topics reviewed here. We can observe systematic variation across the articulatory continua of tongue tip height and degree of raising, which are not evident in the acoustic signal. Further, an exploration of individual differences reveals the potential for patterns of covert allophony as evidenced by a multimodal distribution of articulatory outcomes for some speakers.

2.2 Materials and Methods

2.2.1 Procedure

Synchronised acoustic and articulatory recordings were collected using an NDI Wave Electromagnetic Articulograph and a microphone, through NDI's native software NDI Wavefront. Electromagnetic Articulography (EMA) sensor coils were adhered to key oral articulator points at the tongue tip (TT), tongue dorsum (TD), and lower lip (LL), as well as a reference point on the upper incisors (UI) using a non-toxic high viscosity cyanoacrylate oral adhesive. Three further reference sensors were aligned to each participant's left and right

mastoids and bridge of nose using a lensless spectacle frame that was held in place with surgical tape. The spectacle-mounted sensors were used to define a participant's sella-nasion plane at each frame of time, and new axes (inferior-superior, posterior-anterior, left-right) were created according to this plane to correct for participant head movement. The origin of each new axis was then aligned to the UI sensor to improve interpretability.

All 5 participants are native speakers of Mainstream American English. Each one performed several tasks designed to elicit naturalistic speech. These were a Map Task (participants describe a route on a map so that an interlocutor can draw it), a Semantic Differential Task (participants explain the difference between near-synonyms), two Reading Passages, and finally a Wordlist. Tasks were consistently ordered in the way presented here under the assumption that this creates a continuum of style such that each task evokes a higher degree of metalinguistic awareness than the last. This follows some classic sociolinguistic methodology (Labov, 1972b) in which researchers attempt to tightly control and gradually increase speakers' degree of self-monitoring across the duration of the experiment. This is an important consideration given the observation that articulatory methodologies are likely to already induce a high level of self-monitoring and metalinguistic awareness. Stimuli for all tasks were designed so as to require participants to produce as many critical items as possible, where a critical item is a word with an underlying word-final stop following another consonant, with no other adjacent coronal segments.

2.2.2 Data Manipulation

As previously mentioned, some by-speaker and by-token normalisation is required in order to compare observations from different speakers. Sensor positions cannot be precisely equivalent because the size and shape of each speaker's body is not the same. For each speaker, a measure of the greatest TT height (mm from plane at UI) for a coronal stop closure was recorded and represented a speaker-specific maximum (MAX). For every token, the TT tangential velocity minima immediately preceding and following (A and B, respectively) are defined as coordinates in Time (s) and TT height (mm), and used to define the baseline

AB. The TT tangential velocity minimum corresponding to coronal stop articulatory target for this token (*T*), where the tongue has been raised, is also recorded in terms of TT height and Time. These values are then used to calculte the distance from *AB* to *T* (*h*), which is then redefined as a proportion of the distance from *AB* to that speaker's *MAX* (*H*). This normalised measure of raising is always ≤1, since it is calculated from *h*/*H*. Figure 2.1 is a schematic with the component parts for this normalisation procedure, for a TT height trajectory across time corresponding to a hypothetical coronal stop closure and release.

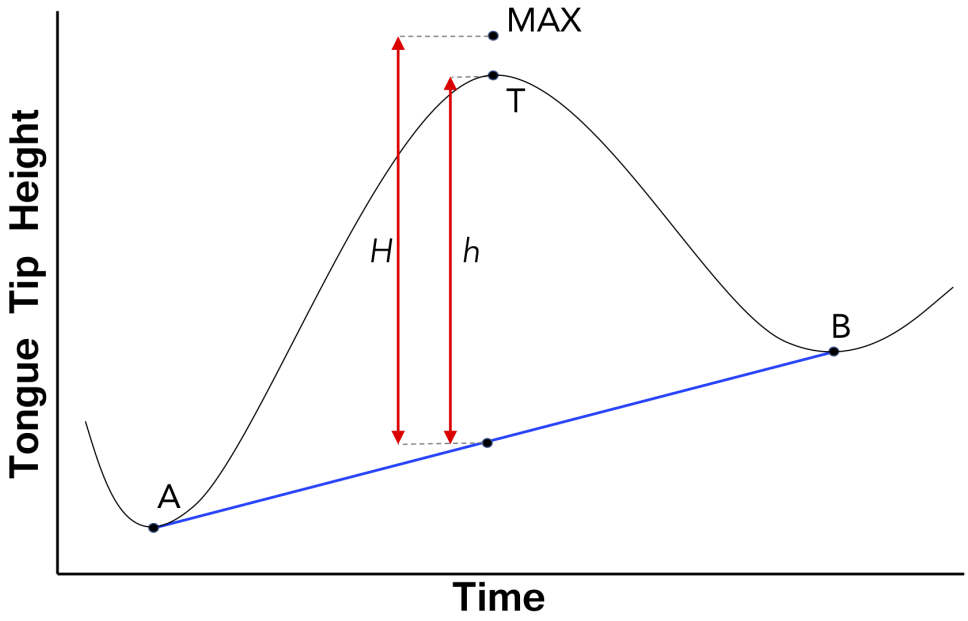

Figure 2.1: Schematic for calculating normalised measure of tongue tip raising.

The normalised measure that is described here could be thought of as a measure of degree of effort expended to reach a speaker-specific maximum TT height. This is exceptionally relevant to CSD as evaluated as a lenition phenomenon, especially when lenition is narrowly construed in terms of a reduction in the magnitude of potential articulatory movement. Values of this measure are no longer particularly informative about absolute tongue tip height, because they are strongly affected by the height of the baseline *AB* from which TT raising takes place. In other words, we may observe very little raising to reach a relatively

high peak if the corresponding baseline is already high. At the same time, this method provides a reliable criterion for identifying complete deletion in that values ≤ 0 denote a complete absence of tongue tip raising from baseline *AB*. This means that measures of both *raising*—a normalised measure of the proportion of maximum articulatory movement from a baseline—and raw *height*, are crucial and crucially different for the present study[1].

2.3 Evaluating 'deletion'

2.3.1 Acoustic-impressionistic coding

This chapter is primarily an investigation of the articulatory reality of word-final coronal stops. However, in order to speak to the previous literature on CSD, the data must first be evaluated in these traditional terms. Table 2.1 shows the rates of CSD in each of the basic morphological classes this chapter considers (monomorphemes, 'complex' regular passive and preterite forms, and semiweak past forms). These tokens were coded according to auditory and spectrographic cues to the presence or absence of a coronal stop. In terms of phonetic environment, only tokens that were immediately followed by another coronal stop or an interdental fricative, since these are almost always neutralising (Temple, 2009).

	Retained	**Deleted**	**Total**
Mono	316 (52%)	286 (48%)	602
Complex	297 (74%)	101 (26%)	398
Semi	24 (72%)	9 (28%)	33
Total	637 (63%)	396 (37%)	1033

Table 2.1: Auditory/acoustic coding of coronal stop retention and deletion.

The well-attested effect of morphological class on CSD is also found in this data, such that CSD ostensibly occurs at a significantly higher rate in monomorphemes than morpho-

[1]Therefore, 'raising' and 'height' are not used interchangeably in this dissertation.

logically complex words ($\chi^2 = 49.5$, $p < 0.00001$). Indeed, according to these results, CSD occurs almost twice as frequently in monomorphemes as in passive or preterite forms. The semiweak past forms appear to pattern with complex forms, but the sample size for this subset is too small to be particularly informative.

2.3.2 Articulatory zeroes

The bulk of the analysis in this chapter concerns tokens of underlying coronal stops with no adjacent coronal segments. The rate of perceived CSD in this subset, based on auditory and spectrographic cues, is a respectable 24%. However, it is not clear from this evidence whether any instance of apparent CSD actually constitutes an absence of any attempt to produce a stop, as the classic analysis implies. In the present analysis, there were 15 tokens in which there was no evidence of tongue tip raising. In each of these, the TT trajectory during interval of time corresponding to an underlying coronal stop stayed level with AB or moved downwards. None of these featured audible stops and would have been judged to be 'deleted' in a traditional CSD analysis. Figure 2.2 is one such token.

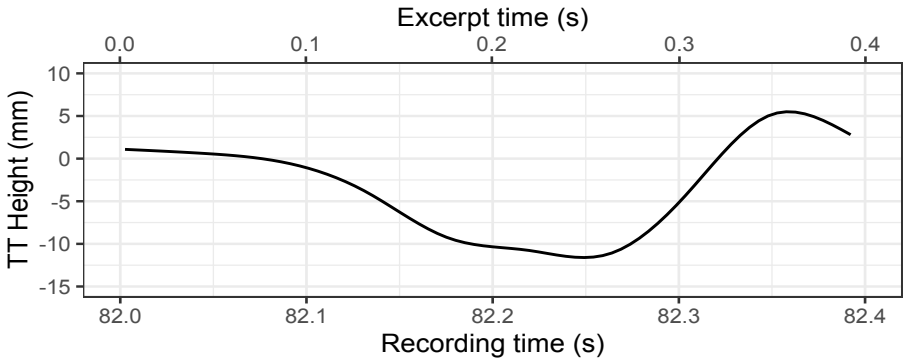

Figure 2.2: TT height trajectory for Speaker 3 producing the sequence *'striped cat'* in a map task.

Table 2.2 shows the rates of coronal stops in each grammatical class that were audible or inaudible, and within the inaudible class how many did or did not exhibit TT raising (raising$\leq$ 0).

Out of 87 tokens that were inaudible, just 15 (17%) appear to be true 'articulatory

	Audible	**Inaudible**		**Total**
		+ *Raising*	− *Raising*	
Mono	123 (79%)	29 (19%)	4 (3%)	156
Complex	139 (72%)	42 (22%)	11 (6%)	192
Semi	13 (93%)	1 (7%)	0 (0%)	14
Total	275 (76%)	72 (20%)	15 (4%)	362

Table 2.2: Articulatory categorisation of perceived coronal stop retention and deletion.

zeroes' with no evidence of TT raising to create a coronal closure. If similar articulatory profiles belie traditional acoustic research on CSD, it may mean that the rates of deletion taken as the output of phonological processes are vastly overestimated. Generalising these results by morphological class, we might predict that only 12% of monomorphemes and 26% of *-ed* suffixed forms where deletion has been observed acoustic-impressionistically can actually be described as such.

There is no statistically significant asymmetry in the distribution of articulatory zeroes amongst morphological classes. We might expect one if we were to attribute the absence of audible coronal stops to different processes with specific inputs. Such a pattern may become evident given a larger sample size, but a sufficient sample will be a challenge to obtain given the nature of articulatory data and the apparent rarity of articulatory zeroes as a phenomenon. It should also be noted that, unlike the analysis that included all non-neutralising phonetic environments in §2.3.1, the analysis that is limited to environments with no adjacent coronal segment does not show the normal morphological conditioning on rate of inaudible stops. That is, there is not a greater proportion of inaudible stops in monomorphemes compared to complex words. This, too, may be a function of the relatively small sample size. It is not cause for too much concern, given that the expected pattern of apparent CSD does obtain in the auditory analysis, and in the articulatory analysis there is evidence of some deletion. Therefore, there is no reason to believe that the use of EMA has precluded the implementation of normal CSD.

2.3.3 Idiosyncratic covert allophony

While only a small portion of inaudible stops exhibited no TT raising whatsoever, there are other relevant facts about the articulatory detail of coronal stop implementation. Specifically, we can conceive of a CSD process whose output is not an articulatory zero, but a discrete category of undershot [t]. Therefore, it is prudent to explore the distributions of individual speakers' coronal stop TT heights for evidence of multiple categories. For some speakers, there is a clear unimodal distribution of coronal stop TT heights, with no evidence of discrete categories. Distributions for TT height at the coronal stop target (T) are displayed for these unimodal speakers in Figure 2.3, with separate polygons for audible and inaudible tokens for the reader's convenience. Here, 0 is the height of the Upper Incisor reference sensor.

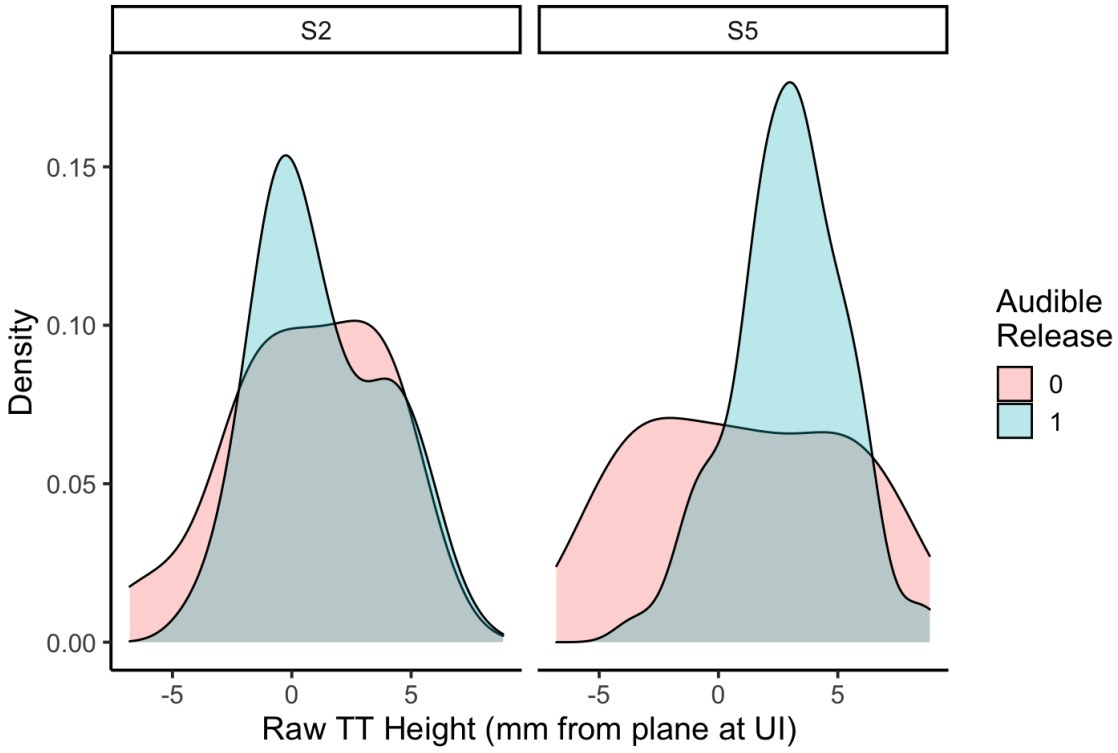

Figure 2.3: Raw TT heights at T for unimodal speakers 2 and 5.

While some speakers display unimodal distributions of TT height, with no evidence for

discrete categories of target, other speakers show a different pattern. Figure 2.4 shows distributions of for TT height for the three remaining speakers, whose overall profiles (audible and inaudible combined) are much more bimodal.

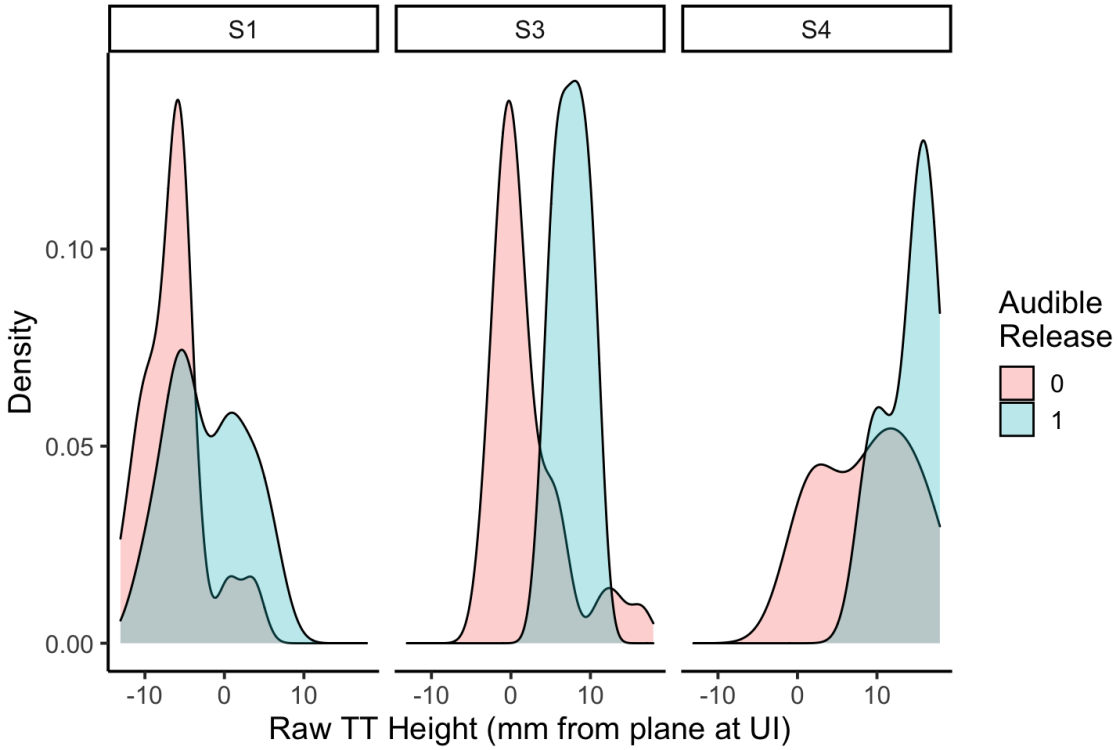

Figure 2.4: Raw TT heights at T for bimodal speakers 1, 3 and 4.

In both Figure 2.3 and Figure 2.4, each speaker exhibits a fairly wide range of TT height measurements for audible stops. This might be unexpected given that a coronal stop is canonically made through TT contact at alveolar ridge, which is thought to act as a biomechanical constraint that limits movement with a, quite literal, ceiling effect. Some of the variability in this group of tokens may be due to unavoidable noise in the signal that the Electromagnetic Articulograph records. However, it is also true that the surface on which a coronal closure can potentially be made spans a much larger area than just the alveolar ridge. Indeed, contact with a large portion of the hard palate, the alveolar ridge, or even the upper teeth will produce something that is recognisably a coronal stop. Thus,

a speaker could feasibly create a dental closure at the bottom of their incisors that would have a correspondingly low TT height but nonetheless be perfectly audible. Moreover, there is a further possibility that speakers could make laminal closures while producing relatively little TT raising. For the present study, none of these potential issues appear to be fatal. There was a characteristic TT height trajectory in the direction of a closure in almost every case, and no speaker had a majority of audible tokens at the low end of their TT height continuum or a majority of inaudible tokens at the high end.

All of speakers 1, 3 and 4 produce a somewhat bimodal distribution of TT heights in their implementation of underlying word-final coronal stops. This suggests that these individuals may have multiple tongue tip targets for coronal stops in this position. There is some variation still within these three speakers and how audible and inaudible tokens are distributed between each of their two peaks. Speaker 3 exhibits the cleanest divide such that almost all of their lower peak is comprised of inaudible tokens, and almost all of their higher peak is comprised of audible tokens. This is the basic pattern that we might expect under the assumption that the required tongue tip movement for a coronal stop is raising along an inferior-superior axis and that insufficient raising will not result in a closure (and therefore be inaudible). The picture for speakers 1 and 4 is a little more complicated. Speaker 1 also has two clear peaks but the lower category is populated with several audible tokens as well as most of the inaudible ones. This suggests that speaker 1 variably produces coronal stops with a low tongue tip strategy (e.g. laminal or dental), and this is non-deterministically correlated with acoustic and auditory categories. As such, perhaps traditional CSD analyses have been indirectly approximating categories that correspond to these kinds of strategies for some speakers. Similarly, all three speakers, but especially speaker 4, produce several inaudible tokens with very high TT heights. This is consistent with Browman and Goldstein's (1990) analysis that stops can be rendered inaudible by the temporal overlap of closures for adjacent segments and nonetheless be fully articulated.

2.4 Systematic lenition

The previous section served to evaluate the potential for a CSD process as traditionally conceived in light of potential articulatory evidence for and against a widespread process of that kind. We can also evaluate the data in terms of a gradient measure of proportional tongue tip raising and examine systematicity within this dimension. Table 2.3 shows the fixed effects for a mixed effects linear regression model predicting magnitude of TT raising[2], with a random slope for log-transformed gesture duration by speaker and a random intercept for word.

	Estimate	Std. Error	DF	t value	Pr(> |t|)	
(Intercept)	0.8653	0.1087	6.01	7.958	2.31e-04	***
log.freq	0.0040	0.0052	12.92	0.768	0.456	
t.d - [t]	0.0178	0.0405	43.53	0.291	0.773	
log.dur	0.3091	0.0766	4.45	4.035	0.013	*
task - Script	-0.0669	0.0366	75.37	-1.829	0.071	
task - SemDiff	-0.0857	0.0450	49.59	-1.903	0.337	
task - Wordlist	-0.3021	0.0497	44.85	-6.078	2.42e-07	***
preplace - Dors	0.0954	0.0291	27.49	3.277	0.003	**
postplace - Dors	-0.0148	0.0531	28.06	-0.278	0.783	
postplace - Open	0.0783	0.0499	151.7	1.569	0.119	
preman - Son	0.0008	0.0476	27.35	0.017	0.987	
postman - Vwl	-0.0054	0.0450	242.6	-0.120	0.905	
postman - Paus	0.0586	0.0473	271.5	-1.237	0.217	
gram - Mono	-0.0680	0.0325	23.78	-2.091	0.047	*
gram - Semi	0.0181	0.0648	91.77	0.273	0.786	

Table 2.3: Fixed effects of mixed effects linear regression predicting magnitude of TT raising.

Table 2.3 shows a number of significant effects on magnitude of TT raising. A particularly large effect shows that the log-transformed duration of the raising gesture interval between points A and B (*log.dur*) affects TT raising such that a longer interval results in higher raising towards a speaker-specific maximum height. The significance of this effect is greatly diminished by the inclusion of a random slope of gesture duration by Speaker, allowing for variation in terms of speakers' physiological capacity to produce articulatory

[2]The same effects were found from a subsequent regression model fit to by-speaker z-scored TT height values.

movements in different amounts of time. There are also effects of task, phonetic environment, and morphological class. Speakers produced significantly less TT raising in the Wordlist reading compared to the Map Task, more TT raising following a dorsal segment compared to a labial segment, and less TT raising in monomorphemes compared to complex words. Some factors that, unexpectedly, did not yield significant results are lexical frequency (*log.freq*) and whether the token in question is canonically realised as [t] or [d] (*t.d*).

2.4.1 Gesture duration

The largest effect observed on the magnitude of TT raising is that of the articulatory gesture duration. For this dissertation, I definte this as the time in seconds that elapses between point A (immediately preceding T) and point B (immediately following T). The roles of points A and B are illustrated in Figure 2.1. The simple correlation between gesture duration, log-transformed, and TT raising in demonstrated in Figure 2.5, with hollow points for inaudible tokens. The dotted line that intercepts the y-axis at 0 indicates the threshold below which tokens were considered articulatory zeroes.

When not controlling for random slopes by speaker, this effect of gesture duration is extremely strong, and it remains whether or not the measure is log-transformed. The data are also nicely partitioned such that there is a threshold below which no token is audible and above which inaudible tokens are in the minority. This effect is not unexpected. Indeed, it is very reminiscent of classic undershoot effects observed by Lindblom (1963) for variation in vowel quality, whereby shorter vowels were more centralised and less peripheral. It should be noticed that there is a less clear correlation for the absolute values of TT raising. That is, tokens in which the TT was substantially lowered from the baseline at line AB have some of the shortest gestures. We could interpret the effect of gesture duration as a phenomenon in the domain of phonetics whereby speech rate conditions TT raising and when speakers allot less time to articulate a coronal stop, they achieve less raising relative to their maximum TT height. However, this does not necessarily constitute evidence against

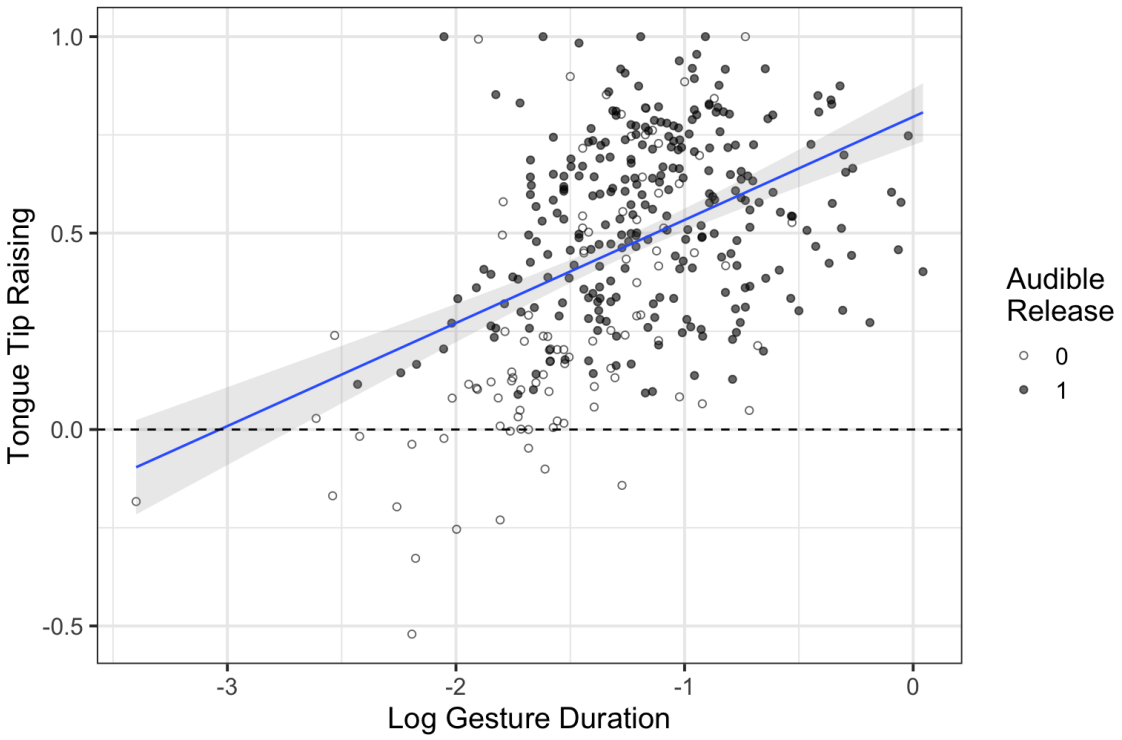

Figure 2.5: Magnitude of TT raising by log-transformed gesture duration.

a phonological deletion process. There is no reason that speakers planning to execute a different articulatory target that requires less TT raising should not allot this target a shorter gesture duration. When we observe the patterns of TT raising across gesture durations for each speaker separately, as in Figure 2.6, there are some interesting findings in terms of this issue of the phonetics-phonology interface and how categoricity can be a diagnostic tool.

The results from speakers 1 and 3 reinforce the idea that some speakers may have something resembling a categorical CSD process, with outputs that lie primarily in the articulatory domain, but that are indirectly and non-deterministically perceived in acoustic analyses. Speaker 4, who exhibited the third bimodal raw TT height distribution, presents no discernible pattern in terms of TT raising. However, this speaker also produced the least data. The clusters of inaudible tokens with low TT raising (or even lowering) in speaker 5 and especially speaker 2 suggest that there may also be idiosyncracies in representation at

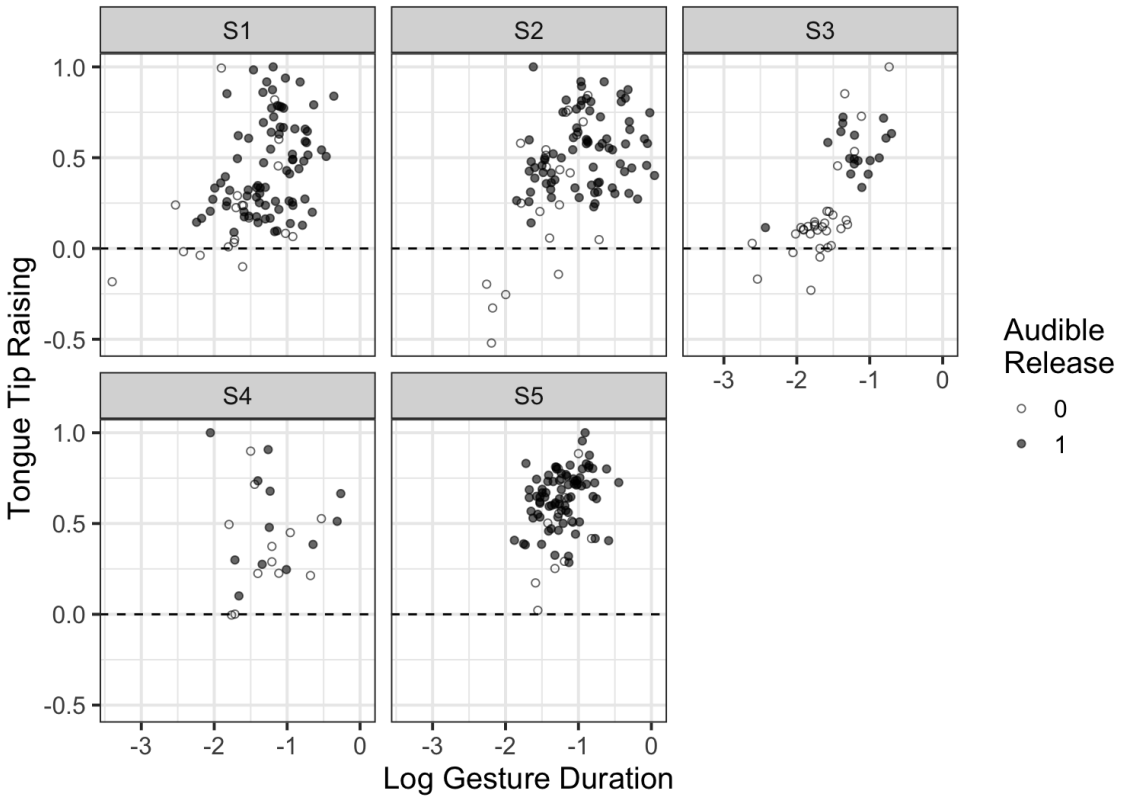

Figure 2.6: Magnitude of TT raising by log-transformed gesture duration for each speaker.

play. In other words, it could be that some speakers store articulatory *targets* that correspond to segmental information and that they attempt to achieve in real-time, while others may store *vectors* along which articulators are to be moved. Under this analysis, several of speaker 2's tokens could be considered to have undergone some process of categorical CSD in that they feature TT lowering rather than raising, even though they are still quite high in terms of raw TT height.

2.4.2 Task

One of the more unexpected results from the regression model summarised in table 2.3 is that of Task. When considering how speakers may have behaved differently in terms of magnitude of TT raising across the various tasks that were designed to elicit speech, we find that speakers produced the least TT raising in the Wordlist task, and that this was

significantly different from the Map task, which contributes to the intercept of the regression model. The distributions of TT raising across each task, collapsed across speakers, is presented in a boxplot in Figure 2.7.

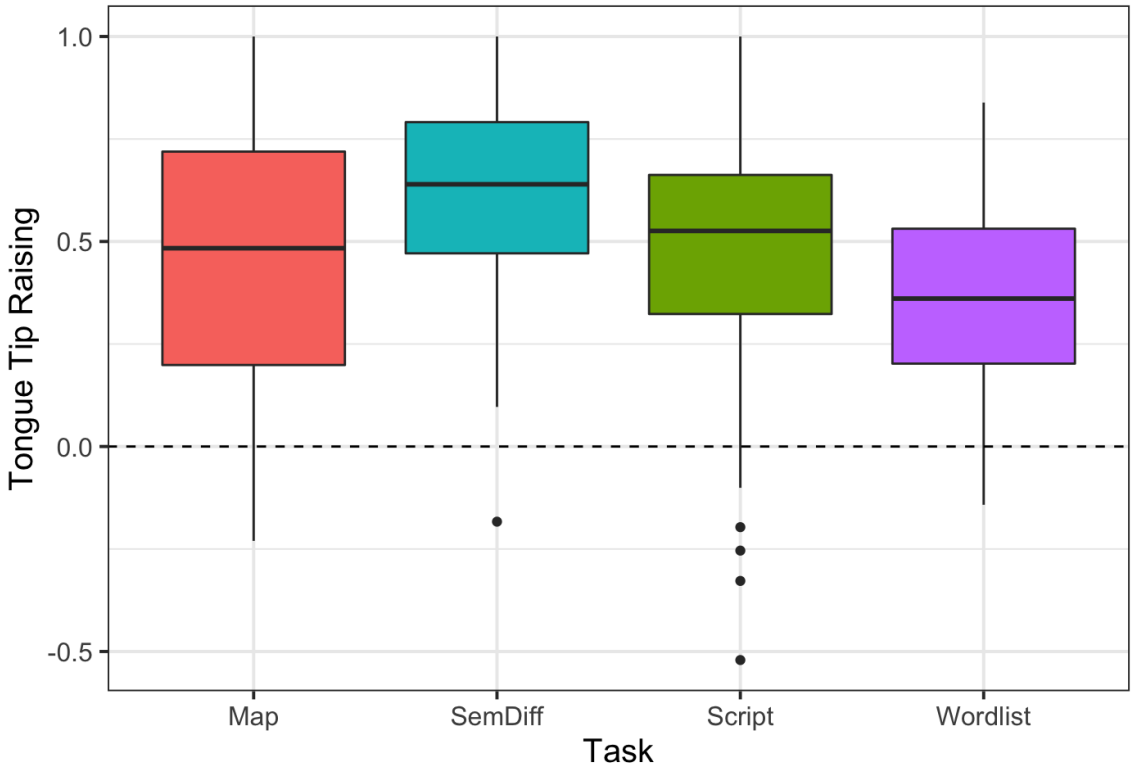

Figure 2.7: Distributions for magnitude of TT raising in tokens from each task.

This effect is somewhat surprising. Under the classic view in which different tasks should prompt different degrees of self-monitoring Labov (1972b), the word-list is expected to inspire the greatest amount of metalinguistic awareness. As such, we might expect speakers to most noticeably eschew features of casual speech and favour careful and precise speech in this context. CSD, and especially the gradient measure of TT raising used in this chapter, are excellent examples of a lenition-type phenomenon where lenition can be very narrowly construed in terms of the magnitude of articulatory movement towards a canonical target. Further, Eckert (2008) and Podesva et al. (2006) attribute some prestige and formality to fully articulated and audible coronal stops such that they index social

meanings of a high level of competence and education. Therefore, TT raising is a prime candidate as a variable where we would expect style shifting and the highest degree of TT raising in the Wordlist context.

A potential explanation for this effect is somewhat 'functional' in nature (Kiparsky, 1972). That is that out of all the tasks there is the least pressure to communicate the stimuli in the Wordlist context clearly to an interlocutor. In all of the other tasks, the participant read or spontaneously produced words in sentences with a researcher listening in. The Wordlist task is the only one in which participants produced words in isolation, with no particular meaning to convey. However, if we are to appeal to such an 'information theoretic' analysis, it remains to be explained why the Map task does not also significantly outperform the Semantic Differential and Script reading tasks in terms of TT raising. Conversely to the situation for the Wordlist, the Map task is the only context in which the speaker is explicitly giving instructions for the researcher to follow. Therefore, we might expect the greatest amount of pressure to communicate clearly in this task, which is not observed. Figure 2.8 shows magnitude of TT raising by gesture duration, with points categorised by task.

