Title
Phonological encoding and phonetic duration

Author
Fricke, Melinda

Publication Date
2013

Peer reviewed|Thesis/dissertation

Phonological encoding and phonetic duration

by

Melinda Denise Fricke

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Linguistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Susanne Gahl, Co-chair
Professor Keith Johnson, Co-chair
Professor Fei Xu

Fall 2013

Phonological encoding and phonetic duration

Abstract

Phonological encoding and phonetic duration

by

Melinda Denise Fricke

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Susanne Gahl, Co-chair

Professor Keith Johnson, Co-chair

Studies of connected speech have repeatedly shown that the contextual predictability of a word is related to its phonetic duration; more predictable words tend to be produced with shorter duration, when other factors are controlled for (Aylett & Turk, 2004, 2006; Bell et al., 2003; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Gahl, 2008). Speaker-oriented accounts of phonetic variation posit that the probability of a word in its context is related to its accessibility for planning processes (Jurafsky, Bell, Gregory, & Raymond, 2001; Bell et al., 2003; Gahl, Yao, & Johnson, 2012), with greater accessibility giving rise to shorter durations. However, the mechanism by which accessibility relates to phonetic duration has not been fully elaborated.

This dissertation presents data relevant for understanding the connection between lexical accessibility and phonetic duration, and it argues that the relevant mechanism rests at the level of phonological encoding. The motivation and interpretation of the studies assume an interactive spreading activation model of speech production (Dell, 1986, 1988) that includes both sequential encoding of phonological segments (Sevald & Dell, 1994) and cascading activation flow.

The primary hypothesis is that examining the effects of phonological neighborhood structure on the phonetic duration of words and segments can shed light on the connection between lexical activation and duration. Three experiments are undertaken to better explain the relationship between neighborhood structure and phonetic duration. Experiment 1 pursues a novel word learning study, in which preschool children are taught words that create minimal pair relationships with already known words. The results indicate that the specific phonological relationships between the words in children's lexicons do not have an appreciable effect on articulatory duration. However, the significant effect of phonotactic probability on children's articulation suggests that the relatively great importance of practice effects (at either the phonological or motor level) may have overshadowed any influence of lexical activation.

Experiment 2 provides a reanalysis of data from a previous study that found an effect of minimal pair neighbors on phonetic duration in adults' single word productions. Baese-Berk and Goldrick (2009) found that words with a voicing-initial minimal pair neighbor (e.g. *cod*, which has a neighbor *god*) were produced with longer voice onset time (VOT) than words without such a neighbor (e.g. *cop*). The reanalysis presented in the dissertation suggests that this effect may not be related to minimal pair status *per se*, but rather to the number of neighbors in the lexicon that differ in their initial consonant segment. In addition to the findings for VOT, when segmental content is taken into account, words with more total phonological neighbors are produced with shorter rime duration. These findings support the idea that positional overlap between phonologically related neighbors is facilitatory for phonological encoding, with facilitation being reflected in the time to articulate a given segment.

Experiment 3 asks whether the results from the study of adult single word productions can be extended to spontaneous speech. An analysis of the Buckeye Corpus of Conversational Speech (Pitt et al., 2007) indicates that they can; monosyllabic English words beginning with voiceless stop consonants are produced with longer VOT when they have relatively more neighbors differing in the initial consonant. The relationship between total neighborhood density and rime duration is not found to be significant, but the numerical pattern is in the predicted direction.

To account for the observed relationship between positional phonological overlap and shorter phonetic duration, the Articulate As Soon As Possible Principle (AASAPP) is proposed. The AASAPP posits that the articulatory plan for a given segment is initiated and executed as quickly as possible, and that the time course for the production plan is related to the activation level of the target segment at the time of selection. Positional competition between co-activated segments is argued to be associated with longer articulatory duration because it slows the process of phonological encoding, while positional overlap is associated with the facilitation of encoding and therefore, articulation.

To transitory bursts of aperiodic noise:
may they always give way to more interesting onsets.

Contents

List of Figures

List of Tables

Acknowledgments

They say it takes a village to raise a child. Apparently it also takes a village to write a dissertation. The scary part about writing acknowledgements is that I'm almost certainly going to forget someone, and I hope that that person or those people will not take offense. If you are reading this: thank you, because you've probably contributed to this work in some way.

