UNIVERSITY OF CALIFORNIA
Los Angeles

Phonetic Evidence for a Feed-forward Model: Rounding and Center of Gravity of English [ʃ]

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Arts in Linguistics

by

Z.L. Zhou

2019

ABSTRACT OF THE THESIS

Phonetic Evidence for a Feed-forward Model:
Rounding and Center of Gravity of English [ʃ]

by

Z.L. Zhou
Master of Arts in Linguistics
University of California, Los Angeles, 2019
Professor Patricia Keating, Co-chair
Professor Megha Sundara, Co-chair

This thesis uses empirical data to make the case that the architecture of Grammar is feed-forward. There are three thematic parts: first, I discuss the separation of different phonetically important variables into modules and how a few prominent planning and production models have regarded the order of those modules. Then, I give a novel, statistical operationalization of the concept of feed-forwardness. Finally, I present the results of a production experiment — targeting the center of gravity (COG) of American English [ʃ] — wherein multiple phonetically important variables are simultaneously manipulated. The aforementioned models are translated into linear mixed effects models and then compared based on their ability to predict the experimental data; I find that the model which corresponds to Keating & Shattuck-Hufnagel's 1989 model and Chomsky's 1965 T-model most parsimoniously fits the data, thereby providing phonetic evidence for a feed-forward model of the Grammar. Discussion of some implications for phonetic and phonological research ensues.

The thesis of Z.L. Zhou is approved.

Bruce Hayes

Megha Sundara, Committee Co-chair

Patricia Keating, Committee Co-chair

University of California, Los Angeles

2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Here, I first thank Megha Sundara and Pat Keating, my committee co-Chairs. Out of their many virtues, I list three: they have been insightful; they have provided wise counsel; they have given my ideas an appropriate amount of attention, whether that be substantial or little. I thank Bruce Hayes, my third committee member, for his help and for his many fascinating thoughts and considerations. And, as a group: the original idea for this thesis developed from a synthesis of their three Spring 2018 courses. I could not have done this without them.

Outside of my committee, many thanks are due to my fellow graduate students. In particular, I thank Hironori Katsuda, Rachel Vogel, and Meng Yang for their wildly different but nonetheless significant contributions and support — intellectual, emotional, and procedural.

Finally, I thank Amir Gold, Dylan Ross, and Angela Xu, my assiduous research assistants, for their diligent work and many hours in the lab.

CHAPTER 1

Introduction

1.1 Grammatical architecture

Conceptually, different aspects of language can be related to different grammars, or *modules*: word order is the output of the syntax; abstract sound manipulation is the domain of the phonology; acoustics and articulations, the result of the phonetics.[1] The relationship between modules is examined at the *interfaces* which, since at least Chomsky (1965), have generally been theorized to be one-way, hierarchical, or *feed-forward*: representations pass from one module to the next, but never backwards. The feed-forward relationship has the specific consequence that an *upstream* module will have direct impact on a *proximal downstream* module, but no direct impact on a *distal downstream* module. Moreover, a downstream module must be unable to influence the processes and representations of its upstream neighbors. In this way, feed-forward models are a game of telephone — but one wherein each participant is required to change the message.

From the perspective of speech production and speech planning, the order of modules with respect to one other (and indeed, the existence of order at all), the nature of information flow at interfaces, and the placement of specific processes within a module are all open to empirical insight. However, these two literatures do not much interact with one another and tend to differ in crucial ways. Many planning models are essentially feed-forward, but production models tend to be flatter — is it possible to tell which models are more accurate?

Of course, there are internal justifications. On the basis of speech error and reaction time stud-

[1] The modules are systems of their own, collections of processes operating on representations, and so the concept of linguistic modules is Fodor's (1983) thesis writ small. Just as different subsystems of the cognitive system (e.g., the visual subsystem, the olfactory subsystem) as a whole are "informationally encapsulated", so too are linguistic subsystems.

ies, models of speech planning such as Dell (1986) and Levelt et al. (1999) provide what are essentially timelines for stages of planning. This perspective of planning as a timeline has an intuitive appeal but does necessarily imply feed-forwardness. Models of speech production such as TADA (Saltzman et al., 2008) and Generative Phonetics (e.g., Flemming and Cho 2017) assume that certain types of relevant information (e.g., lexical frequency) preexist — that some forms of information are strictly inputs to their model. This is tantamount to claiming that those inputs come from upstream modules.

To compare these models to one another requires some commonality, a evaluable topic that every model can weigh in on and can make empirical predictions about. As the speech production models do not weigh in on planning — and cannot be coerced into doing so — the obvious choice is to look at production. Thus, in this paper, I investigate the implications of feed-forward and flat grammars. I do so by using the ability of these models to predict speech production data as my evaluation metric.

1.2 Models of the (architecture of the) grammar

In general, work which has probed the architecture of the grammar (or made suggestions thereon) fall into two categories: planning models and production models. These classes of model are capable of explaining much, but the motivations behind their existence are admittedly rather disparate.

Here, I attempt to compare planning and implementation models to tease out the particular assumptions that they make about the order/lack of order of the modules/stages of the grammar. This is not so easily done, as they use the same terms in different ways. Therefore, I operationalize the *lexicon* as being the collection of abstract facts about a word or phrase including properties such as word frequency and neighborhood density, and as retrieval of any information from memory (e.g., lemmas, wordforms). *Prosody* is defined as phrasal positional prominence (i.e., phrase-initiality/finality) and informational-presentational prominence (e.g., phrasal stress, pitch accents) and the building thereof, but not word-level positional prominence (e.g., if a syllable is stressed). For reasons that will become clear, I also classify global speech rate as a prosodic variable. I define *phonology* as phonological encoding, encodable static phonology (e.g., contrastive features,

segmental order, if a syllable is stressed), and the application of phonological processes. Finally, *phonetics* will be defined as the specifics of acoustic realization and articulation. Rationale for some of these choices are provided in §3.5.2.

1.2.1 Planning models

Although there are many planning models, I will discuss two of the most influential planning models: Dell's (1986) spreading activation model and Levelt et al.'s (1999) Weaver++. In addition, I will discuss the objections to Weaver++ presented in Keating and Shattuck-Hufnagel (1989) — although theirs is not a full planning model, their proposal does represent a point of view that we will be able to empirically evaluate.

1.2.1.1 The Dell lexical activation model

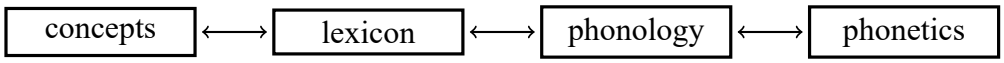

Figure 1.1: Dell's model, adapted from Dell and O'Seaghdha (1992). Each box represents a level of activation nodes which can transport activation both upwards and downwards. Terms have been changed to fit our definitions

A classic planning model is Dell's (1986) activation model, schematically illustrated in Fig. 1.1. This model is a spreading activation model which distinguishes between semantic units (concepts), lexemes/lemmas (lexicon), and phonological units (phonology). In this model, semantic units are first activated externally through conceptual preparation — these semantic units can be considered prelinguistic, representing actuation of thought or of the desire to speak. Lexemes are selected according to the activation of these semantic units; phonological units[2] are activated in a similar manner, being selected according the activation of the lexemes. These phonological units are encoded by being filled into different varieties of slots (i.e., syntactic frames) which are "independently created" separately from the planning process (Dell and O'Seaghdha, 1992).

[2]In the original paper, these units are segments, though there is no principled reason that they must be.

Crucially for us, nodes that are activated but not selected can still have an impact on other nodes: at each level activation is spread both "upstream" and "downstream" by all nodes. However, each level requires a different threshold of combined activation to be considered activated. The thresholds for activation in each level differ to such an extent that the activation of an upstream node by downstream ones is generally rare. This renders the model non-hierarchical in concept, but essentially hierarchical in execution. Note that prosody as we have defined it is not considered by this model.[3]

1.2.1.2 Weaver++: Levelt's lexical activation model

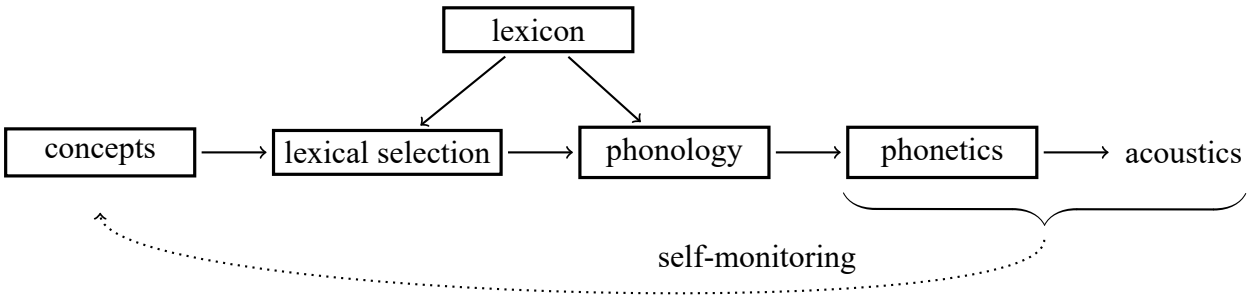

Figure 1.2: Levelt's model, adapted from Levelt et al. (1999). Each box represents a level of activation nodes which can propogate activation only downwards. Included is a self-monitoring loop, which checks to ensure that the right concepts have been chosen. Terms have been changed to fit our definitions; morphological and phonological encoding have been combined

Another classic, activation-based model is Levelt et al.'s (1999) Weaver++ model, schematically illustrated in Fig. 1.2. Weaver++ too distinguishes between conceptual preparation, lexeme selection, and phonological encoding. It also formally adds a morphological encoding stage, which occurs after lexeme selection but before phonological encoding (though for our purposes, we merge it with phonological encoding), and includes phonetic encoding and articulation as further steps in

[3]Unfortunately, the Dell model is vague about the building and effects of prosody. In his 1986 paper, Dell does discuss how his proposed model might be extended to have a role for syntax and morphology. The effect of doing so, beyond accounting for syntactic/morphological speech errors, is not discussed. That said, he does theorize in a footnote that morphology should exert influence on "what sounds are simultaneously active", with the observable effect being the prosodic grouping of words. The production of prosodic structure remains unaddressed by the model and the structure onto which prosodic phrasing is projected (viz., syntactic structure) is pre-built.

the planning process.

Weaver++ and the Dell (1986) model are similar in their overall architecture, but there are two major differences relevant to us. One is that Weaver++ explicitly accesses the lexicon during the stages of lexical selection, morphological encoding, and phonological encoding — in Dell's model, the nature of the lexicon and of lexical access is somewhat amorphous, since all of the nodes which might be later activated are assumed to already exist by the time that conceptual preparation has finished. The second is that Weaver++ is strictly serial and feed-forward: activation does not propogate upstream; phonological encoding *cannot* occur until after morphological encoding has finished, and likewise for the other stages. There is also a self-monitoring feedback loop, the implications of which lie beyond the scope of this paper.

Although Weaver++ does discuss the process of *prosodification*, defined as the incremental generation of the prosodic word given the retrieved segmental and metrical structures of that word, its conceptualization of prosody is different from the utterance-level prosody we are interested in. However, the prosodification process has the general theme of "smaller phonological objects incrementally combine to make larger ones": segments combine to form syllables, and syllables, words. If we extend this premise, then words combine to form phrases, so phrasal prosody cannot be determined before phonological wordhood is determined. Thus, Weaver++ must be a model wherein utterance-level prosodic structure-building must occur last. Again, this premise is not explicitly supported by Weaver++, but is in line with both its general principles and follows the discussion in Roelofs (2000).