Figure 2.8 shows that tokens in the Wordlist task were produced with the longest interval and the lowest TT raising. The overall trend, as demonstrated Figure 2.5, is for tokens with longer intervals to have more TT raising. This direction is maintained between the ellipses for the Map task, the Script Reading task and the Semantic Differential task, but not the Wordlist. Some potential explanations that take this shape into account are a prosodic explanation and a fatigue explanation. The prosodic explanation is that in the Wordlist participants were required to produce the same word in isolation three times. This means that these words could be considered to all have a strong following phrasal boundary. In addition, each group of three words tended to form an intonational contour across which we might expect a gradual weakening effect. This accounts for the fact that the wordlist is longer and lower because speakers will tend to slow down across an utterance. An even more basic explanation is that participants may experience increasing fatigue across the duration

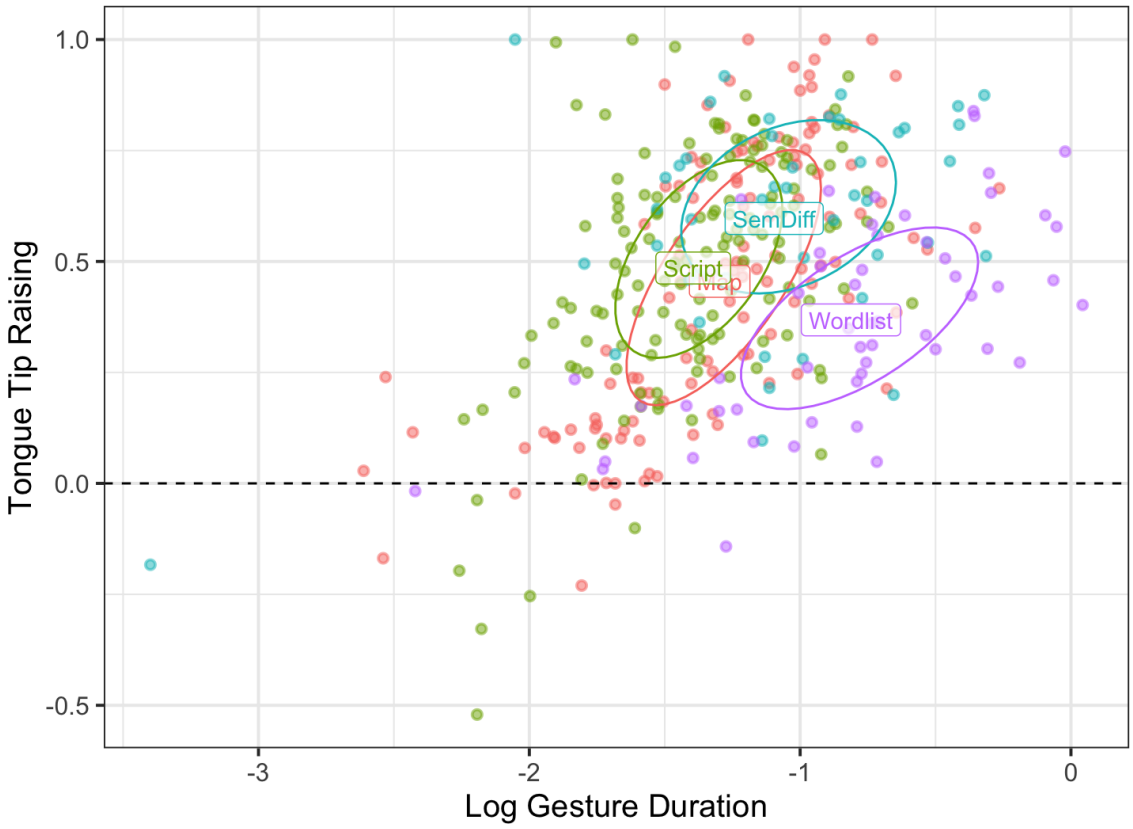

Figure 2.8: TT raising by gesture duration and task, ellipses for 40% confidence interval.

of the experiment. Since the Wordlist task was always conducted last, participants are likely to be at their most fatigued at this point. Therefore, it could be that speakers simply expend less effort to produce coronal stops as they become tired.

2.4.3 Phonetic Environment

It is certainly worth exploring whether different articulatory strategies for coronal stops correspond to different phonetic environments. In particular, the regression model in Table 2.3 demonstrates that speakers produced more TT raising for coronal stops following a dorsal segment than coronal stops following a labial segment. Figure 2.9 shows each speaker's TT raising by Log Gesture Duration again, with each token coloured for the place of articulation of the segment immediately preceding the relevant coronal stop.

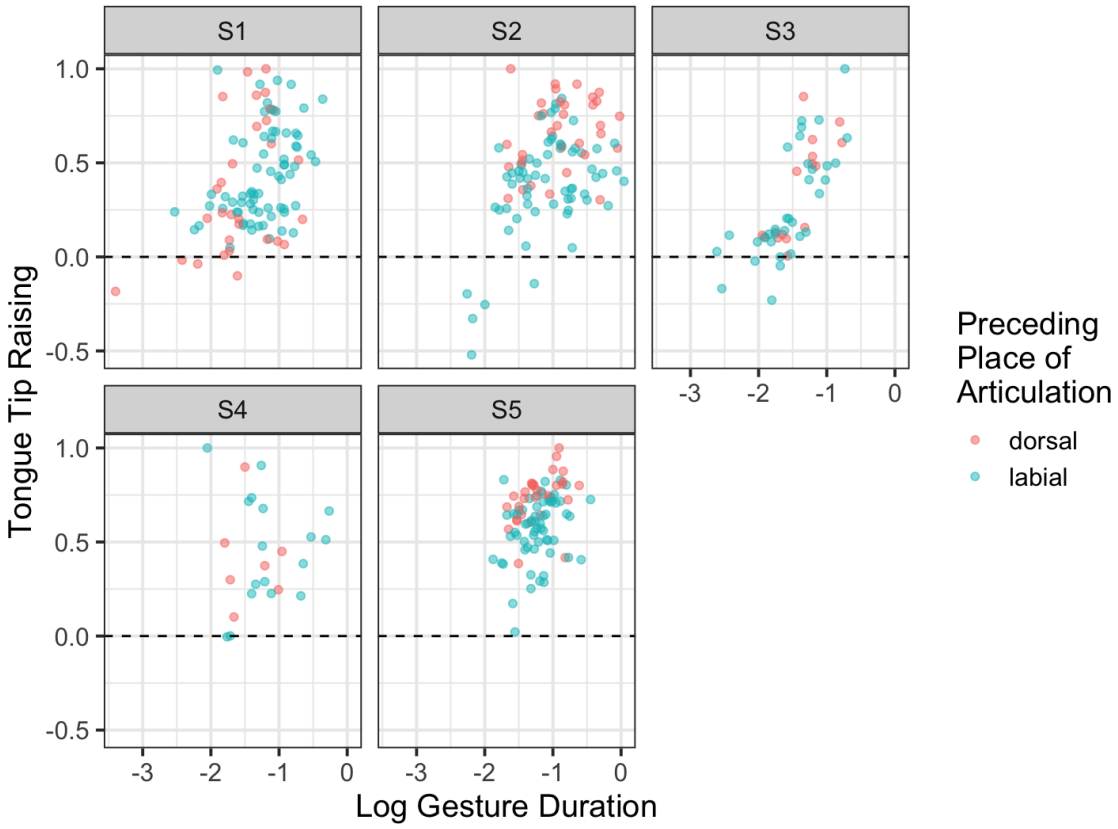

Figure 2.9: TT raising by gesture duration for each speaker. Point colours show preceding place of articulation.

The effect of preceding place of articulation appears to be largely driven by the behaviour of speakers 2 and 5. Both of these speakers distinctly have two overlapping groups for tokens such that the tokens following labial segments have noticeably less TT raising. In addition, the articulatory zeroes produced by speakers 2 and 3 (which feature TT lowering) all follow labial segments. However, this pattern is not shared by all speakers. Speaker 1's articulatory zeroes all follow dorsal segments. Whichever way around, the labial/dorsal effect is not generally documented in the CSD literature. One effect that is commonly described for CSD, but does not have an analogue in the regression model on TT raising is the following segment effect. Specifically, CSD occurs more frequently with a following consonant than a following vowel. In the CSD literature, this is commonly attributed to the

potential resyllabification of the coronal stop as an onset to a following vowel, which bleeds CSD (Guy, 1991a; Kiparsky, 1993; Reynolds, 1994). While there is no significant effect in the regression model, Figure 2.10 shows that for Speaker 3 almost every token in their lower cluster is followed by a consonant, and the majority of tokens in the higher cluster are followed by vowels or pauses.

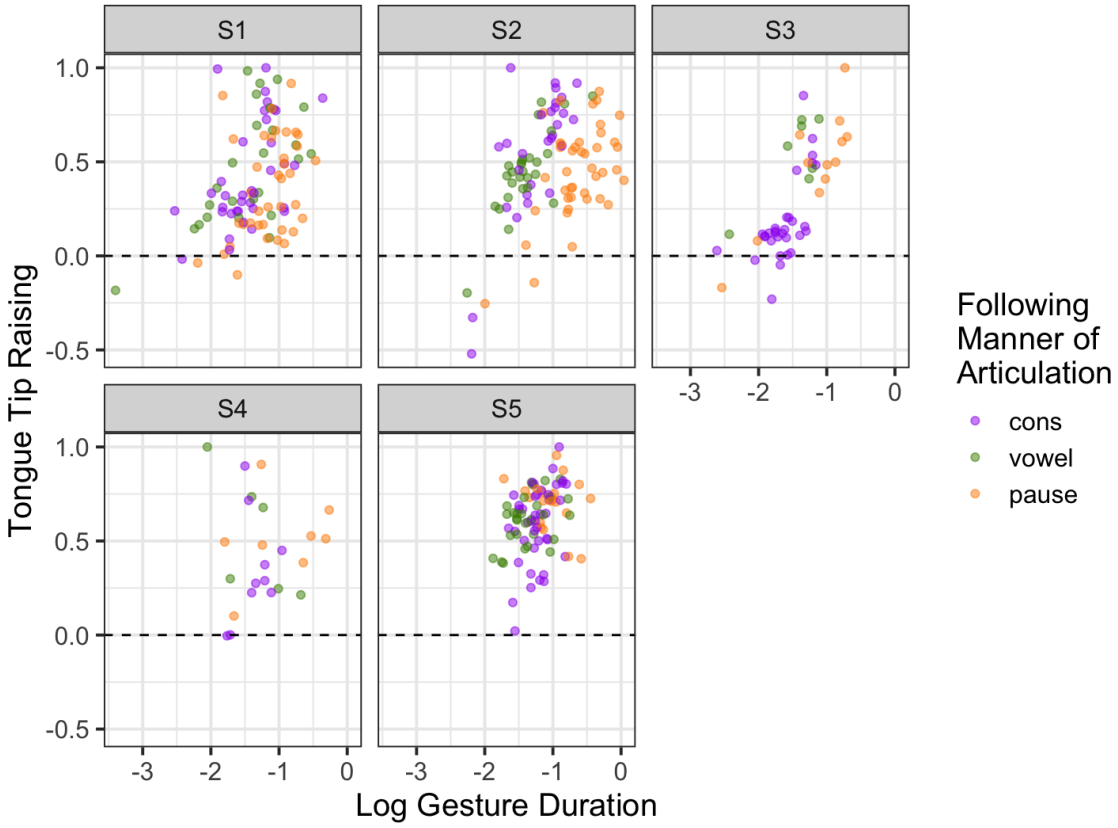

Figure 2.10: TT raising by gesture duration for each speaker. Point colours show following manner of articulation.

2.4.4 Morphological class

One of the most interesting results from the linear regression on magnitude of TT raising concerns the morphological class of tokens. We observe that speakers produce significantly less TT raising for coronal stops at the end of monomorphemic words as compared to coronal stops that constitute an *-ed* suffix at the end of passive or preterite forms. The distributions

for TT raising magnitudes in each morphological class are displayed in a boxplot in Figure 2.11, with significance levels from a basic two sample t-test.

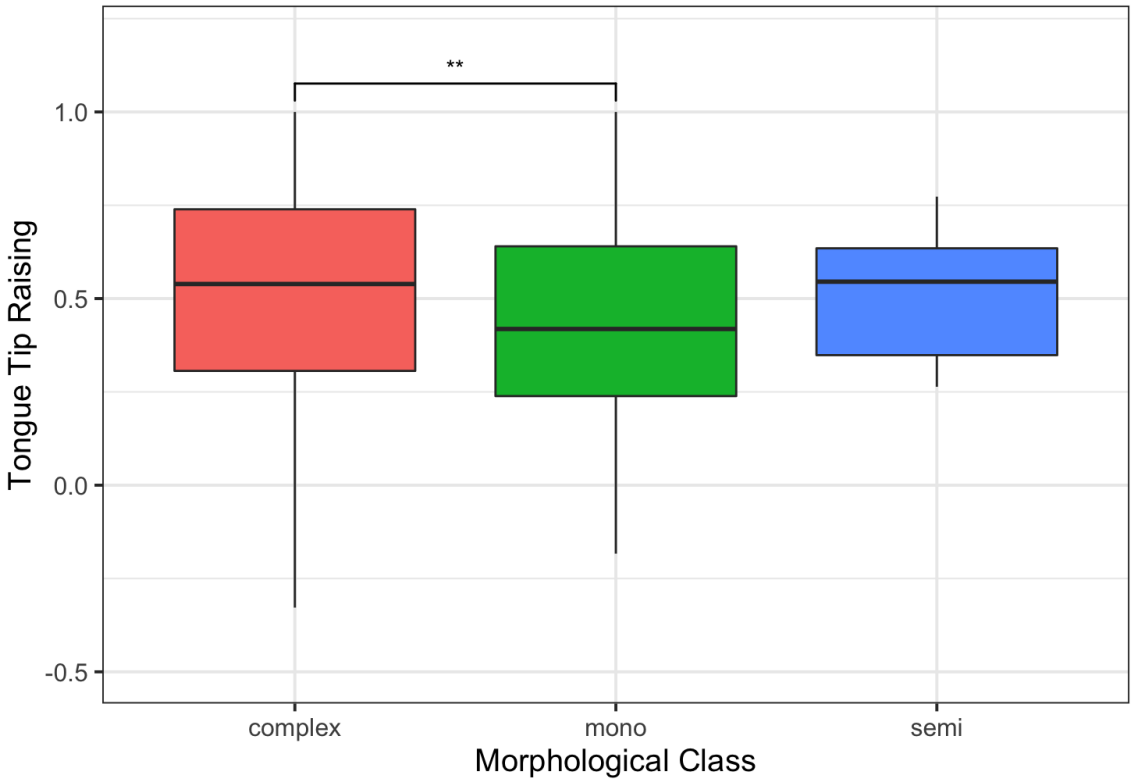

Figure 2.11: Distributions for magnitude of TT raising in tokens from each morphological class.

The results in Figure 2.11 are especially interesting because they are in the same direction of the robustly attested effect of morphological class on CSD. That is, we expect a higher rate of CSD in monomorphemes than complex words, and less TT raising corresponds to less articulatory movement towards a canonical coronal stop closure. This corroborates a similar finding from Purse and Turk (2016), who observe less TT raising for coronal stops in monomorphemes than complex words for a subset of their data: specifically, the speakers of a Southern Standard British English dialect. However, these results pose a problem for the idea that speech production should be strictly modular and that morphology and phonetics should not share an interface. TT raising is a gradient phonetic measure that we do not

expect to be conditioned by morphological variables without the mediation of categorical phonology.

Similarly to the results for phonetic environment, we also see robust individual differences in the implementation of conditioning according to morphological class. This time, Figure 2.12 shows that the majority of Speaker 1's lower cluster of tokens in terms of TT raising is comprised of monomorphemes, whilst the majority of the higher cluster is comprised of complex forms. This pattern is also in the expected direction according to the morphological conditioning on rates of CSD. This is because less TT raising is non-deterministically correlated with inaudible stops, since presumably some of the time this results in the speaker undershooting their target of a coronal closure at the alveolar ridge. Conceptually, this picture consistent with the idea that less TT raising towards a coronal closure is a less successful execution of a canonical coronal stop.

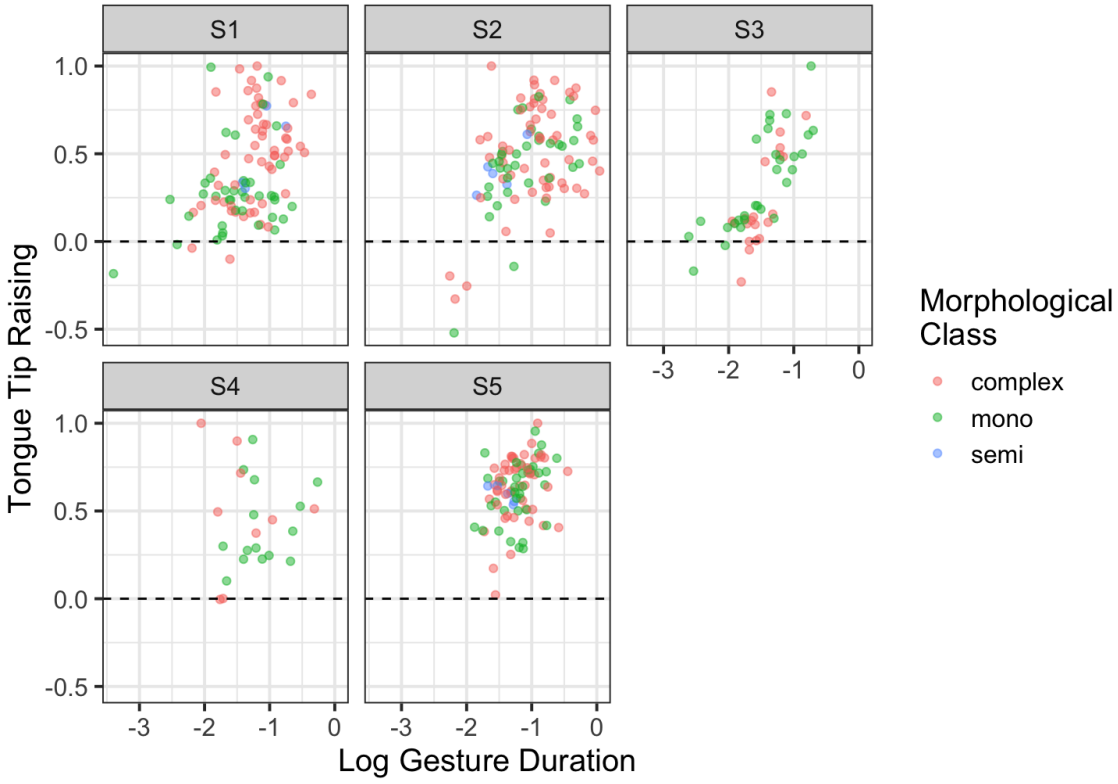

Figure 2.12: TT raising by gesture duration for each speaker. Point colours show morphological class.

2.5 Accounting for categoricity and gradience

Despite a wealth of literature on CSD in several varieties, few studies have investigated the question of categoricity versus gradience with regard to the phenomenon. The assumption suggested by the phenomenon's name and its traditional, phonological, formulation is that CSD creates a discrete alternation between a regular coronal stop and the complete absence of a stop. The data in this chapter reveal that there do indeed appear to be some tokens exhibiting no evidence of an articulatory movement towards a coronal closure, which Browman and Goldstein's (1990) investigation did not find. Almost every individual speaker produced at least one token of this type. However, it is clear that this type of 'articulatory zero' is rare, accounting for just 17% of tokens that were inaudible and would be considered to have undergone CSD in a traditional acoustic analysis of the phenomenon. If researchers are committed to the idea that CSD should result in the complete absence of any seeming attempt to produce a coronal stop, these data suggest that previous work on CSD vastly overestimate the rate at which this kind of true deletion occurs, at least in Mainstream American English. Moreover, we cannot be certain that these articulatory zeroes constitute a separate category of implementation rather than just one extreme end of a continuum of lenition. The latter perspective, that we may be able to do without allophonic deletion, aligns with some more radical suggestions that allophonic variation is altogether redundant (Liberman, 2018).

Some evidence in favour of the opposite analysis, that articulatory zeroes are indeed the outcome of an attempt to reach an entirely different target, could come from the fact that several of these tokens feature noticeable tongue tip movement downwards and away from a coronal closure. For a target of a coronal closure at the alveolar ridge, a continuum of success in executing an articulatory plan to reach this target should span from a complete raising movement that reaches the target to the absence of raising altogether. It is not obvious that such a continuum should extend to include movement in the opposite direction, away from the target, too. For this to be the case, the most lenited end of the continuum would have to

behave as though the stop, and corresponding target, is truly absent. As such, the speaker could be licensed to produce movement in any direction as best serves the gestural demands of the surrounding targets that remain. Such an explanation could feasibly account for the data under the perspective that apparent CSD is simply the result of gradient phonetic phenomena (e.g. Temple, 2014). One prediction from this explanation is that there should be a margin of error around the baseline for articulatory movement in *both* directions. In the present study, only tokens where there was $\leq$0 TT raising were considered true articulatory zeroes. However, if a zero means that the speaker is unconstrained by the coronal stop target, it is conceivable that movement towards this target will still be gesturally optimal for some tokens. Therefore, the cluster of tokens around the baseline that is produced by Speaker 3 (e.g., in Figure 2.6) may all constitute this type of zero. This conception of 'zero' is tenable as a separate category of implementation or as one extreme end of a continuum.

Beyond this consideration, there remain some factors in the data that are not easily explained by either picture of apparent CSD considered so far. That is, neither an alternation between full coronal stop and the absence of articulatory movement towards such a target, nor a continuum of coronal stop phonetic implementation are fully explanatory for this data. For one thing, some speakers exhibit multimodality in their articulatory profiles that do not obviously match what we expect from an alternation between zero and a full coronal stop. The best example of this is Speaker 1, whose pattern of TT raising in Figure 2.6 contains two clear clusters. However, the cluster in which there is less TT raising is mostly comprised of tokens that are raised from the baseline, and actually contains a considerable number of tokens in which a coronal stop was audible and therefore would not be considered deleted in a traditional analysis. The non-deterministic correlation between TT raising and audibility of coronal stops holds to a lesser extent for all speakers, since a token with very high TT raising can still be masked by surrounding segments, and an audible stop could be produced with very little raising if the speaker makes contact with the teeth or produces a hyper-laminal stop. Further, for Speaker 1 the likelihood that a given token will be produced as part of the higher or lower cluster with regard to TT raising appears

to be strongly conditioned by morphological class. As such, Speaker 1 actually produces something resembling the robustly attested effect of morphological class on CSD, where coronal stops for monomorphemes are the majority of members of the cluster with less TT raising, which are more likely to be inaudible, while the majority of tokens in the cluster with higher TT raising are coronal stops in complex forms. However, this morphological effect would not be nearly as evident in an acoustic analysis of this speaker's data. Perhaps the patterns of CSD that have been observed in the acoustic signal since Labov et al.'s (1968) first description have, at least for some speakers, only indirectly accessed a pattern that exists in the articulatory domain. These observations, clustered in terms of the articulatory detail of their implementation, appear to constitute the kind of allophony that we expect to be a result of a categorical CSD process despite the fact that they are not consistently centred around zero or comprised of only inaudible stops. Therefore, we should entertain the possibility that CSD could be acquired as an alternation between a regular coronal stop and a systematically undershot coronal stop, with a significantly lower target.

As well as interesting patterns of multimodality, these data provide evidence for a number of systematic effects that condition the magnitude of TT raising across speakers. Not all of these are easily captured, even under the view that apparent CSD is the result of a phonetic continuum of implementation. One such effect is that a preceding dorsal segment leads to significantly higher TT raising than a preceding labial. This is not particularly surprising on its own, however, the effect is only really observable in the data from Speakers 2 and 5 (Figure 2.9). It is not clear why this effect would be isolated to particular speakers in this way. One potential explanation is that these individuals particularly favour a laminal articulation of coronal stops. The blade of the tongue cannot be isolated from the dorsum as easily as the tip. Therefore, if a speaker produces a dorsal constriction, it may provide some particular facilitation a laminal coronal stop that leads to more maximal tongue tip raising when following a dorsal.

Another effect on the magnitude of TT raising is found with regard to the morphological class of words in which coronal stops are found. Like the difference between Speaker

1's TT raising clusters, the mixed effects regression model presented in table 2.3 found a significant effect of morphological class such that speakers produce less TT raising for monomorphemes than complex forms. This resembles the well-attested effect of morphological class on CSD—higher rates of CSD in monomorphemes than complex forms—in that less TT raising can be interpreted as a less complete articulation of a coronal stop. However, a strictly modular view of speech production stipulates that an effect of morphology should have its reflex in the categorical phonology, not the gradient phonetics. A possible explanation for this effect could be found in the usage-based phonology literature (e.g Pierrehumbert, 2002; Bybee, 2002). Under this kind of approach, the effects of morphology are emergent from effects of particular words, each of which have their own representations in phonetic space. Thus, an apparent effect of morphological class on a phonetic dimension like TT raising could be a function of an effect of particular words that happen to be distributed among morphological classes in a certain way. However, this explanation finds opposition in the fact that lexical frequency, which Bybee (2002) claims should be a more direct measure of propensity to undergo a lenition phenomenon like CSD, does not appear to significantly affect TT raising in this data.

An alternative approach to the morphological effect in tongue tip raising might be found if we connect it to work on phonetic lengthening at morphological boundaries, described in Chapter 1. That is, a number of studies report that suffixes (Walsh and Parker, 1983; Lociewicz, 1992; Seyfarth et al., 2018), stem-final consonants (Schwarzlose and Bradlow, 2001), and stem-final rhymes (Sugahara and Turk, 2009) are produced with slightly longer durations in morphologically complex words than in monomorphemic counterparts. A leading explanation for these findings is that any roots and Level I suffixes to which a Level II suffix can be attached form a prosodic constituent preceding said Level II suffix. As such, prosodic lengthening processes target material around the domain-final boundary that accompanies Level II suffixes. If we are to accept this analysis, regular *-ed* suffixes that indicate passive or preterite forms should also be the site of prosodic lengthening processes. As such, perhaps the morphological effect whereby coronal stops in *-ed* suffixed

are executed with tongue tip raising of greater magnitude than coronal stops at the end of monomorphemes is actually a product of an increased duration, on average, of phonological material around a prosodic boundary that accompanies an *-ed* suffix. In support of this, we observe a strong correlation between gesture duration and TT raising in Figure 2.5, such that a greater gesture duration allows for higher TT raising. This means that longer times for executing articulatory movements allows for movements of greater magnitudes to be achieved. This would be consistent with Parrell and Narayanan's (2018) account of coronal reduction as a result of an invariant articulatory target executed under different prosodic conditions. Incidentally and in addition to a straightforward link between gesture duration and gesture magnitude, longer intervals to articulate sequences should also allow for their execution with reduced overlap, minimising the rate at which coronal stops are rendered inaudible through concurrent noncontinuant constrictions (Browman and Goldstein, 1992).

If we can place the locus of the morphological effect on the magnitude of coronal stop tongue tip raising in the prosody, we need no longer think of it as exceptional in its apparent violation of modularity. But while prosodic lengthening is a neat hypothetical explanation for morphological differences in the magnitude of tongue tip raising, it must not be overlooked that such an effect is not directly observed in this chapter. That is, the coronal stop gestures measured here do not appear to differ in their duration according to the morphological class of the larger word. This is in line with previous null findings for durational differences in *-ed* suffixes versus coronal stops at the ends of monomorphemes (Mousikou et al., 2015; Seyfarth et al., 2018), contrasted with more robust findings of durational differences in *-s* suffixes. This may be due to a certain relative inelasticity in the timing constraints on stop consonants. In order to further probe an association between longer durations and the morphological boundary preceding *-ed* suffixes, Chapter 4 reports the results of an relevant experiment in the perceptual domain. In it, listeners demonstrate that a wordform with a long duration is indeed more likely to be pick out an *-ed* suffixed form (over a monomorphemic homophone) than a wordform with a short duration.

2.6 The distribution of categorical deletion

In this chapter, I have identified three potentially distinct types of apparent CSD. There are instances of inaudible stops that feature high TT raising and presumably have their acoustic masked by perseveratory or anticipatory closures. There are tokens that comprise a cluster in which there is very little TT raising, and which may share a different target than the canonical target at the alveolar ridge. And there are tokens that look to be true articulatory zeroes in that there is no evidence of TT raising towards a closure at the alveolar ridge and, in some cases, TT movement away from such a closure. In the data at hand, there was no evidence that any particular type of token favoured any particular type of CSD across all speakers. However, we might predict a larger sample to reveal that different types of apparent CSD may be distributed unevenly between different morphological classes.

Guy's (1991b) account of the effect of morphological structure on CSD involves a variable CSD process at each of the many levels of a stratified morphophonology. Monomorphemes, which are fully formed from the beginning of this stage of derivation, are eligible to undergo each of these processes, significantly increasing the likelihood that one such process will result in the deletion of a relevant coronal stop. Complex forms, on the other hand, only receive the Level II *-ed* suffix at the final level of a stratified morphophonology. Therefore, coronal stops in complex words are only the potential target of one variable CSD process and are far less likely to be deleted. Bermúdez-Otero (2010) and Myers (1995) draw from this account, explaining that instances of apparent deletion in monomorphemes should be comprised of far more instances of categorical deletion than in complex forms, since the former type of word should be eligible for several times the number of potential applications of a phonological CSD process compared to the latter. Myers (1995) in particular invokes something like Zsiga's (1993) 'Lexical-Categorical Hypothesis', that the output of lexical phonology must be categorical and cannot be gradient. This implies a relaxed assumption of categoricity at the postlexical level, where a final phonological CSD process may target monomorphemes and complex forms alike. Such a position is informed by an assumption

that the phonology must operate on strings of discrete symbols, and so levels of phonology whose output becomes the input to a subsequent level of phonology cannot have a gradient output. The perspective that postlexical phonology may not have a categorical output makes an even stronger prediction with regard to the distribution of categorical CSD across morphological classes, because if the postlexical CSD process is gradient complex forms may not be subject to any categorical phonological CSD process at all. Thus, both Bermúdez-Otero (2010) and Myers (1995) expect more categorical CSD in monomorphemes than complex forms.

On the other hand, Tamminga's (2016) work makes a different prediction based on evidence from persistence, an effect of naturalistic priming whereby a speaker is more likely to apply a variable process that they have just applied immediately beforehand. She finds that apparent CSD in monomorphemes (e.g. *pact*) primes CSD in the exact same word (*pact*), but not in a different monomorpheme (e.g. *soft*) nor in a complex word (e.g. *packed*). However, apparent CSD in a complex word (e.g. *packed*) primes CSD in the same word (*packed*) and in other complex words (e.g. *cracked*). From this, Tamminga (2016) surmises that at least some apparent CSD in complex words should be attributed to zero-allomorphy. The selection of this zero-allomorph in the place of the canonical *-ed* suffix then makes its subsequent selection for another complex form more likely. Since a zero-allomorph will have entered the phonological derivation without phonological content, it should necessarily present itself as an instance of categorical deletion. Therefore, zero-allomorphy provides an avenue not available to monomorphemes through which complex forms could look at though they have undergone categorical CSD.

Interestingly, these predictions are not necessarily mutually exclusive. Both variable allomorphy and lexical phonology could be sources of what looks like categorical CSD. However, the output of these processes may not be identical. Under an assumption that phonology is not constrained to be 'structure-preserving', a coronal stop token with residual TT raising is compatible with having undergone a categorical process in the phonology. However, a zero-allomorph must not involve TT raising towards a canonical coronal stop

target. Saliently, both categorically distinct tokens that feature residual TT raising, and tokens that appear to have no trace of residual TT raising, are present in the data for this chapter. The endeavour to investigate the distribution of apparent CSD tokens from different morphological classes among these types of categorical deletion poses a particular challenge for data collection. Researchers must contend with the rarity of categorical CSD and what appears to be variation in speakers' representation and implementation of the process when it occurs.

2.7 Chapter summary

This chapter constitutes one of the first investigations into the articulatory reality of CSD, despite a wealth of studies that assume it to have various properties. As its central finding, there is widespread evidence that inaudible coronal stops are typically not implemented in terms of categorical 'true zero' deletion. This kind of categorical CSD is far rarer than previously thought, constituting 17% of inaudible coronal stops in the dataset. However, some speakers exhibit categoricity in their distribution of degree of TT raising beyond what is captured by a dichotomy between presence and absence of movement in the direction of a alveolar ridge target.

Beyond the issue of categoricity, individual differences still play a key role in that each speaker appears to exhibit strong conditioning on their articulation of coronal stops according to one factor traditionally associated with CSD. But no single speaker appears to be affected by all the relevant factors at once. These results give rise to difficult questions about how variable phenomena like CSD should be represented, and acquired, if any one kind of representation and acquisition can even account for the differences between different speakers.

Chapter 3

Perception of morphologically-sensitive articulatory variation

In the previous chapter I demonstrated that the vast majority of cases of apparent CSD feature residual tongue tip raising. However, several tokens have no detectable tongue tip raising and look like 'true zeroes', where any coronal gesture has been deleted. Moreover, there is systematic variation in terms of tongue tip raising for coronal gestures that is conditioned by phonological and morphological factors. These findings raise questions regarding the perception and, in turn, acquisition of CSD.

3.1 Perception of CSD

The classic conceptualisation of CSD as a discrete alternation corresponds to some general properties of the relationship between articulation and acoustics and the nature of stop consonants. Specifically, if we consider the continuum of tongue tip raising magnitudes that were under discussion in the previous chapter, it is only tokens at one extreme end—where the tongue tip actually makes contact with the teeth or hard palate and blocks oral airflow—that represent a canonical coronal stop. If the tongue tip raises to any point below that which is necessary to make this contact, the speaker will not achieve the characteristic block and release of oral airflow; a canonical stop consonant will not be produced. This means that a large difference in tongue tip height between two tokens that do not achieve a stop closure can have very little acoustic consequence, while a token that achieves said stop closure is qualitatively different from one that is just shy of it. The potential for temporal

overlap of articulatory gestures then introduces further complexity to this picture. Traditional analyses of CSD use an acoustic–impressionistic methodology that relies heavily on the researcher's perception, and judgments that stops are produced (and not deleted) correspond to the qualitative acoustic difference in tokens where all the conditions are right for a successful canonical stop, as opposed to when sufficient articulatory undershoot or overlap mean that coronal raising does not result in a canonical stop. Very rarely are CSD judgments made in terms of a finer-grained scale than 'retained' versus 'deleted', and very rarely are they elicited from a larger participant group than a single researcher. The goals of this chapter are threefold: (1) To corroborate my own judgments of the audibility of coronal stops, (2) to explore listener sensitivity to the articulatory detail observed in Chapter 2, and (3) to explore listener bias in terms of context.