First and foremost, thank you to my cohort, and to everyone in the Phonology Lab, for interesting discussions and valuable questions and encouragement over the years. This work quite literally would not have been possible without you. Even though they all moved on a while back and may never know how much I appreciate them, I also owe some special thanks to my big brothers and sisters in the lab who always made me feel welcome, and who provided me with excellent examples of how to be impressive, intelligent researchers, but still really nice people, too: Yao Yao, Molly Babel, Shira Katseff, Charles Chang, Reiko Kataoka, Sam Tilsen, and Grant McGuire, thank you guys for being so cool and making me feel cool by association. I still want to be like you when I grow up.

One of the greatest things about being at Berkeley has been the amazing "supporting cast" of faculty members and staff we have in our department. I have been very lucky to be surrounded by people who are incredibly bright, gracious with their time, and give excellent feedback. Thank you to Sharon Inkelas for giving consistently excellent advice, and for being so good at figuring out what I'm trying to say even when I haven't quite figured it out myself yet. Thank you to Terry Regier for teaching me how to give a talk, and write an abstract, and structure an argument, and write a basic computer program. For someone who isn't technically an advisor to me, I sure learned a lot from you! Thank you to Ron Sprouse for copious technical assistance over the years, and for always answering the most embarrassingly basic questions as though they're perfectly reasonable. And thank you to Belen Flores for always knowing how to get things done and just generally taking care of all of us.

I am hugely grateful to the parents, children, and teachers of the Harold Jones Child Study Center for making all of my weird picture naming experiments possible. Ann Wakeley and Lisa Branum deserve special thanks for helping me to coordinate and run multiple projects, for being patient when I wasn't always sure what I was doing or how I should go about it, and for always doing it with a smile and a kind word.

The past two years or so have been hugely challenging for me in many ways. I owe a very special thank you to everyone who has stepped it up as I try to deal with some major life changes *while* putting together a dissertation project. I can't say that I really recommend it, but if you've got to do it, you should be so lucky as to have friends as wonderful as mine. John and Emily Sylak-Glassman, Stephanie Farmer, Greg Finley, Florian Lionnet, Emily Cibelli, Melanie Redeye, Matt Goss, Nico Baier, Clara Cohen, Christine Sheil, and everyone else who has shown me so much support and love: thank you, guys. I needed it.

Thank you especially especially to Jevon Heath, for everything I just listed times a thousand. Times a *million*. Times so many times that all I can do is give you your own

paragraph and trust that you know how much I appreciate you.

Thank you to my fantastic, rock-star collaborator and very dear friend Marisa Casillas, and also to her wonderful husband Shawn "easily deserves an honorary Ph.D. in linguistics" Bird. Speaking of people I want to be when I grow up... Even though we spent the past five years across the Bay from each other, I couldn't have imagined it without Middy. Thank you for listening to me, brainstorming with me, and of course, somehow managing to draw picture stimuli for me when you had your own dissertation to worry about!

Thank you to my frisbee team (who are mostly thanked above, but who deserve an extra paragraph as such) for many lovely hours of running around in the sunshine with friends. Jevon, Steph, Greg, Emily, Randy, Vivian, Marc, Alex, and Nik: you guys are the huckin'-est wugs a gal could ask for. And an extra special thank you to Randy for helping me with my picture stimuli, too!

I would be absolutely remiss if I didn't thank Melissa Baese-Berk and Matt Goldrick for generously sharing their data with me. Chapter 4 would quite obviously not have been possible without you!

I am also hugely grateful to Judy Kroll and Giuli Dussias for giving me a great reason to put the pedal to the metal and finish this thing. It's been a bit of a whirlwind, but I am very much looking forward to the next chapter of my life and career, and I can't imagine a nicer place to be heading off to.

Thank you to my family and to the Woodley family for their love and support over these many years, even (especially?) when they're still not really sure what I *do* all day. I also have to thank Roger for having enough faith in me to follow me all the way to Berkeley, California.

And finally: thank you, thank you, from the bottom of my heart, to my fantastic advisors. I simply don't have enough good things to say about them, but I will certainly try. Thank you for your kind words and thoughtful critiques over these five years. If I manage to convince anyone that the work presented here is theoretically interesting, it is thanks to both of you, to your judiciously timed but appropriately meted eyebrow raises, and to your insistence that I always think about the boxes and arrows.