1.2.1.3 An additional view on planning models: Keating and Shattuck-Hufnagel

Figure 1.3: The Keating and Shattuck-Hufnagel (1989) model, which is feed-forward. Each box represents a collection of processes — in essence, a module. Note that while the model discusses accessing a lexicon, that lexicon is not of abstract word-level facts and is thus left out

The logical alternative to Weaver++'s prosody-last approach is presented by Keating and

Shattuck-Hufnagel (1989), who argue that prosodic encoding must occur before phonological encoding. Their model is schematically illustrated in Fig. 1.3. This model differs significantly from both the Dell (1986) model and from Weaver++ in that it is concerned with the origins and generation of utterance-level prosody: they assume that utterance-level prosodic structure is first read off of syntactic structure, a view in line with, e.g., Nespor and Vogel (1986) and Selkirk (2011).

Keating and Shattuck-Hufnagel argue that, considering evidence that "all aspects of word form encoding, including phonetic encoding, must refer to prosodic structure", prosodic structure is built from the largest units to the smallest units — e.g., the intonational phrase is built before phonological phrases are built; the phonological phrase is built before prosodic words are built. Only after prosodic wordhood has been determined can phonological encoding and phonetic encoding occur. This model presents a bifurcated lexicon. As syntactic structure must exist before prosodic structure and syntactic structure is not empty of lexical content, their model asserts that lexical access, at least in the form of lemma retrieval, occurs very early. There is a secondary stage of lexical access where phonological material is retrieved — following my earlier operationalization, I classify phonological material retrieval as part of phonology.

1.2.2 Implementation models

There exist a plethora of phonetic implementation models, but few are of interest to us. This is because few models use much more than articulatory data, and so for our purposes can be treated as the same model. Here, I'll discuss Saltzman et al.'s (2008) TADA and Beckman and Pierrehumbert's (1986)'s ToBI. I will also discuss the class of Generative Phonetics harmonic grammar models represented by work such as Flemming and Cho (2017) and Lefkowitz (2017).

1.2.2.1 Task Dynamics Application: an Articulatory Phonology model

The Task Dynamics Application (TADA; Saltzman et al. 2008) model is an implementation model of the prosody-phonetics-phonology interface, schematically illustrated in Fig. 1.4. TADA is a model in the Articulatory Phonology (AP; Browman and Goldstein 1986, 1992) framework and as such assumes a phonological specification very close to the phonetic output. This is accomplished

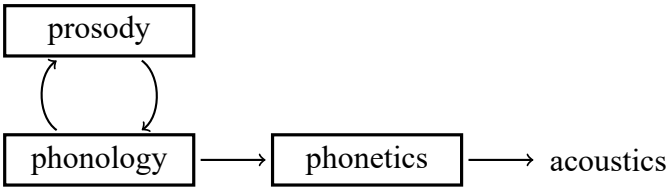

Figure 1.4: The TADA model, which is a specifically phonological implementation model. Each box represents a collection of processes. The relationship between prosody and phonology is deeply intertwined, such that it may be more accurate to consider prosody a subprocess of phonology

by positing a phonological representation of gestures, which specify the state of articulators, and the relative timing of those gestures. These representations are all collected in a gestural score (Fig. 1.4's phonology), which is then implemented, outputting articulatory trajectories which can be directly passed on to the vocal tract (Fig. 1.4's phonetics).

The relationship between prosodic and non-prosodic gestures is somewhat unclear. TADA accounts for prosodic effects such as initial strengthening and final lengthening through the application of μ-gestures,[4] oscillators dependent on nested phrase, foot, and syllable oscillators. This suggests that μ-gestures must be built from pre-existing representations, strongly implying that syllables are organized from pre-existing gestures, feet from pre-existing syllables, etc., and thus that the TADA model, similarly to Weaver++, builds prosodic structure after phonology. That said, Goldstein et al. (2007) suggest that syllables arise naturally from the coupling of gestures to one another. It is not unthinkable to assume this approach could be in principle extended to higher prosodic groups, so I assume that TADA does not itself make any distinction between phonological and prosodic implementation. Furthermore, since the implementation of TADA has μ-gestures coexisting simultaneously with other gestures, prosodic structure implementation must be, in some sense, considered cooccurent with phonological implementation.

There is no room in TADA for any lexical effects as AP does not model the pre-phonological. The creation of the gestural score comes immediately from the linguistic gestural model, i.e., broadly, phonology (Browman and Goldstein, 1990). This is not an accidental failing, as the ex-

[4] μ-gestures are the generalization of the π-gestures developed in Byrd (2000); π-gestures are named as such because π stands for **p**rosodic; μ-gestures because μ stands for **m**odulation.

plicit goal of this model is to "explain a number of phonological phenomena, particularly those that involve overlapping articulatory gestures" through thorough consideration of articulatory organization. But in essence, the start of any TADA utterance *is* the articulatory score. Lexical access, etc., must happen elsewhen.

1.2.2.2 Tones and Break Indices: an autosegmental-metrical model of intonation

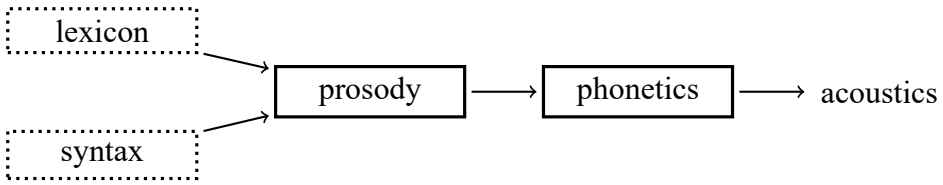

Figure 1.5: The ToBI model, which is a specifically intonational implementation model. Each box represents a module; syntax and the lexicon are dotted to show the assumption that they occur before prosody, even though their presence is not focal to the theory

The Tones and Break Indices (ToBI; Beckman and Pierrehumbert 1986; Beckman and Hirschberg 1994; schematically illustrated in Fig. 1.5) model of intonation is not generally thought of as an implementation model, although it demonstrably is — cf. Anderson et al. (1984). As an autosegmental-metrical (Goldsmith, 1976) model, it makes the strong assertion that the phonological representation of tones corresponds directly to the phonetic realization of those tones: tones which can be phonetically observed are tones which are phonologically specified.

ToBI's strength lies in its fleshed-out predictions about the prosodic phonology-phonetics interface. Lengthening of segments is expected before intermediate and intonational breaks, with greater lengthening before intonational breaks, and syllables associated with a pitch accent are likewise expected to be longer. That said, ToBI is not concerned with other modules of the grammar, though it makes the assumption that lexical access and syntactic structure-building both occur before prosody. This assumption is supported by work such as Pierrehumbert and Hirschberg (1990), Grice et al. (2017), and Zhou and Ahn (2019), which have discussed the interpretative nature of pitch accents and pitch alignments — effects which are presumably the result of prosodic structure arising from syntactic structure.

1.2.2.3 Generative Phonetics models

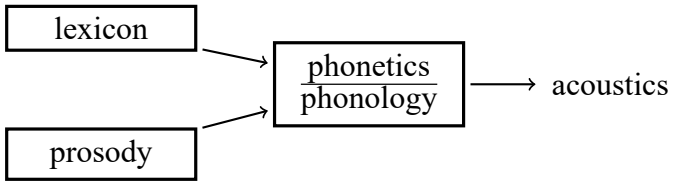

Figure 1.6: A Generative Phonetics model which includes lexical and prosodic effects. Note that these effects are not present in all Generative Phonetics models, but do take this form when they are

Finally, Generative Phonetics harmonic grammars such as the ones pioneered by Flemming (2001) and pursued by Flemming and Cho (2017) and Lefkowitz (2017) are also implementation models of the phonetics-phonology interface. These grammars use traditional phonological models of process to determine acoustic/phonetic output: Flemming and Cho use a harmonic grammar to determine the realization of Mandarin third tones; Lefkowitz uses a Maximum Entropy grammar to output the duration of American English vowels.

While these Generative Phonetics grammars do not intrinsically account for lexical phenomena, nothing bars such models from doing so. Indeed, much work has been done on how lexical effects should be incorporated into harmonic grammars — e.g., Coetzee and Kawahara (2013) add scaling factors to their constraints, causing the weight of a constraint violation to increase and decrease with varying lexical frequency. Prosody is even more straightforwardly incorporated: Lefkowitz (2017) does so by including a constraint which assesses penalties if a pitch accented vowel is not of an appropriate duration. This, of course, assumes that pitch accents are already assigned, meaning that prosody must come before phonology.

It is important to note that Generative Phonetics models are not inherently flat in that the theory behind them does not require the inclusion of any lexical/prosodic phenomena. However, if such properties are to be considered by a harmonic grammar, then that model is not strictly compatible with a feed-forward architecture. This is, in fact, highlighted by Flemming and Cho (2017), who note that compromise must be achieved between phonetic constraints, else the highest weighted constraint would simply dictate the output entirely. A generative phonetic harmonic grammar which

incorporates lexical scaling and constraints which refer to prosody is necessarily one where lexical effects and prosodic effects compete and compromise with each other and all such considerations are evaluated simultaneously.

1.2.3 Commonalities among production and implementation models

Regarding the ordering of modules, we see that there is strikingly little agreement. Indeed, there are only two across the board commonalities. All models either explicitly or implicitly have a lexical access/conceptual access component as upstream from other modules. Similarly, all models have phonetic/acoustic output as the most downstream output.

Beyond those, most models have a phonological component (encoding or procedural) separate from the phonetic component, although ToBI does not consider segmental phonology and Generative Phonetics grammars merge the two. Except for TADA, models that have both an utterance-level prosodic component and a phonological component place prosody before the phonology.

1.3 Effects of modules on modules

The disagreement in these different models sets the scene for our inquiry: which are more right? To answer this question, we must formulate predictions of these models. Correspondingly, we must be able to localize phenomena to specific modules and interfaces. The groupings that I used in §1.2 inherently identify phenomena with modules, but we still must discuss interfaces. Here, I discuss the intertwining effects of the lexicon, prosody, phonology, and phonetics.

1.3.1 Lexical effects

1.3.1.1 ... on phonetics

The effects of lexical properties on phonetic realization have been widely studied (see Vitevitch and Luce (2004) for an overview of phonetic neighborhood density effects). In general, lexical items which are more frequent, more predictable, or have fewer phonological neighbors tend to

be reduced. These effects have been found on many different phonetic variables: when words are more frequent, predictable, or phonologically lonely, their vowels are more centralized (Gahl et al., 2012), their consonants assimilate to neighbors in voicing (Ernestus et al., 2006), their segments are more likely to be deleted (Cohen Priva, 2015), they are shortened (Gahl, 2008), and coarticulation increases (Baker and Bradlow, 2009; Scarborough, 2013).[5]

1.3.1.2 ... on phonology

Similarly, the interactions between lexical factors and phonology are also well-documented. Earlier work such as Itô and Mester (1993) established the notion that there exist lexical strata, groups of words in the lexicon that native speakers treat as exceptional with regard to certain generalizations. Later work has extended this concept, suggesting that even individual words may behave exceptionally in these respects. These effects typically change the application rates of phonological processes: Coetzee and Kawahara (2013) propose that phonological constraint weight should be scaled based on the frequency of the input; lexically-indexed constraints have been also argued for on the basis of descriptive and explanatory adequacy (e.g., Zymet 2018; Moore-Cantwell and Pater 2016)

1.3.1.3 ... on prosody

Very little work has been done on how the lexicon or lexical properties interact with prosody. However, Schweitzer et al. (2015) show that in German, high absolute frequency of a word increases the variability in that word's pitch accent shape, but high relative frequency of that word in its context decreases variability. This result suggests that the collocation of a pitch accent with a word is stored in memory, and that the decision of which pitch accent to use on a word may be (at least partially) lexically determined.