3.2 Methods

3.2.1 Stimuli

In order to specifically evaluate the naturalistic articulatory variation observed in Chapter 2, the recordings from that chapter were adapted to form perceptual stimuli. Audio for 250 tokens of word-final, post-consonantal coronal stops were extracted from recordings of five speakers of Mainstream US English (S1–5 from Chapter 2) completing tasks in an EMA procedure, to create stimuli for a perception experiment. All stimuli were taken from the EMA tasks eliciting connected speech (Map Task, Semantic Differential, Reading Passage) and all were extracted in the context of a larger noun phrase or prepositional phrase (e.g. *'towards the **striped** cat'*). Stimuli were chosen from among tokens where this larger phrase was clearly intelligible, creating a stimuli list that was roughly balanced in terms morphological and phonological context of the underlying coronal stop, and in terms of how many excerpts I judged to have 'audible' coronal stops in each context. Stimuli extracted from each speaker represented a wide range of tongue tip heights at the apex of the coronal gesture, according to the distribution produced by that speaker.

3.2.2 Procedure

44 native listeners of English who grew up in North America and had no diagnosed hearing or reading problems were recruited from the psychology subject pool at the University of Pennsylvania. Before participating, listeners were informed that they were going to hear connected speech in which some /t/s and /d/s are pronounced more clearly than others. As part of this explanation, they were presented with four example stimuli: one in which a canonical /t/ was produced with a high tongue tip, one in which there was no tongue tip raising and no acoustic evidence of a release burst, and two 'ambiguous' cases with low tongue tip raising and quiet or absent release burst. These example stimuli did not appear later in the experiment.

During the experiment, listeners heard audio stimuli along with an orthographic representation of the phrase produced in each case. The word containing the word-final, post-consonantal coronal stop was in bold, and listeners were asked, "How clearly was the [t] or [d] at the end of the word in **bold** produced?" Stimuli were automatically played once, but listeners were able to replay the audio as many times as they wished, before responding. Listeners responded on a 6-point scale from 'very unclear' to 'very clear'. The experiment was conducted online using PCIbex.

Following all ratings, participants were asked to complete a basic demographic questionnaire that included questions to confirm that listeners met the pre-requisites for participation and were not distracted during the experiment.

3.2.3 Analysis

Listener ratings were scaled using a by-speaker z-score. Two listeners were excluded before statistical analysis because they rated every stimulus the same. The z-scored listener ratings were then analysed using mixed effects linear regression models. Demographic factors were not found to significantly affect listener ratings and were removed from the model. The model was first fit to the whole dataset, and then to subsets according to my own binary judgment of coronal stop audibility. This enabled me to test listener ratings within, for

example, the stimuli where I judged there to be an audible stop. Models were fit with fixed effects of tongue tip height (by-speaker z-score), gesture interval duration (log-transformed), trial number (centred), word frequency (log-transformed), target coronal stop ([t] or [d]), morphological class, preceding environment (sum-coded), and following environment (sum-coded). Random intercepts were included for stimuli, listener, and speaker.

3.3 Results

The first regression model was fit to the entire dataset of listener ratings. The predictor estimates, shown in Table 3.1, demonstrate that both articulatory measures do predict listener ratings to some extent. That is, higher tongue tips and longer coronal gestures are associated with 'clearer' word-final coronal stops. This is not surprising, as both parameters are associated with my own impression of where said coronal stops were audibly produced with a characteristic release burst. Listeners also rated stops with following vowels and pauses—contexts that classically favour coronal stop retention—as particularly clear compared to other following contexts. This suggests that these contexts don't just disprefer coronal stop deletion, but retained coronal stops are more salient when they occur in these contexts. A more surprising result can be seen in terms of the morphological class of words containing the coronal stops; listeners rated coronal stops at the end of semiweak past (e.g. *kept*) and especially monomorphemic (e.g. *soft*) forms as clearer than stops at the end of regular past (e.g. *claimed*) forms. Unlike the results for following vowels and pauses, the results for morphological class go in the opposite direction to classic conditioning of CSD. That is, semiweak past and monomorphemic forms are classically associated with higher rates of CSD than regular past forms, and in Chapter 2 we see that they are correspondingly produced with lower tongue tips across the board. A final significant effect is found in trial number, such that stimuli presented later in the experiment were judged to have clearer coronal stops than stimuli presented earlier in the experiment.

In order to better examine the role of the articulatory measures beyond their relationship to the basic impressionistic classification of the audibility of stops, the same model structure

	Estimate	Std. Error	DF	t value	Pr(> |t|)	
(Intercept)	-2.434e-01	3.028e-01	1.296e+02	-0.804	0.423025	
TT Height (z)	7.184e-02	3.065e-02	2.221e+02	2.344	0.019965	*
log.dur	7.703e-02	3.311e-02	2.229e+02	2.326	0.020893	*
trial #	2.085e-02	9.636e-03	7.758e+03	2.164	0.030503	*
log.freq	-2.006e-02	1.337e-02	2.228e+02	-1.501	0.134760	
t.d - [t]	6.956e-02	2.285e-01	2.221e+02	0.304	0.761117	
gram - Mono	3.349e-01	1.001e-01	2.217e+02	3.345	0.000967	***
gram - Semi	3.004e-01	1.515e-01	2.204e+02	1.983	0.048573	*
preceding - [p]	-1.367e-01	3.390e-01	2.224e+02	-0.403	0.687135	
preceding - [f]	-1.805e-01	3.423e-01	2.223e+02	-0.527	0.598280	
preceding - [v]	-3.499e-02	2.729e-01	2.222e+02	-0.128	0.898092	
preceding - [m]	-1.785e-01	2.765e-01	2.225e+02	-0.646	0.519205	
preceding - [k]	-1.664e-02	3.365e-01	2.222e+02	-0.049	0.960593	
preceding - [g]	1.418e-01	3.158e-01	2.219e+02	0.449	0.653866	
preceding - [ŋ]	-3.261e-01	3.058e-01	2.222e+02	-1.066	0.287403	
following - [p]	-1.457e-01	1.641e-01	2.233e+02	-0.887	0.375770	
following - [f]	8.082e-03	1.858e-01	2.236e+02	0.044	0.965336	
following - [m]	1.696e-01	1.968e-01	2.245e+02	0.913	0.362386	
following - [k]	9.762e-02	1.616e-01	2.229e+02	0.604	0.546388	
following - [h]	3.583e-01	1.828e-01	2.240e+02	1.960	0.051278	.
following - [w]	1.606e-01	1.959e-03	2.241e+02	0.820	0.413303	
following - pause	3.007e-01	1.490e-01	2.239e+02	2.019	0.044731	*
following - vowel	4.232e-01	1.306e-01	2.251e+02	3.239	0.001380	**

Table 3.1: Predictor estimates for mixed effects linear regression predicting listener ratings to all stimuli

was fit to subsets of the data corresponding to my own binary judgment of whether coronal stops were or were not deleted. The predictor estimates from the model fit only to 'audible' coronal stops are shown in 3.2. In this model, there are no significant effects for articulatory measures or for phonological context. However, the effect whereby monomorphemes are judged to be clearer than regular past forms remains, as does the effect of trial number.

The predictor estimates for the model fit only to 'inaudible' stops is shown in 3.3. Here, all significant effects vanish, including the effects of morphological class and trial number that remained significant in the other subset model. The fact that articulatory measures to not predict listener ratings in these subset models suggests that listeners are not sensitive to the fine-grained articulatory variation beyond its association with whether a stop is produced with a canonical release burst or not.

	Estimate	Std. Error	DF	t value	Pr(> |t|)	
(Intercept)	2.155e-01	2.978e-01	1.126e+02	0.724	0.470704	
TT Height (z)	7.005e-03	3.439e-02	1.484e+02	0.204	0.838860	
log.dur	9.229e-03	3.215e-02	1.482e+02	0.287	0.774445	
trial #	3.522e-02	1.077e-02	5.404e+03	3.271	0.001077	**
log.freq	-1.887e-02	1.170e-02	1.483e+02	-1.612	0.109009	
t.d - [t]	2.859e-02	2.273e-01	1.465e+02	0.126	0.900080	
gram - Mono	3.435e-01	8.795e-02	1.473e+02	3.906	0.000143	***
gram - Semi	1.142e-01	1.405e-01	1.441e+02	0.813	0.417806	
preceding - [p]	-1.208e-01	3.262-01	1.469e+02	-0.370	0.711726	
preceding - [f]	-3.855e-01	3.336e-01	1.466e+02	-1.156	0.249699	
preceding - [v]	-2.146e-01	2.664e-01	1.470e+02	-0.806	0.421636	
preceding - [m]	-3.020e-01	2.676e-01	1.470e+02	-1.128	0.260981	
preceding - [k]	-1.467e-01	3.304e-01	1.467e+02	-0.444	0.657790	
preceding - [g]	-6.492e-02	2.913e-01	1.468e+02	-0.223	0.823946	
preceding - [ŋ]	-1.621e-01	3.499e-01	1.470e+02	-0.463	0.643877	
following - [p]	1.933e-01	1.952e-01	1.484e+02	0.990	0.323765	
following - [f]	-9.767e-02	1.806e-01	1.468e+02	-0.541	0.589573	
following - [m]	8.655e-02	2.003e-01	1.478e+02	0.432	0.666272	
following - [k]	2.988e-02	1.676e-01	1.480e+02	0.178	0.858781	
following - [h]	1.570e-01	1.763e-01	1.478e+02	0.891	0.374582	
following - [w]	-4.229e-02	1.891e-03	1.478e+02	-0.224	0.823372	
following - pause	1.910e-01	1.488e-01	1.477e+02	1.284	0.201163	
following - vowel	1.705e-01	1.369e-01	1.478e+02	1.246	0.214741	

Table 3.2: Predictor estimates for mixed effects linear regression predicting listener ratings to stimuli with author-judged 'audible' stops.

In these regression models fit to listener clarity ratings, there is no evidence that listeners are sensitive to fine-grained articulatory variation *within* those stimuli that are retained or deleted according to traditional acoustic-impressionistic criteria for judging CSD. While there is a large amount of individual variation, by-and-large listener ratings exhibit a bimodal distribution that corresponds closely to my own binary judgments of coronal stop audibility. Figure 3.1 shows the overall distribution of listener ratings for the clarity of coronal stops, coloured in terms of my own judgments. The bimodal distribution that is evident in the figure instills some confidence that listeners performed the task properly and responded to salient cues to coronal stop production. It also goes some way to corroborating my own acoustic impressionistic judgments in §2.3.1. However, a goal of this chapter is to probe whether listeners are sensitive to variation in CSD tokens beyond a coarse binary

	Estimate	Std. Error	DF	t value	Pr(> |t|)	
(Intercept)	-1.925e+00	7.907e-01	5.470e+01	-2.435	0.0182	*
TT Height (z)	-4.765e-02	4.634e-02	5.401e+01	-1.028	0.3084	
log.dur	1.306e-01	6.605e-02	5.399e+01	1.977	0.0532	.
trial #	-1.239e-02	1.803e-02	2.310e+03	-0.687	0.4920	
log.freq	9.105e-02	6.059e-02	5.399e+01	1.503	0.1387	
t.d - [t]	3.139e-01	3.543e-01	5.400e+01	0.866	0.3796	
gram - Mono	1.881e-01	5.591e-01	5.399e+01	0.336	0.7378	
gram - Semi	7.742e-01	7.399e-01	5.399e+01	1.046	0.3000	
preceding - [p]	5.443e-01	6.729e-01	5.400e+01	0.809	0.4221	
preceding - [f]	7.120e-02	8.048e-01	5.399e+01	0.088	0.9298	
preceding - [v]	7.726e-01	4.323e-01	5.399e+01	1.787	0.0795	.
preceding - [m]	8.337e-01	4.778e-01	5.399e+01	1.745	0.0867	.
preceding - [k]	3.939e-01	7.148e-01	5.399e+01	0.551	0.5839	
preceding - [ŋ]	1.243e+00	7.958e-01	5.399e+01	1.562	0.1241	
following - [p]	-7.003e-02	4.959e-01	5.399e+01	-0.141	0.8882	
following - [f]	-4.934e-01	5.975e-01	5.399e+01	-0.826	0.4126	
following - [m]	2.396e-01	3.518e-01	5.399e+01	0.681	0.4987	
following - [k]	9.102e-02	5.570e-01	5.400e+01	0.163	0.8708	
following - [h]	-3.765e-01	6.915e-01	5.399e+01	-0.544	0.5883	
following - [w]	1.398e-01	6.714e-01	5.399e+01	0.208	0.8358	
following - pause	-4.213e-04	5.905e-01	5.399e+01	-0.001	0.9994	
following - vowel	7.208e-01	5.315e-01	5.399e+01	1.356	0.1807	

Table 3.3: Predictor estimates for mixed effects linear regression predicting listener ratings to stimuli with author-judged 'inaudible' stops

categorisation. As such, this result reinforces the need to consider the canonically 'audible' and 'inaudible' subsets separately, in order to explore what conditions variation in the ratings under each peak.

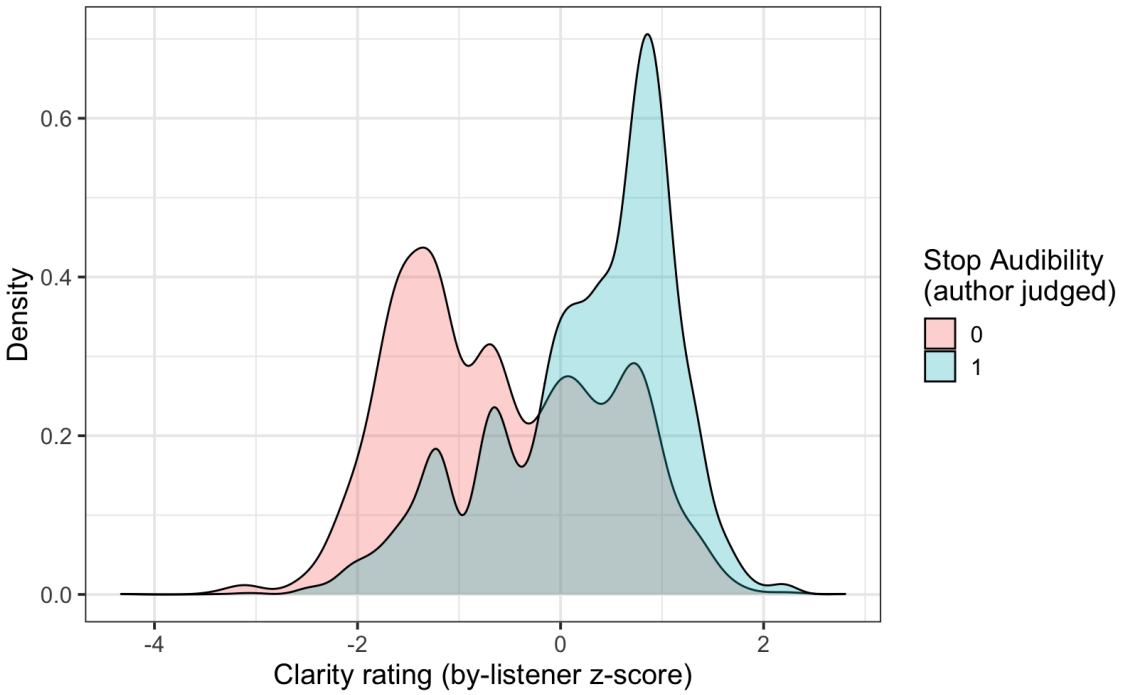

Figure 3.1: Density of z-scored listener ratings for clarity of coronal stops

3.3.1 Effect of morphology on coronal stop clarity ratings

One of the key effects observe in some of the regression models on listener ratings concerns the morphological class of the word containing the coronal stop. Specifically, within those coronal stops I judged to be 'audible', listeners rated the stops at the end of monomorphemic words to be significantly clearer than stops at the end of regular past forms. This is surprising because it goes against what we would expect both in terms of (1) the classic CSD conditioning whereby coronal stops at the end of monomorphemes are deleted more frequently than coronal stops at the end of regular past forms, and (2) Chapter 2's finding that tongue tip raising for coronal stops at the end of monomorphemes is generally smaller in magnitude than for coronal stops at the end of regular past forms. A potential explanation for this effect is that listeners have some knowledge of the morphological conditioning on CSD that influences their expectations. Specifically, listeners may expect a higher rate

of deletion in monomorphemes, so when a coronal stop is audible in this context it is surprising and is perceived as particularly 'clear'. In other words, the different rates of CSD according to morphological context may give rise to correspondingly different thresholds for what constitutes a clear stop. Interestingly, however, the morphological effect is not present among ratings for the subset of coronal stops I judged to be 'inaudible'. Figure 3.2 demonstrates this asymmetry, showing listener ratings in terms of my binary judgments and the morphological class of the word.

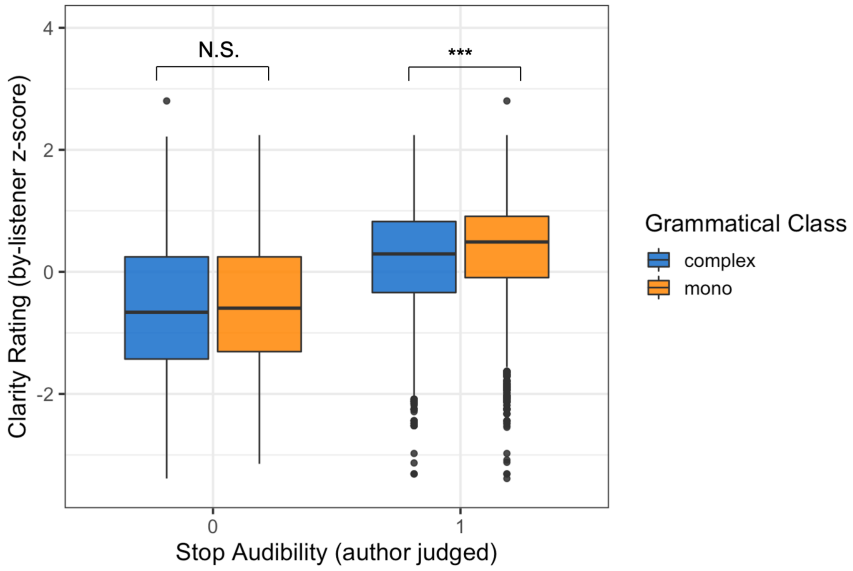

Figure 3.2: Z-scored listener ratings by morphological class and author-judged stop audibility

The fact that the morphological conditioning on listener ratings is limited to the 'audible' subset of coronal stop stimuli further suggests that listeners are not straightforwardly responding to articulatory measures like tongue tip height. The tongue tip height variation according to morphological class is consistent across tokens that are canonically 'audible' *and* 'inaudible'. Figure 3.3 shows the tongue tip height behaviour in the monomorphemic and regular past forms used as perceptual stimuli in this chapter, grouped according to my

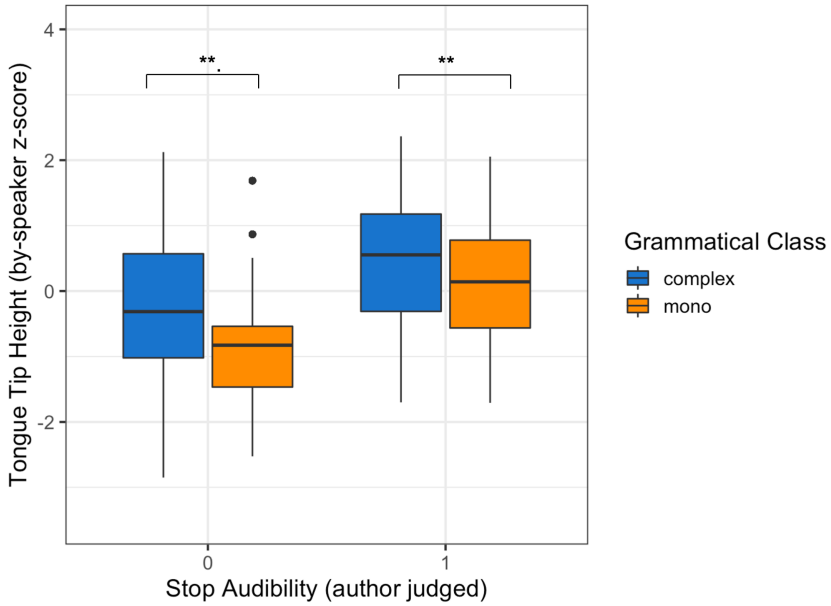

Figure 3.3: Z-scored tongue tip height at the apex of underlying coronal stop trajectories by morphological class and author-judged stop audibility

binary audibility judgments for comparison.

3.3.2 Effect of trial number on coronal stop clarity ratings

The final effect to be considered is that of trial number. Listeners rated 'audible' coronal stops earlier in the experiment as less clear than 'audible' coronal stops later in the experiment. The same effect is not found within the subset of coronal stop stimuli I judged to have 'inaudible' stops—tokens that were not produced with a canonical release burst. A potential explanation for this effect is that listeners rate 'audible' stops to be clearer as they have more experience encountering 'inaudible' coronal stops. In other words, listeners take some time to gain a proper impression of the range of coronal stop clarities that exist among the stimuli. This effect is extremely small, as shown by the very shallow incline in ratings of 'audible' stops in Figure 3.4.

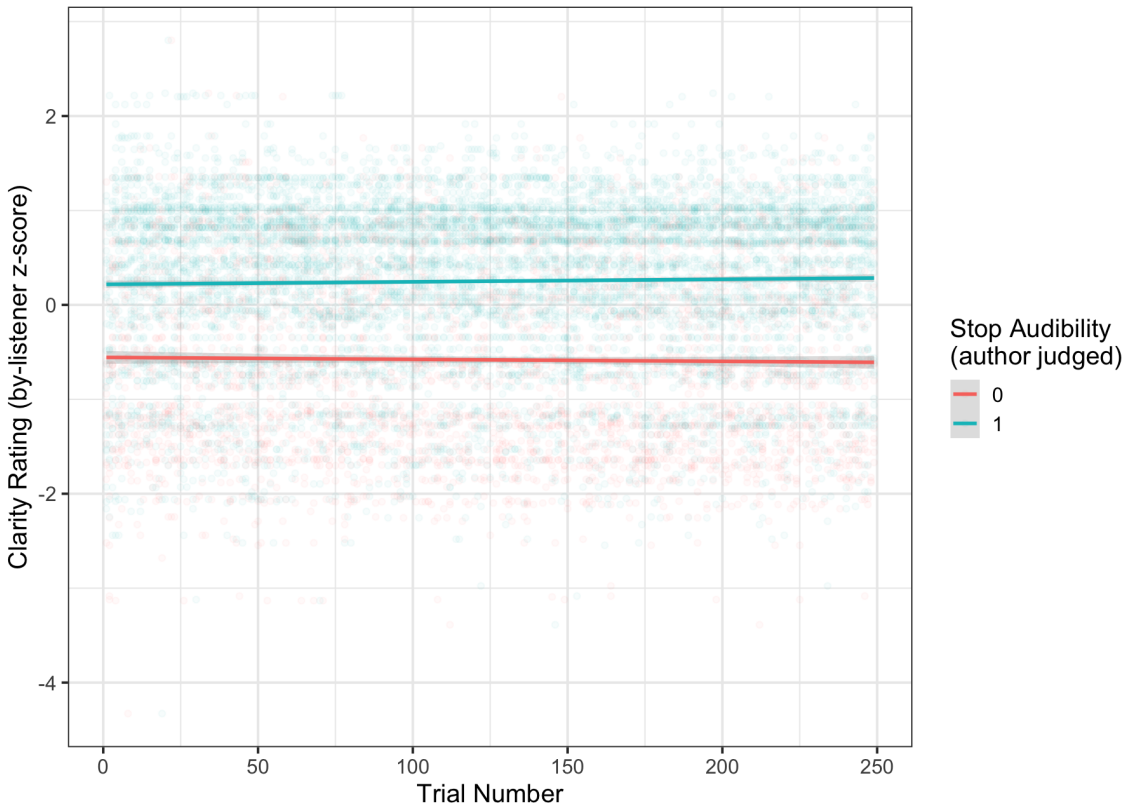

Figure 3.4: Z-scored listener ratings across trials

3.4 Discussion

3.4.1 Perceptual non-sensitivity to fine-grained articulatory variation

While, to a first approximation, listener ratings for the clarity of coronal stops are conditioned by the articulatory measures of tongue tip height and gesture duration, responses are generally bimodal and correspond with classic binary coding in terms of deletion and retention of stops. Despite the fact that listeners in this study were given a larger scale on which to make their ratings, their responses resemble transcriber insensitivity to the gradient articulatory aspects of American English flapping when deciding between discrete symbol options de Jong's (1998). Moreover, subsetting the data according to my own binary

coding (as a proxy for the cues to which listeners seem most sensitive) reveals that there is no evidence of listener sensitivity to articulatory measures *within* each subset. That is, underlying coronal stops I judged to be 'inaudible' are not significantly *less* clear if they are articulated with a lower tongue tip or a shorter gesture, and coronal stops I judged to be 'audible' are not significantly *more* clear if they are articulated with a higher tongue tip or a longer gesture, as we might expect. Thus, listener judgments of coronal stop clarity, by and large, support my own coarse judgments of coronal stop audibility (retention versus deletion) using acoustic-impressionistic criteria. We can, therefore, characterise the articulatory variation observed in Chapter 2 as 'covert'.

We should consider whether, if listeners do not perceive it, covert variation is particularly meaningful. At least for sociolinguists, variable phenomena of interest are typically those that are shared across a speech community. Individual speakers can perform all sorts of idiosyncratic variation but what is key is that certain variables are picked up and propagated, exhibit widespread systematic conditioning, garner shared social meanings and are generally understood to convey something of a speaker's self and their place in the world. If variation is not perceived, it is unclear how any of this can occur. In this sense, we should take heart that classic acoustic-impressionistic studies of CSD may reflect something important about the evaluation of the variable at the level of the listener. Indeed, acoustic-impressionistic studies basically amount to coarse perception experiments with a small pool of participants. However, this does not detract from the fact that covert variation gives us key insights into the representation of CSD, and the central finding of rare or non-existent categorical implementations of CSD was shared across all five speakers in Chapter 2. Putting the results from both chapters together, an asymmetry between production and perception raises new puzzles for the representation of CSD.

3.4.2 Morphology and listener expectations

One surprising pattern in the listener ratings of coronal stop clarity is that monomorphemes were judged to be less clear than *-ed* suffixed forms (within the 'audible' subset of stimuli).

This finding seems to go in the opposite direction to the robustly attested finding that CSD is most common in monomorphemes and least common in regular past forms, as well as the finding in this dissertation that coronal stops at the end of monomorphemes are articulated with lower tongue tips at their apex than those in *-ed* suffixed forms. This difference in articulatory magnitude is also present in the 250-token subset that were used to create perceptual stimuli for this chapter. As such, this effect serves to reinforce the observation that listeners do not seem sensitive to the articulatory measures.

The explanation for the morphological effect on listener ratings that I find most compelling is in terms of listener expectations, and some degree of surprisal when they are not met. If listeners have knowledge of standard patterns of CSD such that coronal stops in monomorphemes are expected to be more susceptible than those in *-ed* suffixed forms, perhaps the non-deletion (audible retention) of coronal stops is less expected in monomorphemes than in *-ed* suffixed forms. Therefore, when listeners encounter an audibly retained coronal stop at the end of a monomorpheme, it is not just clear but *unusually* clear.

There is a sense in which the expectation reasoning is conceptually problematic for functional accounts of CSD, which might otherwise form a convenient alternative to avoid postulating a morphology-phonetics interface. Functional accounts attribute the fact that coronal stops are most frequently retained when they constitute *-ed* suffixes to a general imperative to convey important grammatical information such as the past tense. But if listeners correspondingly adjust their listening behaviour to downgrade *-ed* suffixed forms, they are essentially rendering the conscientious speaker's behaviour less effective. It may be that this circularity is not present when these functional accounts are more fleshed out; listeners may adjust ratings of coronal stops according to a containing word's morphological class, but there are more factors involved in optimising the communicative efficiency of these stops. In Chapter 5 of this dissertation, I explore a more complex version of this kind of account in terms of morphologically-informed predictability.

3.4.3 Possible implications for acquisition

As mentioned in Chapter 2, the articulatory finding whereby speakers employ diverse strategies in order to produce underlying coronal stop tokens that would traditionally be coded as deleted creates questions about the nature of acquiring CSD as a process. Perception preceding acquisition, the results from the current chapter compound the problem. The evidence I have presented now points to (1) multiple strategies for producing apparent CSD, the majority of which look more like gradient lenition than categorical deletion, and (2) listener non-sensitivity to the differences between these strategies. The combination of these points has important implications for acquisition. If listeners cannot perceive the fine-grained articulatory detail in CSD beyond whether a stop is canonically produced or not, this detail must not be directly acquired through imitation.

CSD has, in the past, been a crucial source of evidence on the the acquisition of variable phenomena, and specifically in terms of probability matching mechanisms where learners produce a variant at an equivalent rate to what they perceive in the input. Labov's (1989) study on a middle class family in the suburbs of Philadelphia was among the first to focus on the question of how CSD is acquired by children learning English. He concludes that children acquire CSD with all the same conditioning factors as adults by the age of 4, and that the actual probability of CSD in each context matches the adult probability by the age of 7. This makes it one of a number of studies where it is claimed that variable phonological processes like CSD are acquired in tandem with, and sometimes even before, categorical phonological processes (Roberts and Labov, 1995; Foulkes et al., 2005; Smith et al., 2009). In terms of CSD, this process of acquisition tends to involve increasing the number of *retained* coronal stops and introducing constraints on their omission, rather than the other way around. The broader phenomenon of cluster simplification is well documented during children's acquisition of their first language's phonology (Shriberg and Kwiatkowski, 1980). The rate of simplified variants of clusters drops from around 70% between 2;0 and 3;4 (McLeod et al., 2001) to around 10% by the age of 4;0 (Waring et al., 2001). However, the pathway from this aggressive cluster simplification to adult-like conditioning on CSD

is complicated. For example, the regular marking of preterite and passive forms is absent in young childrens' speech in English (Radford, 1992) and only appears after 40 months (Brown, 1973). Relatedly, multiple studies report that semiweak past forms show near-obligatory CSD in the speech of young children (Guy and Boyd, 1990; Roberts, 1997), but interpret these results differently: these forms could have no underlying stop in early acquisition, or they could pattern with monomorphemes until children learn that the stop constituted a suffix. If the latter analysis is to be believed, it is not clear why the regular *-ed* suffix should also be so frequently absent at an early stage of acquisition. Studies also differ as to the order in which constraints are acquired—sociostylistic constraints first in some (Labov, 1989), and phonetic constraints first in others (Roberts, 1997; Smith et al., 2009)—and the effect size of these constraints—e.g. the effect of the preceding segment, which is relatively weak in some studies (Guy, 1980; Labov, 1989), but strong in others (Bayley, 1994; Santa Ana, 1996). But the end result is always a CSD pattern that closely matches that found in the speech of caregivers, suggesting the child learner imitates what they hear.

The results from this and the previous chapter problematise any assumptions of the acquisition of CSD as imitating rates of categorical deletion. Apparent cases of inaudible stops are not produced in a unitary process of deletion, and listeners do not rate underlying coronal stop cases with tongue tip raising but no obvious acoustic cues like a release burst (suggesting phonetic undershoot or gestural overlap) as any clearer than cases without tongue tip raising (suggesting categorical phonological deletion). All of these stimuli are given very low average scores for clarity of stop production. Applying a hypothetical straightforward mechanism of CSD imitation to these facts, we would expect a learner to acquire all such cases as equally 'unclear' or absent, and attempt to produce CSD (with a single strategy) at an equivalent rate. However, since many coronal stops are rendered inaudible by phonetic effects like lenition and coarticulation, in addition to what looks like more categorical deletion, we would expect these learners to actually produce an elevated rate of inaudible stops compared to their input. That is, they would match the rate of

inaudible coronal stops in their input with an equivalent rate of phonological CSD, and augment it with coronal stops rendered inaudible through gradient connected speech processes. This elevated rate of inaudible stops would then presumably give rise to an even higher rate of CSD in a subsequent generation of learners, and so on. Since, as has been widely attested, CSD is a stable variable whose rate of implementation has not increased over time (Roberts and Labov, 1995; Baranowski and Turton, 2020), we can *a priori* dismiss the notion that this kind of straightforward imitation is the only mechanism involved in the acquisition of CSD.

Since we do not see rapid generational change CSD rates, and it does not appear to be the case that learners maintain this stability by directly perceiving and imitating the different articulatory strategies for rendering coronal stops inaudible, there are limited remaining possible for the acquisition of CSD patterns. One is that learners employ complex inference and fine-grained control over their own phonetic behaviour in order to render an appropriate proportion of underlying word-final coronal stops inaudible, and this rate can be augmented as necessary with a categorical deletion process. This may be consistent with the previously described tendency for children to initially overproduce cluster simplification before pulling back to an adult-like CSD rate. However, this does not help to explain the robust patterns of conditioning on CSD or, more to the point, on the articulatory detail in the execution of underlying coronal stops. A second possibility is that these patterns are acquired indirectly, and covary with some other property that can be more straightforwardly learned. A good candidate for this kind of indirect learning is the kind of timing properties that may be associated with morphological structure, as discussed in Chapters 1 and 2. In other words, listeners may associate morphologically complex words with longer durations, which in turn results in higher tongue tip raising and less apparent deletion of coronal stops. This idea will is explored a little further in Chapter 4.

Chapter 4

Duration as a perceptual cue to morphological complexity

In the previous chapter I demonstrated that while listeners show some interesting patterns in their ratings of how clearly an underlying word-final coronal stop was pronounced, they do not appear to be sensitive to fine-grained articulatory parameters beyond the relationship of these parameters to the traditionally-recognised binary categorisation of these coronal stops as retained or deleted (audible, or inaudible). This finding casts doubt on the idea that CSD, implemented through a variety of strategies as shown in Chapter 2, is acquired through imitation alone since it seems unlikely that the learner can perceive the different implementation strategies in order to learn them. One possible explanation to reconcile these findings is that the systematic variation in tongue tip behaviour might be an indirect consequence of some other process. Specifically, timing differences in the production of different words may give rise to differences in the magnitude of articulatory movement and, ultimately, different rates of apparent CSD. In this chapter, I explore an existing idea that morphological complexity is associated with phonetic lengthening, approaching it from a perceptual angle. I show that listeners can utilise duration differences associated with word frequency and morphological complexity for homophone disambiguation.