To Susanne: When I arrived in Berkeley, I never imagined I would write a "production dissertation". It is a testament to your consistently clear and convincing argumentation that I have become interested in this set of research questions, and that I have begun to understand the many ways in which speech production is *not* the reverse of speech perception! Thank you for your support, for sharing my excitement when I think I've found something exciting, and for always knowing when and how hard to push.

To Keith: I have learned so much from you, about how to navigate academic life and "real" life both. Thank you for your help, your encouragement, your patience, your kindness, and for never seeming to mind when I come knocking on your door unannounced.

I couldn't have asked for two better people to help me figure out where I'm going, and I find myself wishing I could take them with me, but I know that's not how advising works.

Chapter 1

Introduction

Language unfolds in time. For talkers as well as listeners, efficient communication often involves being in two places at once, psychologically speaking; at any given moment, we are simultaneously processing what's going on in the present and trying to anticipate what's coming next. For listeners, the ability to constantly update our predictions allows us to understand speech more quickly and easily than if each individual sound were somehow perceived in isolation from its greater phonetic and syntactic context. For talkers, the ability to plan the speech that will immediately follow our current articulatory movements allows us to speak more rapidly and to make smooth transitions between constituents. In both cases, keeping an eye toward the future helps ensure that meaningful communication can proceed quickly and smoothly.

In this way, talkers and listeners are each engaged in tracking the greater linguistic context for their own purposes. And because their purposes are generally overlapping, some linguistic phenomena are likely to be the result of talkers' and listeners' convergent goals. Phonetic reduction is one phenomenon that is subject to forces from both directions. The reduction of frequent words and phrases has long been recognized on both synchronic and diachronic time scales. At a given stage in a language's development, more frequent words will tend to be produced with more reduced pronunciations than less frequent words, and the phonetic forms of highly frequent words and phrases will also tend to shorten over diachronic time. In extreme cases, this can result in new lexical items. *Going* and *to* occur together so frequently that their reduced form, *gonna*, is now generally accepted as the marker of the English future tense.

But why *gonna*? What happened to the vowel? What happened to the /t/ sound? One possibility is that because listeners know that *going* is extremely likely to be followed by *to*, speakers can afford to be less careful with their pronunciation. If the topic of conversation takes place in the future, both the speaker and the listener know that *going* will be followed by *to*, so the speaker is free to conserve articulatory effort. This conservation of effort is hypothesized to result in phonetic reduction; the vowels become schwas, and the stop becomes a sonorant.

A second possibility is that phonetic reduction occurs when planning is easier for the

speaker. That is, because *going* is so likely to be followed by *to*, the process of accessing the first word in memory essentially "primes the pump" for accessing the second word. Under this scenario, frequently co-occurring words become associated with one another, and keeping track of such associations allows speakers to access the words they need more quickly.

It is important to keep in mind that *both* of these processes are likely to operate to a certain extent, in certain situations. This dissertation focuses primarily on the second process, addressing an issue of great importance for speaker-oriented theories of phonetic reduction: assuming that easier planning from the speaker's point of view is associated with shorter word duration, why should this be the case? While it seems intuitive that easier to plan words would be articulated more quickly, models of speech production have not elaborated an explicit mechanism linking the ease of planning to the speed of articulation. The goal of this dissertation will be to examine the assumptions necessary to incorporate such a mechanism into currently existing models. What is the evidence that relative planning difficulty is associated with articulatory speed? And if we accept this premise, to what extent can it account for differences in articulatory duration in different speaking contexts?

1.1 Structure of the dissertation

Three experiments, each examining phonetic duration in a different speaking context, will be presented in turn. For words produced in each context, measures of total word duration – the primary dependent measure in previous research related to production planning and articulatory duration – are compared to measures of voice onset time and rime duration in monosyllabic English words. In this respect, the studies presented here bring a unique set of data to bear on the relation between planning difficulty and duration, in that previous studies examining production-internal processes in planning have largely focused on total word duration. By comparing word duration to the duration of units smaller than the word, it will become clear whether previously observed effects of planning difficulty on duration are primarily operative at the lexical level, or at a level of representation smaller than the word.