[5] Scarborough's discussion centers around "hard" and "easy" words, but these are synonyms for high relative neighborhood density and low relative neighborhood density, respectively. Note that she finds that coarticulation increases for "hard" words.

1.3.2 Prosodic effects

1.3.2.1 ... on phonetics

Shattuck-Hufnagel and Turk (1996) present the view that prosody feeds directly into phonetics. While this view is not dominant, it may be ascendant: much work finds that prosodic variables have significant effects on phonetic realization. Fougeron and Keating (1997) and Cho and Keating (2001) find evidence of articulatory strengthening at prosodic edges. Aylett and Turk (2004) finds that prosodic prominence in English is strongly associated with increased syllable length, in line with Cho's (2016) discussion of syntagmatic and paradigmatic prosodic enhancement. The effect of speech rate has been of long-standing interest to phoneticians: Lindblom (1963) finds that vowel reduction is associated with high speech rate; Byrd and Tan (1996) finds that increased consonant coarticulation is correlated with the same.

1.3.2.2 ...on phonology

It has been long acknowledged that many phonological phenomena commonly refer to prosodic constituency and prosodic position: many, many processes have been found which refer to the syllable, foot, word, phrase, and utterance (Nespor and Vogel, 1986; Selkirk, 1986). Other prosodic categories such as the colon have also been argued for, though the evidence is admittedly scant (Lionnet, 2019). These processes run the gamut from intonation-driven lengthening to phrasal-position-driven featural changes (Hayes and Lahiri, 1991; Jun, 1993), and it has been argued that when a process varies in application rate due to speech rate, it is the result of speech rate affecting the size of prosodic constituents (Jun, 1993).

1.3.3 Phonological effects on phonetics

It is perhaps a little redundant to discuss how phonology and phonetics relate. After all, it is widely agreed upon that phonetics in some sense relies on phonology. Furthermore, in all previously discussed models, phonology has a close relationship with phonetics, either serving as direct input or being concurrently processed. However, one small but crucial point must be made: the interpre-

tation of phonological features by the phonetics is language-specific (Keating, 1985). This is to say that there is no default phonetic implementation of a phonological representation: phonetics must be considered a full linguistic module, as phonological representations must be processed in language-specific ways.

1.4 The statistical evaluation of architectures

Having shown that there is extensive disagreement between planning and implementation models, and having discussed effects of module-module interaction, we have enough information to be able to conduct an experiment where the data provide evidence in favor of flat vs feed-forward organization.

The key is in how a feed-forward model restricts possible interactions between modules: an affect of the nth module on some representation of the $n+2$th module is only possible if there is an effect of the nth module on a representation of $n+1$th module *and* the representation of the $n+1$th module has an effect on the representation of the $n+2$th module. In other words, upstream modules affect output only if they transmit their effects through downstream modules. All interactions are allowed so long as at least one variable from every module, from the topmost module which any variable (in the interaction) belongs to, to the bottommost module, is included

For example, consider a feed-forward grammar where the order of the modules is lexicon, syntax, phonology — and consider what we would predict if we want to examine the effect of three variables, one lexical (*lex*), one syntactic (*syn*), and one phonological (*phon*), on the output of phonetics, an acoustic measure (*acu*). A feed-forward model will allow a direct (i.e., main) effect of *phon* on *acu*, but it cannot allow a main effect of *syn* on *acu*. Instead, it predicts that the effect of *syn* on *acu* is to be seen though the effect of *syn* on *phon*, and then on *phon* on *acu*. Extending this logic, any effect of *lex* on *acu* must be an effect of *lex* on *syn* on *phon* on *acu*.

To concretize, consider a situation where syntactic category (e.g., noun, verb) has an effect on word duration and lexical frequency also has an effect on vowel duration. Perhaps an experiment asks participants to use words like *table*, *counter*, and *alarm* as both nouns and verbs. Imagine we find a lexical effect and a syntactic effect. The feed-forward model corresponding to this is one

where the syntactic effect is a main effect (the type of syntactic category has a direct effect on the duration of the vowel) but the effect of lexical frequency is only seen through an interaction. In this example, the effect of lexical frequency would vary based on the syntactic category — this is the same as saying that lexical frequency adjusts the effect of syntactic category.[6] If we add a phonological variable such as intrinsic vowel length, we can apply the same logic and specify a model where syntactic category has a moderating effect on intrinsic vowel length, but there is no direct effect of syntactic category, and so on.

The preceding example corresponds to a regression model where *acu* is being predicted by *phon*, the two-way interaction of *phon* and *syn*, and the three-way interaction of *phon* and *syn* and *lex*.[7] This has the effect of transmitting effects forward from one module to another in specific way: in essence, the only thing that *ever* affects *acu* is *phon*, but when *syn* has a particular value, *phon* changes and so *acu* is changed. Generalizing, a feed-forward model can be represented by a regression model where:

1. variables are grouped according to module: m_0, m_1 ... m_n
2. variables in the furthermost downstream module m_n are included as main effects; no other main effects are allowed
3. any interaction, if it includes a variable in module m_{x-1}, must include at minimum one variable in its immediately downstream module m_x

This is not a perfect representation of a feed-forward model, although it is very close. Regression models are incapable of implementing any notion of effect directionality, which is to say that an effect of *syn* on *phon* cannot be distinguished from an effect of *phon* on *syn*. However, given how we constructed our toy model, I assert that the most coherent interpretation is the one I have provided.

[6] It is hypothetically possible that strength of the lexical frequency effect is the same for nouns and verbs; if these are the only syntactic categories considered, then the model approximates one where lexical frequency is a main effect.

[7] In R regression formula, $acu \sim phon + phon : syn + phon : syn : lex$.

With this established, it is thus possible to conduct model comparison on acoustic data to see if the data favors or disfavors a feed-forward model. We simply require an acoustic measure which is known or predicted to be affected by variables that can be localized to specific modules.

CHAPTER 2

Experimental background

2.1 Why rounding?

For this paper, I wish to test the empirical predictions of different grammatical models. I also wish to present results novel in a few different ways: I want to investigate whether lexical, prosodic, and phonological factors affect non-contrastive features in the same way that they affect contrastive features and I will do so with a feature that has not been greatly studied.

I choose, then, to study the [round] feature, which Keyser and Stevens (2006) argues to be a non-contrastive helping feature on many English segment types. In addition to being gradiently realized, [round] fulfills both of the additional criteria listed above. Moreover, the literature discussed in §1.3 clearly supports the fact that phonetic realization may be affected by multiple modules, so if we believe that non-contrastive features strengthen like contrastive ones, we predict that lip rounding degree will be affected by multiple modules. Note that none of these predictions have been previously shown: they are all extrapolations of the literature, and the results of my experiment will confirm or disprove these extrapolations.

For effects of the lexicon, we expect higher frequency words to have less extreme articulation, less lip rounding; we expect words with more phonological neighbors to have more extreme articulation, so more rounding (Gahl, 2008; Gahl et al., 2012).

For effects of prosody, we might expect utterance-initial words to have more extreme articulation and more lip rounding (van Lieshout et al., 2014). Alternatively, we could also predict utterance-final (and thus lengthened) words to have more lip rounding, since a longer realizational duration would allow more time for the lips to round. The prediction here is not immediately clear, though we do expect there to be likely some effect. Following the same logic, we do expect higher

global speech rate to result in less extreme articulation and less rounding (Byrd and Tan, 1996; Pouplier et al., 2017).

For effects of phonology/the phonetics-phonology interface, we expect similar strengthening and lengthening effects: that if a rounded sound is within a stressed syllable, it will be strengthened and be more rounded (Cohn, 1990); if a rounded sound is word-initial, it will be strengthened (Fougeron and Keating, 1997); if a rounded sound is word-final, it will be lengthened and be more rounded. We also expect clear coarticulatory effects — if a rounded sound is adjacent to a rounded vowel, the rounded sound will have more time to realize its rounding gesture and therefore be more rounded **(cite from 203)**.

2.2 Why [ʃ]?

Having motivated rounding as a phonetic feature of interest, we must decide how to study it. Ladefoged and Maddieson (1996) report that rounding appears on four classes of segments in American English: back vowels, [w], [ɹ], and post-alveolar obstruents. If the goal is to examine rounding, why examine [ʃ]?

Of these classes, it is least feasible to examine [ɹ]. This is because [ɹ] is notoriously produced with different articulations by different speakers, and even by the same speaker in different contexts (Mielke et al., 2016). As Nieto-Castanon et al. (2005) find, acoustic properties tend to be relatively preserved — even when articulatory obstacles are present (Mayer and Gick, 2012); as an [ɹ] with a tongue body position which produces a lower F3 would be predicted to have less rounding, it would be difficult to predict degree of rounding with any accuracy on [ɹ] without also knowing the tongue body position. A study focusing on [ɹ]s would require an imaging technique such as ultrasound or EMA.

Rounded vowels pose a different problem: Goldstein (1991) finds that the baseline amount of lip rounding for vowels varies based on vowel height — the degree of rounding of [u] is different from that of [o̞ʊ̟], is different from that of [ʊ]. Furthermore, vowel rounding overall varies both individually and regionally (De Jong, 1995). A study involving rounded vowels would necessitate

video capture, which would complicate data collection and data analysis.[1]

The post-alveolars therefore require the least complex setup to study, since we expect that rounding and COG to track each other well. Among the post-alveolars, it is best to look at fricatives. Whalen and Gick (1998) establishes that tongue position affects spectral resonance of English fricatives; since Keating et al. (1999) find that [ʧ] shows prosodically-driven changes in tongue contact degree, we would expect that a trading relationship between rounding and position might be found for the affricates. These changes in articulation are not observed in the fricatives, removing a potential confound. Among the fricatives, [ʃ] appears in more words: the limited number of words which contain [ʒ] would make it difficult to observe lexical effects.

2.3 Why center of gravity?

I operationalize rounding on [ʃ] as measurable by the center of gravity of that segment. The center of gravity (COG; also referred to as spectral mean, spectral centroid, and first spectral moment) of a sound is an acoustic measure, the weighted sum of the frequencies present in that sound. In general, COG can be impressionistically related to the "pitch" of a sound, even when that sound is voiceless and therefore can carry no f0.

Jongman et al. (2000) find that COG is a useful acoustic measure for the discrimination of English fricatives; all four places of articulation at which fricatives can be made are distinguishable from each other by examining the spectral mean of the middle 40 ms of the fricative. And indeed, cross-linguistically, COG appears to be a robust indicator of fricative place of articulation (Gordon et al., 2002).[2] COG is therefore an important characterization of fricatives and expected to be relatively stable between realizations of the same phone.

However, COG does vary with lip rounding degree because lip rounding lowers frequencies across the board (Lindblom and Sundberg, 1971). This, Keyser and Stevens (2006) argue, is why

[1] I am not aware of any research on variation in [w] rounding degree, but that very fact makes it incautious to run an experiment on [w] without first establishing if variation exists.