4.1 Word duration

Speech unfolds over time, but exactly how much time it takes to say something is a difficult question. This is, in part, due to the nested structure of prosodic constituents. While there

are specific timing constraints on the execution of different phones (i.e. a range of times in which a given kinematic movement is comfortable/possible according to the laws of physics), the timings that are implemented for the same phone vary according to the identity of the larger syllable, word, and phrase in which it appears. In general, the more sub-units within a given prosodic constituent, the less time is allotted to each one. In addition, speech timing is modulated according to the presence of prosodic boundaries and prominences, both of which tend to induce local lengthening effects. Then, on top of all of this, we must allow for the influence of a global speech rate parameter such that speakers can execute the same utterance faster or slower. Some theoretical frameworks, like Articulatory Phonology (Browman and Goldstein, 1985; Saltzman et al., 2008) position many of these aspects of speech timing as intrinsic to phonological representation (also see Fowler et al., 1980), while others relegate timing to mechanisms of the phonetic interpretation of atemporal strings of phonological symbols (Henke, 1966; Keating, 1990; Fujimura, 1992; Guenther, 1995; Levelt et al., 1999). This aspect of speech timing an active area of debate that is beyond the scope of the current chapter (see Turk and Shattuck-Hufnagel 2021 for overview).

Whatever their source and representation, timing effects are ubiquitous. In terms of the timing of individual words, even when ostensibly the same set of phones are arranged in the same order (i.e. homophones), a number of parameters seem to condition their duration in speech production. One such parameter, as is the focus of this dissertation, concerns morphological structure. A number of studies find that morphologically complex words are, all else equal, produced with some degree of lengthening compared to morphologically simplex words (Walsh and Parker, 1983; Lociewicz, 1992; Schwarzlose and Bradlow, 2001; Sugahara and Turk, 2009; Seyfarth et al., 2018). One explanation that has been suggested for this is that some morphological boundaries are encoded as prosodic boundaries and induce the same kind of local lengthening found at prosodic phrase boundaries (Sugahara and Turk, 2009). Another way of approaching the problem is to cite paradigm uniformity, such that a morphologically complex word (e.g. *baking*) is influenced by the production of the same stem in other contexts, including a morphologically simplex word (*bake*) which

features word-final lengthening at the site corresponding the morphological boundary (Seyfarth et al., 2018).

Another factor that effects word duration is word frequency. It is commonly observed that frequent words are shorter in duration than infrequent words, both in terms of the whole wordform (Wright, 1979) and individual matching segments (Kawamoto, 1999). While some studies fail to replicate these effects in the laboratory (Damian, 2003; Mousikou et al., 2015), they are consistently reported for corpus studies on spontaneous speech (Aylett and Turk, 2004; Gahl, 2008). A number of explanations for this type of effect have been put forward. Firstly, the impact of frequency on pronunciation is commonly cited as evidence for usage-based frameworks like Exemplar Theory (Pierrehumbert, 2002; Bybee, 2002). In this perspective words are represented as separate clouds of episodic traces, even if they are nominally homophones, and word frequency is a measure of how quickly these are accumulated. Other accounts attribute frequency effects in pronunciation to online mechanisms in speech production, either in terms of a speaker's differential ease of access to words that are more or less familiar (e.g. Baese-Berk and Goldrick, 2009) or in terms of accommodating listeners' ease of lexical access along the same lines (e.g Lindblom, 1990; Aylett and Turk, 2004). These different explanations for frequency effects, and their relation to the question of morphology–phonetics interactions, are explored further in Chapter 5.

Finally, and relevantly for the experimental design in this chapter, there has been some suggestion that the orthographic representation of words affects word production and—ultimately—duration. Research in this area has predominantly focused on delays in the onset of word production, with very mixed results. Some studies suggesting that a word prime can speed the production of words with shared orthography but not shared phonology (Damian, 2003), while others fail to find any effect of shared orthography (Chen et al., 2002), and others still claim that any such effect is attributable to shared phonology after all (Alario et al., 2007). Roelofs (2006) argue that these different effects are the result of different experimental paradigms, and orthographic effects are only found when participants are asked to read or memorise orthographic representations. As for the effects of orthography

on pronunciation, words that are elicited using real orthographic representations containing more graphemes are produced with a longer duration (Warner et al., 2004; Brewer, 2008; Grippando, 2021). The same effect is found for the duration of individual consonants and the corresponding graphemes in the orthographic representation (e.g. the final /k/ in *clique* is produced longer than that in *click*). These effects are also found, to some extent, in corpus spontaneous speech (Brewer, 2008), but not in non-words or novel orthographies for real words. However, the effect of orthography is not always effectively disentangled from that of word frequency or morphological complexity. In this study, I control for all three factors by exploring the effects of word duration on homophones in the perceptual domain.

4.2 Methods

4.2.1 Stimuli

A list of pairs of English homophones was prepared such that each word in a pair had a different orthographic representation but their canonical phonological representation was identical. For 60 of the pairs, both words were monomorphemic (e.g. *time*, *thyme*), with no clear affixes to decompose. For this pairtype, the word whose recording was ultimately chosen and extracted (described below) was designated as Word 1. For 33 of the pairs, Word 1 ended in an *-ed* while Word 2 was monomorphemic (e.g. *packed*, *pact*). For 33 more of the pairs, Word 1 ended in an *-s* suffix (marking plural or 3SG agreement) while Word 2 was monomorphemic (e.g. *laps*, *lapse*). This amounted to 126 English homophone pairs from three pair types. All words were monosyllabic and listed alongside their frequency (extracted from $\text{SUBTLEX}_{\text{US}}$ and log-transformed) and orthographic length (number of letters in standard US English spelling).

A 36-year-old upper-middle class white male speaker of Mainstream American English from Southern New Jersey recorded each word in the list of 126 homophone pairs. The speaker was asked to speak clearly, at a measured pace, to repeat each word three to five times, and to precede each instance with the phrase “The word is...”. The speaker also

recorded a further 10 words with no homophonous counterparts, to serve as check stimuli. These 10 words were paired with words that contrasted only in the final consonant (e.g. *pet* was paired with *peck*).

All the recordings were visually inspected in Praat, and for each of the 136 pairs, a single recording of one of the words was selected and extracted without the carrier phrase. For the monomorphemic pairs, this was the instance of either word with the closest to median duration, that was also judged to be sufficiently clear in terms of modal voice quality and 'neutral' affect. For the pairs with one *-ed* suffixed or *-s* suffixed form, this recording was always among the instances of the morphologically complex word (containing the suffix), and was similarly selected to optimise clarity while aiming for the median word duration across instances of both the complex and monomorphemic words in the pair. Similarly, one recording of each of the 10 words with no homophonous counterparts was selected to be closest to the median duration for that word and also sufficiently clearly pronounced. Finally, a single instance of the carrier phrase "The word is...", that impressionistically had a steady pace and sufficient clarity, was selected and extracted. All recordings were trimmed to remove silence before and after the relevant utterance. This resulted in 136 recordings of words in isolation and a single recording of the carrier phrase "The word is...".

Each of the 126 recordings of words with homophones was opened in Praat and converted to two manipulation objects. The pitch thresholds for the manipulations were tweaked on a word-by-word basis in order to ensure proper interpolation of the glottal pulses when editing. The manipulation objects were used to stretch and compress the duration of the whole rhyme in each word (e.g. [akt] in *packed*). For each word, one manipulation was used to increase the duration of the rhyme to 120% of the original recording, and the other manipulation was used to decrease the duration of the rhyme to 80% of the original recording. All manipulations were resynthesised using the overlap-add method. This resulted in three versions of each of the 126 homophone recordings at three different durations: long, medium, and short.

4.2.2 Participants

98 listeners who were native speakers of American English and reported spending most of their childhood in the United States and Canada were recruited using the University of Pennsylvania Psychology Subject Pool. All listeners were undergraduate students at the University of Pennsylvania, reported no diagnosed hearing or reading difficulties, and reported that they were not distracted during the experiment.

Upon starting the experiment, listeners were randomly assigned to one of six stimulus lists. Each list contained 136 unique items, 126 of which were critical homophone pairs. A third of the homophone pairs were presented with a long duration, a second third with a medium duration, and a final third with a short duration. The 10 test items were presented unmanipulated to all participants. Each third was balanced within its pair type (monomorphemic, one *-ed* suffixed word, or one *-s* suffixed word) for average difference in log-frequency, summed differences in log-frequency, average difference in orthographic length, and summed differences in orthographic length. For the pairs with one suffixed word (*-ed* or *-s*), differences were calculated by subtracting Word 2's (monomorphemic) value from Word 1's (complex) value. For the monomorphemic pairs, absolute values were used. Groups of pairs with one suffixed word were also balanced in terms of the number of pairs where Word 1 was greater, equal, or lesser than Word 2 in terms of frequency and orthographic length. Lists were counterbalanced for stimulus duration according to a Latin square design.

4.2.3 Procedure

Listeners were asked to wear headphones and avoid distractions. For each of 136 trials, an identical carrier phrase, "The word is...", was played using the same recording, and immediately followed by a word from the stimulus list. Simultaneous with the onset of this word, listeners were presented with two orthographic representations on the left and right sides of the screen. In the case of the 126 critical homophone trials, these orthographic representations corresponded to the two homophones with different spellings that made up

the original list and from which stimuli recordings were made. In the case of the 10 check trials, one orthographic representation matched the word that was played, while the other was a word that differed in terms of the final consonant (e.g. *pet*, *peck*). The position of these orthographic representations on the left or right was randomised and the order in which trials were presented was randomised. Check trials were coerced to be presented at even intervals throughout the experiment, to ensure continued attention.

In each trial, listeners were asked to click on the orthographic representation of the word that matched what they thought they heard. Once listeners made a selection, there was a one second interval before the subsequent trial began. There was no time limit for selections, but listeners were encouraged not to deliberate for more than a few seconds for each trial. Prior to beginning the experiment, listeners were given instructions about the task that included some example trials. These example trials included both non-homophone and homophone pairs. Listeners were warned that the majority of trials would resemble the homophone pairs and it might be 'difficult to tell' what they had heard.

4.2.4 Analysis

16 listeners responded incorrectly to more than one of the check trials and were excluded from analysis. In addition, no listener selected *quay* as a possible orthographic representation for the *key–quay* homophone pair, so this item was excluded entirely. The remaining critical homophone data, from 82 listeners, was analysed using mixed effects logistic regression modeling. Three models were fit, one each for data from the three pair types. The fixed and random effect structure in each model was identical and corresponded to the hypothesised relationships between stimulus duration and morphological complexity, word frequency, and orthographic length. Each model was fit with main fixed effects of stimulus duration (long; medium; short), frequency difference (the log-transformed Word 1 frequency minus the log-transformed Word 2 frequency), and orthographic length difference (the number of letters in Word 1 minus the number of letters in Word 2). Models also included interaction terms for stimulus duration with frequency difference, and for stimulus

duration with orthographic length difference. Finally, all three models included random intercepts for each listener and each pair. Figures were made using the SJPlots package in R to visualise predicted effects directly from logistic regression models.

4.3 Results

4.3.1 Monomorphemic homophone pairs

Before tackling the the morphologically complex word classes, it seems prudent to explore listener behaviour when presented with homophone pairs where both words are monomorphemic (e.g. *time*, *thyme*). Since these words feature no morphological boundary to induce lengthening effects, the potential effects of word frequency and orthographic length are all that remain to investigate. Starting with frequency, monomorphemes show both a main effect of word frequency and an interaction between word frequency and stimulus duration. Figure 4.1 is a visualisation of both the main effect of the difference in word frequency between homophones, and the interaction between word frequency and stimulus duration, in terms of probability for participants to select Word 1 over Word 2.

The more frequent a word is compared to the homophone it is presented with, the more likely participants are to select that word [$z=4.423$, $p<0.001$]. Put differently, participants exhibit a general bias towards selecting the more frequent of two words when presented with two viable options for what they hear. As we will see, this bias pervades across all word classes in the experiment. However, the frequent word bias is modulated by the duration of the stimuli. Specifically, the frequency bias is *stronger* for short stimuli than for long stimuli [$z=2.099$, $p<0.05$]. In other words, participants are more willing to select the infrequent word in a pair when they hear the long stimulus. This effect is in the same direction as the association between word frequency and word duration that is observed in production; speakers generally produce frequent words faster than infrequent words. The results in monomorphemes seem to indicate that this association can also be utilised in speech perception. The effect of word frequency difference is not significantly different in

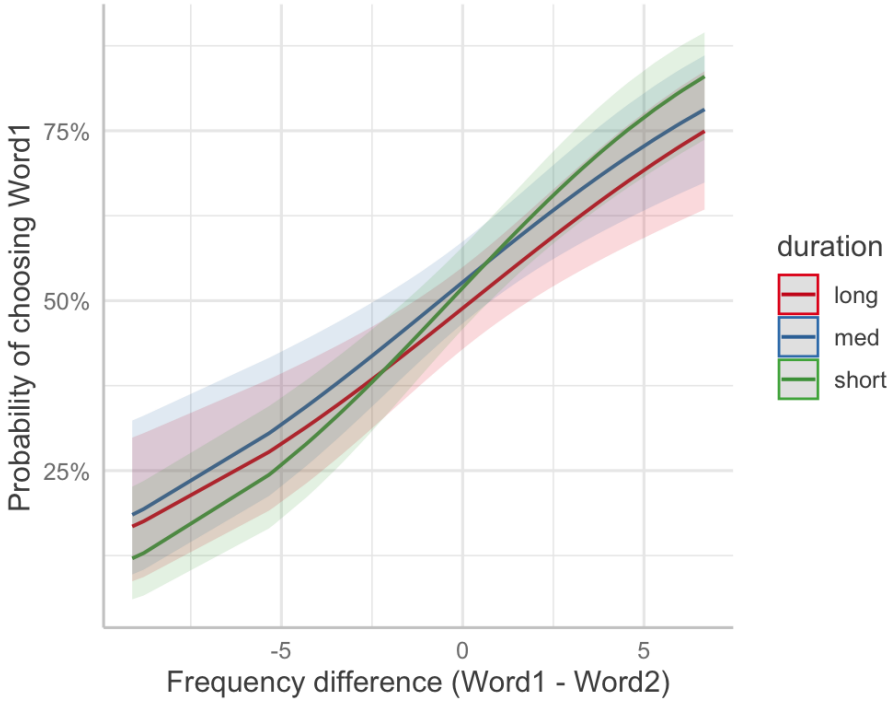

Figure 4.1: Probability of choosing Word 1 by (1) difference in word frequency between Word 1 and Word 2 and (2) stimulus duration, for monomorphemic homophone pairs

medium duration stimuli compared to long stimuli.

Participant responses to monomorphemic homophone pairs show no significant main effect of orthographic length or any interaction between orthographic length and stimulus duration. Figure 4.2 is a visualisation of both of these parameters. On average, participants selected Word 2 more often when it was both orthographically longer than Word 1 (negative orthographic difference score) and presented with an auditory stimulus that was long in duration, but this effect did not achieve significance in the model. This means that the data do not show evidence of an association between orthographic length and stimulus duration in terms of homophone selection, nor do participants show a general preference for words with more or fewer letters.

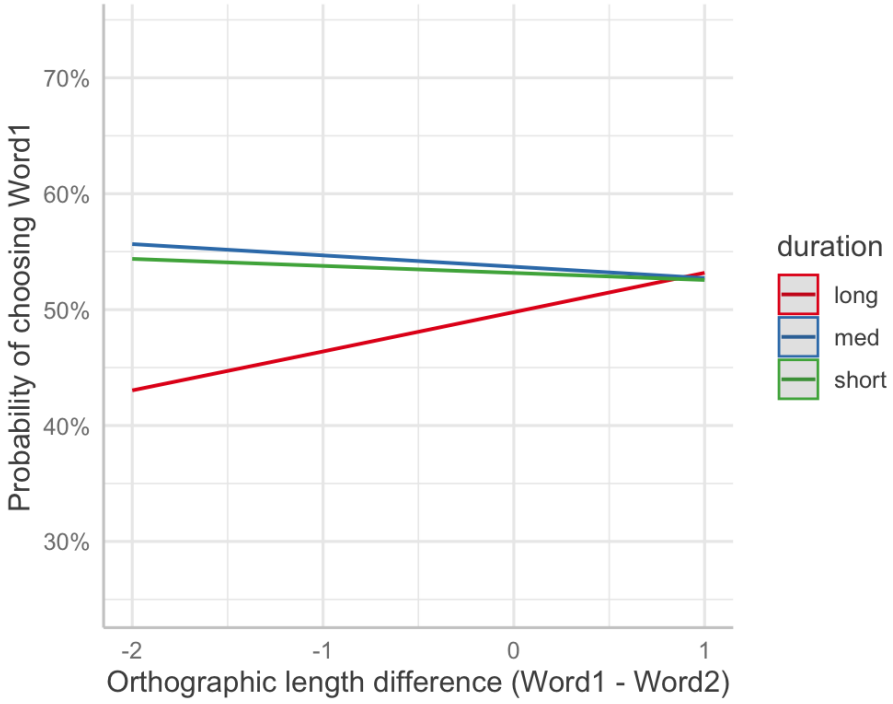

Figure 4.2: Probability of choosing Word 1 by (1) difference in orthographic length between Word 1 and Word 2 and (2) stimulus duration, for monomorphemic homophone pairs

4.3.2 *-ed* suffixed/monomorphemic homophone pairs

Taking a similar approach, the class of homophone pairs are those containing one *-ed* suffixed word and one monomorphemic word (e.g. *packed*, *pact*). Figure 4.3 first shows the main effect of word frequency difference and its relationship to stimulus duration in terms of participant selections. In these pairs, Word 1 is always the complex *-ed* suffixed word. Once again we see a main effect of the difference in word frequency between homophones such that participants are generally more likely to select the more frequent word, and this likelihood increases with a greater difference in word frequency between homophones [$z=3.965$, $p<0.001$]. In addition, pairs with one *-ed* suffixed word exhibit the same interaction between word frequency and stimulus duration as monomorphemic homophone pairs. That is, the effect of word frequency difference is stronger in short stimuli than long stimuli [$z=2.437$, $p<0.05$]. Figure 4.3 shows this interaction is primarily driven by differences in

responses when Word 2 (no *-ed* suffix) is more frequent than Word 1; participants are less likely to choose a frequent Word 2 when it is long in duration than when it is short. As for monomorphemic pairs, the effect of word frequency difference is not significantly different in medium duration stimuli compared to long stimuli for pairs with one *-ed* suffixed word.

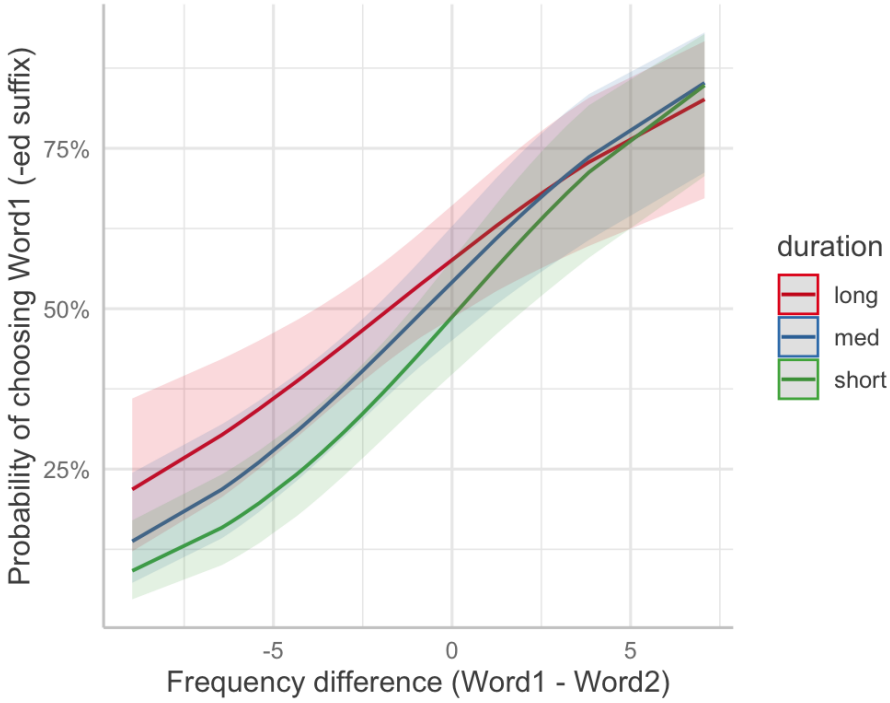

Figure 4.3: Probability of choosing Word 1 by (1) difference in word frequency between Word 1 and Word 2 and (2) stimulus duration, for *-ed* suffixed/monomorphemic homophone pairs

Secondly, Figure 4.4 shows a main effect of stimulus duration, to test the association between stimulus duration and morphological complexity, as the presence of an *-ed* suffix was deliberately manipulated in these homophone pairs. Participants show a main effect of stimulus duration in homophone pairs with one *-ed* suffixed word. Participants were less likely to select Word 1 (*-ed* suffixed) over Word 2 (monomorphemic) when they heard a short stimulus than when they heard a long stimulus [$z=-2.206$, $p<0.05$]. This effect is in the same direction as observations of associations between morphological complexity and phonetic lengthening in speech production, and suggests that these durational effects can

be utilised to some extent for disambiguation of homophones in speech perception. While responses to medium duration stimuli are intermediary, they are not significantly different from responses to long stimuli.

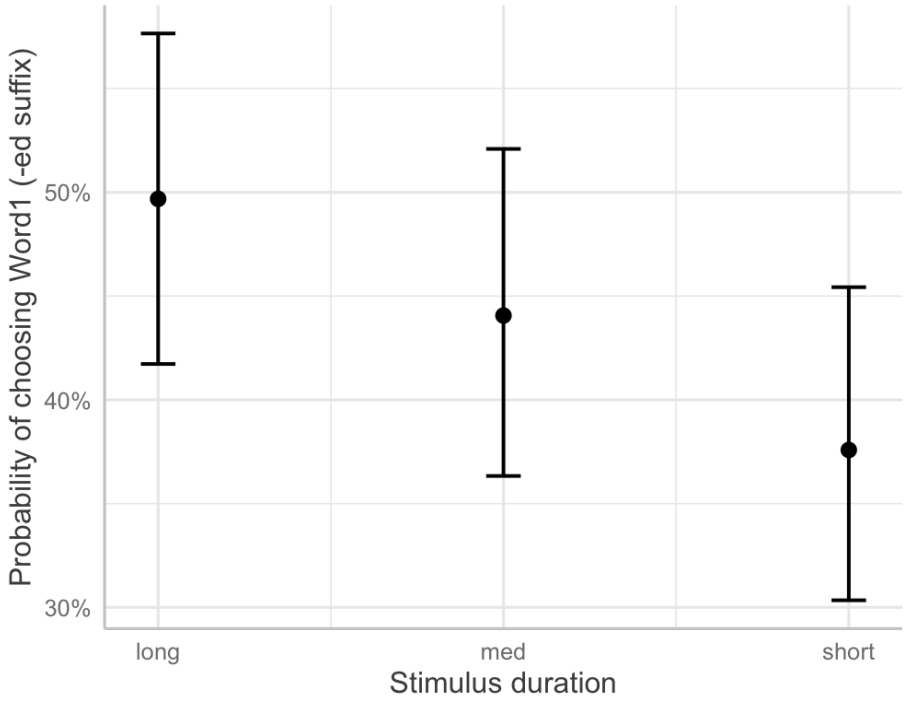

Figure 4.4: Probability of choosing Word 1 by stimulus duration, for *-ed* suffixed/monomorphemic homophone pairs

Homophone pairs with one *-ed* suffixed word show no main effect of orthographic length, nor an interaction between orthographic length and stimulus duration, on participant selections. Figure 4.5 is a visualisation of these parameters in terms of participant probability to select Word 1 (*-ed* suffixed).

4.3.3 *-s* suffixed/monomorphemic homophone pairs

The third and final class of homophone pairs considered were those where one word contains an *-s* suffix marking plural (e.g. *laps*) or 3SG agreement (e.g. *frees*), and the other is monomorphemic (e.g. *lapse*; *freeze*). Figure 4.6 the relationship between word frequency

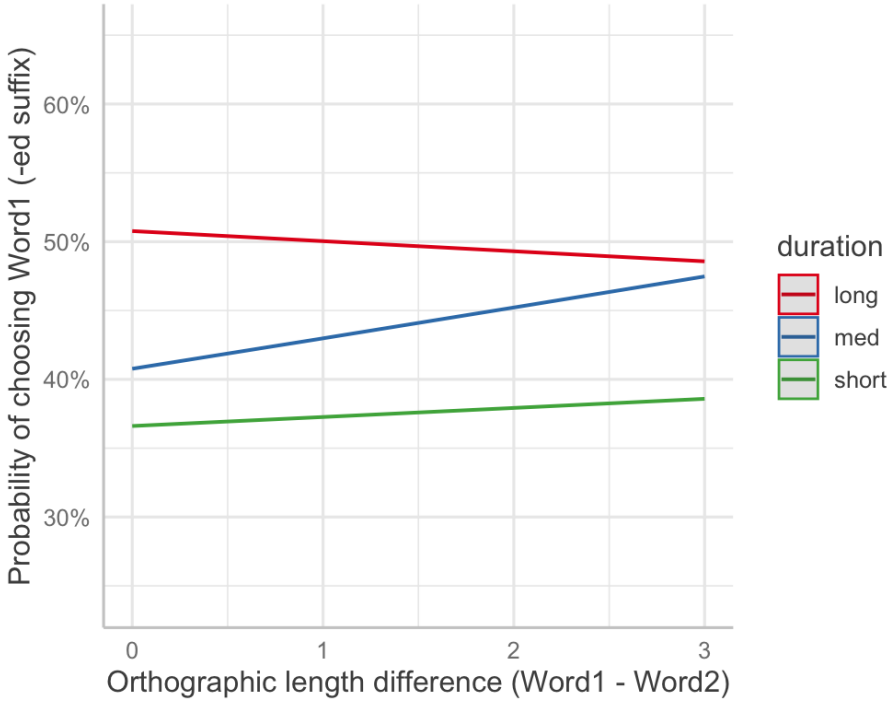

Figure 4.5: Probability of choosing Word 1 by (1) difference in orthographic length between Word 1 and Word 2 and (2) stimulus duration, for *-ed* suffixed/monomorphemic homophone pairs

difference and stimulus duration in terms of participant selections. In these pairs, Word 1 is always the complex *-s* suffixed word. Once again, participants show a bias for selecting whichever word is more frequent when presented with homophone pairs [$z=4.773$, $p<0.001$]. However, while participants on average seem to show a weaker frequency bias for long stimuli than medium or short stimuli (the same direction as the effects for the other pair categories) this interaction is not significant. It is also worth noting that participants were generally less willing to select Word 1 (*-s* suffix) over Word 2 (monomorphemic) than in the other pair categories. This can be observed in the floor effect whereby Word 2 was almost exlusively selected when it was more frequent than Word 1. In addition, when Word 1 was more frequent than Word 2 participants exhibited a very high level of variance in their selections compared to in other pair categories.

Figure 4.7 investigates the effect of stimulus duration on probability for participants to

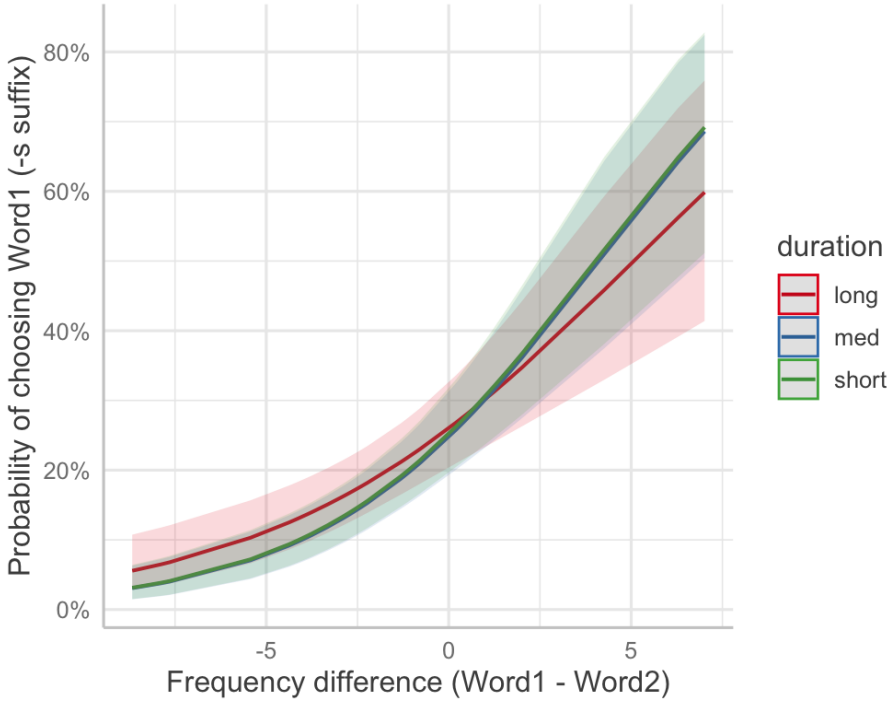

Figure 4.6: Probability of choosing Word 1 by (1) difference in word frequency between Word 1 and Word 2 and (2) stimulus duration, for *-s* suffixed/monomorphemic homophone pairs

select Word 1 (*-s* suffix). As before, this was a test of the association between stimulus duration and morphological complexity, since the presence of an *-s* suffix was deliberately manipulated in these homophone pairs. While participants, on average, selected Word 1 more frequently in long stimuli than in medium or short stimuli (just as in pairs with one *-ed* suffixed word), this effect was not significant. This means that there is no statistical evidence of an association between morphological complexity and stimulus duration for homophone disambiguation with *-s* suffixed words.

Finally, homophone pairs with one *-s* suffixed word show an unexpected main effect of orthographic length. Specifically, participants were more likely to choose Word 1 (*-s* suffix) when Word 1 had more letters compared to Word 2 [$z=2.196$, $p<0.05$]. I will discuss the possibility that this points to some other property covarying with orthographic length difference in homophone pairs with one *-s* suffixed word in the next section. The

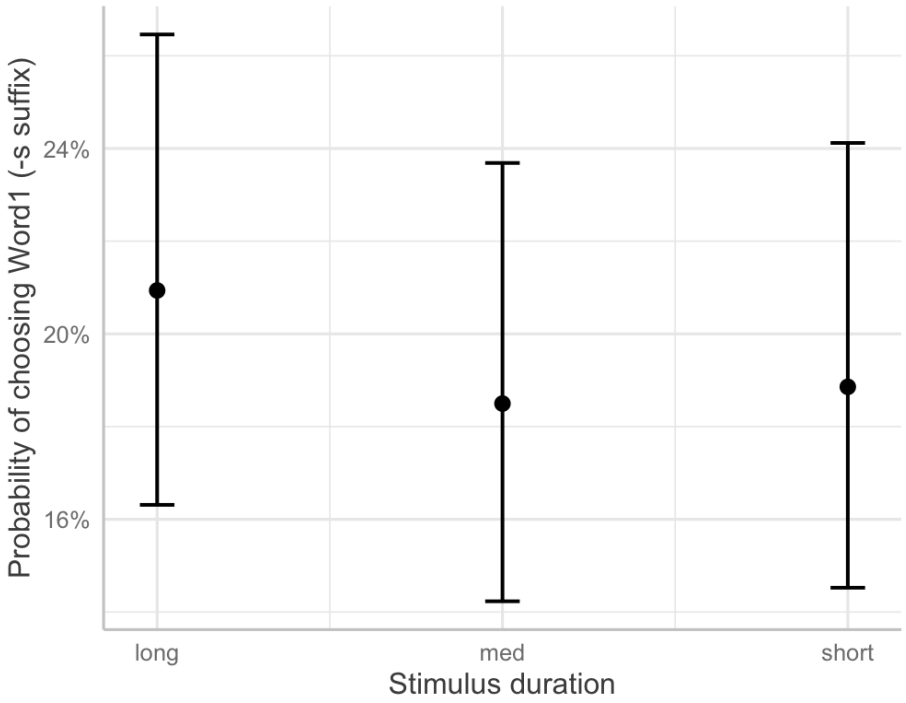

Figure 4.7: Probability of choosing Word 1 by stimulus duration, for *-s* suffixed/monomorphemic homophone pairs

effect of orthographic length was not significantly modulated by stimulus duration, meaning no evidence was found that participants associated greater stimulus durations with longer orthographic length. Figure 4.8 shows the relationship between orthographic length and stimulus duration for participant selections.

4.4 Discussion

4.4.1 Frequent word bias

The strongest and most pervasive effect found in this experiment is a general bias for listeners to select the more frequent word in a pair. Similar biases to default to frequent words are observed in various experimental paradigms. Most notably, in ambiguous phoneme perception, listeners are more likely to categorise ambiguous phonemes in order to form fre-

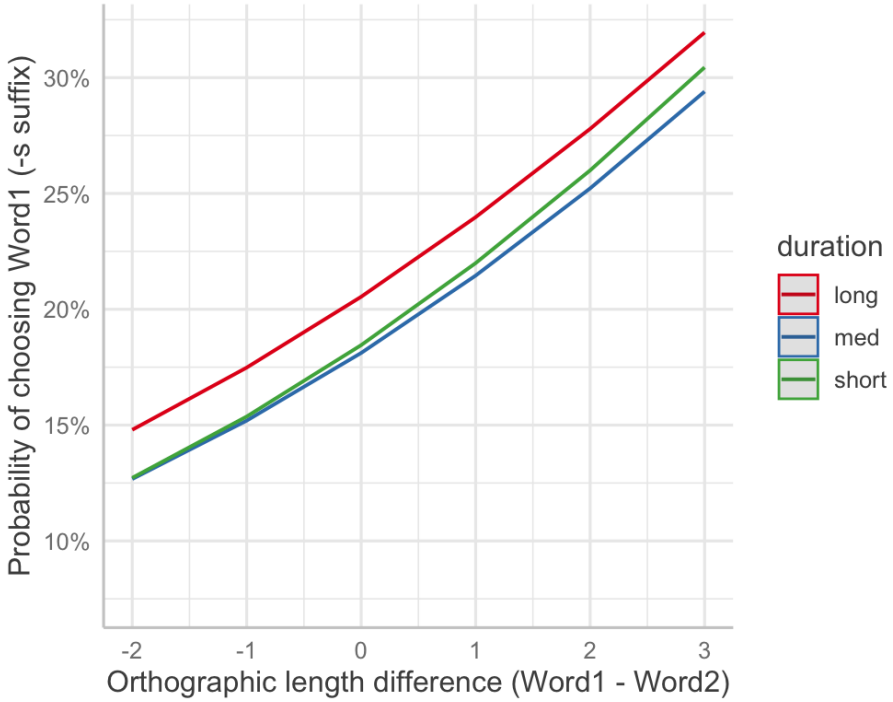

Figure 4.8: Probability of choosing Word 1 by (1) difference in orthographic length between Word 1 and Word 2 and (2) stimulus duration, for *-s* suffixed/monomorphemic homophone pairs

quent words than infrequent words. For example, given an onset that is ambiguous between /p/ and /b/, listeners are more likely to categorise it as /b/ if presented a choice between *best* and *pest* (i.e. they select *best*, the more frequent of the pair), and by the same token are more likely to categorise it as /p/ if presented a choice between *pot* and *bot* (Connine et al., 1993). An extreme version of this is commonly referred to as the 'Ganong effect'. This is when listeners perform phoneme categorisation by choosing between real words and non-words with an effective frequency of zero. Listeners are far more likely to select the real word than the non-word when the percept is phonetically ambiguous (Ganong, 1980; Fox, 1984).