Experiment 1 (described in Chapter 3) uses a word learning paradigm with preschool-aged children to ask whether and to what extent the specific phonological relationships between words in the lexicon affect articulatory durations in children's speech. To preview the results, the learning of novel words does not affect articulatory durations in the expected way; no evidence is found to suggest that introducing new lexical items has a specific effect on the pronunciation of already known items. Consistent with previous studies, however, the frequency of phonotactic patterns is a significant predictor of articulatory duration. Less frequent patterns are associated with longer articulatory durations, and this is true for known words as well as novel words. These findings support the idea that relative planning difficulty is associated with longer articulatory duration, although in this case it is unclear whether the locus of the effect is at the level of phonological planning or motor planning. As children accrue practice with a given sequence of sounds, they are able to plan and execute

the associated articulatory gestures more quickly.

Experiment 2 (described in Chapter 4) examines single word productions in adult speech. While phonotactic frequency is not found to be associated with adults' articulatory duration, phonological neighborhood structure is. The number and type of a word's phonological neighbors is argued to impact articulatory duration by way of feedback between lexical and phonological representations, combined with sequential encoding of phonological segments. Feedback causes lexical neighbors to become active, which in turn affects the relative activation levels of the target phonological segments. When many segments compete for a given position in the word, the encoding of the target segment is more difficult, resulting in longer articulatory duration.

Experiment 3 (described in Chapter 5) extends the investigation of adults' articulatory duration to spontaneous, conversational speech. Consistent with results in the literature, higher contextual probability is found to be associated with shorter articulatory duration. Expanding on previous results, however, effects of contextual probability are found to be quite local in scope; greater probability given the previous word is associated with shorter voice onset time, while greater probability given the following word is associated with shorter rime duration. Additionally, the findings from the study of single word production in Chapter 4 are replicated for conversational speech: the number and type of a given word's phonological neighbors have a significant effect on its voice onset time in spontaneous speech.

The results of these studies are argued to be consistent with a model of speech production that incorporates both feedback between lexical and phonological representations, and sequential encoding of phonological segments (Sevald & Dell, 1994). To make the link between planning difficulty and articulatory duration more explicit, Chapters 5 and 6 introduce and develop the Articulate As Soon As Possible Principle, which proposes that the duration of a given phonological segment is affected primarily by the ease with which it is encoded, and by the ease with which the following segment is encoded. It is hypothesized that the articulatory gestures associated with easy to encode segments are initiated and executed more quickly, all else being equal.

The Articulate As Soon As Possible Principle is argued to account for the present set of results, and to be consistent with previous findings in the literature. It is not argued that listener-oriented explanations of variation in articulatory duration are not useful or valid. Rather, listener-oriented and speaker-oriented phenomena should be seen as operating at the same time, on the very same articulations; this is why factors such as prosody and phrase-final lengthening must be controlled in order to observe the effects reported here.

The primary goal of the dissertation is to make the link between ease of planning and articulatory duration more explicit, and it does so by grounding effects of articulatory reduction in the interactive and sequential nature of phonological planning. When segments are easy to encode, their articulations are planned and executed more quickly, leaving speakers free to move on to the planning and execution of the segments that follow. The effect of phonological planning on phonetic duration follows naturally from the fact that language unfolds in time.

Chapter 2

Background

2.1 Duration in language

The durations of words and segments can vary for a great number of reasons. Duration is one of the many cues that can be used to signal phonological contrast. In English, for example, the most reliable cue distinguishing a phonologically voiced stop from a voiceless stop in word initial position is voice onset time, the amount of time that elapses between the stop burst and the onset of vocal fold vibration. When stops occur at the ends of words, the duration of the preceding vowel is an important cue to stop voicing. Phonologically voiced stops are largely cued by longer preceding vowels, while voiceless stops are preceded by relatively shorter vowels.

Duration is relevant at higher levels of linguistic structure, too. In addition to signaling phonological contrast, duration can be used to invoke a contrast at the lexical level, as when words are lengthened for emphasis or contrastive focus. Katz and Selkirk (2011) recently found that speakers used differences in word duration to create a three-way distinction involving contrastively focused elements and new versus old information.