[2] Cepstral measures are generally better at uniquely identifying fricatives, but Spinu et al. (2018) do find that midpoint COG is one of the best spectral identifiers.

American English post-alveolar consonants [ʃ, ʒ, ʧ, ʤ] are rounded — rounding is co-opted as a feature to enhance the paradigmatic contrast between American English [ʃ] and [s]. Within speakers, Keating et al. (1999) have shown that the tongue positions for English sibilant fricatives are generally invariant and Kim (2001) reports the same result for Korean fricatives. It is therefore likely that the majority of variation in [ʃ] COG is the result of changes in lip rounding degree.

2.4 Interim summary

Thus far, I have laid out the justification for an experiment where COG is measured on [ʃ] as a proxy for lip rounding degree. This is feasible because there are two major articulatory variables that might impact COG, tongue position and lip rounding. As there is reason to expect that tongue position is essentially invariant for [ʃ], we can use differences in COG as evidence for different degrees of lip rounding.

I have also discussed how [round] and COG are expected to vary as a function of some factors. A crucial point here is that decreased rounding degree represents a type of reduction, whereas increased rounding degree represents strengthening. As such, higher COG is predicted whereever reduction is expected: in high frequency words and at high speech rates. Lower COG is predicted where strengthening is expected: in words with high neighborhood density, in words at the edges of utterances, at the edges of words, and in stressed syllables. Lower COG is also predicted next to rounded vowels, as the adjacent rounding gestures are predicted to allow more time for rounding to be fully realized.

For two reasons have I presented the choice of [round] as the object of study: novelty and extension. This experiment will contribute novel results to the literature, as rounding has not been much studied as a feature that conceivably undergoes strengthening/reduction. It will also show if non-contrastive features, in fact, do undergo strengthening and reduction, as previous research has mostly focused on contrastive features.

CHAPTER 3

Experiment

3.1 Participants and exclusions

Participants were recruited from the undergraduate student population of UCLA. 69 participants were recruited and demographic information was collected in a survey administered at the beginning of the experiment.

Participant data was entirely excluded from the final analysis if they indicated in the self-report that they did not speak English as a native language ($n = 1$), if they did not follow the instructions (for example, if they only spoke the target instead of the frame; $n = 5$), if they did not complete at least 90% of the experiment ($n = 7$), or if they produced poor quality recordings (for example, if their hair repeatedly brushed up against the microphone; $n = 2$). This left us with 56 speakers (female $n = 44$, queergender $n = 1$). Ages ranged from 18 to 32, with a mean of 20.4.

3.2 Stimuli

A wordlist consisting of 269 target words containing /ʃ/ was assembled. To avoid possible influences of roots on derived words (e.g., Sugahara and Turk (2009)), semantically decomposable polymorphemic words were not included. Words with non-/ʃ/ post-alveolars were excluded, as well as any words containing /ɹ/ or /ɚ/. A full list of target words can be found in Appendix A.

Items were created by concatenating the words with two carrier sentences: *"Target" is the word I have just said* and *Now I will say the word "target"*. This created a total of 538 items.

3.3 Procedure

Prior to data collection, participants were shown a list of 88 low-frequency target words and asked if there were any which they were unfamiliar with. These words were then spoken to the participant by the experimenter.[1]

Participants were then fitted with a head-mounted SM10A SHURE microphone and seated in a sound booth. Once the gain had been adjusted to be as high as possible without causing clipping, participants were instructed to first silently read any sentences they were presented with and then speak those sentences out loud. Items were presented in a random order, and each item was presented once. Because of the number of items, participants took a 3 minute break in the middle of the task, at which point the task resumed. The entire task took approximately 45 minutes.

3.4 Data annotation and exclusion

To ensure reproducibility, productions were forced-aligned with the Montreal Forced Aligner McAuliffe et al. (2017). After forced-alignment, the data was inspected by research assistants in Praat (Boersma and Weenink, 2018) using a custom Praatscript (see Appendix C). From each item, we automatically extracted the COG of the middle 40 ms of the segment. Recall that Jongman et al. (2000) found that English fricatives could be distinguished from each other on the basis of this exact measure. This is therefore the most conservative (i.e., likely to be invariant) measure of COG that was supported by the literature. We also extracted the duration of the utterance in order to calculate syllables per second as a proxy for speech rate.

Individual observations were excluded from analysis if the target word was misarticulated (e.g., adjacent vowels were produced incorrectly, stress was incorrect) or if the target word had been misaligned ($n = 3766$). We further discarded utterances whose COG, intensity, or duration measures were more than three standard deviations from the mean. After all exclusions, we obtained a final dataset of 26229 observations.

[1] The choice of 88 target words was because because a pilot experiment demonstrated that showing the participants the entire list of target words was overwhelming and therefore failed to familiarize participants with low frequency words. These 88 targets were chosen because they had a $\text{SUBTLEX}_{\text{US}}$ frequency of < 1; see Appendix B.

3.5 Variables

3.5.1 Independent variables

Items were coded for:

Frequency, which I operationalized as the $\log_{10}$ lexical frequency of the target, as determined from and reported by the $\text{SUBTLEX}_{\text{US}}$ corpus (Brysbaert and New, 2009). This measure ranged from 5.2796 to 0.301 ($\bar{x} = 2.12$). Log frequency was used because of the observation by Hay (2001) that log frequency more accurately tracks frequency effects when considering a large frequency range.

Density, which I operationalized as the adjusted phonological neighborhood Density of the target, as determined by Levenshtein distance, calculated using the tool developed by Vitevitch and Luce (2004). Adjustments were made in order to more closely represent the phonological inventory of California English speakers: the neighborhood densities of target words with /ɔ/ or /ɑ/ were looked up for both phonemes and then combined (words in the neighborhood of both forms were only counted once). Similarly, measures were combined for targets with [ʌ, ə, ɨ] (looked up with [ʌ, ə, ɨ], plus [n̩, m̩, l̩] as appropriate).[2] This measure ranged from 0 to 31($\bar{x} \approx 5.95$).

FirstWord, which I operationalized as a dummy-coded variable which represents if the target word is utterance-initial or utterance-final within the item. "1" represents utterance-initial and "0" represents utterance-final.

SpeechRate, which I operationalized as the global speech rate measured in syllables per second. This measure was calculated by dividing the duration of the utterance by the number of syllables within that utterance; it ranged from 0.56 to 5.52 syl/s ($\bar{x} = 3.02$).

WordPosition, which I operationalized as the position of the [ʃ] in the target word. This was represented as a categorical variable with three levels, word-initial, word-medial, and word-final;

[2] Vitevitch and Luce's tool treats [ʌ, ə, ɨ] as different segments for the purposes of calculation. It also treats [n̩, m̩, l̩] as different from [ən, əm, əl, ...]. This leads to some strange results: the density for *ocean* is different depending on if the word is entered as [oʃən], [oʃɨn], or [oʃn̩]. To deal with this, we calculated the neighborhood densities of every reasonable variant — there was no need to look up [oʃʌn], for example, since [ʌ] was only an option for stressed vowels.

word-medial was the reference level.

InStressedSyl, which I operationalized as a dummy-coded variable which represents if the syllable containing [ʃ] was stressed. "1" represents that the [ʃ] was in a stressed syllable and "0" represents that it was not. Stress degree was not considered due to a lack of potential target words with secondary stress.

RoundedAdj, which I operationalized as a dummy-coded variable which represents if a preceding or following vowel was /u/, /ʊ/, or /oʊ̯/. "1" represents that the [ʃ] was adjacent to at least one rounded vowel and "0" represents that it was adjacent to none.

3.5.2 The assignment of variables to modules

It is clearly critical for us to agree on which modules variables are located in, as it is this organization which provides us with the grounding for an empirical evaluation of these models.[3]

My experiment sorted the previously discussed variables into 3 groups: lexical (Frequency, Density); prosodic (FirstWord, SpeechRate); and phonological (WordPosition, InStressedSyl, RoundedAdj). These groups correspond to the definitions given in §1.2.

I have said earlier that my definition of phonology includes static phonology. This is why WordPosition, InStressedSyl, RoundedAdj are included as phonological variables: these are facts about the word's featural specifications and stress pattern, facts which are tied to the identity of the word.

Some might wonder why Density is considered a lexical variable and not a phonological one. It is true that (Levenshtein) neighborhood density must be calculated with reference to some sort of phonological object. However, Vitevitch and Luce (2016) defines neighborhood density as the "set of similar-sounding form-based representations that are activated in memory", which does not necessitate that those representations are phonologically active. Rather, those representations could be just phoneme representations with no granularity. Thus, as defined, Density and Frequency are of the same type — broadly, non-encoded information about words residing in the lexicon.

[3]Note that for our purposes, this is identical to asking where different pieces of information are stored.

The two prosodic variables are clearly different from the phonological and lexical variables. WordPosition is a phrasal positional prominence variable, but what about global speech rate? Local speech rate is, of course, correlated with, among other factors, segmental context — should global speech rate be considered a separate type of variable entirely? I do not know of any work which seeks to answer this question. However, I suggest that it is reasonable to assume that global speech rate as a prosodic variable, as global speech rate depends less on the segmental content of an utterance and more on something like style. It can also be consciously varied, as parenthetical statements frequently have faster speech rate and a different pitch range (Local, 1992).[4] Both of these suggest that global — or at least non-local — speech rate should be considered a prosodic variable.

3.5.3 Other variables

Minimal models were compared to determine the maximal random effects structure possible. As a result, Word and Speaker were included as random effects; it was not possible to include more random effects without causing convergence errors.

In addition, Intensity was included as a covariate in every model because a pilot study found a significant effect of intensity on COG. Visual analysis of the non-pilot data showed a continued strong correlation between the two, prompting its inclusion. Note that intensity is predicted to vary as a function of other variables of interest, e.g., FirstWord. However, intensity also varies totally randomly, as speakers may spontaneously decide to use vary their vocal effort at any point. The inclusion of Intensity as a covariate weakens the effect of variables like FirstWord, but allows us to be more confident in the effects that are found to be significant.

[4]Indeed, just as pitch range is constant within an intermediate phrase, so too is speech rate: imagine a parenthetical "although it's not like she even likes marmalade" where speech rate changes halfway through.

3.6 Results

3.6.1 Analysis

The data were analyzed with linear mixed effects models using the `lme4` package in R (Bates et al., 2015). Model comparison was conducted by using the `anova()` function to generate χ^2 values, as all models were in strict subset relationships.

3.6.2 Main effects models

variable	effect on COG	p value
Frequency	—	$p = .918$
Density	decrease	$p < .001$
FirstWord	increase	$p < .001$
SpeechRate	increase	$p < .001$
WordPosition		
word-initial	decrease	$p < .001$
word-final	decrease	$p < .001$
InStressedSyl	decrease	$p < .001$
RoundedAdj	decrease	$p < .001$

Table 3.1: Summary of single variable model results

I wished to first examine the effect of each variable on COG. In effect, I wanted to extend the findings of previous research on how these variables affect phonetic production to COG. To this end, I created a series of models, each of which only included one fixed effect, the random effects, and the covariate. I further included random slopes of variable|Speaker and variable|Word for each model except for the phonological variables, which could not have a variable|Word random slope. The results for these models are shown in Table 3.1.

These models show that, when considered one by one, every variable expected to effect COG was significant except for lexical frequency. In addition, every significant variable had an effect in

the predicted direction. These effects were significant even when controlled for multiple comparisons ($\alpha = .007$). We therefore have grounds to believe that our choice of variables was correct, and that we were correct in believing that non-contrastive features — or at least [round] — also strengthen in the same way as contrastive features.