The effect observed in this chapter is similar in that listeners tend to select the more frequent word to correspond with an ambiguous stimulus. However, it is different in that there is, in theory, no threshold of duration manipulation at which one homophone is no

longer a possible option. All auditory stimuli for the 126 homophone pairs were perfectly well-formed instances of either option. In other words, while manipulations of duration may perturb the frequent word bias, we can only observe this on a between-listener basis. There is no stimulus for which no listener chose the frequent word, and in all likelihood there is no manipulation of duration that would lead to this result.

4.4.2 Word frequency and duration

For monomorphemic homophone pairs and homophone pairs containing one *-ed* suffixed word, the effect whereby participants tended to choose the more frequent word is modulated by the duration of the auditory stimulus they heard. For these pair types, when participants heard a stimulus with a long duration they were slightly more willing to choose the less frequent word in a pair than when they heard a stimulus with a short duration. Even the results for homophone pairs containing one *-s* suffixed word trended in the same direction as this interaction, although it was not a significant effect. The upshot of this interaction between word frequency and stimulus duration is that participants appear to associate short durations with frequent words and long durations with infrequent words.

The association between short durations and frequent words, and between long durations and infrequent words, is the same general pattern that has been robustly observed in speech production (Wright, 1979; Kawamoto, 1999; Aylett and Turk, 2004; Gahl, 2008). It is interesting that listeners appear to be able to utilise this effect in perception for the purposes of homophone disambiguation. The most straightforward way to conceptualise this might be to invoke representations with intrinsic timing, such that different word durations more closely resemble the actual representation of different words. This is easily modelled in an Exemplar Theoretic framework. For example, the sequence [taɪm] with a long duration might more closely resemble a greater number of memory traces of the word *thyme* than the word *time*.

Alternatively, and if we account for frequency effects in speech production in terms of online mechanisms of lexical access, there is no reason to think that listeners couldn't

have tacit knowledge of various timing pressures experienced by speakers, and account for them in perception. Put differently, perhaps listeners understand that on average infrequent words are accessed and, in turn, produced more slowly than frequent words (Baese-Berk and Goldrick, 2009), and this can factor into their decision-making process for choosing between words with identical representations. There is already evidence to suggest that listeners compensate for certain speaker behaviours like coarticulation (Elman and McClelland, 1988).

The picture is slightly more complex if the primary mechanism at play in inducing frequency effects on duration in speech production is to accommodate the processing capacity of listeners (Lindblom, 1990; Aylett and Turk, 2004). If we consider the notion that speakers primarily produce infrequent words with long durations because they have tacit knowledge that these words take longer for listeners to process, it is interesting that listeners then seem to be able to use that effect for a different process than it was intended—to disambiguate between homophones. In a way, this parallels the morphological conditioning result on CSD perception in Chapter 3. Many explanations of the morphological conditioning on CSD production argue that deletion occurs at a lower rate in *-ed* suffixed forms because these suffixes are unpredictable and encode important morphosyntactic information. However, listeners appear to reflect an expectation of these different deletion rates in a way that suggests they are aware of them. It would not be unreasonable to think that the idea of speakers hyperarticulating infrequent words to accommodate listener processing time suggests that a hyperarticulated infrequent word and a hypoarticulated frequent word are perceptually equivalent in some sense. Words of different frequencies are tailored, in production, to the task of processing them, in perception. However, the results from this experiment suggest that listeners are aware of these production differences to some extent and may modulate their listening strategies to take them into account.

4.4.3 Morphological complexity and duration

In homophone pairs with one *-ed* suffixed word, participants were more likely to select this *-ed* suffixed word (Word 1) when they heard a stimulus with a long duration than when they heard a stimulus with a short duration. This parallels the observed effect whereby morphologically complex words are produced with some degree of lengthening around the morphological boundary (Walsh and Parker, 1983; Lociewicz, 1992; Schwarzlose and Bradlow, 2001; Sugahara and Turk, 2009; Seyfarth et al., 2018). As expected, this same effect does not obtain in monomorphemic pairs. However, it is unexpected that this same effect does not obtain in pairs with one *-s* suffixed word. Laboratory tests of morphological lengthening generally find more robust effects with *-s* suffixes than *-ed* suffixes (e.g. Sugahara and Turk, 2009; Seyfarth et al., 2018). However, there is some disagreement about whether *-s* suffixes are longer or shorter than tautomorphemic word-final /s/ based on corpus research (Song et al., 2013; Plag et al., 2017), which could cause some confusion or disagreement in listener responses. Alternatively, the fact that an association between morphology and duration was not found in the pairs with one *-s* suffixed word, as well as the general participant dispreference for *-s* suffixed words, could be a result of the fact that these pairs were mixed as to the dominant morphosyntactic interpretation of the suffix. That is, Word 1 in some of these pairs was most naturally interpreted as plural (e.g. *laps*), while in other words it was most naturally interpreted as 3SG agreement (e.g. *frees*). This mix may have led to difficulty on the part of some listeners in accessing an appropriate meaning in some cases, and the ultimate treatment of these words as non-words.

The demonstration of an association between duration and morphological complexity in *-ed* suffixed words, isolated from related effects of word frequency and orthographic length, is also an important piece of evidence for the story of English Coronal Stop Deletion (CSD). In Chapter 2 I showed that speakers exhibit systematic variation in tongue tip raising for coronal stops according to the morphological class of the word containing it. This tongue tip variation seems to stand out among other examples of morphology-sensitive phonetic variation, which seem to be almost exclusively durational in nature. However, I

also demonstrate that the same tongue tip duration is positively correlated with the duration of the coronal gesture and, in Chapter 3, that listeners do not exhibit evidence of directly perceiving this tongue tip variation. Together, these results suggest that the anomalous tongue tip variation according to morphological class (and indeed the morphological effect in CSD as it is traditionally analysed) may be a byproduct of variation in word duration. While morphological lengthening effects have been more spottily reported for *-ed* suffixes than *-s* suffixes, the present results reinforce that morphological lengthening associated with *-ed* suffixes is a real phenomenon and listeners have some knowledge of it in order to utilise it for homophone disambiguation. This reinforces the idea that lengthening in the production of *-ed* suffixed words may be harder to pin down because of the relative durational inflexibility of coronal stops compared to that of sibilants.

4.4.4 Orthographic length and duration

I predicted that participants might associate long stimulus durations with greater orthographic length and vice versa. There is some indication that monomorphemic pairs trend in this duration, with orthographically long Word 2s chosen slightly more frequently when paired with long stimuli than with short or medium stimuli. However, this is not a significant effect, and neither other pair type shows any sign of the same pattern. This is somewhat unexpected in light of results showing that orthographic length is associated with word duration in production (Warner et al., 2004; Brewer, 2008; Grippando, 2021), especially since listeners were presented with orthographic representations throughout the experiment, which is found to be a key factor influencing in the presence of orthographic effects in production (Roelofs, 2006).

The absence of an effect of orthographic length is particularly important for the pairs with one *-ed* suffixed word. This is because orthographic length is a potential confounding factor with morphological complexity. That is, there is no *-ed* suffixed word spelled with fewer letters than its homophonous monomorphemic counterpart. As such, a hypothetical tendency for participants to associate orthographic length with stimulus duration might

cast doubt on any observed association between stimulus duration and morphological complexity. However, since there is no such observed association of orthographic length and stimulus duration in this or either of the other pair types, we can be more confident that the association between duration and morphological complexity is real.

An unexpected main effect of orthographic length was observed in homophone pairs with one *-s* suffixed word. Participants were significantly more likely to choose the word with more letters in its orthographic representation, compared to a baseline probability of around 20% Word 1 (*-s* suffix) when both words were spelled with the same number of letters. No such main effect of orthography was observed in either other homophone pair type. This may be in part to do with issues in how much balancing is possible in the English language. Specifically, orthographic length difference in these pairs covaries loosely with word frequency difference. Out of eleven test pairs where Word 1 (*-s* suffix) is spelled with more letters than Word 2, six (more than half) have a Word 1 that is also more frequent than Word 2 according to the $\text{SUBTLEX}_{\text{US}}$ corpus. In contrast, out of thirteen test pairs where Word 1 (*-s* suffix) is spelled with fewer letters than Word 2, only four (less than a third) have a Word 1 that is also more frequent than Word 2. Moreover, the subset of pairs with one *-s* suffixed word, where that *-s* suffixed word is orthographically shorter than its monomorphemic counterpart, contains some of the words with principally 3SG interpretations that may be most challenging to parse in isolation (e.g. *chews*; *frees*; *sees*), and are correspondingly dispreferred by most participants.

Chapter 5

Morphologically-informed frequency as a predictor of variation

In Chapter 4, I presented evidence for an association between morphological boundaries at *-ed* suffixes and phonetic duration. This association is a potential source of explanation for the typologically unusual phonetic variation found in Chapter 2. That is, perhaps differences in the magnitude of coronal gestures are driven by differences in timing due to lengthening effects at morphological boundaries. In this chapter[1], I present an alternative source of explanation by presenting evidence that a large portion of Coronal Stop Deletion rates are accounted for by online pressures to optimise communicative efficiency. While previous accounts of Coronal Stop Deletion in these terms have been contentious, I show that when measures of word frequency are built to capture morphological structure they perform well as predictors of Coronal Stop Deletion rates. This suggests that we may be able to position an online pressure like word-end predictability as an intervening factor between morphology and phonetics to account for some of phonetic variation that otherwise looks exceptional.

5.1 Multiple measures of frequency

Taking on the topic of "word frequency" requires resolving the question of *which* frequency measures ought to be used, and how these measures relate to issues of theoretical concern. For example, in many psycholinguistic models the unit of lexical representation is the lemma, which contains the syntactic properties of a word and is shared by morphological relatives

[1]This chapter is based on collaborative work with Josef Fruehwald and Meredith Tamminga

with the same root (Roelofs, 1992). In such models, *jump* and *jumped* and *jumping* would all have the same lemma, with inflection added outside the lexicon. If frequency information is stored lexically and the lemma is the unit of lexical representation, then lemma frequency (i.e. the summed frequency of all words containing the *jump* lemma, or STEM FREQUENCY, as I call it here, might be more relevant to word recognition than surface whole-word frequency (where *jump*, *jumped*, and *jumping* would have distinct frequency values).

In sociophonetic research, it is more common to see the effect of WHOLE-WORD FREQUENCY on variable performance investigated. The use of whole-word frequency has theoretical underpinnings in more austere forms of Exemplar Theory which proposed that morphological abstraction was not a stored component of speakers' knowledge, but rather online analogisation of word-forms in an associative network (Bybee, 2002). As such, the whole word form is the most reliable linguistic unit on which to hang frequency estimates. There is also a methodological convenience to whole-word frequency: it is easily estimated from corpus data without the need for lemmatisation. The frequency of the word forms derived from the same corpus the data is drawn from has been argued by some to more accurately capture the localised and subjective experiences that speakers have with words and therefore word frequencies (Hay et al., 2015). While there is some controversy around the use of within-corpus versus corpus-external whole-word frequency estimates, adjudicating this issue is not a goal of this paper, and I will be using the whole-word frequency norms from $\text{SUBTLEX}_{\text{US}}$ (Brysbaert and New, 2009) (see §5.4).

Another possibility is that the mechanism by which frequency affects speech production is driven by the predictability of words. Higher frequency words are more predictable, and therefore may be subject to greater compression and reduction (Lindblom, 1963; Aylett and Turk, 2004; Turnbull, 2015) (see §5.2). While both lemma and whole-word frequency may contribute to the predictability of a word, so too may the *relative* frequency of a word form within its inflectional and derivational paradigm, which I call CONDITIONAL FREQUENCY. There are, of course, many other contextual factors over which predictability could be computed (both by language users and by researchers). Here I focus on the predictability

of the suffix (or lack thereof) given the stem because this is an area of active research in psycholinguistics whose connection to the literature on sociolinguistic variation is relatively underexplored (Kuperman et al., 2007; Cohen, 2015; Tomaschek et al., 2019). Another reason to focus on conditional frequency here rather than some other, simpler contextual predictability measures (such as probability given the previous or subsequent word) is that such measures have been investigated previously and not found to strongly predict outcomes in the variable I focus on in this study (Jurafsky et al., 1998, 2001).

This chapter is an investigation into how these three frequency measures (stem frequency, whole-word frequency, and conditional frequency) relate to patterns of Coronal Stop Deletion (CSD). This investigation is relevant in that it provides yet another potential avenue through which an effect of morphological structure on phonetic variation (of the type observed in Chapter 2) may be explained. Specifically, there is no theoretical requirement that online pressures for speakers to optimise communicative efficiency be limited to manipulations of specific phonetic parameters. If, as I argue, a significant portion of CSD patterns can be explained in terms of speakers producing more lenition where word endings are more predictable, we can sidestep some issues of modularity that have been implicated so far.

5.2 Frequency and variation in form

Frequency, specifically whole-word frequency, is associated with variation in phonetic and phonological form in many cases. In general, frequent whole-words tend to be pronounced faster, and in more lenited or reduced forms, than infrequent whole-words. This is relevant insofar as we conceive of CSD as an example of lenition, and we generally expect phonetic reduction and lenition to be intimately related to duration (Lindblom, 1963). However, in laboratory studies, evidence for the precise details of the relationship between duration and frequency is somewhat mixed. Wright (1979) claims that rare words are spoken as much as 24% slower than common words, but some subsequent studies have failed to replicate these effects between matching segments (Damian 2003; Mousikou et al. 2015; cf. Kawamoto

1999). Laboratory studies are also not entirely aligned in terms of how phonetic reduction and lenition are sensitive to frequency. In one articulatory study, Lin et al. (2014) find that tongue tip activity is generally reduced in highly frequent words, but Tomaschek et al. (2013, 2014) find that the magnitude of vowel gestures is highly sensitive to segmental context and may only be compressed for frequent words with phonologically short vowels. In contrast with these laboratory studies, corpus studies on more spontaneous speech reliably find that frequent whole-words are produced with shorter durations (Aylett and Turk, 2004; Gahl, 2008), and with more centralised vowels (Munson and Solomon, 2004) than infrequent whole-words.

Beyond gradient phonetic properties like duration, there exist a number of variables where the apparent rate of discrete variants[2] is correlated with lexical frequency. This is particularly well exemplified by work on varieties of Spanish. Highly frequent Spanish whole-words are more likely to exhibit intervocalic /d/ deletion (Bybee, 2002; Diaz-Campós and Gradoville, 2011), /r/ deletion (Diaz-Campós and Carmen, 2008), vowel coalescence (Alba, 2006), /s/ lenition and deletion (Brown and Cacoullos, 2003; Brown, 2009; File-Muriel, 2009), and less likely to feature /ʒ/ devoicing than infrequent whole-words. In English, too, schwa deletion (Hooper, 1976), yod retention (Phillips, 1981, 1984), and alveolar word-final *-in'* for the ING variable (Tamminga, 2016; Forrest, 2017), have all been found to be more common in frequent whole-words than infrequent whole-words. In a more general approach that is not limited to specific sociolinguistic variables, Turnbull (2018) compares phonological and phonetic transcriptions in corpora of English and Japanese and computes the segment deletions necessary between the underlying and surface forms. He finds that whole-word frequency (among other predictability measures) conditions the rate of segment deletion, and concludes that these patterns mirror those of phonetic reduction.

Investigations into the effect of frequency and other predictability measures on binary-coded CSD have had slightly more mixed results. While a handful of these studies do find that frequent whole-words have slightly higher rates of CSD than infrequent whole-words

[2]While they are categorised in discrete terms, for many of these variables the question of whether they arise in the phonetics or phonology is not settled.

(Bybee 2002; Jurafsky et al. 2001; Tamminga 2016, cf. Walker 2012), other studies report that whole-word frequency has an inconsistent effects across different subsets of data (Myers and Guy, 1997; Guy, 2019). Perhaps more striking is that contextual measures of word and biphone probability do that are typically good predictors of reduction do not seem to predict CSD outcomes (Jurafsky et al., 1998, 2001). Once again, CSD is positioned as something of an outlier with respect to other phonetic and phonological variables based on previous methods of measuring frequency and predictability more generally.

Outlying results notwithstanding, it seems generally true that frequent words are more susceptible to compression and 'weakening' of their pronunciations. Explanations for this kind of reduction phenomenon fall into three main theoretical camps (Clopper and Turnbull, 2018), two of which link production effects to robust results that frequent words are recognised more quickly and accurately in perception experiments (e.g. Howes, 1957; Savin, 1963; Connine, 1990; Dupoux and Mehler, 1990; Taft and Hambly, 1986). (1) 'Listener-oriented' accounts (e.g Lindblom, 1990; Aylett and Turk, 2004) explain production effects in terms of word predictability, to which I have already alluded, and the optimisation of the speech signal in order to maximise communicative efficacy while minimising effort. In other words, speakers use tacit knowledge that frequent words are easier to perceive and attenuate the articulatory effort spent on them. (2) For 'talker-oriented' accounts (e.g. Baese-Berk and Goldrick, 2009), frequency effects arise as part of the cognitive mechanisms of speech production. Just as in perception, infrequent word forms have a higher threshold for activation during production, and properties of timing and magnitude of activation during retrieval are passed on to properties of timing and articulation in the phonetic implementation. (3) Finally, there are 'passive' perspectives, in which word frequency directly shapes the mental representation of words, rather than creating on-line production pressures. A notable example of this kind of perspective is Exemplar Theory (Pierrehumbert, 2002; Bybee, 2002), in which a persistent leniting bias affects all words, but high frequency words—which are encountered most often—most quickly accumulate exemplars with compressed and 'weakened' pronunciations. While the frequency measures I discuss in more detail in §5.4.2 are

correlated with each other (e.g. a word with a high stem frequency is likely to also have a high whole-word frequency), each one is likely more indicative of one of these theoretical mechanisms being at play than the others. For example, an effect of stem frequency is more likely to be indicative of a talker-oriented account than a listener-oriented or a passive account. This is discussed in more detail below.

5.3 Morphology and frequency

I now turn to a brief examination of the relationship between frequency and morphological structure, with reference to both sociolinguistic and psycholinguistic results that highlight possible frequency–morphology interactions. There is already some reason to believe that frequency and morphological structure interact in how they condition CSD itself. Myers and Guy (1997) report, based on data from two Philadelphian speakers, that there is a robust effect of whole-word frequency among monomorphemes, but no such effect among *-ed* suffixed words. Similarly, Bayley (2014) finds a small effect of whole-word frequency that is limited to monomorphemes in San Antonio Chicano English CSD. Interactions between lexical frequency and grammatically-defined conditioning contexts in sociolinguistics have also been reported for morphosyntactic variables. Erker and Guy (2012) find that lexical frequency has an 'amplification' effect on the grammatical conditions influencing null subjects in Spanish: effects of verb regularity, verb semantics, subject person/number, and utterance tense/mood/aspect are small or nonexistent among low frequency verbs, but very significant among high frequency verbs. An interesting question I return to in my discussion in Section 5.6 is whether reported frequency/grammatical context interactions are the same kind of effect for CSD and null subjects.

The relevance of morphological structure for word processing has led to more widely recognised interactions in this domain. There is some evidence that morphologically complex words are generally recognised faster than monomorphemic words of equal length and frequency (Fiorentino and Poeppel, 2007), but highly frequent complex words are disadvantaged if the suffix is also highly frequent (Balling and Baayen, 2008). Moreover, it has

been suggested that the frequency-bearing unit most appropriate to capture variance in word recognition latencies depends on the morphological complexity of the word (Vannest et al., 2011). Morphologically complex words are recognised at speeds that vary according to lemma—or "base"—frequency, while monomorphemic words' recognition speeds are best accounted for with whole word—or "surface"—frequency.

In addition to basic frequency/morphology interactions in behavioral reaction times, there is also a growing body of work making inferences about what level of representation is active at a given point in the timecourse of spoken word recognition based on what kind of frequency measure correlates best with neural activity during processing. Specifically, a number of MEG studies find neurological activity to be most strongly correlated with measures of morphological structure, including lemma frequency and the transition probability between stem and suffix, at around 170ms (Solomyak and Marantz, 2009, 2010; Lewis et al., 2011; Zweig and Pylkkänen, 2009; Fruchter et al., 2013) and again at around 350ms (Solomyak and Marantz, 2009) following exposure to visual word stimuli. These results are taken as evidence for word recognition making reference to smaller morphological units, since these frequency measures associated with activation levels reflect the frequency of those sub-word units. While this literature has typically discussed these sub-lexical units in terms of decomposition (see also Embick et al. forthcoming on the nature of decomposition), I do not believe it is necessary to endorse a particular view on whether morphologically-complex words are decomposed *per se* in order to draw similar inferences about the relevance of sub-lexical structure for variation in morphologically-complex words. Even models that posit whole-word episodic storage in the lexicon allow for morphological relationships to emerge from patterns of phonetic and semantic overlap (Bybee, 2002); these relationships may in principle influence variable outcomes.

Among the studies that do apply this strategy of comparing frequency measures to explore the role of morphological structure in production, one interesting result that has emerged is evidence of 'paradigmatic enhancement' effects. These are also discussed in Chapter 1 in the context of other types of phonetic variation that are conditioned by mor-

phology. As well as the basic effect whereby frequent items are realised (and recognised) faster as a result of their predictability or ease of retrieval, some words with a high frequency compared to morphologically related words within the same paradigm are reinforced and pronounced with *less* phonetic reduction. An intuitive way to conceptualise this idea is in terms of speaker confidence, such that speakers are reassured that they are 'correct' in selecting the most relatively frequent form and do not hold back in production (Kuperman et al., 2007). Originally, paradigmatic enhancement was proposed to explain effects in 'pockets of uncertainty' between functionally equivalent forms that directly compete for use in the same position, like Dutch compound linking morphemes (Kuperman et al., 2007) and variable Russian agreement suffixes (Cohen, 2015). This 'pocket of uncertainty' aligns fairly closely with mainstream variationist conceptions of the linguistic variable, and indeed we see parallel results in the effect of variant frequency on variant duration in French variable schwa (Bürki et al., 2011). More recently, however, research on paradigmatic probability has been extended to explain variation in pronunciation across paradigmatically related words that are not in direct competition, with evidence for both the enhancement (Schuppler et al., 2012; Tucker et al., 2019; Tomaschek et al., 2019, 2021; Bell et al., 2021) and reduction (Hanique and Ernestus, 2011; Smith et al., 2012; Ben Hedia and Plag, 2017; Plag and Ben Hedia, 2017) of more relatively frequent forms. The present study represents, among other things, a contribution to this literature that may help reconcile these seemingly contradictory results.

5.4 Data and methods

5.4.1 Corpus and coding

For this chapter, data are taken from the Philadelphia Neighborhood Corpus of LING560 Studies (PNC) (Labov and Rosenfelder, 2011). This corpus is comprised of sociolinguistic interviews conducted by students in a graduate-level sociolinguistics course at the University of Pennsylvania. Recordings were made between 1973 and 2012, and generally last about an

hour. This study uses a sample of interviews from 118 white speakers found in working-class Irish-American and Italian-American neighborhoods. Speaker birth years span from 1888 to 1991, and the speakers are roughly balanced in terms of binary gender (66 women, 52 men). All interviews have been transcribed and had this transcription forced-aligned with the corresponding audio file using the FAVE suite (Rosenfelder et al., 2011).

While the articulatory evidence presented in Chapter 2 indicates that apparent CSD is often realised in the gradient phonetics, Chapter 3 suggests that the classic binary coding of CSD outcomes accesses something real about how CSD is evaluated. Moreover, just as the conditioning of the articulatory variation in Chapter 2 mirroring classic CSD conditioning gives us confidence that it is part of the same phenomenon, we can have confidence that binary CSD coding is a good working proxy the gradient phonetic patterns that give rise to it. For this reason, and to investigate production data on a larger scale, this chapter uses binary coding of CSD outcomes.

CSD outcomes were hand-coded according to auditory and spectrographic cues. A Praat script[3] was used to search for tokens and play a short corresponding excerpt for researcher evaluation. A number of decisions were made in order to restrict the dataset to straightforward cases that are consistently found to be eligible for CSD across its vast literature. Only words whose underlying forms end in coronal stops that are immediately preceded by consonants were considered. Instances of glottalisation and palatalisation were counted as /t,d/ retention[4]. Tokens preceding a stop, non-sibilant fricative, or affricate with a coronal place of articulation (i.e. /t,d,θ,ð,t͡ʃ,d͡ʒ/) were excluded, as well as tokens with both a preceding /n/ and following /s/. These contexts are particularly susceptible to processes that would neutralise the distinction between deleted and undeleted word-final coronal stops[5]. Words in which a final coronal stop was preceded by /r/ (e.g. *part*, *card*)

[3]Code available at `https://github.com/JoFrhwld/FAAV/blob/master/praat/handCoder.praat`

[4]This is the usual decision for CSD studies on American English. It has recently been suggested that British English glottal replacement of /t/ blocks CSD (Baranowski and Turton, 2020), but the exclusion of glottalised cases should only enhance the morphological effect since the contexts most favouring glottalisation (/nt#/, /lt#/) do not occur in *-ed* suffixed forms.

[5]For example, quasi-gemination across word boundaries makes it very difficult to distinguish between *last time* and *las' time*.

were excluded as it has been suggested that these stops are ineligible for deletion, at least in Philadelphia English (Cofer, 1972). The word *and* was excluded entirely, since it has been analysed as an exceptional case with multiple underlying representations (Neu, 1980; Guy, 2007). Irregular past forms (e.g. *kept*) and negative contraction forms (e.g. *wasn't*) were also excluded to focus on a more straightforward comparison between the most common morphological categories. In addition, I follow MacKenzie and Tamminga (ming) in further restricting the 'monomorphemic' category in this chapter to include only true monomorphemes, excluding superlative forms (e.g. *biggest*), agentive forms (e.g. *specialist*), and deverbal nominalised forms (e.g. *management*), among others. This brings the dataset in line with the monomorphemes in other chapters, which also do not include these multimorphemic word types. These methods yielded 8,912 word-final /t,d/ tokens, coded as belonging to monomorphemic (e.g. *act*) or regular past[6] (e.g. *jumped*) word forms.

5.4.2 Frequency measures

In concrete terms, the goal of this study is to evaluate how different frequency-related measures may be associated with variable CSD. This is an exploration into how possible communicative pressures like optimising the predictability of the signal affect CSD in a way that may resemble a direct effect of morphology. In particular, I will compare whether the frequency of the whole word, the frequency of some smaller constituent, or indeed the frequency relationship between the whole word and its component parts, predict CSD outcomes. To that end, I compare how well three different measures, calculated from values in the $\text{SUBTLEX}_{\text{US}}$ Corpus, account for variance in the CSD variable. These three measures do not exhaust all possible relationships between the frequency of different strings or units and CSD, but they do capture several distinct perspectives on how frequency measures might be relevant to the variable at hand.

My first such measure, whole-word frequency, is extracted from the FREQlow values in $\text{SUBTLEX}_{\text{US}}$: the raw number of times that a word appeared in the corpus in lower

[6]The 'regular past' category includes all preterite, perfect, and passive forms featuring an *-ed* suffix.

case. This measure, or a similar one, is the most widely used in linguistics, but it has some quirks. For example, in $\text{SUBTLEX}_{\text{US}}$, as in other corpora, frequency norms are calculated according to orthographic strings. This means that homographs have the same FREQlow value whether or not they are phonologically or morphologically related. However, whole-word frequency basically approximates the frequency of a surface phonological form. This measure was natural log-transformed and centred with the mean at zero.

I call my second measure stem frequency.[7] For this measure, I manually extracted and calculated the sum of all the whole-word frequencies for words that share the same stem as words in the data. I was careful to only add the frequency of the relevant parts of speech. For example, the calculation of the stem frequency for monomorphemic directional *left* does not include occurrences of verbal *left* or its morphological relatives such as *leftovers*. The stem frequency for monomorphemic *left* was calculated from its own, part-of-speech-corrected whole-word frequency, plus the whole-word frequencies for *lefty*, *lefties*, *lefts*, *leftist*, *leftists*, and *lefter*. This measure was also log-transformed and centred.

The third measure is conditional frequency. Conditional frequency is computed from the other two measures; the whole-word frequency is divided by the Stem frequency. Quantitatively speaking, conditional frequency is a proportion, bounded by 0 and 1. In other words, Conditional frequency approximates the frequency of a particular word among its morphological relatives.

5.4.3 Statistical modeling

The primary methodology used in this paper is comparison of mixed effects logistic regression models using the `lme4` package (Bates et al., 2015) in R (R Core Team, 2015). I set up a baseline model, which included fixed effects for following segmental context (pause; vowel; consonant eligible for stop resyllabification, e.g. /t#r/; or consonant ineligible for stop resyllabification, e.g. /t#l/, sum coded), grammatical class (monomorphemic versus regular

[7] Similar measures to my stem frequency measure have been called lemma frequency in previous literature. However, lemma frequency typically only includes inflectionally related words that share a stem. Since I count both inflectionally and derivationally related words that share a stem, I opted for a different name.

past, sum coded) and speech rate (vowels-per-second in a 7 word window, by-speaker z-score normalised), and a random intercept for speaker. I retained all these predictors in all subsequent models. Fixed effects for preceding phonological context could not be included without inducing a convergence error. From the baseline, I constructed models with all possible combinations of the three lexical frequency measures as fixed effects, including a model that included all three measures. I then performed paired likelihood ratio tests on nested models, and compard the AIC and BIC of each model. I rely on these global goodness-of-fit criteria as they are more robust to the multicollinearity between the frequency measures than coefficient estimates are.

5.5 Results

A central goal of this article is to compare multiple measures which are not only arithmetically related, but also attempt to capture similar (if not identical) aspects of how words are represented and processed. Therefore, before assessing the relative contributions of each of these frequency measures on CSD outcomes, we must explore the relationship between them. Figure 5.1 shows scatterplots indicating how words in each morphological class are distributed across the frequency measures, taken pairwise. Each plot, and the Pearson's correlation test results with which it is labelled, are generated from a 'dictionary' version of the data, with one entry for each unique word along with its values for each frequency measure according to $\text{SUBTLEX}_{\text{US}}$.

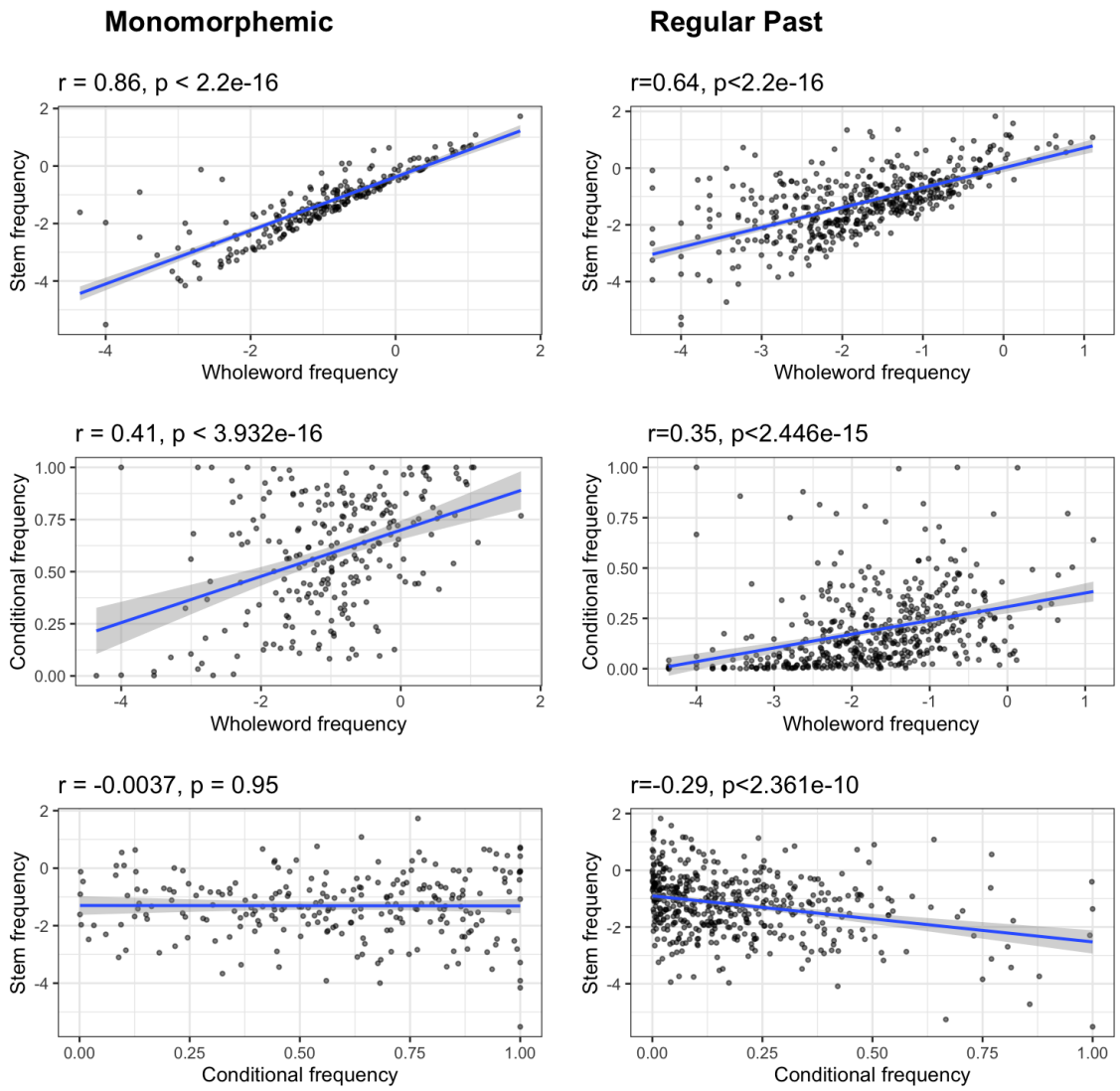

Figure 5.1: Relationship between frequency measures for monomorphemic (left) and *-ed* suffixed (right) words.

As is evident from Figure 5.1, monomorphemic words have quite different frequency properties to regular past *-ed* suffixed words. In both word types, there is a positive correlation between whole-word and stem frequency, and a hard border where a word's stem frequency must, by definition, be greater than or equal to its whole-word frequency. Each word's whole-word frequency value itself contributes to the stem frequency value,

along with the whole-word frequencies of morphologically related forms. This means that the stem frequency cannot be lower than the whole-word frequency. It is also linked to the positive correlation between whole-word and stem frequency, which is especially strong in monomorphemes. As whole-word frequency increases, the corresponding component part of stem frequency also increases. On one hand, this correlation means that it will be difficult to compare how well whole-word and stem frequency predict CSD outcomes, especially for monomorphemes. On the other hand, from a practical methodological perspective, it is useful to know that whole-word and stem frequency can, at least for monomorphemes, be used more or less interchangeably.