Duration helps provide structure at the level of linguistic discourse. The correlation between longer word duration and new information has been recognized for some time, and documented in several studies. Fowler and Housum (1987)'s experiments focusing on naturalistic read speech showed that speakers' first production of a word was reliably longer than subsequent productions, and that listeners seemed to use this new/old distinction to retrieve information about a word's previous context.

It is not entirely clear whether this use of duration primarily creates a binary distinction between "new" and "previously mentioned" information, or whether it carries more precise informational content than that. Aylett and Turk (2004)'s more recent study of spontaneous conversational speech indicated that new versus old information might be encoded in a scalar rather than binary way; in that study, syllable durations followed a decreasing trend with the number of times a given word had been produced previously.

In addition to signaling new versus old information, the modulation of word durations

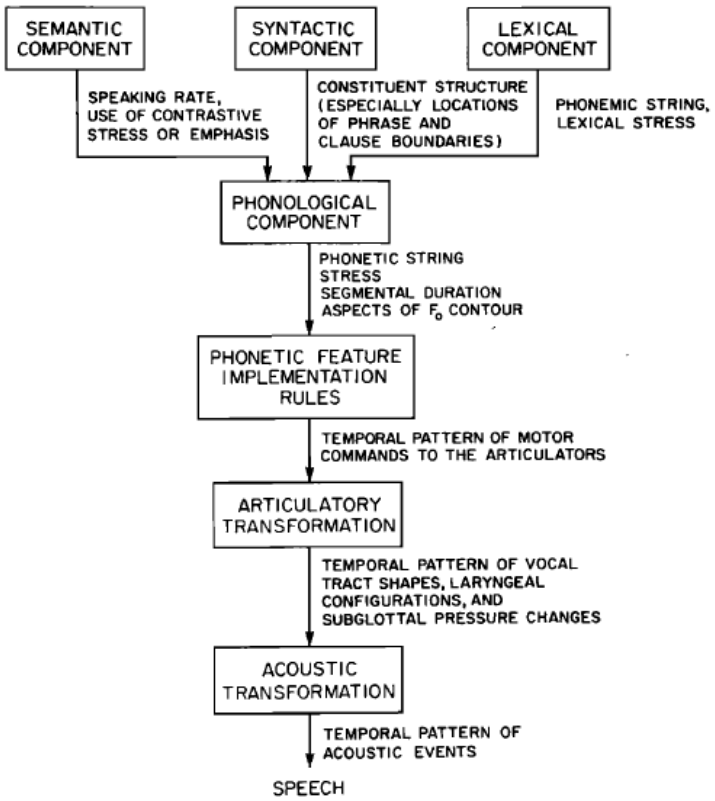


Figure 2.1: A schematic representation of durational information in language, reproduced from Klatt (1976).

can also help to organize the speech stream into phrases and sentences. Phrase-final lengthening is a well-known phenomenon that may be relevant here: Klatt (1976) reports that on average, phrase-final vowels and consonants are longer than their phrase-medial counterparts by approximately 30% to 35%. An earlier study by Klatt and Cooper (1975, cited in Klatt, 1976) indicated that listeners expect such phrase-final lengthening, and the authors suggested that it may serve as a primary perceptual cue for parsing syntactic structure.

Figure 2.1 is reproduced from Figure 2 of Klatt (1976), and summarizes the many processes that are known to affect the duration of words and segments. Klatt argues that durational information can make it in to the speech signal through many routes. The "phonological component" comprises information regarding the durational values needed to convey linguistic information of many types; the intrinsic duration of segments, their phonotactic context, and the location of stress in the word are all represented here. The phonological component also receives instructions from higher levels of linguistic structure, such as the syntactic constituency of the present sentence, and whether the speaker wishes to convey emphasis.

Below the phonological component are a set of phonetic implementation rules that trans-

form the phonological information into phonetic information: presumably a set of precise durational targets, perhaps specified in milliseconds rather than relative values, for example. These phonetic instructions are then transformed into actual motor commands specifying the target speed and location of the articulators, and the motor commands themselves produce the acoustic information that we interpret as speech. Klatt's main point, and the one that is emphasized here, is that the duration of a given segment is affected by a series of complicated, coordinated processes, the result being that the interpretation of durational differences in linguistic data is far from straightforward.