I will mention two results which might be surprising. First, lexical frequency was found to have have a non-significant effect on COG ($p = .898$). Second, the effect of FirstWord on COG was positive. Although it may seem counterintuitive that the effect of FirstWord on COG should be positive, this is in fact what we expect given that FirstWord is dummy-coded. The model shows that the COG of a word is higher in utterance-initial position *than in utterance-final position*. Recall that this was the one variable where the predicted direction was unclear — this result supports that the hypothesis that initial-strengthening effects affect fewer segments than final-lengthening effects.[5]

variable	effect on COG	p value
Frequency	—	$p = .430$
Density	—	$p = .684$
FirstWord	increase	$p < .001$
SpeechRate	increase	$p < .001$
WordPosition		
word-initial	decrease	$p < .001$
word-final	decrease	$p < .001$
InStressedSyl	—	$p = .837$
RoundedAdj	decrease	$p < .001$

Table 3.2: Summary of model with all variables as main effects

I also created a baseline model which included every variable as a main effect, the random effects, and the covariate. This model serves as the second part of my extension of previous re-

[5]Unfortunately, the frame sentence introduces a confound in that utterance-final, word-initial [ʃ] was always adjacent to an [ɔ˞], which is rounded. They were not strictly adjacent, as a [d] always intervened between the vowel and [ʃ], but the proximity might be a problem. To assuage fears that the effects of FirstWord might be partially driven by vowel quality, a second model was examined which only contained words with non-word-initial [ʃ] ($n = 19313$). This model found an effect with the same positive direction ($p < .001$).

search: few experiments have simultaneously manipulated as many variables as mine. This model thus serves as a safeguard against finding spurious results, as the effects of variables which may seem significant in isolation may be in fact better explained by other variables once all variables are included in one model. The results for this model are shown in Table 3.2. This model found no effect of Frequency ($p = .430$), Density ($p = .684$), or InStressedSyl ($p = .837$); see §4.2 for some discussion of this results. The variables which were found to be significant had effects in the predicted directions.

3.6.3 Models

In order to conduct model comparison, I first created a flat model which includes every possible main and interaction effect, with no limit on the number of variables in an interaction effect. This model, which is atheoretical and corresponds to no proposal previously outlined, is a completely flat model in that every variable is free to interact with every other variable.

The flat model has two desirable properties: first, it is a strict superset of every model I examined. Second, it is the best possible model in the sense that it includes every possible predictor. These properties allow us to conduct direct comparisons of the flat model to the theoretically-motivated models using χ^2 values. A result of $\chi^2 > .05$ will indicate that the flat model is not significantly better than the other model, highlighting those models as better than the flat model in a different sense: they explain the data as well as the flat model, but with fewer parameters.

Note that not all of the models discussed in §1.2 can be statistically distinguished from one another. Given the organizational schema given in §1.4, the Dell planning model and Weaver++ have the same statistical counterpart. The same is true for TADA and ToBI.

For full model specifications, see Appendix D.

3.6.3.1 Weaver++: Lexicon on phonology on prosody

The statistical model which corresponds to the Dell (1986) model and Weaver++, henceforth simply the Weaver model, only permits prosodic variables to be main effects, though it allows phonological variables to be in interaction effects so long as at least one prosodic variable is also included.

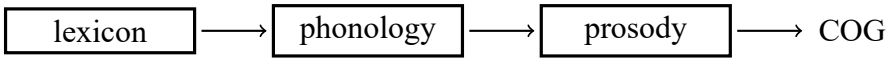

Figure 3.1: The Dell (1986) model and Weaver++, recast as a statistical model. Lexicon=Frequency, Density; Prosody=FirstWord, Speechrate; Phonology=WordPosition, InStressedSyl, RoundedAdj

Similarly, lexical variables may be included so long as at least one phonological and one prosodic variable is also included. Its model specification is of the form

```
COG ~ Prosody + Prosody:Prosody +
      Prosody:Phonology + Prosody:Phonology:Phonology + ...
      Prosody:Phonology:Lexicon + Prosody:Phonology:Lexicon:Lexicon + ...
      Intensity + (1|Word) + (1|Speaker)
```

The flat model is a better model than the Weaver++ model ($\chi^2 < .0001$), so we will not return to this model.

3.6.3.2 Keating and Shattuck-Hufnagel: Lexicon on prosody on phonology

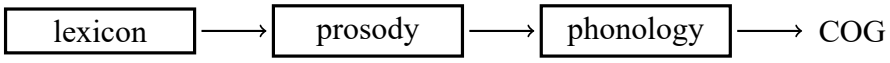

Figure 3.2: A schematic recasting of the Keating and Shattuck-Hufnagel (1989) model as a statistical model. Lexicon=Frequency, Density; Prosody=FirstWord, Speechrate; Phonology=WordPosition, InStressedSyl, RoundedAdj

The statistical model which corresponds to the Keating and Shattuck-Hufnagel (1989) model, henceforth the KSH model, only permits phonological variables to be main effects, though it allows prosodic variables to be in interaction effects so long as at least one phonological variable is also included. Similarly, lexical variables may be included so long as at least one phonological and one prosodic variable is also included. Its model specification is of the form

```
COG ~ Phonology + Phonology:Phonology + Phonology:Phonology:Phonology +
```

```
Phonology:Prosody + Phonology:Prosody:Prosody + ...
Phonology:Prosody:Lexicon + Phonology:Prosody:Lexicon:Lexicon + ...
Intensity + (1|Word) + (1|Speaker)
```

The flat model is not a better model than the KSH model ($\chi^2 = .1921$), so we will return to this model later.

3.6.3.3 TADA: Lexicon on phonology and prosody

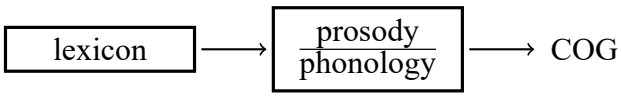

Figure 3.3: The TADA model and ToBI, recast as a statistical model. Lexicon=Frequency, Density; Prosody/Phonology=FirstWord, Speechrate, WordPosition, InStressedSyl, RoundedAdj

The statistical model which corresponds to the TADA and ToBI models, henceforth the TADA model, permits both prosodic and phonological variables to be main effects, though it allows lexical variables to be in interaction effects so long as at least one prosodic or phonological variable is also included. Its model specification is of the form

```
COG ~ Prosody + Phonology + Prosody:Phonology + ...
Prosody:Lexicon + Phonology:Lexicon + Prosody:Lexicon:Lexicon + ...
      Intensity + (1|Word) + (1|Speaker)
```

The flat model is not a better model than the TADA-inspired model ($\chi^2 = .1194$), so we will return to this model later.

3.6.3.4 Generative Phonetics: Lexicon and prosody on phonology

The statistical model which corresponds to the Generative Phonetics models, henceforth the GP model, only permits phonological variables to be main effects, though it allows lexical and prosodic variables to be in interaction effects so long as at least one phonological variable is also included. Its model specification is of the form

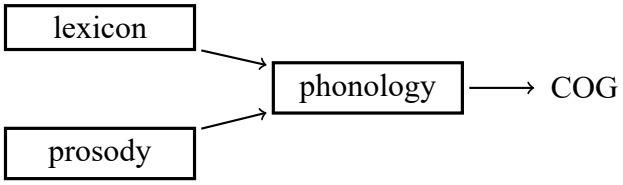

Figure 3.4: Generative Phonetics models, recast as a statistical model. Lexicon=Frequency, Density; Prosody=FirstWord, Speechrate; Phonology=WordPosition, InStressedSyl, RoundedAdj

```
COG ~ Phonology + Phonology:Phonology + Phonology:Phonology:Phonology +
      Phonology:Prosody + Phonology:Prosody:Prosody + ...
      Phonology:Lexicon + Phonology:Lexicon:Lexicon + ...
      Intensity + (1|Word) + (1|Speaker)
```

The flat model is a better model than the GP model ($\chi^2 < .0001$), so we will not return to this model, as this shows that the KSH and TADA models better explains the data than this one.

3.6.3.5 KSH vs. TADA

Only two models, the KSH and TADA model, had $\chi^2 > .05$ when compared to the flat model, meaning that they explained the data as well as the flat model even though they have fewer predictors.

The KSH and TADA model are in a subset relationship: the TADA model permits prosodic main effects, whereas the KSH model does not. The TADA model also permits lexical-prosodic interaction effects without the inclusion of a phonological variable. However, the TADA model is not a better model than the KSH model ($\chi^2 = .2904$), so the KSH model explains the data as well as the TADA model even though it has fewer predictors. By standard reckonings of model performance, the KSH model therefore is our preferred model.

CHAPTER 4

Discussion

4.1 General discussion

In §3.6, we saw that the best statistical model[1] was the one corresponding to the Keating and Shattuck-Hufnagel model. And, as discussed in §1.2.1.3, the KSH model has an essentially hierarchical, feed-forward organization. We must examine two questions here, both regarding hierarchy.

First, there is the question of sequence. Both the KSH and Weaver models are feed-forward and differ only in how their variables were arranged; the KSH model places prosodic variables before phonological ones, and the Weaver model, phonological before prosodic. In the statistical analysis, the essential difference between these models was the decision of which main effects to include. Hierarchically-minded models, such as the KSH and Weaver models, are forced to not include certain main effects as a result of the schema outlined in §1.4. In general, while some the excluded main effects were ones that were not significant — for example, neither of those models included lexical main effects, which we found to be nonsignificant in the baseline model — both of these models necessarily did not include main effects which *had* been found to be significant in the baseline. As the KSH model places prosody before phonology, it (ironically) cannot include prosodic main effects of speech rate or of utterance initiality. And of course, the Weaver model can include those prosodic main effects, but since it places prosody *after* phonology, it cannot include phonological main effects of position in word, being in a stressed syllable, or being adjacent to a rounded vowel.

This comparison between the KSH and Weaver models indicates that phonological variables

[1] Best, here, is used in the sense that the best model is one which uses fewer predictors than the most complex model but explains the data approximately as well.

have more direct influence on [ʃ] COG than prosodic ones. More generally, if we consider phonological variables to be specifications which are to be implemented, then that specification intuitively should be more important than other factors which only modify specifications.

Second, there is the question of feed-forwardness. The KSH model, which is feed-forward, is a better model than the flat model. I claim this because the KSH model is more parsimonious than the flat model. Here, in the data generated by this production experiment, upstream main effects are not found to be sufficiently meaningful predictors of COG, nor "upstream" interaction effects (e.g., Frequency×FirstWord). Model comparison penalizes the inclusion of insufficiently meaningful predictors, because most any predictor can be added to any model and improve model fit. The success of one model over another is therefore a balance between the desire to have the best fit as possible, and the desire to include as few predictors as possible. The flat model simply contains too many extraneous predictors.

The relative successes and failures of the other models can be similarly related to the factors that they were not able to take into account. The TADA model, for example, attempts to account for prosodic main effects as well as phonological ones. It also predicts and allows lexical-prosodic interaction effects. It turns out that this is not necessary to explain the data.

4.2 Implications for phonetics research

At the broadest level, my results show that non-contrastive features prosodically strengthen in the same way as contrastive features. My results also suggest that, at least for [ʃ], the relationship between rounding and COG is as predicted: the direction of the effects that were found matched the results we would expect if rounding both undergoes strengthening and is a direct predictor of COG. Finally, recall that previous research has shown that tongue position is invariant when it comes to postalveolar fricatives, and it is usually assumed that the featural difference between /s/ and /ʃ/ is [anterior], a feature whose articulatory reflex is in tongue configuration. My results therefore show that a non-contrastive feature may strengthen even when a contrastive feature does not.