Monomorphemes and regular past forms differ in particular in their conditional frequency distributions. While monomorphemes are distributed fairly evenly, the majority of regular past forms have a very low conditional frequency. Reflecting on the properties of these word types, this might not be entirely unexpected. By definition, regular past forms are verbal, and implicate a whole paradigm of differently-inflected verb forms whose whole-word frequencies contribute to the stem frequency value. As a result, the regular past form often makes up only a small part of the stem frequency. On the other hand, the monomorphemic class includes words from a number of parts of speech that differ in the types of morphological relatives that occur. Since whole-word and stem frequency form the numerator and denominator in the calculation of conditional frequency, respectively, we can expect a positive relationship between conditional and whole-word frequency and a negative relationship between conditional and stem frequency. Sure enough, the directions of these relationships is borne out, but the correlations between conditional and stem frequency are far shallower than the other cases. In fact, a Pearson's correlation test finds no relationship between conditional and stem frequency for monomorphemes; the line is practically flat. These results parallel the non-correlation between the closely related measures of lemma frequency and "paradigmatic probability" found by Tomaschek et al. (2021).

The investigation of correlations between the different frequency measures gives us confidence that it is reasonable to include both conditional frequency and stem frequency as

predictors in a single model. Conversely, we should be wary of multicollinearity effects in models with other pairs of frequency measures. For the sake of completeness, I include all possible combinations of frequency measures in my model comparison analysis, but note that some improvements to model fit are likely to be artifacts of the relationship between measures.

5.5.1 First approach

In order to probe which frequency measure best captures variance in CSD, I compared a series of logistic regression models predicting CSD outcomes. The baseline model does not contain any frequency measures but does include the fixed effects for speech rate, grammatical class, and following segmental context, plus a random intercepts for speaker. The subsequent models add all possible combinations of the three frequency measures to this baseline model. I use likelihood ratio tests to assess whether each additional level of complexity (i.e. each additional frequency measure) was warranted as a significant improvement over the nested smaller models.[8].

In addition to likelihood ratio tests, each model's fixed AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) and log-likelihood statistics were recorded. While the log-likelihood is inevitably improved by adding additional complexity to a model, the AIC and BIC penalise model complexity at the same time as evaluating a model's ability to account for variance. This is especially true of the BIC, whose penalty for additional complexity is proportional to the number of observations, and frequently disagrees with the AIC in favour of a simpler model. Together, these information criteria provide the clearest evaluation of these models, indicating in particular where multiple frequency measures do not account for a sufficient amount of variance to justify their inclusion. Figure 5.2 shows the degree to which models with various combinations of frequency measures reduce the AIC and BIC, compared to a baseline model with no frequency measures.

Including each of the three frequency measures, individually, yields information criteria

[8]The full results of model comparison can be found in Appendices B and C

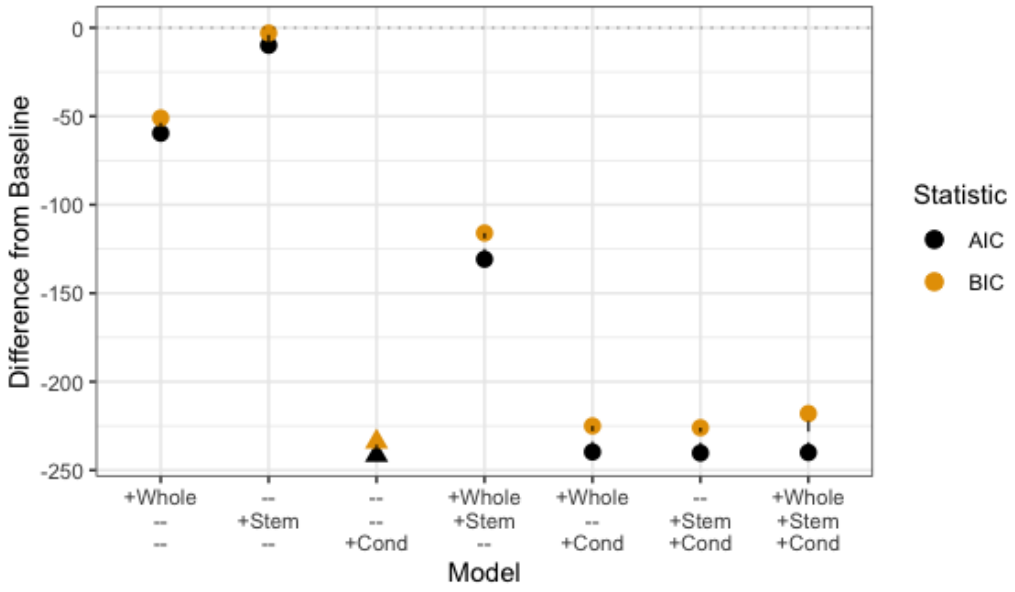

Figure 5.2: Information criteria reduction from baseline comparing models of full dataset (triangles = most reduced)

statistics that are somewhat reduced compared to the baseline model. This result is reinforced by significant likelihood ratio tests ($p<.001$) in each case. However, the reduction in both AIC and BIC that is attained from the addition of conditional frequency far outstrips that of the other measures. In fact, the addition of conditional frequency provides a large reductions in both AIC and BIC regardless of any other frequency measures already included in a model. The model comparison also suggests that the combination of stem frequency and whole-word frequency in a single model is a significant improvement over just one of these measures. However, we cannot rule out the possibility that this is an artifact of the strong correlation between these measures causing enhancement of their estimated effects. In addition, neither stem nor whole-word frequency significantly improves any model that already includes an effect of conditional frequency. This is demonstrated by likelihood ratio tests ($p>.05$), and the fact that these measures do not account for enough additional variance to counteract the penalty for model complexity that occurs in either the AIC or the BIC.

The initial model comparison results point to a need to reconsider how frequency is

accounted for in linguistic variation. In particular, the success of conditional frequency over other measures in terms of accounting for variance suggests that the interplay between word frequency and morphological structure within the lexicon is important and underexplored. Morphological structure is particularly relevant for a variable like coronal stop deletion, since it has repeatedly been reported that coronal stops at the end of monomorphemes are more likely to be deleted than coronal stops that constitute *-ed* suffixes (Guy, 1980, 1991b). This basic difference is controlled for with the main effect of grammatical category in each of the models in Figure 5.2. However, the effect of morphology may be more complicated still, as Figure 5.3 shows. In the top-left panel of the figure, I replicate previous findings that only monomorphemes are sensitive to whole-word frequency, and not regular past forms (Guy, 2019). This result strengthens my confidence in the interaction between frequency and morphological category for CSD, because compared to previous reports it is based on a significantly larger dataset with more narrowly defined morphological categories. Unsurprisingly, this interaction also holds for the closely-related stem frequency in the top-right panel. In addition to replicating previous reports for CSD, I note that these results resemble the "amplification" effect described by Erker and Guy (2012), in which the effect of grammatical class is stronger at high frequencies, and may not exist at all between low frequency words. However, in all of the CSD studies, including ours, the slope of the line for regular past forms does not significantly deviate from 0, suggesting that any apparent amplification of the morphological effect does not affect the morphological categories evenly, but rather is driven by differences between high- and low-frequency monomorphemes[9].

Compared to the results for whole-word and stem frequency, the results for conditional frequency are striking. Here, not only is there an effect for both monomorphemes and regular past forms, but the lines are almost parallel. This helps to explain why conditional frequency was so highly favored by the model comparison for the combined data in Figure 5.2 with sum-coded morphological categories. Furthermore, recall from Figure 5.1 that the relationship between conditional frequency and the other measures was weaker than

[9]Interactions of morphological class with wholeword frequency and stem frequency are fairly significant when they are added to models, but they are always heavily penalised in model comparison.

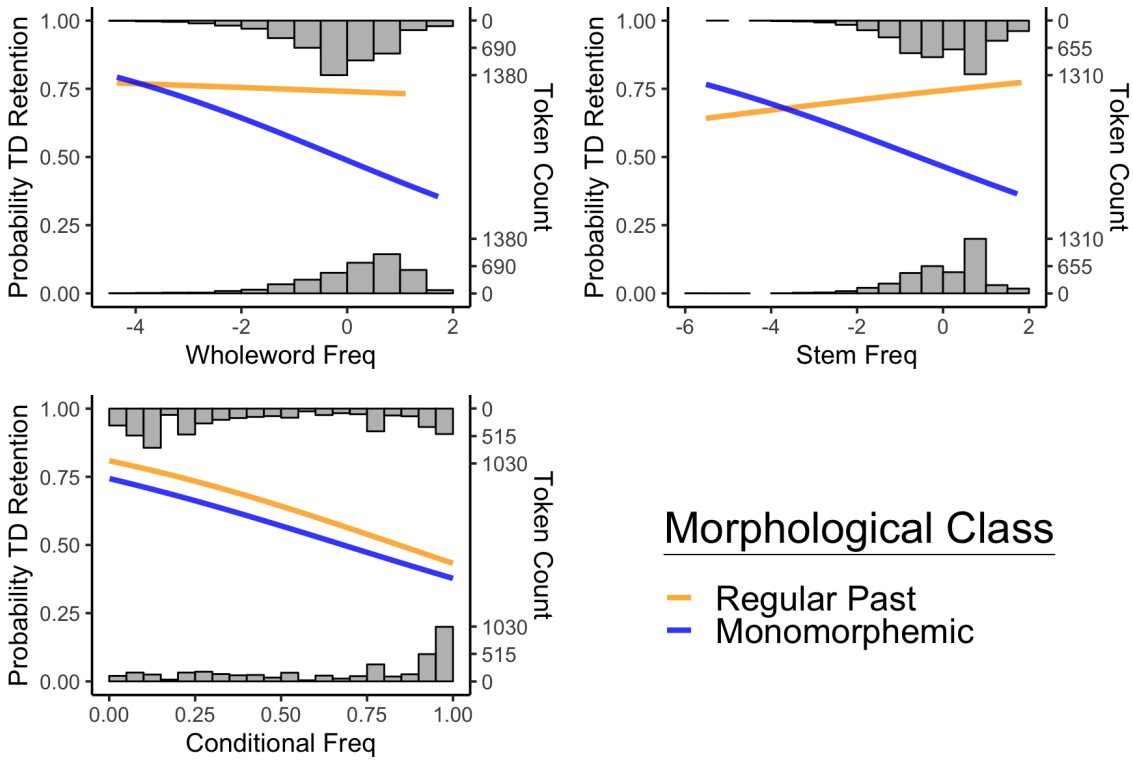

Figure 5.3: Observed CSD outcomes according to each frequency measure and morphological class.

the relationship between whole-word and stem frequency; the robust conditional frequency effect observed here accounts for a portion of the variance in CSD outcomes that is virtually untapped by controlling for just whole-word or stem frequency.

The differences in the effects of the frequency measures between morphological categories is not captured by the regression models I have been discussing, because they do not include any interaction terms targeting the non-independence of frequency and grammatical category. As a result, the best models I have presented so far, which combine regular past and monomorphemic words (sum-coded), will compromise between the two. In other words, a frequency measure that might be best for one group of words will be penalised if it is inappropriate for another. This raises questions about the performance of frequency measures within morphological categories, which are not addressed by the models I have presented so far. Therefore, in the following subsections I divide the data by morphological

class and test the different frequency predictors within each word type.

5.5.2 Monomorphemes

I begin by adopting the same method of model comparison as described for the full dataset, implemented over a subset of the data containing only monomorphemes. Once again, all models include fixed effects for speech rate and following segmental context, and a random intercept for speaker, but since all the words are monomorphemic no morphological category predictor is included.

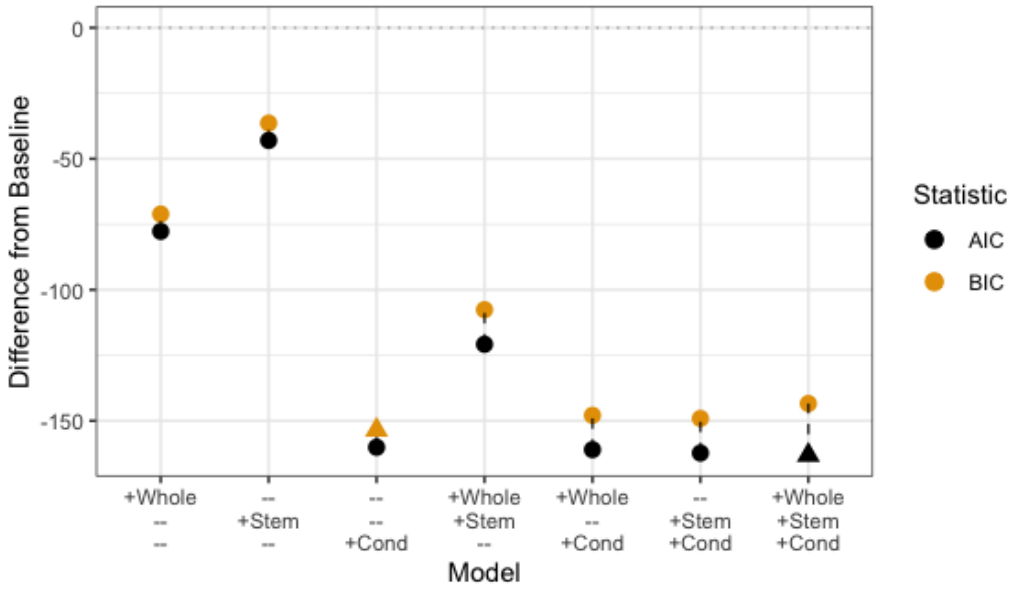

Figure 5.4: Information criteria reduction from baseline comparing models of monomorpheme subset (triangles = most reduced)

In Figure 5.4, we can see that the picture for monomorphemes alone is very similar. As in the models for the full dataset, all the frequency measures significantly improve model fit over the baseline when they are added individually, but conditional frequency outperforms the other measures and improves every model to which it is added. These results are reinforced by likelihood ratio tests ($p<.001$). In this case, the addition of stem frequency provides a slightly more obvious reduction in AIC and BIC than was observed for the full dataset, but it is still the smallest in magnitude out of the three frequency measures. Once

again, we see that the combination of stem and whole-word frequency outperforms either measure on its own, but this is very likely an artifact of the especially strong multicollinearity between these measures for monomorphemes.

In terms of the models that best reduce the information criteria, the results for the monomorpheme models are slightly less straightforward than for the full dataset in that the AIC and BIC disagree. Once again, the BIC is lowest for the model with just conditional frequency in addition to the baseline effects. However, the AIC is lower in the models containing at least one other frequency measure in addition to conditional frequency, and lowest in the model with all three measures. This suggests the other measures do capture enough variance in monomorphemes to outperform the relatively small penalty for additional model complexity that is applied in the computation of AIC. This seems especially true for stem frequency, which significantly improves the fit of every model it is added to according to likelihood ratio tests ($p<.05$). This includes all models with conditional frequency and/or whole-word frequency already present. In contrast, likelihood ratio tests do not show whole-word frequency to significantly improve models with conditional frequency already present. This is likely due, in large part, to the complete absence of a correlation between conditional and stem frequency for monomorphemes, such that they do not compete to account for the same variance.

5.5.3 Regular past

Just like for monomorphemes, I conducted the same method of model comparison for the regular past forms alone. Again, all models include a fixed effect of speech rate and following segmental context, a random intercepts for speaker. According to Figure 5.3, only conditional frequency appears to have the expected frequency effect for this group, but model comparison allows us to observe the interplay between the different measures when they are included in different combinations. The AIC and BIC values for each of the regular past models are plotted in Figure 5.5.

Unsurprisingly, conditional frequency once again introduces a large reduction in both

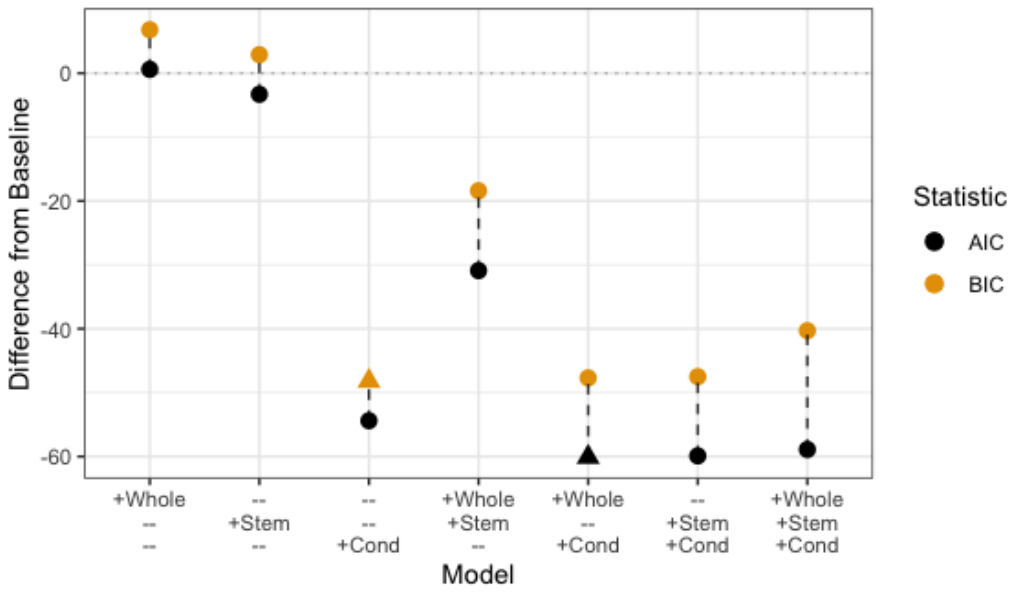

Figure 5.5: Information criteria reduction from baseline comparing models of complex form subset (triangles = most reduced)

the AIC and BIC of every model it is added to, as well as a significant improvement in terms of likelihood ratio tests ($p<.001$). Unlike for the full and monomorpheme datasets, not all of the frequency measures improve the baseline model when they are added individually. The addition of whole-word frequency does not account for enough variance to overcome the penalty for model complexity in either the AIC or BIC, and does not significantly improve model fit according to a likelihood ratio test ($p>.1$). Stem frequency, on the other hand, does marginally reduce the AIC and significantly improve model fit according to a likelihood ratio test ($p<.05$), but the magnitude of its improvement is still less than the penalty applied by the BIC for introducing additional complexity to the baseline model. Once again, the combination of both whole-word and stem frequency apparently reduces both the AIC and BIC by a fair amount compared to the baseline model. Even though the correlation between whole-word and stem frequency is weaker for regular past forms than for monomorphemes, it is still strong enough that this effect is likely to be an artifact of multicollinearity, especially given how poorly both whole-word and stem frequency perform individually.

Like for the monomorpheme models, the AIC and BIC disagree as to the optimal model for regular past forms. For the third time, the model with conditional frequency alone is favored by the BIC, and additional frequency measures are penalised for unnecessary complexity. However, this time, the AIC is minimised in the model with both conditional and whole-word frequency. This is despite the fact that whole-word frequency performed poorest when it was added to the baseline model individually, *and* the fact that it is more strongly correlated with conditional frequency than stem frequency is, for the regular past forms.

5.6 Discussion

There are two clear results presented in this chapter, each of which this section will discuss in turn. First, whole-word frequency (and to a lesser extent, stem frequency) is a significant predictor of CSD in monomorphemes but not in regular past tense forms. The direction of the effect within monomorphemes is as expected for reduction phenomena in general, with more CSD in higher-frequency whole-words. Second, both monomorphemes and past tense forms are highly sensitive to conditional frequency, again in the direction of more CSD with higher conditional frequency. Conditional frequency, therefore, has both a stronger and more pervasive across-the-board effect on CSD than the more familiar whole-word frequency measure. While this effect does not replace a main effect of morphological class on CSD outcomes, it does diminish it somewhat. In the following subsections, I discuss these results in light of some of their theoretical implications.

5.6.1 Interaction between whole-word/stem frequency and morphology

Whole-word frequency and stem (or 'base' or 'lemma') frequency are the measures of frequency most commonly incorporated into studies in contemporary sociolinguistics and psycholinguistics. CSD is no exception, and to my knowledge all existing investigations of word frequency effects in CSD are based on these measures. While a handful of these studies do find that frequent whole-words have slightly higher rates of CSD than infrequent whole-

words (Bybee 2002; Jurafsky et al. 2001; Tamminga 2016, cf. Walker 2012), other studies report that whole-word frequency has an inconsistent effects across different subsets of data (Myers and Guy, 1997; Guy, 2019).

For this data, it first of all turns out that whole-word and stem frequency are very highly correlated, and correspondingly predict extremely similar patterns of CSD across different subsets of the data. On the assumption that these frequency measures would also correlate this strongly throughout the lexicon (not just for CSD words), I offer the methodological recommendation that whole-word frequency, which is considerably more straightforward to implement than stem frequency, will be at least as effective as stem frequency for capturing frequency-related variance in other linguistic variables. In other words, for researchers who simply want to incorporate a reasonable frequency control into studies that are primarily aimed at investigating other phenomena, it will not be worth the effort to operationalise a stem frequency measure.

With regard to the specific pattern found for these two frequency measures, I observe a main effect of whole-word and stem frequency on CSD outcomes for the monomorphemes—coronal stops are more likely to be deleted at the end of frequent monomorphemes than infrequent monomorphemes—but not for regular past forms. An equivalent interaction between morphological category and whole-word frequency has also been reported for other CSD datasets (Myers and Guy, 1997; Bayley, 2014), but never at such a large scale, or corroborated with the same finding for stem frequency. In this dataset, as in these previous studies, the effect of frequency within monomorphemes is similar to that which has been observed for a number variable lenition and reduction phenomena, specifically that highly frequent and therefore highly predictable and/or highly practised words are pronounced with more reduced and lenited forms. However, additional explanation is required for why the same effect is not straightforwardly found for whole-word or stem frequency among regular past forms.

A potential avenue for explanation comes from Erker and Guy (2012), who report a similar interaction between whole-word frequency and grammatical category in the rate of

subject personal pronoun omission in Spanish. In their data, the effects of verb regularity, verb semantics, subject person/number, and utterance tense/mood/aspect are small or nonexistent among low (whole-word) frequency verbs, but large among high frequency verbs. Thus, whole-word frequency is taken to 'amplify' the effect of these grammatical categories. The proposed reason for this is that speakers need a certain amount of experience with a word in order for the effects of its grammatical category to be learned and reproduced, either as emergent from the particular contexts in which words of that category appear or as a more abstract property that entails a particular rate of some variant. This aligns with a 'passive' perspective on frequency effects in that it is the mental representation of words that is implicated, rather than any on-line mechanism. My results are also consistent with this idea of amplification: among high frequency words, the rate of CSD is far higher in monomorphemes than in regular past forms, but there is very little difference between low frequency monomorphemes and regular past forms in terms of rates of CSD.

On the other hand, a deficiency of the amplification story is that, at least for CSD, grammatical categories are treated more or less like arbitrary labels for words. In reality, monomorphemes and regular past forms differ in terms of morphological complexity, which may explain what we observe in terms of sensitivity (or lack thereof) to measures of frequency. Morphological complexity has two relevant properties as pertains to frequency. The first is that of informativity: while coronal stops at the end of monomorphemes are often highly predictable and contain no additional disambiguating information about the word, coronal stops at the end of regular past forms constitute a suffix that marks past tense. Moreover, when this suffix is deleted, regular past forms are always homophonous with a present or infinitival form of the verb. These are some of the primary concerns of linguists who ascribe a 'functional' motivation to grammatical patterns of CSD, arguing that deletion is avoided in cases where it would eliminate important past tense information (e.g. Kiparsky 1972).

The second relevant property of morphological complexity is that it entails pieces (whether independently-represented or emergent from shared phonology and semantics)

being shared across words. That is, not only does CSD target an informative suffix when it applies in regular past forms, it targets the *same* suffix identity for every lexical item in the grammatical category. Given that I am asking about the frequency of different linguistic units, I am forced to consider whether the relevant frequency measure for this kind of word might not be of the whole-word or the stem, but the *-ed* suffix itself. Of course, a raw measure of this kind would amount to a single (high) frequency value, and would not be particularly useful for explaining the basic effect of grammatical category, never mind differences between words within a single category. Therefore, instead of considering the frequency of a suffix overall, it may be more fruitful to consider the frequency of a suffix (or the absence of a suffix) under certain conditions. This is what is achieved by my conditional frequency measure.

5.6.2 Main effect of conditional frequency

What I have called 'conditional frequency' is the proportion of instances of a stem that are realised as a certain whole-word. Unlike for whole-word and stem frequency, I find strong effects of conditional frequency on predicting CSD outcomes in all of the regression models. This is in contrast to the small or mixed effects previously found for frequency effects on CSD. This suggests that we must consider morphological structure as a resource that language users may rely on for calculating the relative frequency of words and parts of words. Once we build this consideration into our measures, we find results that suggest a portion of the variation we see may be attributable to pressures like optimising communicative efficiency (with reference to morphological structure) rather than just a direct effect of morphology.

For regular past forms, conditional frequency corresponds to the decontextualised probability of the *-ed* suffix as an ending for a given stem. *-ed* suffixed forms that are common, relative to other reflexes of the same stem, are less likely to retain the coronal stop that marks this suffix. We can consider this result in light of the functionalist framing that is sometimes used to describe the main effect of grammatical category, which states that

coronal stops are less likely to be deleted when they encode important grammatical information, i.e. an *-ed* suffix. As previously mentioned, the functional analysis has a great deal in common with listener-oriented perspectives of reduction/lenition phenomena (Lindblom, 1990; Aylett and Turk, 2004; Jurafsky et al., 2001), which also describe the preservation of unpredictable and therefore informative structure, but the functional account makes specific reference to the grammatical information encoded by an *-ed* suffix rather than more general properties of word or segment probability. My findings show that even a functionally important coronal stop, representing the past tense suffix, is frequently deleted when that suffix is a highly paradigmatically frequent—and therefore highly predictable—ending to the stem.

While *-ed* suffixed words lend themselves to an intuitive interpretation of the conditional frequency measure, it may be more difficult to conceptualise a similar effect in monomorphemes. If the effect of conditional frequency is to be explained in terms of how predictable a suffix is given a stem, why would we see the same effect for no suffix at all? Indeed, as well as the frequency of an *-ed* suffix given a stem, conditional frequency in regular past forms is also equivalent to the frequency of an underlying coronal stop in this context. The conditional frequency of *kicked* is both the rate at which $\sqrt{\text{KICK}}$ is used in the past or passive and the rate at which /kɪk/ is followed by an underlying word-final /t/ with no intervening word boundary. In other words, both morphological and phonological levels of representation are captured with the same measure. Conversely, for the monomorphemes in this study, all or most of the words that are morphologically related to them also have underlying representations that contain the relevant coronal stop. The conditional frequency of *act* does not capture the rate at which a /t/ appears with the stem $\sqrt{\text{ACT}}$, because that /t/ is part of the reflex of the stem itself and therefore also occurs in *acts*, *acting* and *actor*. Instead, conditional frequency in monomorphemes only corresponds to the rate at which the stem occurs with a coronal stop in word-final position, as opposed to being followed by additional phonological material within the word. If we are to reconcile this with the predictability view that works well for *-ed* suffixed forms, we can think of the conditional frequency effect

in monomorphemes in terms of edge marking. Stems that do not commonly appear with a word-final coronal stop, relative to other possible forms where the stem is combined with various suffixes, are more likely to retain this coronal stop due to hyperarticulation at the word edge. In other words, stem-conditionally predictable word endings promote deletion, just as stem-conditionally predictable suffixes favor deletion.

The results from this chapter, that high conditional frequency corresponds to a high rate of coronal stop deletion, conflict with some recent findings of 'paradigmatic enhancement' effects. This is the class of results where the most common reflexes of a particular word or morpheme are found to be phonetically reinforced rather than reduced. These effects are framed from both speaker-oriented and passive perspectives. They are commonly interpreted in terms of speakers articulating common reflexes of a morpheme with increased confidence, suggesting an on-line pressure to reduce in cases where the speaker is unconfident. At the same time, speaker confidence itself has been explained as the result of extensive motor practice, allowing these words to be executed with enhanced kinematic skill (Tomaschek et al., 2018), suggesting an evolution of the specific representation or implementation associated with a word that is not generated on-line. However, comparison between paradigmatic enhancement findings and my own results is not straightforward. As I have already discussed in this section, the conditional frequency measure captures different facts about the coronal stops in monomorphemes versus *-ed* suffixed forms. The results in this chapter for monomorphemes lend themselves to a comparison with findings regarding the pronunciation of stem vowels with various suffixes (Tucker et al., 2019; Tomaschek et al., 2021), and perhaps even more pertinently with those concerning the pronunciation of consonants in the component pieces of compound nouns (Bell et al., 2021). All three of these studies find evidence of reinforcement when a stem or word is followed by a common ending; the present study finds the opposite. Instead, I show that monomorphemes whose stems typically occur in that form, with no additional suffix, are more likely to exhibit CSD than monomorphemes whose stems are more commonly suffixed.

In the case of *-ed* suffixed forms, my results can be more straightforwardly compared to

research on the pronunciation of affixes themselves in terms of their relationship to a given stem (Kuperman et al., 2007; Hanique and Ernestus, 2011; Smith et al., 2012; Schuppler et al., 2012; Cohen, 2015; Ben Hedia and Plag, 2017; Plag and Ben Hedia, 2017; Tomaschek et al., 2019). In these studies and in terms of the regular past forms in ours, the frequency of the affix itself, as attached to a given stem, is what is compared to the frequency of the same word/stem with other affixes or with no affix at all. While some studies of this type look at functionally equivalent affixes in direct competition (Kuperman et al., 2007; Cohen, 2015), the past tense form study in this chapter aligns with the many others comparing the frequency of one affix as an ending to a stem to the frequency of the whole paradigm (Hanique and Ernestus, 2011; Cohen, 2015; Tomaschek et al., 2019). Like these studies, using a different suffix in place of *-ed* will no longer denote the past tense. This means the conditional frequency of an *-ed* suffixed form does not capture the same 'pocket of uncertainty', where a language user could use more than one form to convey the same thing, that was considered to be so important in many early paradigmatic enhancement results. Correspondingly, the results in this chapter indicate more reduction when *-ed* is a frequent ending to a stem, aligning in particular with Hanique and Ernestus's (2011) result of greater reduction and deletion of word-final /t/ in Dutch irregular past verb forms when it is frequent within the paradigm, as opposed to Schuppler et al.'s (2012) result of greater word-final /t/ *retention* in Dutch 3SG present verb forms when this form is more frequently used than the 1SG form of the same verb. However, a similar pocket of uncertainty is surely to be found at every site of a sociolinguistic variable. Certainly, Bürki et al. (2011) find a comparable enhancement effect such that French variable schwa (e.g. *fenêtre* [f(ə)nɛtʀ] 'window') is longer in words that appear relatively more frequently with schwa compared to without it. In other words, even though I find no evidence of paradigmatic enhancement effects in this study, we might predict that future studies would find such effects corresponding to *variant frequency*, e.g. enhancement that is negatively correlated with the rate of CSD for a given word, such that more commonly retained stops have a reinforced pronunciation when they are retained. What my present results for conditional frequency do point towards is an

effect of suffix predictability, and a corresponding effect of word edge predictability, that are correlated with CSD rates. These effects are ultimately different reflexes of the same, listener-oriented, goal to signal that the listener should not expect another suffix.

5.7 Chapter Summary

I have interpreted conditional probability in terms of the predictability of either an *-ed* suffix (for morphologically complex CSD words) or a word boundary (for monomorphemes), given the stem. Under that interpretation, coronal stops are more likely to be retained when they are associated with word endings that have low stem-conditional predictability. The relatively high importance of conditional probability therefore suggests an important role for listener-oriented considerations in the explanation of the frequency/lenition relationship. However, the results presented in this chapter go beyond basic functional accounts that involve avoiding the omission of grammatical information by showing that even key grammatical information can be elided when it is highly predictable. At the same time, the robust interaction I find between whole-word and stem frequency measures and morphological category indicates that basic word predictability measures may be insufficient for cases where phonetic or phonological variation extends across morphological boundaries. It appears that, at least for a phenomenon like CSD, the predictability measures that matter most are ones that are relative to the internal structure of words and their morphological relatives. Exactly how speakers and listeners make predictions across word forms, and how far explanations appealing to the consequences of this kind of predictive behavior can take us in understanding pronunciation variation, remains to be seen.

As a larger point, I argue that these results should lead us to understand the different frequency measures as different in kind, capturing different mechanisms that may affect variability in speech production. More specifically, if I adopt the theoretical interpretations that I have already briefly suggested, in which lemma frequency approximates variable ease of lexical access, whole word frequency captures the long-term accumulation of reduction in the word form, and conditional frequency is a proxy for the variable predictability of

word forms, the finding of a strong and consistent influence of conditional frequency points to an important role for predictability in CSD. However, this interpretation also suggests that there is no one simple effect of "word frequency" that can be expected to have a uniform influence on different phenomena; in other words, this chapter's results should not be interpreted as showing that conditional frequency is the "correct" frequency measure to use in the study of variation across the board. Rather, I conclude that the question of how different frequency measures relate to any given phenomenon is an empirical one: different variable phenomena may turn out to be more or less sensitive to the different mechanisms or structural properties that these measures tap into. As a methodological issue, then, the selection of a frequency measure to use in quantitative analysis ought to be a considered one.

Chapter 6

Discussion and Conclusions

6.1 Major contributions

Classic and contemporary architectures of the grammar alike are often stratified in terms of which levels of structure interact with one another. This is typically conceptualised in terms of 'modularity': that different components of the grammar should be self-contained and strictly ordered. An area of ongoing debate about these architectures is centred on the phonetics and the phonology and the boundary between them, particularly as concerns their sensitivity to morphological structure. Even as we problematise a simple distinction between the phonology as an invariant, categorical, abstract representation and the phonetics as a variable, gradient, physical implementation, something must be preserved in the insensitivity of gradient parameters to morphosyntactic categories. This argument is sometimes made from a typological perspective; egregiously unmodular behaviours, like specific morphosyntactic categories (e.g. grammatical gender) with non-contrastive manipulations of phonetic parameters as primary exponents (e.g. slightly longer VOT), are not attested (Bermúdez-Otero, 2010). Even those phenomena where gradient phonetic parameters do seem to be influenced by morphology, such as lengthening at morphological boundaries and phonetic uniformity with other words in a morphological paradigm, are relatively rare. What's more, they are generally limited to manipulations of duration, rather than any other phonetic parameters.