4.2.1 Factors affecting [ʃ] COG

In the model where all of the variables are included as main effects, I have found evidence for some phonological effects: [ʃ] in word-medial position has a higher COG and less rounding than [ʃ] in either word-initial or word-final position; [ʃ] next to a rounded vowel has a lower COG and more rounding than when not. I did not find an effect of stress — as [round] is not a contrastive feature in English, this finding follows Cho et al.'s (2015) observation that stress-related enhancement appears to target paradigmatic contrasts.

As predicted, I found that increased speech rate results in higher COG and less rounding. I also found that [ʃ] in an utterance-initial word has a higher COG and less rounding than [ʃ] in an utterance-final word. I suggest that this is the result of asymmetries between strengthening effects, which affect few segments, and lengthening effects, which affect more: the effect of being in an utterance-initial word may be more salient for words with word-initial [ʃ].[2]

I found no evidence for a direct effect of lexical frequency, nor of neighborhood density. It appears that rounding degree on [ʃ] is unaffected by lexical factors.

4.2.2 Lack of lexical effects

In the baseline model, which contained every variable as a main effect, I did not find any effects of lexical frequency or neighborhood density. At first blush, this is a highly surprising result, as previous research has repeatedly found that lexical frequency (Gahl, 2008; Ernestus et al., 2006; Baker and Bradlow, 2009) and neighborhood density (Gahl et al., 2012; Scarborough, 2013) affect phonetic realization.[3] Why might my results indicate otherwise?

First, it might be that the design of the experiment precludes finding significant lexical effects. I note that the random effect of Word is somewhat confounded with Frequency and Density, although

[2]Unfortunately, I am unable to test this hypothesis directly with the data I have collected. If I run a model on only words with word-initial [ʃ], I still find the same result, although with a reduced effect size. However, this model has the opposite problem in that it may systematically underrepresent the effect of being utterance-final. To best test this, an utterance-medial condition is necessary, or baring that, a dataset comprised only of very short words with initial [ʃ].

[3]While it is true that no research has examined the effects of these variables on [ʃ] COG, the breadth of the effects which *have* been found suggest that lexical variables might have an effect on nearly every aspect of production.

it is not entirely so. 112 out of 268 words share a Frequency value with at least one other word, and only 4 out of 268 words have unique neighborhood density values. If the problem is one of overconservative estimation — that the random effect of Word explains so much of the variation that Frequency and Density are extraneous — then that problem should be greater with Frequency than Density, but as we find no effect of Frequency or Density, this is not a satisfactory explanation. It might also be that the effects of Frequency and Density on COG are simply too small to be found. There is no easy way to conduct power analysis on a linear mixed effects model, but with over 26000 observations and 17 estimators, this possibility is unlikely.

Some insights might be gleaned by examining the KSH model. The KSH model does not include lexical main effects, so there are of course no significant main effects. However, it does find significant interaction effects which include lexical factors — for example, it finds that the effect of being next to a rounded vowel is lessened when both neighborhood density and speechrate are high (Density×SpeechRate×RoundedAdj). One possibility is therefore that when lexical effects have been found in previous research, the explanatory power of that effect might have been more parsimoniously understood as an interaction effect between that lexical factor and some other factor.

If the above is true, then the success of the KSH model also implies that where prosodic "main" effects have been found, they might have been also more parsimoniously understood as an interaction between prosodic and phonological variables. As a potential example of this, Cho and Keating (2009) examine how a number of different acoustic and physiological measures might vary as a function of boundary strength, stress degree, and accentedness. For every measure but for one, where a main effect of boundary strength is found, either an interaction between boundary strength and stress degree or one between boundary strength and accentedness was also found. Although their study employed anovas instead of lmers and had different goals than this one, the results are suggestive — were their study to be replicated on the same scale as this one, perhaps they would find that interaction effects alone would be adequate to explain the data.

4.3 Implications for phonological models/Generative Phonetics

Generative Phonetics models use phonological grammars to predict phonetic output. Goldwater and Johnson (2003) and Becker et al. (2017) note that logistic regression models are a subset of maximum entropy models, with predictors corresponding to constraints. As such, the models explored in this paper are all Generative Phonetics models with externally imposed constraints on the information which can be incorporated into the grammar and how. The results of this experiment can therefore been seen as a recommendation for the types of constraints that should be included in future Generative Phonetics models.

As the only difference between the KSH and the GP models is that the KSH model does not allow lexical variables to directly interact with phonological ones, my results suggest that Generative Phonetics grammars should not directly incorporate lexical scaling of the type proposed in Coetzee and Kawahara (2013). Instead, this experiment suggests that lexical scaling is more complex than previously assumed, and that future models should necessitate that lexical effects also scale with prosodic ones.

That said, one major difference between this paper's models and other such models, e.g., Flemming and Cho's (2017), is that the models herewithin are required to have every possible predictor(/constraint) that is permitted, where others select the predictors to be included. It could be argued that the models in this paper do not strictly follow the conventions of a proper Generative Phonetics analysis, as model comparison, when done on maximum entropy models, is typically to remove extraneous constraints(/predictors).

However, in the case of [ʃ] COG, there is no *a priori* reason to not include every possible predictor in a Generative Phonetics analysis, as it is not possible to know which predictors are likely to be meaningful. To the contrary, there is in fact every reason to include them, as a goal of this project was to understand which theories predict the best predictors. I therefore argue that models presented here represent systems which approximate full grammars. To put it differently, this project presents different Cons predicted by different architectures; it then tries to use the results from those Cons to evaluate those architectures. This novel goal motivates the different model construction procedure.

4.4 Implications for models of grammar

Figure 4.1: One branch of the Chomskyian T-model, adapted from Halle and Marantz (1993)

Although this paper presents only one experiment focusing on one phonetic output, the results do suggest that the structure of models of grammar should be not flat, but rather hierarchical and feed-forward. More generally, the structure of the KSH model can be mapped onto the structure of the most prominent feed-forward model that has been proposed, Chomsky's (1965) *T-model*, a portion of which is shown in Fig. 4.1.[4] Going forward, I will discuss the T-model with the understanding that my results supporting the KSH model support the T-model.

In essence, the T-model captures the fact that certain modules seem to rely on the outputs of other modules, and that this relationship is asymmetric. It makes the prediction that there are interfaces between adjacent modules and that there is no interface between non-adjacent modules. To a large extent, this seems to bear out: to grossly oversimplify, prosodic structure can be well-explained by a posited interface between syntax and phonology, but there seems to be no glaring need for a semantics-phonology interface. More specifically, this experiment's results accord with the T-model's predictions. This is not without its problems. Given the strictest interpretation of the T-model, Jackendoff (1997) argues that whatever information downstream modules need, that information must first be "invisibly dragged" through upstream modules — that that information must exist within a module but not be used by it.

Because of this and other objections, Sadock (2012) offers an *automodular model*, which has the same properties as the flat model which I compare other models to in §3.6.3. His book focuses on syntax, but the principles behind his model are generalizable; I present such a model in Fig. 4.2. In this model, each primary module receives the same conceptual/lexical information and processes it without reference to each others's output. A crucial element is the inclusion of an explicit interface

[4]It is true that, while the FirstWord prosodic variable is clearly the output of syntax, this is not true of SpeechRate. However, I return to the discussion in §refintro:assignment and the observation that speech rate is bound by parentheticals, which are also clearly the output of syntax.

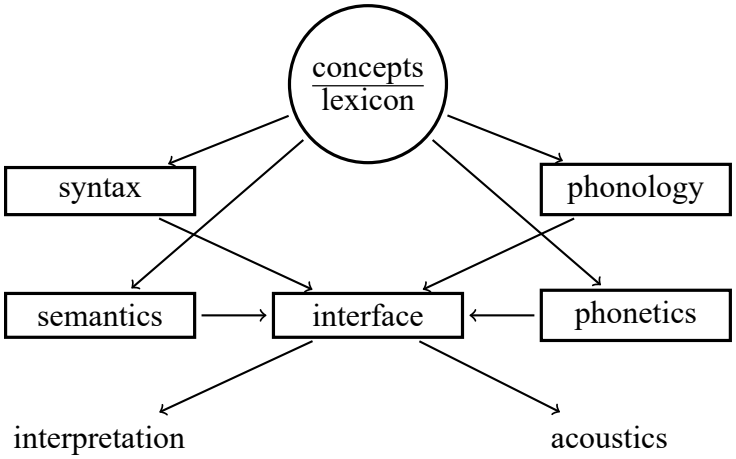

Figure 4.2: A generalization of Sadock's model. Note the interface component which checks the representations outputted from each module to ensure that they are compatible with the same utterance

module which checks the representations of each of the standard modules and ensures that they are compatible with each other. This interface is conceived of as a filter which has specific constraints that must be satisfied. The effect of such a component is that each module of the Grammar is able to influence every other module — this structure, interpreted as a regression model, is the same as the flat model. Through the interface, a particular type of correspondence can be forced between, e.g., syntactic and phonetic representations without needing for phonological representation to be included.

Although an automodular model avoids the issue of Jackendoff's informational dragging, the problems that the interface module poses are substantial. Where are the semantics-phonology interface phenomena, the semantics-phonetics phenomena? The interface module is capable of producing them, but they have not been observed. Of course, there is the syntax-phonetics interface, which this project instantiates as prosodic-phonetic phenomena. The results of my experiment argue against a direct effect of prosody on phonetics, and thus against that interface as well.

Due to the limited scope of this project, these results are not definitive evidence for either type of model. However, there has been little research into the empirical predictions of different grammatical architectures, which I assert to be both important to the field and a valuable director of future research: if the T-model (or something like it) is the correct model of grammar, we must

understand the nature of Jackendoff's informational dragging; if the automodular model is correct, we must reckon with the effects of the unexplored interfaces.

CHAPTER 5

Conclusion

In this paper, I have attempted to discern if feed-forwardness is a desirable trait in models of grammar. I have done so by noting that different models of grammar predict different systems of interactions between different modules and then conducting an experiment which allowed me to differentiate between some of those models. My results have confirmed many of the predictions that are made about how various prosodic and phonological factors should affect [ʃ] center of gravity. More broadly, they show that a hierarchical architecture better explains the variation of this variable than a flat architecture.

This experiment evinces a need for thorough, data-driven analysis of linguistic phenomena. It also carries many implications for different fields of research: I have discussed the implications of this experiment for future research in phonetic variation, also engaging in some speculation on the nature of lexical and prosodic effects — *viz.* that they are indirect and act through their effects on other variables. In addition, the end model implies that lexical scaling in Generative Phonetics grammars should not be done directly, but through interactions with prosodic variables as well, which suggests that the same might be true of Maxent grammars in general.