In this dissertation I have aligned the literature on morphologically-sensitive phonetics with the literature from variationist sociolinguistics on morphologically-conditioned vari-

able phenomena. I pay special attention to English Coronal Stop Deletion, which appears to largely exist as a morphologically-conditioned phonetic phenomenon in terms of the magnitude of tongue movement. In Chapter 2 I demonstrate that English Coronal Stop Deletion is implemented with articulatory diverse strategies, but the majority of instances of inaudible stops do not look like categorical deletion of the coronal gesture. Moreover, the gradient phonetic measures of tongue tip raising (and height) show sensitivity to the morphological class of the containing word even when tokens with no tongue tip raising were removed from consideration. Coronal stops at the end of monomorphemes were articulated with a smaller magnitude of tongue tip raising than coronal stops constituting *-ed* suffixes. This reinforces the notion that binary acoustic-impressionistic evaluations of Coronal Stop Deletion belie, in large part, a more subtle manipulation of gradient phonetic parameters. This is because the findings in terms of magnitude of tongue tip raising parallel the robust observation throughout the variationist literature that coronal stops at the end of monomorphemes are more frequently deleted than coronal stops constituting *-ed* suffixes, patterns that are not reflected in the distribution of the handful of tokens where it does look like categorical deletion of the coronal gesture may have taken place. The discovery of widespread gradience in the implementation of Coronal Stop Deletion, even in terms of its morphological conditioning, creates new puzzles for the representation of the variable that I have attempted to address in the rest of the dissertation.

In Chapter 3 I explored listener knowledge about Coronal Stop Deletion in light of the new articulatory evidence, using audio from the articulatory procedure to create stimuli. This was at once an attempt to corroborate my own binary, acoustic-impressionistic categorisation of tokens and an exploration of the potential for perceiving gradient differences in the articulation and ultimately acquiring patterns of articulatory variation through imitation. In the end, listener ratings were distributed bimodally in a way that adhered fairly closely to my own binary judgments, suggesting that they primarily responded to cues in terms of canonical stop production. Moreover, I found no evidence that listeners ratings reflected the fine-grained articulatory parameters of tongue tip height or gesture

duration. This suggests that the fine-grained articulatory variation, if it is learned, must be learned indirectly. Listeners did, however, generally rate coronal stops at the end of monomorphemes to be clearer than those constituting *-ed* suffixes. This was unexpected since coronal stops at the end of monomorphemes are most frequently deleted according to traditional Coronal Stop Deletion analyses, and the monomorphemic stimuli used in this study were correspondingly contained coronal stops produced with lower tongue tip heights than those in *-ed* suffixed forms.

In Chapter 4 I explore duration as a perceptual cue to morphological complexity, building on findings of phonetic lengthening at morphological boundaries in production that have been more spottily reported for *-ed* suffixes than other suffixes. I presented listeners with artificially stretched and compressed stimuli and required them to choose between orthographic representations of homophones, controlling for differences in word frequency, orthographic length, and morphological complexity. As well as strong effects of frequency—a general bias to choose the orthographic representation for more frequent words and an increased willingness to choose infrequent words when presented with stimuli of long durations—there was an effect of morphological complexity for *-ed* suffixed words. Listeners were more likely to choose orthographic representations for *-ed* suffixed words (e.g. *packed*) over monomorphemic homophones (*pact*) when presented with long duration stimuli than when presented with short duration stimuli. This suggests that listeners do associate morphological complexity (in *-ed* suffixes) with phonetic lengthening, in line with some reports of similar results in production. This association may be the primary source of the difference in tongue tip raising magnitude found in the articulatory study, bringing Coronal Stop Deletion typologically in line with other morphologically-conditioned phonetic phenomena that primarily concern durational parameters. This is important because there exist compelling explanations for this kind of durational manipulation and its sensitivity, namely in terms of sublexical prosody or paradigm uniformity.

Finally, in Chapter 5, I take a different tack, and explore the potential for accounting for patterns in Coronal Stop Deletion with online processing mechanisms. This chapter

constitutes a large-scale corpus study based on acoustic-impressionistic data (reinforced by the finding in Chapter 3 that this type of data is relevant to the listener). In it, I explore the potential for different frequency measures, capturing different potential communicative pressures on speakers, to account for Coronal Stop Deletion outcomes. I find that while whole-word frequency and stem frequency measures are weakly related to rates of Coronal Stop Deletion in a monomorphemic subset of tokens, conditional frequency—P(whole-word|stem)—is a strong predictor of word-final coronal stop behaviour in both monomorphemic and *-ed* suffixed forms. This is borne out through extensive logistic regression model comparison, and amounts to new evidence for a morphologically-informed predictability account of Coronal Stop Deletion. That is, coronal stops are less likely to be deleted when they constitute suffixed (or word-edges) that are infrequent, and therefore less readily predictable, given the stem. This builds on contentious accounts of Coronal Stop Deletion as a 'functional' process that reflects speakers' desire to convey important grammatical information like the past tense marked by *-ed* suffixes. In reality, it looks like speakers' response to this kind of communicative pressure may be more complex and depend on the identity of a word and its morphological relatives. In this way, these findings bring Coronal Stop Deletion into line with another cluster of morphologically-sensitive phonetic phenomena whereby the phonetic forms of words are found to be affected by properties of the larger morphological paradigm to which the word belongs.

The results presented in this dissertation taken together, Coronal Stop Deletion does indeed behave, in large part, like a gradient phonetic variable that is sensitive to morphology. And, to a first approximation, the locus of this variation in production is typologically unusual compared to other such processes; it is found in the magnitude of tongue movements (which are sensitive to gesture duration) rather than directly in a durational parameter. However, perceptual results suggest that this articulatory detail is not salient to listeners, who instead show the expected association between morphological complexity and duration. Plus, new findings of frequency effects, when we account for morphological complexity within the frequency measure itself, give fresh support to an account of Coro-

nal Stop Deletion that is partly driven by online communicative pressures like optimising the speech signal in terms of predictability. Therefore, it seems that Coronal Stop Deletion may not be exceptional among morphologically-conditioned phonetic variables after all, and rather represents a confluence of prosodic and predictability effects at sites that are particularly striking examples of the non-linear relationship between articulation and acoustics (Stevens, 1972, 1989).

6.2 New and remaining puzzles

Rather than a straightforward roadmap towards solving a problem from one perspective, this dissertation tests the waters of multiple perspectives on the role of morphological structure in phonetic variation, through the lens English Coronal Stop Deletion. It is my hope that these initial steps into articulatory, perceptual, prosodic, and complex frequency analyses provide fertile ground for future research. Along with that, however, comes the concession that the findings of this dissertation generate at least as many questions as they answer. In this section, I select a few puzzles that have newly arisen or that present data falls short of addressing, that I see as particularly interesting and ripe for investigation. Fortunately, many of them can be directly addressed with the collection of more articulatory data.

6.2.1 Individual differences in production

While a key finding from Chapter 2 is that examples of apparent Coronal Stop Deletion that are *not* categorically implemented (in terms of a missing coronal gesture) are widespread, I also report tentative evidence for discrete categories in terms of the magnitude of tongue tip raising. These categories are only evident in the data from some speakers, and do not consistently align with the same conditioning factors that are associated with Coronal Stop Deletion. However, evidence for multimodality does complicate the conclusion that apparent Coronal Stop Deletion is a matter of the morphologically-conditioned variation of a gradient phonetic parameter.

On the other hand, if we are to treat what look like covert categories in the magnitude

of tongue tip raising as allophonic categories, it would mean that some speakers have postulated a variable but discrete alternation between slightly raised and fully raised coronal gestures to articulate coronal stops. If we assume that variation in the production of these coronal stops is perceived, evaluated, and acquired principally in terms of the presence or absence of canonical cues to stop closure and release (as the results of Chapter 3 point to), this kind of covert allophony is highly inefficient. That is because requires that speakers create a structure non-preserving target (e.g. $[t_{reduced}]$, where a structure preserving one (i.e. zero) would accomplish the same acoustic task (i.e. fewer audibly retained coronal stops) with a *higher* rate of success and with *less* muscular effort. However, Coronal Stop Deletion may be peculiar as a variable in exactly this way; so much of the detail of its implementation is covert, tongue behaviour that does not amount to complete constrictions and/or is masked by adjacent consonants within a cluster. As such, it is exactly the sort of microcosm where were might expect speakers to exhibit different representations and strategies for achieving equivalent surface results. Along similar lines, (de Jong, 1998) suggests that American English flapping takes a certain acoustic outcome as its output and speakers are otherwise fairly unconstrained in terms of implementation.

As a possible alternative, rather than covert allophony in tongue tip raising, we might actually be observing underlying zeroes but the tongue is then raised for other reasons. A potential cause of residual tongue tip raising is akin to the phonetic paradigm uniformity account of incomplete neutralisation. To recapitulate, this account says that phonologically neutralised contrasts differ because of the influence of morphologically-related words. For example, the word-final [t] in German *Rad* is produced with subtle evidence of voicing because it is related to the plural *Räder*, in which the equivalent stop is voiced (Roettger et al., 2014). Similarly, we could imagine that instances of categorical Coronal Stop Deletion would be influenced by related forms, specifically forms with undeleted coronal stops. This influence may lead to what looks like residual tongue tip raising. However, this account has the disadvantage of once again disrupting the typology of morphology-sensitive phonetics from affecting primarily durational parameters to inducing spatial articulatory differences,

which much of this dissertation is spent arguing may be illusory.

Another possibility for explaining clusters in the magnitude of tongue tip raising as equivalent to true zeroes relates to a central theme throughout this dissertation: timing. It seems increasingly apparent that timing differences are at the core of the mechanics of Coronal Stop Deletion. There is research showing that timing delays, especially to allow for additional speech planning time, are associated with additional articulatory movement in otherwise idle articulators (Heyward et al., 2014; Krivokapić et al., 2020). Therefore, perhaps slight tongue tip raising is just an indication that sufficient time is taken to warrant more than a zero. In other words, given enough time, the tongue tip has to do *something* and it might as well raise a little to a relaxed position. However, it seems unlikely that this kind of delay-induced posture would result in a stop closure, as some tokens among the reduced raising categories appear to have. Plus, more work is necessary to figure out how this potential effect of timing would interact with phonetic lengthening as it is associated with morphological complexity. Ultimately, more articulatory data is needed to flesh out the extent of what individual patterns are possible, as well as how they are distributed across speakers.

6.2.2 Morphology and spatiotemporal properties of articulation

In my review of phonetic phenomena that are sensitive to morphological structure, I have taken care to highlight what I see as a broad typological generalisation in terms of not only the what morphological structures can influence phonetic implementation, but what phonetic parameters seem to be responsive to them. Phonetic lengthening at morphological boundaries; incomplete neutralisation of obstruent voicing manifesting in preceding vowel length differences; enhancement and reduction of words and part-words according to properties of their morphological paradigm; all of these effects are realised chiefly in terms of timing and durational parameters. This sets up the larger framing of this dissertation, in which Coronal Stop Deletion is initially presented as an exceptional example of a morphologically-conditioned phonetic variable, since the key patterns of variation primarily

manifest in the 'spatial' magnitude of tongue tip raising rather than a 'temporal' phonetic parameter (Chapter 2). The exceptional status of the variable is then problematised by demonstrating listener non-sensitivity to the finer details of tongue tip height (Chapter 3) and a more reliable association between the presence of an *-ed* suffix and a 'temporal' parameter of word duration. That being said, the evidence that I have presented is not sufficient to conclusively demonstrate that timing is an important locus of Coronal Stop Deletion effects. To spell this out more explicitly: I show that the articulation of *-ed* suffixes (as opposed to coronal stops at the end of monomorphemes) is associated with large tongue tip raising motions to a high apex, and these large raising motions are associated with long gesture durations. At the same time, listeners show an association between the presence of an *-ed* suffix and longer word durations (compared to a monomorphemic homophone), paralleling lengthening effects reported in production. I do *not* find direct evidence that gesture duration is influenced by the morphological category in which coronal stops are articulated.

The absence of a straightforward morphological effect in the gesture duration of word-final coronal stops may be a simple matter of statistical power. This would align with previous speculations that phonetic lengthening may be more sparsely reported in the production of *-ed* suffixes than in *-s* suffixes because of the relative durational elasticity of sibilants compared to stops (Seyfarth et al., 2018). That is, stops are relatively short by nature, and a proportional modulation of their duration will be correspondingly small. Moreover, the minimum duration of a successful stop constriction is bounded in terms of the time required to build up sufficient air pressure. The most straightforward way to address this shortcoming is to expand the articulatory dataset to be larger and more varied in order to capture what may be a small but consistent effect.

In addition to this, the search for durational effects should be widened in scope from the investigation of a single gesture. In fact, if we take the suggestion that phonetic lengthening is induced by the presence of prosodic boundaries (or their presence in paradigmatically related words) seriously, the domain of lengthening ought radiate from the boundary itself.

Models of prosodic lengthening in Articulatory Phonology are particularly explicit about the properties of this domain, formalised as an abstract π-gesture centred on the boundary that slows the timecourse of all coactive constriction gestures (Krivokapić, 2020). Future articulatory investigations into morphological lengthening at *-ed* suffixes should correspondingly be directed at multiple simultaneous articulatory tiers, and comparing their behaviour to lengthening effects at stronger prosodic boundaries. This kind of investigation will necessarily engage more thoroughly with the phenomenon of gestural overlap, which this dissertation merely points out as a potential cause for the inaudibility of some coronal stops where the tongue tip is raised sufficiently high as to potentially form a complete constriction. That is, longer time windows not only allow for articulatory movements of larger magnitudes to be achieved (less undershoot), but for multiple sequential movements to be performed with less overlap. Therefore, longer time windows for movements of all articulators makes the classic overlap-induced inaudibility of coronal stops observed by (Browman and Goldstein, 1990) less likely. In this way, the potential lengthening of movements made by other articulators than the tongue tip at these *-ed* suffix boundaries is still highly relevant to understanding apparent Coronal Stop Deletion.

However, a closely related puzzle that is not so easily addressed in this way is that of /l/. While stems ending in /l/ have seemed generally impervious to previous investigations of lengthening at morphological boundaries (Sproat and Fujimura, 1993; Strycharczuk and Scobbie, 2016, 2017), /l/ before morphological boundaries is more likely to be dark (Sproat and Fujimura, 1993; Lee-Kim et al., 2013; Strycharczuk and Scobbie, 2016, 2017; Turton, 2017), and dark /l/ is produced with longer durations on average than light /l/ (Sproat and Fujimura, 1993). This situation resembles the one presented in the present dissertation in that a direct effect of morphology on duration seems hard to pin down, but otherwise duration-sensitive parameters do seem sensitive to morphological structure. Crucially, though, there is no reason for /l/ not to be elastic in its duration, and any lengthening effects should be far less subtle than for /t/ or /d/. In light of this, accounting for the absence of evidence of a morphological effect on the production of /t/, /d/, and /l/, may yet

require a more radical view on the relationship between the spatial and temporal properties of these articulatory gestures such that a reduction (or enhancement) could be realised in either one.

6.3 Final remarks

This dissertation constitutes a diverse set of studies probing the relationship between morphological structure and phonetic variation through the lens of Coronal Stop Deletion. While no single avenue of investigation is exhausted, the stage is set for continued research on multiple promising fronts. In particular, the findings from the articulatory study, while a compelling reassessment of the widespread assumption of categoricity in Coronal Stop Deletion, seem to only scratch the surface of possible patterns in implementation. This is in addition to further work triangulating the relationship between spatial and temporal phonetic parameters, in terms of morphologicall-conditioned variation. As techniques for collecting kinematic data become more accessible, the expansion of this kind of study should only grow more feasible.

Appendix A

Homophone pairs

Word1	Word2	Word1	Word2	Word1	Word2
heir	air	bear	bare	cent	scent
cast	caste	base	bass	dear	deer
sight	cite	cede	seed	faze	phase
hall	haul	cell	sell	gate	gait
horde	hoard	course	coarse	hair	hare
hoarse	horse	creak	creek	hear	here
meet	meat	due	dew	key	quay
might	mite	die	dye	knead	need
mousse	moose	flee	flea	maze	maize
wring	ring	great	grate	moat	mote
slay	sleigh	knight	night	pail	pale
stake	steak	pair	pear	pain	pane
tear	tier	plain	plane	peace	piece
thyme	time	row	roe	peek	peak
tyre	tire	seam	seem	pole	poll
vein	vain	suite	sweet	rain	reign
vile	vial	tale	tail	role	roll
wait	weight	waste	waist	sale	sail
waive	wave	weak	week	sole	soul
wrap	rap	yolk	yoke	steel	steal

Table A.1: Homophone pairs containing two monomorphemic words

Word1	**Word2**	**Word1**	**Word2**	**Word1**	**Word2**
banned	band	bawled	bald	picked	pict
bored	board	bowled	bold	brewed	brood
billed	build	charred	chard	crewed	crude
chased	chaste	ducked	duct	grayed	grade
missed	mist	foaled	fold	mowed	mode
packed	pact	guessed	guest	pried	pride
passed	past	mined	mind	rowed	road
riffed	rift	owed	ode	spayed	spade
sighed	side	paced	paste	swayed	suede
weighed	wade	tied	tide	tracked	tract
whirled	world	trussed	trust	wrapped	rapt

Table A.2: Homophone pairs containing one *-ed* suffixed and one monomorphemic word

Word1	**Word2**	**Word1**	**Word2**	**Word1**	**Word2**
frees	freeze	claws	clause	brews	bruise
boos	booze	crews	cruise	chews	choose
guys	guise	days	daze	grays	graze
hoes	hose	flecks	flex	links	lynx
knows	nose	laps	lapse	loos	lose
lacks	lax	locks	lox	paws	pause
mews	muse	rays	raise	quarts	quartz
packs	pax	pries	prize	sacks	sax
prays	praise	rues	ruse	sees	seize
pleas	please	ewes	use	tacks	tax
sighs	size	whacks	wax	teas	tease

Table A.3: Homophone pairs containing one *-s* suffixed and one monomorphemic word

Appendix B

Frequency measure model comparisons

		AIC	BIC	logLik	p
Baseline	-	10228.5	10278	-5107.2	—
Baseline	+ Whole-word	10170.3	10227	-5077.1	8.531e-15 ***
	+ Stem	10218.6	10275	-5101.3	0.0005554 ***
	+ Conditional	9987.0	10044	-4985.5	<2e-16 ***
Whole-word	+ Stem	10097.7	10162	-5039.9	< 2.2e-16 ***
	+ Conditional	9988.8	10053	-4985.4	<2.2e-16 ***
Stem	+ Whole-word	10097.7	10162	-5039.3	<2.2e-16 ***
	+ Conditional	9988.2	10052	-4985.1	<2.2e-16 ***
Conditional	+ Whole-word	9988.8	10053	-4985.4	.6509
	+ Stem	9988.2	10052	-4985.1	.3688
Whole-word, Stem	+ Conditional	9988.5	10060	-4984.3	<2.2e-16 ***
Whole-word, Conditional	+ Stem	9988.5	10060	-4984.3	0.1304
Stem, Conditional	+ Whole-word	9988.5	10060	-4984.3	.1942470

Table B.1: Comparison of full dataset nested mixed effects logistic regression models for CSD.

		AIC	BIC	logLik	p
Baseline	-	6686.9	6726.2	-3337.4	—
Baseline	+ whole-word	6609.2	6655.1	-3297.6	<2.2e-16 ***
	+ Stem	6643.9	6689.8	-3315.0	1.146e-11 ***
	+ Conditional	6526.8	6572.7	-3256.4	<2e-16 ***
Whole-word	+ Stem	6566.1	6618.6	-3275.0	1.887e-11 ***
	+ Conditional	6525.8	6578.3	-3254.9	<2.2e-16 ***
Stem	+ Whole-word	6566.1	6618.6	-3275.0	<2.2e-16 ***
	+ Conditional	6524.6	6577.1	-3254.3	<2.2e-16 ***
Conditional	+ Whole-word	6525.8	6578.3	-3254.9	0.08174 .
	+ Stem	6524.6	6577.1	-3254.3	0.03937 *
Whole-word, Stem	+ Conditional	6523.8	6582.8	-3252.9	2.856e-11 ***
Whole-word, Conditional	+ Stem	6523.8	6582.8	-3252.9	0.0459 *
Stem, Conditional	+ Whole-word	6523.8	6582.8	-3252.9	0.09 .

Table B.2: Comparison of nested mixed effects logistic regression models for CSD in monomorphemes.

Model1	Model2	AIC	BIC	logLik	p
Baseline	-	3381.1	3418.4	-1684.5	—
Baseline	+ Whole-word	3381.7	3425.2	-1683.9	0.2433
	+ Stem	3377.8	3421.3	-1681.9	0.02124 *
	+ Conditional	3326.7	3370.2	-1656.2	8.602e-16 ***
Whole-word	+ Stem	3350.2	3400.0	-1667.1	7.149-09 ***
	+ Conditional	3321.0	3370.7	-1652.5	2.318e-15 ***
Stem	+ Whole-word	3350.2	3400.0	-1667.1	5.453e-08 ***
	+ Conditional	3321.2	3370.9	-1652.6	1.956e-14 ***
Conditional	+ Whole-word	3321.0	3370.7	-1652.5	0.5913
	+ Stem	3321.2	3370.9	-1652.6	0.8509
Whole-word, Stem	+ Conditional	3322.2	3378.1	-1652.1	4.252e-08 ***
Whole-word, Conditional	+ Stem	3322.2	3378.1	-1652.1	0.3864
Stem, Conditional	+ Whole-word	3322.2	3378.1	-1652.1	0.31652

Table B.3: Comparison of nested mixed effects logistic regression models for CSD in regular past forms.

Appendix C

Frequency model summaries

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	-0.60607	0.06492	-9.335	<2e-16
Morph: Complex	0.59897	0.02591	23.113	<2e-16
Speech rate	-0.22461	0.02690	-8.348	<2e-16
Following syllabifiable C	0.72050	0.08996	8.009	1.16e-15
Following pause	1.27788	0.06899	18.523	<2e-16
Following vowel	1.8560	0.06429	29.017	<2e-16

Table C.1: CSD(full_dataset) ~ Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	-0.62557	0.06516	-9.601	<2e-16
Morph: Complex	0.51650	0.02792	18.497	<2e-16
Speech rate	-0.21471	0.02702	-7.947	1.91e-15
Following syllabifiable C	0.71628	0.09047	7.917	2.43e-15
Following pause	1.25977	0.06926	18.190	<2e-16
Following vowel	1.87434	0.06455	29.038	<2e-16
Wholeword freq	-0.22067	0.02875	-7.674	1.66e-14

Table C.2: CSD(full_dataset) ~ WholewordFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	-0.60076	0.06598	-9.245	<2e-16
Morph: Complex	0.60084	0.02595	23.158	<2e-16
Speech rate	-0.22039	0.02694	-8.180	2.83e-16
Following syllabifiable C	0.71560	0.09010	7.942	1.98e-15
Following pause	1.26411	0.06912	18.288	<2e-16
Following vowel	1.86650	0.06433	29.014	<2e-16
Stem freq	-0.08880	0.02579	-3.443	0.000574

Table C.3: CSD(full_dataset) ∼ StemFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	0.24090	0.08455	2.849	0.00438
Morph: Complex	0.11086	0.04073	2.722	0.00649
Speech rate	-0.21394	0.02736	-7.820	5.26e-15
Following syllabifiable C	0.80341	0.09177	8.755	<2e-15
Following pause	1.34206	0.07049	19.040	<2e-16
Following vowel	1.91623	0.06566	29.184	<2e-16
Conditional freq	-1.79569	0.11698	-15.351	<2e-16

Table C.4: CSD(full_dataset) ∼ ConditionalFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	-0.70110	0.06579	-10.656	<2e-16
Morph: Complex	0.29980	0.03778	7.935	2.11e-15
Speech rate	-0.21552	0.02716	-7.936	2.09e-15
Following syllabifiable C	0.74117	0.09113	8.113	4.19e-16
Following pause	1.30009	0.06984	18.615	<2e-16
Following vowel	1.88953	0.06497	29.082	<2e-16
Wholeword freq	-0.78662	0.07516	-10.466	<2e-16
Stem freq	0.55723	0.06669	8.355	<2e-16

Table C.5: CSD(full_dataset) ~ WholewordFreq + StemFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	0.22636	0.09046	2.502	0.01234
Morph: Complex	0.11275	0.04095	2.754	0.00589
Speech rate	-0.21344	0.02738	-7.796	6.38e-15
Following syllabifiable C	0.80202	0.09184	8.733	<2e-16
Following pause	1.34003	0.07063	18.974	<2e-16
Following vowel	1.91615	0.06566	29.184	<2e-16
Wholeword freq	-0.01465	0.03236	-0.453	0.65086
Conditional freq	-1.76816	0.13181	-13.414	<2e-16

Table C.6: CSD(full_dataset) ~ WholewordFreq + ConditionalFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	0.23376	0.08493	2.753	0.00591
Morph: Complex	0.11546	0.04107	2.811	0.00494
Speech rate	-0.21296	0.02738	-7.778	7.34e-15
Following syllabifiable C	0.80153	0.09182	8.729	<2e-16
Following pause	1.33802	0.07063	18.945	<2e-16
Following vowel	1.91617	0.06566	29.184	<2e-16
Stem freq	-0.02362	0.02628	-0.899	0.36887
Conditional freq	-1.77914	0.11844	-15.022	<2e-16

Table C.7: CSD(full_dataset) ∼ StemFreq + ConditionalFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	0.33606	0.11565	2.906	0.00366
Morph: Complex	0.11828	0.04117	2.873	0.00407
Speech rate	-0.21318	0.02738	-7.785	6.98e-15
Following syllabifiable C	0.80616	0.09186	8.776	<2e-16
Following pause	1.33893	0.07064	18.953	<2e-16
Following vowel	1.91669	0.06567	29.185	<2e-16
Wholeword freq	0.13408	0.10250	1.308	0.19082
Stem freq	-0.12721	0.08343	-1.525	0.12731
Conditional freq	-1.95907	0.18205	-10.761	<2e-16

Table C.8: CSD(full_dataset) ∼ WholewordFreq + StemFreq + ConditionalFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	-1.03980	0.07714	-13.479	<2e-16
Speech rate	-0.16248	0.03267	-4.974	6.56e-07
Following syllabifiable C	0.75293	0.11077	6.797	1.07e-11
Following pause	1.26840	0.08515	14.895	<2e-16
Following vowel	1.37874	0.08041	17.147	<2e-16

Table C.9: CSD(monomorphemes) $\sim$ Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	-0.95555	0.07775	-23.289	<2e-16
Speech rate	-0.14845	0.03301	-4.497	6.56e-07
Following syllabifiable C	0.75293	0.11077	6.797	1.07e-11
Following pause	1.26840	0.08515	14.895	<2e-16
Following vowel	1.37874	0.08041	17.147	<2e-16
Wholeword freq	-0.31437	0.03575	-8.793	<2e-16

Table C.10: CSD(monomorphemes) $\sim$ WholewordFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	-1.03736	0.07735	-13.411	<2e-16
Speech rate	-0.15169	0.03290	-4.611	4.01e-06
Following syllabifiable C	0.76024	0.11132	6.830	8.52e-12
Following pause	1.24971	0.08554	14.609	<2e-16
Following vowel	1.38662	0.08080	17.160	<2e-16
Stem freq	-0.22390	0.03369	-6.646	3.01e-11

Table C.11: CSD(monomorphemes) $\sim$ StemFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	0.20066	0.12376	1.621	0.105
Speech rate	-0.15207	0.03319	-4.583	4.59e-06
Following syllabifiable C	0.81544	0.11278	7.230	4.83e-13
Following pause	1.30481	0.08692	15.011	<2e-16
Following vowel	1.42259	0.08203	17.342	<2e-16
Conditional freq	-1.64958	0.13225	-12.473	<2e-16

Table C.12: CSD(monomorphemes) ~ ConditionalFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	-0.72285	0.08574	-8.431	<2e-16
Speech rate	-0.15040	0.03311	-4.543	5.55e-06
Following syllabifiable C	0.79411	0.11250	7.059	1.68e-12
Following pause	1.28824	0.08656	14.882	<2e-16
Following vowel	1.42217	0.08170	17.408	<2e-16
Wholeword freq	-1.19555	0.14509	-8.240	<2e-16
Stem freq	-0.85818	0.13416	6.397	1.59e-10

Table C.13: CSD(monomorphemes) ~ WholewordFreq + StemFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	0.09724	0.13738	0.708	0.4790
Speech rate	-0.14968	0.03324	-4.503	6.69e-06
Following syllabifiable C	0.81505	0.11290	7.219	5.23e-13
Following pause	1.29881	0.08702	14.925	<2e-16
Following vowel	1.42386	0.08208	17.347	<2e-16
Wholeword freq	-0.07617	0.04380	-1.739	0.0821
Conditional freq	-1.48591	0.16213	-9.165	<2e-16

Table C.14: CSD(monomorphemes) ∼ WholewordFreq + ConditionalFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	0.11919	0.12997	0.917	0.3591
Speech rate	-0.14920	0.03325	-4.488	7.20e-06
Following syllabifiable C	0.81549	0.11292	7.222	5.13e-13
Following pause	1.29720	0.08704	14.903	<2e-16
Following vowel	1.42328	0.08209	17.338	<2e-16
Stem freq	-0.07440	0.03615	-2.058	0.0396
Conditional freq	-1.54172	0.14209	-10.850	<2e-16

Table C.15: CSD(monomorphemes) ∼ StemFreq + ConditionalFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	0.31433	0.17455	1.801	0.0717
Speech rate	-0.14921	0.03326	-4.486	7.25e-06
Following syllabifiable C	0.81798	0.11293	7.243	4.39e-13
Following pause	1.29576	0.08706	14.883	<2e-16
Following vowel	1.41951	0.08214	17.282	<2e-16
Wholeword freq	0.41031	0.24350	1.685	0.0920
Stem freq	-0.40814	0.20136	-2.027	0.0427
Conditional freq	-1.94377	0.27899	-6.967	3.24e-12

Table C.16: CSD(monomorphemes) ~ WholewordFreq + StemFreq + ConditionalFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	-0.24768	0.09665	-2.563	0.0104
Speech rate	-0.33780	0.05010	-6.742	1.56e-11
Following syllabifiable C	0.58247	0.14969	3.891	9.98e-05
Following pause	1.21267	0.11702	10.363	<2e-16
Following vowel	2.86196	0.12134	23.587	<2e-16

Table C.17: CSD(complex_forms) ~ Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	-0.27508	0.09971	-2.759	0.005799
Speech rate	-0.33416	0.05018	-6.659	2.76e-11
Following syllabifiable C	0.57476	0.14992	3.834	0.000126
Following pause	1.20554	0.11721	10.285	<2e-16
Following vowel	2.86181	0.12135	23.584	<2e-16
Wholeword freq	-0.05809	0.04982	-1.166	0.243553

Table C.18: CSD(complex_forms) ~ WholewordFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	-0.25522	0.09658	-2.642	0.00823
Speech rate	-0.34461	0.05032	-6.849	7.44e-12
Following syllabifiable C	0.60042	0.15005	4.001	6.29e-05
Following pause	1.23697	0.11772	10.508	<2e-16
Following vowel	2.86707	0.12153	23.591	<2e-16
Stem freq	0.09845	0.04263	1.948	0.05139

Table C.19: CSD(complex_forms) ~ StemFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	0.16219	0.10981	1.477	0.14
Speech rate	-0.32613	0.05082	-6.417	1.39e-10
Following syllabifiable C	0.69643	0.15341	4.540	5.64e-06
Following pause	1.31982	0.11974	11.022	<2e-16
Following vowel	2.90578	0.12290	23.644	<2e-16
Conditional freq	-2.08971	0.26174	-7.984	1.42e-15

Table C.20: CSD(complex_forms) ~ ConditionalFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	-0.51170	0.10923	-4.685	2.80e-06
Speech rate	-0.33705	0.05064	-6.656	2.82e-11
Following syllabifiable C	0.59772	0.15145	3.947	7.93e-05
Following pause	1.26182	0.11875	10.626	<2e-16
Following vowel	2.88155	0.12218	23.584	<2e-16
Wholeword freq	-0.49199	0.09363	-5.255	1.48e-07
Stem freq	0.44837	0.07928	5.655	1.55e-08

Table C.21: CSD(complex_forms) ~ WholewordFreq + StemFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	0.18119	0.11530	1.571	0.116
Speech rate	-0.32766	0.05092	-6.435	1.24e-10
Following syllabifiable C	0.70205	0.15378	4.565	4.99e-06
Following pause	1.32490	0.12014	11.028	<2e-16
Following vowel	2.90667	0.12293	23.644	<2e-16
Wholeword freq	0.02782	0.05168	0.538	0.590
Conditional freq	-2.12083	0.26826	-7.906	2.66e-15

Table C.22: CSD(complex_forms) ~ WholewordFreq + ConditionalFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	0.158841	0.111220	1.428	0.153
Speech rate	-0.326753	0.050936	-6.415	1.41e-10
Following syllabifiable C	0.697289	0.153481	4.543	5.54e-06
Following pause	1.321277	0.120002	11.010	<2e-16
Following vowel	2.906003	0.122912	23.643	<2e-16
Stem freq	0.008511	0.045145	0.189	0.850
Conditional freq	-2.076236	0.271251	-7.654	1.94e-14

Table C.23: CSD(complex_forms) ~ StemFreq + ConditionalFreq + Morph + SpeechRate + Following + (1|Speaker)

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	0.30253	0.18096	1.672	0.0946
Speech rate	-0.32604	0.05094	-6.400	1.56e-10
Following syllabifiable C	0.71425	0.15448	4.624	3.77e-06
Following pause	1.32752	0.12016	11.048	<2e-16
Following vowel	2.90746	0.12293	23.650	<2e-16
Wholeword freq	0.14327	0.14230	1.007	0.3140
Stem freq	-0.10817	0.12440	-0.870	0.3846
Conditional freq	-2.42078	0.43740	-5.534	3.12e-08

Table C.24: CSD(complex_forms) $\sim$ WholewordFreq + StemFreq + ConditionalFreq + Morph + SpeechRate + Following + (1|Speaker)

Bibliography

Adger, D. (2006). Combinatorial variability. *Journal of Linguistics*, 42:503–530.

Alario, F.-X., Perre, L., Castel, C., and Ziegler, J. C. (2007). The role of orthography in speech production revisited. *Cognition*, 102:464–475.

Alba, M. C. (2006). Accounting for variability in the production of Spanish vowel sequences. In Sagarra, N. and Toribio, A. J., editors, *Selected Proceedings of the 9th Hispanic Linguistics Symposium*, pages 273–285.

Aylett, M. and Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence and duration in spontaneous speech. *Language and Speech*, 47:31–56.

Baese-Berk, M. M. and Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, 24(4):527–554.

Baker, A., Archangeli, D., and Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change*, 23:347–374.

Balling, L. W. and Baayen, H. (2008). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes*, 23(7-8):1159–1190.

Baranowski, M. and Turton, D. (2020). TD-deletion in British English: New evidence for the long-lost morphological effect. *Language Variation and Change*, 32:1–23.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Bayley, R. (1994). Consonant cluster reduction in Tejano English. *Language Variation and Change*, 6:303–326.

Bayley, R. (2014). The role of frequency in phonological and morphological variation. In *New Ways of Analyzing Variation 43*.

Bell, M. J., Ben Hedia, S., and Plag, I. (2021). How morphological structure affects phonetic realisation in English compound nouns. *Morphology*, 31(2):87–120.

Ben Hedia, S. and Plag, I. (2017). Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics*, 62:34–49.

Benua, L. (1995). Identity effects in morphological truncation. *University of Massachussetts Occasional Papers in Linguistics*, 18:77–136.