APPENDIX A

Items

she	condition	shout	population
should	issue	suspicious	vicious
show	social	flash	emotion
shut	vacation	foundation	smash
wish	shoe	bush	shield
shoot	ocean	audition	institution
special	shape	commission	shaft
finish	gosh	shave	potion
ship	shock	species	suspicion
fish	location	sheep	shelf
station	official	session	shovel
english	shift	shell	negotiate
cash	flesh	shove	notion
position	delicious	chef	initial
push	shadow	sheet	establish
patient	shine	dish	compassion
shop	nation	shed	physician
mission	passion	tissue	dash
shame	motion	shaw	shade
chicago	fashion	execution	ammunition
wash	shy	punish	destination
shake	bishop	ash	accomplish

michigan	chute	coalition	sheen
expedition	hash	distinguish	shunt
shack	intuition	amish	succession
constitution	initiate	shabby	absolution
shuttle	plantation	lash	echelon
ambush	bash	shea	ovation
ambition	lotion	swish	extinguish
evolution	vanish	malicious	quiche
shoo	nauseous	pistachio	shah
caution	shin	sham	diminish
publish	deposition	sheik	vocation
stash	shuffle	shilling	cliche
salvation	geisha	aviation	goulash
efficient	initiative	shawl	beneficial
mustache	technician	banish	exposition
edition	shaman	concession	lavish
ambitious	fascist	judicial	shingle
sufficient	slash	shank	dalmatian
vanquish	douche	shindig	shoddy
mash	chic	clash	yiddish
squash	fetish	tush	cushy
sushi	militia	posh	mesh
leash	anguish	socialist	galoshes
petition	cushion	bashful	shogun
shampoo	disposition	whoosh	bushel
splash	commotion	abomination	cavendish
devotion	commotion	sash	fellatio
tuition	shag	insatiable	skittish
schultz	pollution	constellation	cache

fictitious	sedition	specious	slosh
blemish	shanty	aficionado	babushka
chauvinist	fuchsia	panache	gnash
hogwash	socialite	shoal	minutia
deficient	gauche	cachet	slapdash
demolish	schist	acacia	pastiche
audacious	touche	languish	shenanigan
auspicious	omniscient	loquacious	aleutian
chalet	quotient	mosh	phoenician
shun	demotion	munition	sheaf
chablis	embellish	shim	ashkenazi
chiffon	elocution	locomotion	syncopation
sheath	shuck	shekel	losh
bushy	swoosh	slipshod	eschew
abolish	tosh	mackintosh	
astonish	cashew	potash	
bangladesh	quash	shinto	

APPENDIX B

Pre-presented items

fictitious	shogun	phoenician	babushka
shoal	chauvinist	shuck	deficient
skittish	losh	ashkenazi	languish
locomotion	shunt	panache	fellatio
hogwash	goulash	insatiable	cushy
minutia	aleutian	quiche	echelon
shah	cavendish	shun	shenanigan
cashew	pastiche	lavish	sheen
slosh	potash	cachet	sheath
slapdash	munition	shim	astonish
acacia	mosh	elocution	touché
absolution	cache	demotion	slipshod
quotient	shanty	syncopation	tosh
embellish	mackintosh	swoosh	shinto
gnash	chablis	galoshes	shekel
specious	aficionado	quash	bangladesh
sedition	shoddy	auspicious	blemish
cliché	dalmatian	chalet	shingle
abolish	loquacious	sheaf	ovation
succession	chiffon	fuchsia	eschew
schist	bushel	exposition	yiddish
gauche	vocation	audacious	omniscient

APPENDIX C

COG Praatscript

```
#Opens all the files in a folder and an associated
#textgrid folder. Then finds the COG of a specified segment
#and outputs to a file
#source_directory is to be written with "\"s

#Author: Z.L. Zhou
#        UCLA
#        2019-02

#        Cobbled together from a script by Bert Remijsen
#        and one by Katherine Crosswhite



form Calculate COG for a specific segment
comment Speaker ID
word speaker_ID

comment Directory of sound files
text sound_directory wavs\
sentence Sound_file_extension .wav
comment Directory of TextGrid files
text textGrid_directory textgrids\
```

```
sentence TextGrid_file_extension .TextGrid
comment The label of segments to be measured, and the tier in the TextGrid:
word the_label SH
integer the_tier 2
comment Path of file to write results to
text the_directory cogresults.csv
comment Length of window over which spectrogram is calculated:
positive length 0.005
comment Play sound?
choice playit 1
button yes
button no
endform


Create Strings as file list... list 'sound_directory$'*'sound_file_extension$'
numberOfFiles = Get number of strings
for ifile to numberOfFiles
select Strings list
filename$ = Get string... ifile
Read from file... 'sound_directory$''filename$'
sound = selected("Sound")
soundname$ = selected$ ("Sound", 1)
gridfile$ = "'textGrid_directory$''soundname$''textGrid_file_extension$'"
Read from file... 'gridfile$'
textgrid = selected("TextGrid")

counter = 0
select 'textgrid'
```

```
finishing_time = Get finishing time
nlabels = Get number of intervals... 'the_tier'
for label from 1 to 'nlabels'
select 'textgrid'
labelx$ = Get label of interval... 'the_tier' 'label'
if (labelx$ = the_label$)
counter = counter + 1
file_b = Get end point... 'the_tier' 1
file_e = Get starting point... 'the_tier' 'nlabels'
file_length = file_e - file_b
n_b = Get starting point... 'the_tier' 'label'
n_e = Get end point... 'the_tier' 'label'
n_md = ('n_b' + 'n_e') / 2
call measurecog 'n_b' 'n_e' 'n_md' 'name$'
endif
select 'textgrid'
plus 'sound'
endfor

Remove

endfor
select Strings list
Remove

procedure measurecog n_b n_e n_md name$
#first get middle 40ms COG
n_mdplus20 = 'n_md' + 0.02
n_mdmins20 = 'n_md' - 0.02
```

```
select 'sound'
Extract part... 'n_mdmins20' 'n_mdplus20' rectangular 1.0 0
midfortyintensity = Get intensity (dB)
midfortysound_part = selected("Sound")
To Spectrum... 0
midfortyspectrum = selected("Spectrum")
midfortycog = Get centre of gravity... 2.0

#then get middle 60% COG
n_d = ('n_e' - 'n_b') / 10
n_bplus15 = 'n_b' + (1.5 * 'n_d')
n_emins15 = 'n_e' - (1.5 * 'n_d')

select 'sound'
Extract part... 'n_bplus15' 'n_emins15' rectangular 1.0 0
intensity = Get intensity (dB)
sound_part = selected("Sound")
To Spectrum... 0
spectrum = selected("Spectrum")
cog = Get centre of gravity... 2.0

#then get middle 90% COG
n_bplus5 = 'n_b' + (0.5 * 'n_d')
n_emins5 = 'n_e' - (0.5 * 'n_d')
select 'sound'
Extract part... 'n_bplus5' 'n_emins5' rectangular 1.0 0
completeintensity = Get intensity (dB)
completesound_part = selected("Sound")
```

```
To Spectrum... 0
completespectrum = selected("Spectrum")
completecog = Get centre of gravity... 2.0


# display spectrogram.
Erase all
Font size... 14
display_from = 'n_b' - 0.15
if ('display_from' < 0)
display_from = 0
endif
display_until = 'n_e' + 0.15
if ('display_until' > 'finishing_time')
display_until = 'finishing_time'
endif
play_from = 'n_b' - 1
if ('play_from' < 0)
play_from = 0
endif
play_until = 'n_e' + 1
if ('play_until' > 'finishing_time')
play_until = 'finishing_time'
endif
select 'sound'
To Spectrogram... 'length' 4000 0.002 20 Gaussian
spectrogram = selected("Spectrogram")
Viewport... 0 7 0 3.5
Paint... 'display_from' 'display_until' 0 4000 100 yes 50 6 0 no
```

```
Viewport... 0 7 0 4.5
select 'textgrid'
Black
Draw... 'display_from' 'display_until' no yes yes
One mark bottom... 'n_md' yes yes yes
rcog = round('cog')
Text top... no Tracker output -- COG: 'rcog'


## display the spectrum, with Ltas and LPC
select 'spectrum'
Viewport... 0 7 4.5 8
Draw... 2500 6000 0 80 yes
To Ltas (1-to-1)
ltas = selected("Ltas")
Viewport... 0 7 4.5 8
Draw... 2500 6000 0 80 no bars
Marks bottom every... 1 500 yes yes no
Marks bottom every... 1 250 no no yes
select 'sound'
To LPC (autocorrelation)... 18 0.025 0.005 50
lpc = selected("LPC")
To Spectrum (slice)... 'n_md' 20 0 50
Rename... LPC_'name$'
spectrum_lpc = selected("Spectrum")
select 'lpc'
Remove
select 'spectrum_lpc'
Line width... 2
```

```
Draw... 2500 6000 0 80 no
Line width... 1
Text top... no Spectrum, Ltas(1-to-1), LPC(autocorrelation), all three overlaid

# play sound
if (playit = 1)
select 'sound'
Extract part... 'play_from' 'play_until' Hanning 1 no
Play
Remove
endif

beginPause: "Does this token have problems?"
boolean: "Alignment problem", 0
boolean: "Pronunciation problem", 0
clicked = endPause: "All is well", "Continue", 2, 1

if clicked = 1
alignment_problem = 0
pronunciation_problem = 0
endif


# write results to file
if fileReadable (the_directory$)
appendFileLine: the_directory$, speaker_ID$, ",", soundname$, ",", file_length, ",",
else
writeFileLine: the_directory$, "speaker,filename,length,COG,COG40ms,COG90,int,int40ms
appendFileLine: the_directory$, speaker_ID$, ",", soundname$, ",", file_length, ",",
```

```
endif

select 'sound_part'
plus 'spectrum'
plus 'midfortysound_part'
plus 'midfortyspectrum'
plus 'completesound_part'
plus 'completespectrum'
plus 'spectrogram'
plus 'ltas'
plus 'spectrum_lpc'
Remove
endproc
```

APPENDIX D

Model specifications

D.1 Baseline model

This model includes each variable as only a main effect.

```
COG ~ Intensity + (1|Speaker) + (1|Word) +
 Frequency + Density +
 FirstWord + Speechrate +
 WordPosition + InStressedSyl + RoundedAdj
```

D.2 Flat model

This model includes each variable as a main effect, as well as every interaction effect possible.

```
 COG ~ Intensity + (1|Speaker) + (1|Word) +
  Frequency * Density *
  FirstWord * Speechrate *
  WordPosition * InStressedSyl * RoundedAdj
```

D.3 Keating & Shattuck-Hufnagel model

This model represents a feed-forward architecture where the order of the modules is lexicon, prosody, phonology.

```
COG ~ Intensity + (1|Speaker) + (1|Word) +
```

```
 Frequency * Density *
 FirstWord * Speechrate *
 WordPosition * InStressedSyl * RoundedAdj -
(Frequency * Density * FirstWord * Speechrate) -
(Frequency * Density * WordPosition * InStressedSyl * RoundedAdj) +
 WordPosition * InStressedSyl * RoundedAdj
```

D.4 Weaver++ model

This model represents a feed-forward architecture where the order of the modules is lexicon, phonology, prosody.

```
COG ~ Intensity + (1|Speaker) + (1|Word) +
 Frequency * Density *
 WordPosition * InStressedSyl * RoundedAdj *
 FirstWord * Speechrate -
(Frequency * Density * WordPosition * InStressedSyl * RoundedAdj) -
(Frequency * Density * FirstWord * Speechrate) +
 FirstWord * Speechrate
```

D.5 TADA model

This model represents an architecture where lexicon comes before phonology and prosody, which are in the same module.

```
COG ~ Intensity + (1|Speaker) + (1|Word) +
 Frequency * Density *
 FirstWord * Speechrate *
 WordPosition * InStressedSyl * RoundedAdj -
(Frequency * Density)
```

D.6 GP model

This model represents an architecture where lexicon and prosody come before phonology, but lexicon and prosody do not interact.

```
COG ~ Intensity + (1|Speaker) + (1|Word) +
 Frequency * Density *
 FirstWord * Speechrate *
 WordPosition * InStressedSyl * RoundedAdj -
(Frequency * Density * FirstWord * Speechrate)
```

Bibliography

Anderson, Mark D, Janet B Pierrehumbert, and Mark Y Liberman. 1984. Synthesis by rule of English intonation patterns. In *ICASSP'84. IEEE International Conference on Acoustics, Speech, and Signal Processing*. 77–80.

Aylett, Matthew and Alice E Turk. 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.

Baker, Rachel E and Ann R Bradlow. 2009. Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech* 52:391–413.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48.

Becker, Michael, Lauren Eby Clemens, and Andrew Nevins. 2017. Generalization of French and Portuguese plural alternations and initial syllable protection. *Natural Language & Linguistic Theory* 35:299–345.

Beckman, Mary E and Julia B Hirschberg. 1994. The ToBI annotation conventions.

Beckman, Mary E and Janet B Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3:255–309.

Boersma, Paul and David Weenink. 2018. Praat.

Browman, Catherine and Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology* 3:219–252.

Browman, Catherine and Louis Goldstein. 1990. Tiers in Articulatory Phonology, with some implications for casual speech. In John Kingston and Mary E Beckman, eds. *Papers in laboratory phonology I: Between the grammar and physics of speech*. 341–376. Cambridge: Cambridge University Press.

Browman, Catherine and Louis Goldstein. 1992. Articulatory Phonology: An overview. *Phonetica* 49:155–180.

Brysbaert, Marc and Boris New. 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. *Behavior Research Methods* 41:977–990.

Byrd, Dani. 2000. Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* 57:3–16.

Byrd, Dani and Cheng Cheng Tan. 1996. Saying consonant clusters quickly. *Journal of Phonetics* 24:263–282.

Cho, Taehong. 2016. Prosodic boundary strengthening in the Phonetics–Prosody Interface. *Language and Linguistics Compass* 10:120–141.

Cho, Taehong and Patricia Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29:155–190.

Cho, Taehong and Patricia Keating. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37:466–485.

Cho, Taehong, Daejin Kim, and Sahyang Kim. 2015. Prosodic strengthening on consonantal nasality and its asymmetric coarticulatory influence on vowel nasalization in CVN# and #NVC in English. In *Proceedings of the International Congress of Phonetic Sciences 17*.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press. first ed.

Coetzee, Andries W and Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 31:47–89.

Cohen Priva, Uriel. 2015. Informativity affects consonant duration and deletion rates. *Laboratory Phonology* 6:243–278.

Cohn, Abigail C. 1990. *Phonetic and phonological rules of nasalization*. Doctoral dissertation. University of California, Los Angeles.

De Jong, Kenneth. 1995. On the status of redundant features: the case of backing and rounding in American English. In Bruce Connel and Amalia Arvaniti, eds. *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. First ed. Chap. 6, 68–86. New York, NY: Cambridge University Press.

Dell, Gary S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93:283–321.

Dell, Gary S and Padraig G O'Seaghdha. 1992. Stages of lexical access in language production. *Cognition* 42:287–314.

Ernestus, Mirjam, Mybeth Lahey, and Femke Verhees. 2006. Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America* 120:1040–1051.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18:7–44.

Flemming, Edward and Hyesun Cho. 2017. The phonetic specification of contour tones: Evidence from the Mandarin rising tone. *Phonology* 34:1–40.

Fodor, Jerry A. 1983. *The modularity of mind: an essay on faculty psychology*. New York NY: MIT Press. first ed.

Fougeron, Cécile and Patricia Keating. 1997. Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America* 101:3728–3740.

Gahl, Susanne. 2008. Time and Thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84:474–496.

Gahl, Susanne, Yao Yao, and Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66:789–806.

Goldsmith, John. 1976. *Autosegmental phonology*. Doctoral dissertation. MIT.

Goldstein, Louis. 1991. Lip rounding as side contact. In *Proceedings of the XIIth International Congress of Phonetic Sciences*. Aix-en-Provence: Publications de l'Université de Provence. 97–101.

Goldstein, Louis, Ioana Chitoran, and Elisabeth Selkirk. 2007. Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber. In Jürgen Trouvain and William J Barry, eds. *Proceedings of the International Congress of Phonetic Sciences 16*. 241–244.

Goldwater, Sharon and Mark Johnson. 2003. Learning OT constraint rankings using a Maximum Entropy model. In *Procedings of the Workshop on Variation within Optimality Theory*. Stockholm. 111–120.

Gordon, Matthew, Paul Barthmaier, and Kathy Sands. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32:141–174.

Grice, Martine, Simon Ritter, Henrik Niemann, and Timo B Roettger. 2017. Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics* 64:90–107.

Halle, Morris and Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In Kenneth Hale and S. Jay Keyser, eds. *The View from Building 20*. First ed. Chap. 3, 111–176. Cambridge, MA: MIT Press.

Hay, Jennifer. 2001. Lexical frequency in morphology: is everything relative? *Linguistics* 39:1041–1070.

Hayes, Bruce and Aditi Lahiri. 1991. Durationally specified intonation in English and Bengali. In Johan Sundberg, Lennart Nord, and Rolf Carlson, eds. *Music, Language, Speech and Brain*. Chap. 7, 78–91. London: Macmillan Education UK.

Itô, Junko and Armin Mester. 1993. Japanese phonology constraint domains and structure preservation. In John Goldsmith, ed. *the Handbook of Phonological Theory*. First ed. Chap. 29, 817–838. New York, New York: Blackwell.

Jackendoff, Ray. 1997. *The architecture of the language faculty*. Cambridge, MA: MIT Press. first ed.

Jongman, Allard, Ratree Wayland, and Serena Wong. 2000. Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America* 108:1252–2979.

Jun, Sun-Ah. 1993. *The phonetics and phonology of Korean prosody*. Doctoral dissertation. Ohio State University.

Keating, Patricia. 1985. Universal Phonetics and the organization of grammars. In Victoria A Fromkin, ed. *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. First ed. Chap. 8, 115–132. Orlando: Academic Press Inc.

Keating, Patricia and Stefanie Shattuck-Hufnagel. 1989. A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics* 101:112–156.

Keating, Patricia, Richard Wright, and Jie Zhang. 1999. Word-level asymmetries in consonant articulation. In *UCLA Working Papers in Phonetics* 97. 157–173. UCLA.

Keyser, Samuel Jay and Kenneth Noble Stevens. 2006. Enhancement and overlap in the speech chain. *Language* 82:33–63.

Kim, Sahyang. 2001. *The interaction between prosodic domain and segmental properties: domain initial strengthening of fricatives and Post Obstruent Tensing rule in Korean*. mathesis. University of California, Los Angeles.

Ladefoged, Peter and Ian Maddieson. 1996. *The sounds of the world's languages*. Cambridge, MA: Blackwell Publishers.

Lefkowitz, L Michael. 2017. *Maxent Harmonic Grammars and Phonetic Duration*. Doctoral dissertation. University of California, Los Angeles.

Levelt, Willem J M, Ardi Roelofs, and Antje S Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22:1–75.

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

Lindblom, Björn and Johan Sundberg. 1971. Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America* 50:1166–1179.

Lionnet, Florian J. 2019. The Colon as a Separate Prosodic Category: Tonal Evidence from Paicî (Oceanic, New Caledonia). In Maura O'Leary, Richard Stockwell, Zhongshi Xu, and Z.L. Zhou, eds. *Proceedings of the 36th West Coast Conference on Formal Linguistics*. Sommerville, MA: Cascadilla Press.

Local, John. 1992. Continuing and restarting. In Peter Auer and Aldo Di Luzio, eds. *The Contextualization of Language*. Chap. 11, 273–296. Amsterdam, The Netherlands: John Benjamins.

Mayer, Connor and Bryan Gick. 2012. Talking while chewing: Speaker response to natural perturbation of speech. *Phonetica* 69:109–123.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. 2017. Montreal Forced Aligner.

Mielke, Jeff, Adam Baker, and Diana Archangeli. 2016. Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language* 92:101–140.

Moore-Cantwell, Claire and Joe Pater. 2016. Gradient exceptionality in Maximum Entropy grammar with lexically specific constraints. *Catalan Journal of Linguistics* 15:53–66.

Nespor, Marina and Irene Vogel. 1986. *Prosodic Phonology*. New York, New York: Mouton de Gruyter. first ed.

Nieto-Castanon, Alfonso, Frank H Guenther, Joseph S Perkell, and Hugh D Curtin. 2005. A modeling investigation of articulatory variability and acoustic stability during American English /r/ production. *Journal of the Acoustical Society of America* 117:3196–3212.

Pierrehumbert, Janet B and Julia B Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. *Intentions in communication* 271–311.

Pouplier, Marianne, Stefania Marin, Philip Hoole, and Alexei Kochetov. 2017. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* 64:108–126.

Roelofs, Ardi. 2000. WEAVER++ and other computational models of lemma retrieval and word-form encoding. In Linda Wheeldon, ed. *Aspects of Language Production*. Chap. 4, 71–114. Philadelphia, PA: Psychology Press.

Sadock, Jerrold M. 2012. *The modular architecture of grammar*. Cambridge, MA: Cambridge University Press. first ed.

Saltzman, Elliot, Hosung Nam, Jelena Krivokapić, and Louis Goldstein. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of the 4th International Conference on Speech Prosody (Speech Prosody 2008)*. Campinas, Brazil.

Scarborough, Rebecca. 2013. Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. *Journal of Phonetics* 41:491–508.

Schweitzer, Katrin, Michael Walsh, Sasha Calhoun, Hinrich Schütze, Bernd Möbius, Antje Schweitzer, and Grzegorz Dogil. 2015. Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation. *Speech Communication* 66:65–81.

Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3:371–405.

Selkirk, Elisabeth. 2011. The Syntax-Phonology Interface. In John Goldsmith, Jason Riggle, and Alan Yu, eds. *The Handbook of Phonological Theory*. Second ed. Chap. 14, 435–483. New York, New York: Wiley Blackwell.

Shattuck-Hufnagel, Stefanie and Alice E Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25:193–247.

Spinu, Laura, Alexei Kochetov, and Jason Lilley. 2018. Acoustic classification of Russian plain and palatalized sibilant fricatives: Spectral vs. cepstral measures. *Speech Communication* 100:41–45.

Sugahara, Mariko and Alice E Turk. 2009. Durational correlates of English sublexical constituent structure. *Phonology* 26:477–524.

van Lieshout, Pascal H H M, C Woodruff Starkweather, Wouter Hulstijn, and Herman F M Peters. 2014. Effects of linguistic correlates of stuttering on Emg activity in nonstuttering speakers. *Journal of Speech, Language, and Hearing Research* 38:360–372.

Vitevitch, Michael S and Paul A Luce. 2004. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers* 36:481–487.

Vitevitch, Michael S and Paul A Luce. 2016. Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics* 2:75–94.

Whalen, Douglas H and Bryan Gick. 1998. Parsing the contribution of lip and tongue position to fricative noise frequency. *The Journal of the Acoustical Society of America* 102:3164.

Zhou, Z.L. and Byron Ahn. 2019. Is this in the phonology? Examining the intonational phonetics-phonology interface with American English polar questions. In *Proceedings of the International Congress of Phonetic Sciences 19*.

Zymet, Jesse. 2018. *Lexical propensities in phonology: corpus and experimental evidence, grammar, and learning*. Doctoral dissertation. University of California, Los Angeles.