Bermúdez-Otero, R. (2007). Diachronic phonology. In de Lacy, P., editor, *The Cambridge Handbook of Phonology*. Cambridge University Press.

Bermúdez-Otero, R. (2010). Morphologically conditioned phonetics? Not proven. In *On Linguistic Interfaces II*, Belfast. OnLI II.

Bermúdez-Otero, R. and Trousdale, G. (2012). Cycles and continua: on unidirectionality and gradualness in language change. In Nevalainen, T. and Traugott, E. C., editors, *Handbook on the history of English: Rethinking and extending approaches and methods*, pages 691–720. Oxford University Press, Oxford.

Boyd, Z., Elliot, Z., Fruehwald, J., Hall-Lew, L., and Lawrence, D. (2015). An evaluation of sociolinguistic elicitation methods. In *Proceedings of the 18th International Congress of the Phonetic Sciences*. International Phonetics Association.

Braver, A. (2014). Imperceptible incomplete neutralization: Production, non-identifiability, and non-discriminability in American English flapping. *Lingua*, 152:24–44.

Braver, A. (2019). Modelling incomplete neutralisation with weighted phonetic constraints. *Phonology*, 36(1):1–36.

Braver, A. and Kawahara, S. (2016). Incomplete neutralization in Japanese monomoraic lengthening. In *Proceedings of the Annual Meeting on Phonology 2*.

Brewer, J. B. (2008). *Phonetic reflexes of orthographic characteristics in lexical representation*. PhD thesis, University of Arizona.

Browman, C. and Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In Beckman, M. E. and Kingston, J., editors, *Papers in Laboratory Phonology I: Between the grammar and the physics of speech*, pages 341–376.

Browman, C. and Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49:155–180.

Browman, C. P. and Goldstein, L. M. (1985). Dynamic modeling of phonetic structure. In Fromkin, V. A., editor, *Phonetic linguistics*, pages 35–53. Academic Press, New York.

Brown, E. K. (2009). The relative importance of lexical frequency in syllable- and word-final /s/ reduction in Cali, Colombia. In Collentine, J., García, M., Lafford, B., and Marcos Marín, F., editors, *Selected proceedings of the 11th Hispanic Linguistics symposium*, pages 165–178.

Brown, E. L. and Cacoullos, R. T. (2003). Spanish /s/. A different story from beginning (initial) to end (final). In Núñez-Cedeño, R., López, L., and Cameron, R., editors, *A Romance perspective in language knowledge and use*, pages 22–38. John Benjamins, Amsterdam.

Brown, R. (1973). *A first language*. Harvard University Press, Cambridge, MA.

Bruce, G. (1984). Rhythmic alternation in swedish. In Elert, C.-C., Johansson, I., and Strangert, E., editors, *Nordic prosody III*, pages 31–41, University of Umeä.

Brysbaert, M. and New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavioral Research Methods*, 41(4):977–90.

Bürki, A., Ernestus, M., Gendrot, C., Fougeron, C., and Frauenfelder, U. H. (2011). What affects the presence versus absence of schwa and its duration: A corpus analysis of French connected speech. *Journal of the Acoustical Society of America*, 130(6):3980–3991.

Burzio, L. (1994). *Principles of English Stress*. Cambridge University Press.

Bybee, J. (2001). *Phonology and language use*. Cambridge University Press, Cambridge.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14:261–290.

Cedergren, H. J. and Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, 50(2):333–355.

Charles-Luce, J. (1993). The effects of semantic context on voicing neutralization. *Phonetica*, 50:28–43.

Chen, J. Y., Chen, T. M., and Dell, G. S. (2002). Word-form encoding in mandarin chinese as assessed by the implicit priming task. *Journal of Memory and Language*, 46(4):751–781.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22:129–159.

Cho, T. and Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27:207–229.

Chomsky, N. (1965). *Aspects of the theory of syntax*. The MIT Press, Cambridge, MA.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Praeger, New York.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English*. Harper & Row, New York.

Clopper, C. G. and Turnbull, R. (2018). Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B., and Zellers, M., editors, *Rethinking Reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*, pages 25–72. De Gruyter Mouton.

Coetzee, A. W. and Pater, J. (2011). The place of variation in phonological theory. In Goldsmith, J., Riggle, J., and Yu, A., editors, *The Handbook of Phonological Theory. 2nd Edition*. Blackwell, 401-434.

Cofer, T. M. (1972). *Linguistic variability in a Philadelphia speech community*. PhD thesis, University of Pennsylvania.

Cohen, C. (2015). Context and paradigms: two patterns of probabilistic pronunciation variation in Russian agreement suffixes. *Mental Lexicon*, 10(3):313–338. Publisher: John Benjamins Publishing Company.

Cohn, A. C. (2007). Phonetics in phonology and phonology in phonetics. *Working papers of the Cornell Phonetics Laboratory*, 16:1–31.

Connine, C. M. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16:1084–1096.

Connine, C. M., Blasko, D. G., and Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32:193–210.

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29(3):416–431.

Davidson, L. and Stone, M. (2004). Epenthesis versus gestural mistiming in consonant cluster production. In Garding, G. and Tsujimura, M., editors, *Proceedings of WCCFL22*, Somerville, MA. Cascadilla Press.

de Jong, K. (1998). Stress-related variation in the articulation of coda alveolar stops: flapping revisited. *Journal of Phonetics*, 26:283–310.

Delattre, P. and Freeman, D. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics*, 6:29–68.

Diaz-Campós, M. and Carmen, R.-S. (2008). The value of frequency as a linguistic factor: The case of two dialectal regions in the Spanish speaking world. In *Selected Proceedings of the 4th Workshop on Spanish Sociolinguistics*.

Diaz-Campós, M. and Gradoville, M. (2011). An analysis of frequency as a factor contributing to the diffusion of variable phenomena: Evidence from Spanish data. In Ortiz-López, L. A., editor, *Selected proceedings of the 13th Hispanic Linguistics symposium*, pages 224–238. Cascadilla Proceedings Project.

Dinnsen, D. A. and Charles-Luce, J. (1984). Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics*, 12:49–60.

Dmitrieva, O., Jongman, A., and Sereno, J. (2010). Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of Phonetics*, 38:483–492.

Dupoux, E. and Mehler, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory and Language*, 29:316–335.

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4):453–476.

Ellis, L. and Hardcastle, W. J. (2002). Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics*, 30(3):373–396.

Elman, J. and McClelland, J. (1988). Cognitive penetration of the mechanisms of perception: compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, 27(2):143–165.

Erker, D. and Guy, G. R. (2012). The role of lexical frequency in syntactic variability: variable subject personal pronoun expression in Spanish. *Language*, 88(3):526–557.

Ernestus, M. and Baayen, R. H. (2006). The functionality of incomplete neutralization in Dutch: the case of past-tense formation. *Laboratory Phonology*, 8:27–49.

Fasold, R. (1972). *Tense marking in Black English. A linguistic and social analysis.* Number 8 in Urban Language Series. Center for Applied Linguistics.

File-Muriel, R. (2009). The role of lexical frequency in the weakening of syllable-final lexical /s/ in the Spanish of Barranquilla. *Hispania*, 92:348–360.

Fiorentino, R. and Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and cognitive processes*, 22(7):953–1000.

Fisher, W. M. and Hirsh, I. J. (1976). Intervocalic flapping in English. In *Papers from the Twelfth Regional Meeting of the Chicago Linguistics Society.*

Fodor, J. (1983). *The Modularity of Mind.* MIT Press, Cambridge, MA.

Forrest, J. (2017). The dynamic interaction between lexical and contextual frequency: A case study of (ing). *Language Variation and Change*, 29:129–156.

Foulkes, P., Docherty, G., and Watt, D. (2005). Phonological variation in child-directed speech. *Language*, 81(1):177–206.

Foulkes, P., Scobbie, J., and Watt, D. (2010). Sociophonetics. In Hardcastle, W. J., Laver, J., and Gibbon, F. E., editors, *The Handbook of Phonetic Sciences*, pages 703–764. Wiley-Blackwell.

Fourakis, M. and Iverson, G. (1984). On the 'incomplete neutralization' of German final obstruents. *Phonetica*, 41:140–149.

Fowler, C. A., Rubin, P. E., Remez, R. E., and Turvey, M. T. (1980). Implications for speech production of a general theory of action. In Butterworth, B., editor, *Language Production*. Academic Press, New York.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human perception and performance*, 10(4):526.

Fox, R. A. and Terbeek, D. (1977). Dental flaps, vowel duration, and rule ordering in American English. *Journal of Phonetics*, 5:27–34.

Frazier, M. (2006). Output-output faithfulness to moraic structure: Evidence from American English. In *Proceedings of the 36th annual meeting of the north east linguistics society*, pages 1–14.

Fruchter, J., Stockall, L., and Marantz, A. (2013). MEG masked priming evidence for form-based decomposition of irregular verbs. *Frontiers in Human Neuroscience*, 7:798.

Fujimura, O. (1992). Phonology and phonetics—a syllable-based model of articulatory organization. *Journal of the Acoustical Society of Japan (E)*, 13(1):39–48.

Gahl, S. (2008). *Time* and *thyme* are not homophones: the effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3):474–496.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1):110.

Grippando, S. (2021). Japanese orthographic complexity and speech duration in a reading task. *Phonetica*, 78(4):317–344.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural-network model of speech production. *Psychological Review*, 102(3):594–621.

Gut, U. (2007). First language influence and final consonant clusters in the new Englishes of Singapore and Nigeria. *World Englishes*, 26(3).

Guy, G. R. (1980). Variation in the group and the individual: The case of final stop deletion. In Labov, W., editor, *Locating language in time and space*, chapter 1, pages 1–36. Academic Press, New York.

Guy, G. R. (1991a). Contextual conditioning in variable lexical phonology. *Language Variation and Change*, 3:223–239.

Guy, G. R. (1991b). Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change*, 3:1–22.

Guy, G. R. (2007). Lexical exceptions in variable phonology. *University of Pennsylvania Working Papers in Linguistics*, 13(2):109–119.

Guy, G. R. (2019). Variation and mental representation. In Lightfoot, D. W. and Havenhill, J., editors, *Variable properties in language: their nature and acquisition*, chapter 11, pages 129–140. Georgetown University Press.

Guy, G. R. and Boyd, S. (1990). The development of a morphological class. *Language Variation and Change*, 2:1–18.

Hanique, I. and Ernestus, M. (2011). Final /t/ reduction in Dutch past-participles: The role of word predictability and morphological decomposability. In *Interspeech 2011: 12th Annual Conference of the International Speech Communication Association*, pages 2849–2852.

Hansen Edwards, J. G. (2016). Sociolinguistic variation in asian englishes. *English World-Wide. A Journal of Varieties of English*, 37(2):138–167.

Hay, J. (2004). *Causes and consequences of word structure*. Routledge.

Hay, J. B., Pierrehumbert, J. B., Walker, A., and LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, 139:83–91.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. University of Chicago Press.

Hayes, B. (2000). Gradient well-formedness in Optimality Theory. In Dekkers, J., van der Leeuw, F., and van de Weijer, J., editors, *Optimality Theory: Phonology, syntax, and acquisition*, pages 88–120. Oxford University Press.

Hazen, K. (2011). Flying high above the social radar: Coronal stop deletion in modern Appalachia. *Language Variation and Change*, 23:105–137.

Henke, W. L. (1966). *Dynamic articulatory model of speech production using computer simulation*. PhD thesis, Massachussetts Institute of Technology.

Herd, W., Jongman, A., and Sereno, J. (2010). An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *Journal of Phonetics*, 38:504–516.

Heyward, J., Turk, A., and Geng, C. (2014). Does /t/ produced as [?] involve tongue tip raising? Articulatory evidence for the nature of phonological representations. Paper presented at Laboratory Phonology 14.

Holmes, J. and Bell, A. (1994). Consonant cluster reduction in New Zealand English. *Wellington Working Papers in Linguistics*, 6.

Hombert, J.-M., Ohala, J. J., and Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, 55:37–58.

Hooper, J. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In Christie, Jr., W. M., editor, *Current Progress in Historical Linguistics*, pages 95–105. North-Holland, Amsterdam.

Howes, D. H. (1957). On the relation between the probability of a word as an association and in general linguistic usage. *The Journal of Abnormal and Social Psychology*, 54(1):75–85.

Huff, C. T. (1980). Voicing and flap neutralization in New York City English. *Research in Phonetics*, 1:233–256.

Inkelas, S. (1990). *Prosodic constituency in the lexicon*. Garland Publishing, New York.

Itô, J. (1990). Prosodic minimality in Japanese. In Ziolkowski, M., Noske, M., and Deaton, K., editors, *Proceedings of the Chicago Linguistic Society 26*, pages 213–239, Chicago. Chicago Linguistic Society.

Jakobson, R., Fant, C. G., and Halle, M. (1951). *Preliminaries to speech analysis: The distinctinve features and their correlates*. MIT Press.

Joos, M. (1942). A phonological dilemma in Canadian English. *Language*, 18:141–144.

Jurafsky, D., Bell, A., Fosler-Lussier, E., Girand, C., and Raymond, W. (1998). Reduction of English function words in Switchboard. *Proceedings of ICSLP*, 98(7):3111–3114.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. and Hopper, P., editors, *Frequency and the emergence of linguistic structure*, pages 229–254. John Benjamins, Amsterdam.

Kahn, D. (1980). *Syllable-based generalizations in English phonology*. Garland Publishing, New York.

Kaplan, A. (2017). Incomplete neutralization and the (a)symmetry of paradigm uniformity. In *Proceedings of the 34th west coast conference on formal linguistics*, pages 319–328.

Kawamoto, A. H. (1999). Incremental encoding and incremental articulation in speech production: Evidence based on response latency and initial segment duration. *Behavioral and Brain Sciences*, 22:48–49.

Kazanina, N., Phillips, C., and Idsardi, W. J. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences*, 103(30):11381–11386.

Keating, P. (1990). The window model of coarticulation: Articulatory evidence. In Kingston, J. and Beckman, M. E., editors, *Papers in Laboratory Phonology I: Be-*

tween the grammar and the physics of speech, pages 450–469. Cambridge University Press, Cambridge.

Kenstowicz, m. (1995). Base-identity and uniform exponence: Alternatives to cyclicity. In Durand, J. and Laks, B., editors, *Current Trends in Phonology: Models and Methods*. CNRS Paris-X.

Kharlamov, V. (2012). *Incomplete neutralization and task effects in experimentally-elicited speech: evidence from the production and perception of word-final devoicing in Russian*. PhD thesis, University of Ottawa.

Kharlamov, V. (2014). Incomplete neutralization of the voicing contrast in word-final obstruents in Russian: Phonological, lexical, and methodological influences. *Journal of Phonetics*, 43:47–56.

Kingston, J. and Diehl, R. L. (1994). Phonetic knowledge. *Language*, 70:419–454.

Kiparsky, P. (1972). Explanation in phonology. In Peters, S., editor, *Goals of Linguistic Theory*, pages 189–227. Prentice-Hall, Cinnaminson, NJ.

Kiparsky, P. (1982). Lexical morphology and phonology. In Yang, I., editor, *Linguistics in the morning calm*, pages 3–91. Hanshin, Seoul.

Kiparsky, P. (1993). Variable rules. Paper presented at Rutgers Optimality Workshop 1, October.

Kluender, K. R., Diehl, R. L., and Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: an auditory explanation. *Journal of Phonetics*, 16:153–169.

Krauss, M. E. (1975). St Lawrence Island Eskimo phonology and orthography. *Linguistics*, 13(152):39–72.

Krivokapić, J. (2020). Prosody in Articulatory Phonology. In Shattuck-Hufnagel, S. and Barnes, J., editors, *Prosodic Theory and Practice*. MIT Press.

Krivokapić, J., Styler, W., and Parrell, B. (2020). Pause postures: The relationship between articulation and cognitive processes during pauses. *Journal of Phonetics*, 79.

Kuperman, V., Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *The Journal of the Acoustical Society of America*, 121(4):2261–2271.

Labov, W. (1972a). Negative attraction and negative concord in English grammar. *Language*, 48(4):773–818.

Labov, W. (1972b). Some principles of linguistic methodology. *Language in Society*, 1(1):97–120.

Labov, W. (1989). The child as linguistic historian. *Language Variation and Change*, 1:85–97.

Labov, W., Ash, S., and Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change*. Mouton de Gruyter, Berlin.

Labov, W., Cohen, P., Robins, C., and Lewis, J. (1968). A study of the non-standard English of Negro and Puerto Rican speakers in New York City. Final report, Cooperative Research Project 3288, Vols. I and II.

Labov, W. and Rosenfelder, I. (2011). The Philadelphia Neighborhood Corpus of LING 560 studies, 1972-2010. With support of NSF contract 921643.

Ladd, D. R. (2011). Phonetics in phonology. In Goldsmith, J., Riggle, J., and Yu, A. C. L., editors, *Handbook of Phonological Theory*, pages 348–373. Blackwell.

Lawson, E., Scobbie, J., and Stuart-Smith, J. (2011). The social stratification of tongue shape for postvocalic /r/ in Scottish English. *Journal of Sociolinguistics*, 15:256–68.

Lawson, E., Stuart-Smith, J., and Scobbie, J. (2018). The role of gesture delay in coda /r/ weakening: An articulatory, auditory and acoustic study. *The Journal of the Acoustical Society of America*, 143:1646–1657.

Lee-Kim, S.-I., Davidson, L., and Hwang, S. (2013). Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology*, 4:475–511.

Levelt, W., Roelofs, A., and Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22:1–75.

Lewis, G., Solomyak, O., and Marantz, A. (2011). The neural basis of obligatory decomposition of suffixed words. *Brain and Language*, pages 118–127.

Liberman, M. (2018). Toward progress in theories of language sound structure. In Brentari, D. and Lee, J., editors, *Shaping Phonology*, chapter 9, pages 201–217. University of Chicago Press.

Lichtman, K. (2010). Testing articulatory phonology: Variation in gestures for coda /t/. Illinois Language and Linguistics Society 2.

Lim, L. T. and Guy, G. R. (2005). The limits of linguistic community: Speech styles and variable constraint effects. *University of Pennsylvania Working Papers in Linguistics*, 10(2):157–170.

Lin, S., Beddor, P. S., and Coetzee, A. W. (2014). Gestural reduction, lexical frequency, and sound change: A study of post vocalic /l/. *Laboratory Phonology*, 5(1):9–36.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35:1773–1781.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, W. J. and Marchal, A., editors, *Speech Production and Speech Modelling*, pages 403–439. Kluwer Academic Publishers.

Lisker, L. and Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Sciences*, pages 563–567, Prague.

Lociewicz, B. L. (1992). *The effect of frequency on linguistic morphology.* PhD thesis, University of Texas, Austin, Austin, TX.

Lõo, K., Järvikivi, J., Tomaschek, F., Tucker, B. V., and Baayen, R. H. (2018). Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology*, 28(1):71–97.

MacKenzie, L. and Tamminga, M. (forthcoming). New and old puzzles in the morphological conditioning of coronal stop deletion. *Language Variation and Change.*

Magloughlin, L. (2016). Accounting for variability in North American English /r/: Evidence from children's articulation. *Journal of Phonetics*, 54:51–67.

McCarthy, J. J. and Prince, A. (1993). Prosodic morphology: Constraint interaction and satisfaction. Technical Report ROA 482, Rutgers University.

McCarthy, J. J. and Prince, A. (1995). Faithfulness and reduplicative identity. *University of Massachussetts Occasional Papers in Linguistics*, 18:249–384.

McLeod, S., van Doorn, J., and Reed, V. (2001). Normal acquisition of consonant clusters. *American Journal of Speech-Language Pathology*, 10:99–110.

McQueen, J. M., Cutler, A., and Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30(6):1113–1126.

Mielke, J., Baker, A., and Archangeli, D. (2010). Variability and homogeneity in American English /r/ allophony and /s/ retraction. In Kühnert, B., editor, *Variation, detail, and representation. LabPhon 10.* Mouton de Gruyter, Berlin.

Mielke, J., Baker, A., and Archangeli, D. (2016). Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /r/. *Language*, 92(1):101–140.

Mitleb, F. (1981). Temporal coordinates of 'voicing' and its neutralization in German. *Research in Phonetics*, 2:173–191.

Mitterer, H. and Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, 109(1):168–73.

Mousikou, P., Strycharczuk, P., Turk, A., and Scobbie, J. (2015). Morphological effects on pronunciation. In *Proceedings of the 18th International Congress of the Phonetic Sciences*.

Munson, B. and Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of speech, language, and hearing research*, 47(5):1048–1058.

Myers, J. (1995). The categorical and gradient phonology of variable t-deletion in English. Paper presented September 1995 at the International Workshop on Language Variation and Linguistic Theory, University of Nijmegen, Netherlands,.

Myers, J. and Guy, G. R. (1997). Frequency effects in variable lexical phonology. *University of Pennsylvania Working Papers in Linguistics*, 4(1):215–228.

Neu, H. (1980). Ranking of constraints on /t,d/ deletion in American English: A statistical analysis. In Labov, W., editor, *Locating language in time and space*, chapter 2, pages 37–54. Academic Press, New York.

Nicenboim, B., Roettger, T. B., and Vasishth, S. (2018). Using meta-analysis for evidence synthesis: The case of incomplete neutralization in German. *Journal of Phonetics*, 70:39–55.

Parrell, B. and Narayanan, S. (2018). Explaining coronal reduction: Prosodic structure and and articulatory posture. *Phonetica*, 75(2):151–181.

Parrott, J. K. (2007). *Distributed Morphological mechanisms of Labovian variation in morphosyntax*. PhD thesis, Georgetown University, Washington, D.C.

Patrick, P. L. (1991). Creoles at the intersection of variable processes: (td)-deletion and past-marking in the Jamaican mesolect. *Language Variation and Change*, 3(2):171–189.

Patterson, D. and Connine, C. (2001). Variant frequency in flap production. *Phonetica*, 58(4):254–275.

Perkell, J. S. and Matthies, M. L. (1992). Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability. *Journal of the Acoustical Society of America*, 91:2911–2925.

Phillips, B. S. (1981). Lexical diffusion and Southern *tune, duke, news*. *American Speech*, 56:72–78.

Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, 60(2):320–342.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In *Laboratory Phonology VII*, pages 101–139. Mouton de Gruyter, Berlin.

Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimination and robustness. In Bod, R., Hay, J., and Jannedy, S., editors, *Probabilistic linguistics*, pages 177–228. MIT Press, Cambridge, MA.

Plag, I. and Ben Hedia, S. (2017). The phonetics of newly derived words: Testing the effect of morphological segmentability on affix duration. In Arndt-Lappe, S., Braun, A., Moulin, C., and Winter-Froemel, E., editors, *Expanding the Lexicon: Linguistic Innovation, Morphological Productivity, and the Role of Discourse-related Factors*. de Gruyter Mouton, Berlin, New York.

Plag, I., Homann, J., and Kunter, G. (2017). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, 53:181–216.

Podesva, R. J., Roberts, S. J., and Campbell-Kibler, K. (2006). Sharing resources and indexing meaning in the production of gay styles. In Cameron, D. and Kulick, D., editors, *The Language and Sexuality Reader*, pages 141–150. Routledge, London.

Port, R. F., Mitleb, F., and O'Dell, M. L. (1981). Neutralization of obstruent voicing in incomplete. *Journal of the Acoustical Society of America*, 70(13).

Port, R. F. and O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, 13:455–471.

Purse, R. (2019). The articulatory reality of coronal stop 'deletion'. In *Proceedings of the 19th International Congress of Phonetic Sciences*.

Purse, R. and Turk, A. (2016). '/t,d/ deletion': Articulatory gradience in variable phonology. Paper presented at Laboratory Phonology 15, Cornell University, Ithaca, NY.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Radford, A. (1992). The acquisition of the morphosyntax of finite verbs in English. In Meisel, J. M., editor, *The acquisition of verb placement: Functional catefories and V2 phenomena in language acquisition*, pages 23–62. Kluwer, Dordrecht.

Reynolds, W. (1994). *Variation and phonological theory*. PhD thesis, University of Pennsylvania.

Roberts, J. (1997). Hitting a moving target: Acquisition of sound change in progress by Philadelphia children. *Language Variation and Change*, 9:249–266.

Roberts, J. and Labov, W. (1995). Learning to talk philadelphian: Acquisition of short *a* by preschool children. *Language Variation and Change*, 7:101–112.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42:107–142.

Roelofs, A. (2006). The influence of spelling on phonological encoding in word reading, object naming, and word generation. *Psychonomic Bulletin & Review*, 13:33–37.

Roettger, T. B., Winter, B., Grawunder, S., Kirby, J., and Grice, M. (2014). Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics*, 43:11–25.

Rosenfelder, I., Fruehwald, J., Evanini, K., and Yuan, J. (2011). *FAVE Program Suite [Forced alignment and vowel extraction]*. University of Pennsylvania.

Saltzman, E., Nam, H., Krivokapić, J., and Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In Barbosa, P., Madureira, S., and Reis, C., editors, *Proceedings of the Speech Prosody 2008 Conference*, pages 175–184.

Santa Ana, O. (1991). *Phonetic simplification processes in the English of the barrio: a cross-generational sociolinguistic study of the Chicanos of Los Angeles*. PhD thesis, University of Pennsylvania.

Santa Ana, O. (1996). Sonority and syllable structure in Chicano English. *Language Variation and Change*, 8:63–89.

Saussure, F. d. (1916). Cours de linguistique générale. Publié par Charles Bailly et Albert Séchehaye avec la collaboration de Albert Riedlinger. Paris: Payot et Rivages.

Savin, H. B. (1963). Word frequency effects and errors in the perception of speech. *Journal of Verbal Learning and Verbal Behavior*, 9:292–302.

Schuppler, B., Dommelen, W. A. v., Koreman, J., and Ernestus, M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final /t/: Studies at the phonemic and sub-phonemic level. *Journal of Phonetics*, 40(4):595–607.

Schwarzlose, R. and Bradlow, A. R. (2001). What happens to segment durations at the end of a word? *Journal of the Acoustical Society of America*, 109(2292).

Scobbie, J. (2007). Interface and overlap in phonetics and phonology. In Ramchand, G. and

Reiss, C., editors, *The Oxford Handbook of Linguistic Interfaces*. Oxford University Press.

Scobbie, J. M., Segbregts, K., and Stuart-Smith, J. (2009). Dutch rhotic allophony, coda weakening, and the phonetics–phonology interface. *QMU Speech Science Research Centre Working Paper*.

Selkirk, E. (2011). The syntax-phonology interface. In Goldsmith, J., Riggle, J., and Yu, A. C. L., editors, *The Handbook of Phonological Theory, Second Edition*, pages 435–484. Wiley-Blackwell.

Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., and Malouf, R. (2018). Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience*, 33:32–49.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, 42:213–259.

Shattuck-Hufnagel, S. and Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, 18:41–55.

Shriberg, L. and Kwiatkowski, J. (1980). *Natural process analysis*. John Wiley, New York.

Slowiaczek, L. M. and Dinnsen, D. A. (1985). Individual differences and word-final devoicing in Polish. *Journal of the Acoustical Society of America Supplement*, 77:S85.

Smith, J., Durham, M., and Fortune, L. (2009). Universal and dialect-specific pathways of acquisition: caregivers, children, and t/d deletion. *Language Variation and Change*, 21:69–95.

Smith, R., Baker, R., and Hawkins, S. (2012). Phonetic detail that distinguishes prefixed from pseudo-prefixed words. *Journal of Phonetics*, 40(5):689–705.

Smolensky, P. and Goldrick, M. (2016). Gradient symbolic representations in grammar: The case of French liaison. Manuscript, Johns Hopkins University and Northwestern University, ROA 1286.

Solomyak, O. and Marantz, A. (2009). Lexical access in early stages of visual word processing: A single-trial correlational MEG study of heteronym recognition. *Brain and Language*, 108:191–196.

Solomyak, O. and Marantz, A. (2010). MEG evidence for early morphological decomposition in visual word recognition: A single-trial correlational meg study. *Journal of Cognitive Neuroscience*, 22:2042–2057.

Song, J. Y., Demuth, K., Evans, K., and Shattuck-Hufnagel, S. (2013). Durational cues to fricative codas in 2-year-olds' American English: Voicing and morphemic factors. *Journal of the Acoustical Society of America*, 133(5):2931–2946.

Sproat, R. and Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21:291–311.

Stavness, I., Gick, B., Derrick, D., and Fels, S. (2012). Biomechanical modelling of English /r/ variants. *Journal of the Acoustical Society of America Express Letters*, 131.

Steriade, D. (2000). Paradign uniformity and the phonetics-phonology boundary. In Broe, M. B. and Pierrehumbert, J. B., editors, *Papers in Laboratory Phonology V: Acquisition and the lexicon*, pages 13–34. Cambridge University Press, Cambridge.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In Denes, P. B. and David Jr., E. E., editors, *Human communication: a unified view*, pages 51–66. McGraw Hill, New York.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17:3–46.

Strycharczuk, P. (2019). Phonetic detail and phonetic gradience in morphological processes. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press.

Strycharczuk, P. and Scobbie, J. (2016). Gradual or abrupt? The phonetic path to morphologisation. *Journal of Phonetics*, 59:76–91.

Strycharczuk, P. and Scobbie, J. (2017). Whence the fuzziness? Morphological effects in interacting sound changes in southern british english. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(7).

Sugahara, M. and Turk, A. (2009). Durational correlates of English sublexical constituent structure. *Phonology*, 26(3):477–524.

Taft, M. and Hambly, G. (1986). Exploring the cohort model of spoken word recognition. *Cognition*, 22:259–282.

Tagliamonte, S. (2004). Someth[in]'s go[ing] on!: Variable *ing* at ground zero. In B.-L., G., Bergström, L., Eklund, G., Fidell, S., Hansen, L. H., Karstadt, A., Nordberg, B., Sundergren, E., and Thelander, M., editors, *Language Variation in Europe: papers from the Second International Conference on Language Variation in Europe, ICLAVE 2*, pages 390–403. Department of Scandinavian Languages, Uppsala University.

Tagliamonte, S. and Temple, R. (2005). New perspectives on an ol' variable: (t,d) in British English. *Language Variation and Change*, 17:281–302.

Tamminga, M. (2016). Persistence in phonological and morphological variation. *Language Variation and Change*, 28(03):335–356.

Tamminga, M. (2018). Modulation of the following segment effect on English coronal stop deletion by syntactic boundaries. *Glossa*, 3(1):86.

Temple, R. (2009). (t,d): the variable status of a variable rule. *Oxford University Working Papers in Linguistics, Philology and Phonetics*, 12:145–170.

Temple, R. (2014). Where and what is (t,d)? A case study in taking a step back in order to advance sociophonetics. In Celata, C. and Calamai, S., editors, *Advances in Sociophonetics*, pages 97–136. John Benjamins, Amsterdam.

Thomas, E. R. (2008). Instrumental phonetics. In Chambers, J., Trudgill, P., and Schilling-Estes, N., editors, *The Handbook of Language Variation and Change*, pages 168–200. Blackwell.

Tiede, M., Boyse, S. E., Espy-Wilson, C., and Gracco, V. L. (2011). Variability of North American English /r/ production in response to palatal perturbation. In Maassen, B. and van Lieshout, P., editors, *Speech motor control: New developments in basic and applied research*, pages 53–67. Oxford University Press, Oxford.

Tomaschek, F., Plag, I., Ernestus, M., and Baayen, R. H. (2019). Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning. *Journal of Linguistics*, 57(1):123–161. Publisher: Cambridge University Press.

Tomaschek, F., Tucker, B. V., Fasiolo, M., and Baayen, R. H. (2018). Practice makes perfect: The consequences of lexical proficiency for articulation. *Linguistics Vanguard*, 4(S2).

Tomaschek, F., Tucker, B. V., Ramscar, M., and Baayen, R. H. (2021). Paradigmatic enhancement of stem vowels in regular English inflected verb forms. *Morphology*, 31(2):171–199.

Tomaschek, F., Tucker, B. V., Wieling, M., and Baayen, R. H. (2014). Vowel articulation affected by word frequency. In *Proceedings of the 10th ISSP, Cologne*, pages 425–428.

Tomaschek, F., Wieling, M., Arnold, D., and Baayen, R. H. (2013). Word frequency, vowel length and vowel quality in speech production: An ema study of the importance of experience. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*.

Tucker, B. V., Sims, M., and Baayen, R. H. (2019). Opposing forces on acoustic duration. Technical report. Publisher: PsyArXiv.

Turk, A. and Shattuck-Hufnagel, S. (2021). *Speech Timing: Implications for Theories of Phonology, Phonetics, and Speech Motor Control*. Oxford Studies in Phonology & Phonetics. Oxford University Press.

Turnbull, R. (2015). Patterns of individual differences in reduction: Implications for listener-oriented theories. In *Proceedings of the 18th International Congress of the Phonetic Sciences*. International Phonetics Association.

Turnbull, R. (2018). Effects of lexical predictability of patterns of phoneme deletion/reduction in conversational speech in English and Japanese. *Linguistics Vanguard*, 4(S2).

Turton, D. (2017). Categorical or gradient? An ultrasound investigation of /l/ darkening and vocalization in varieties of english. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8:13.

Twist, A., Baker, A., Mielke, J., and Archangeli, D. (2007). Are "covert" /r/ allophones really indistinguishable? *University of Pennsylvania Working Papers in Linguistics*, 13(2):205–216.

Vannest, J., Newport, E. L., Newman, A. J., and Bavelier, D. (2011). Interplay between morphology and frequency in lexical access: The case of the base frequency effecy. *Brain Research*, 1373:144–159.

Walker, J. A. (2012). Form, function, and frequency in phonological variation. *Language Variation and Change*, 24:397–415.

Walsh, T. and Parker, F. (1983). The duration of morphemic and non-morphemic /s/ in English. *Journal of Phonetics*, 11:201–206.

Waring, R., Fisher, J., and Atkin, N. (2001). The articulation survey: Putting numbers to it. In Wilson, L. and Hewat, S., editors, *Proceedings of the 2001 Speech Pathology Australia national conference*, pages 145–151. Speech Pathology Australia.

Warner, N., Jongman, A., Sereno, J., and Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics*, 32:251–276.

Weinreich, U., Labov, W., and Herzog, M. I. (1968). Empirical foundations for a theory of language change. In Lehmann, W. and Malkiel, Y., editors, *Directions for historical linguistics: A symposium*, pages 97–195. University of Texas Press, Austin and London.

Wolfram, W. and Christian, D. (1976). *Appalachian Speech*. Center for Applied Linguistics, Arlington, VA.

Wright, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory and Cognition*, 7(6):411–419.

Zellou, G. and Tamminga, M. (2014). Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics*, 47:18–35.

Zsiga, E. (1993). *Features, gestures, and the temporal aspects of phonological organization*. PhD thesis, Yale University.

Zsiga, E. (1995). An acoustic and electropalatographic study of lexical and post-lexical palatization in American English. In Connell, B. and Arvaniti, A., editors, *Phonology and phonetic evidence: papers in laboratory phonology IV*, pages 282–302. Cambridge University Press.

Zweig, E. and Pylkkänen, L. (2009). A visual M170 effect of morphological complexity. *Language and Cognitive Processes*, pages 412–439.