Further complicating this narrative is the fact that for any given phenomenon or level of structure, it is often difficult to say whether any durational modulations are effected primarily for the benefit for the listener or the speaker. There is one relatively uncontroversial exception to this statement, the speech style known as "clear speech", wherein speakers purposefully modify their articulation to improve intelligibility in adverse listening conditions. In a recent review of the clear speech literature, Smiljanić and Bradlow (2009) note that clear speech shares close ties with other listener-oriented phenomena, such as talking in noise ("Lombard speech"), and speech directed toward infants and non-native listeners. The defining characteristic of research on clear speech is that talkers are explicitly directed to "speak clearly and precisely", such that any differences between clear and normal speech are directly attributable to intentional effort on the part of the talker.

Smiljanić and Bradlow (2009) describe clear speech phenomena (which, it should be noted, are known to comprise durational adjustments at multiple levels of linguistic structure) as forming a part of Lindblom's hypo- to hyper-articulation continuum. H & H Theory (Lindblom, 1990) posits that the output of language production processes is the result of a balancing act between the acoustic clarity needed to effectively communicate one's message, and the constraints placed on the speaker during articulation.

Thus while investigations of clear speech directly address the extent to which speakers can intentionally modify their articulation for the benefit of the listener, other phenomena that could be placed on the H & H continuum remain more difficult to interpret. To take the association between new information and longer duration as an example, one possibility is that speakers intentionally lengthen new words in order to call their listeners' attention to them. An alternative, but related, explanation would be that because such lengthening ultimately adds helpful information to the acoustic signal, over time it has become a part of the grammar, and as such it occurs automatically, with no intentional effort on the part of the speaker. In this scenario, durational lengthening associated with information structure is not computed online by the speaker, but rather is represented along with other abstract grammatical features, such as syntactic category information.

A third possibility, and the one that will be the focus of the dissertation, is that in some cases, speakers may have limited or no control over durational lengthening. So called "speaker-centric" or "accessibility based" accounts of variation in articulatory speed posit that the ease with which a speaker can encode words and segments for production can also have an influence on articulatory duration. Bell et al. (2003), for example, found that function words produced in the context of a disfluency were on average 1.34 times longer than

the same words produced in fluent stretches of speech, when other factors were statistically controlled for. This would suggest that when speakers have difficulty accessing words in memory, their local rate of articulation can be substantially slowed.

Of course, even the interpretation of such local slowing is not without controversy. Some authors have argued that speakers actively use the phonetic cues associated with disfluencies to signal planning difficulty to their listeners (Fox Tree & Clark, 1997). It does seem likely that in some cases, speakers can exploit filled pauses and longer word durations to help hold the floor as they search for words, but it also seems likely that this is not the whole story.

It should be clear that the durations of words and segments are hugely variable for a host of complex reasons. All of the factors mentioned above – and probably others not addressed here – contribute to the continuously fluctuating nature of articulation rate. The focus of this dissertation, however, will be to better understand the differences in duration that result from relative planning difficulty. The remainder of the chapter will be dedicated to summarizing evidence related to the relationship between articulatory duration and higher level planning processes, and to providing the theoretical background necessary to interpret that evidence.

2.2 Plan for the chapter

The sections that follow summarize the available evidence regarding "speaker-centric" influences on the duration of words and segments. To understand the motivation and logic behind the speaker-centric account, it is first necessary to establish a model of speech production that can account for the research findings to be discussed. Section 2.3 argues for the adoption of Dell's (1986, 1988) interactive spreading activation model of speech production. Section 2.4 then reviews empirical studies related to the production of isolated words in a laboratory setting, focusing on findings related to articulatory duration and production latency. In Section 2.5, studies related to articulatory duration in conversational speech are discussed, with a view toward understanding them in terms of the findings from controlled laboratory studies.

Section 2.6 then provides a brief overview of current models of motor planning. While the concepts considered in this section are not directly based on or specifically tied to speech articulation, they will be important for relating models of higher level planning processes in speech production to the lower level motoric processes that are directly relevant to articulatory duration. In this section, a selection of findings from the developmental speech literature that are directly relevant for understanding the distinction between motor planning and phonological planning will also be considered.

The chapter concludes with an overview of the research questions to be addressed in the remainder of the dissertation. Section 2.7 explains the link between the existing literature and the three experiments to be presented here, along with the preliminary predictions associated with each of the experiments.

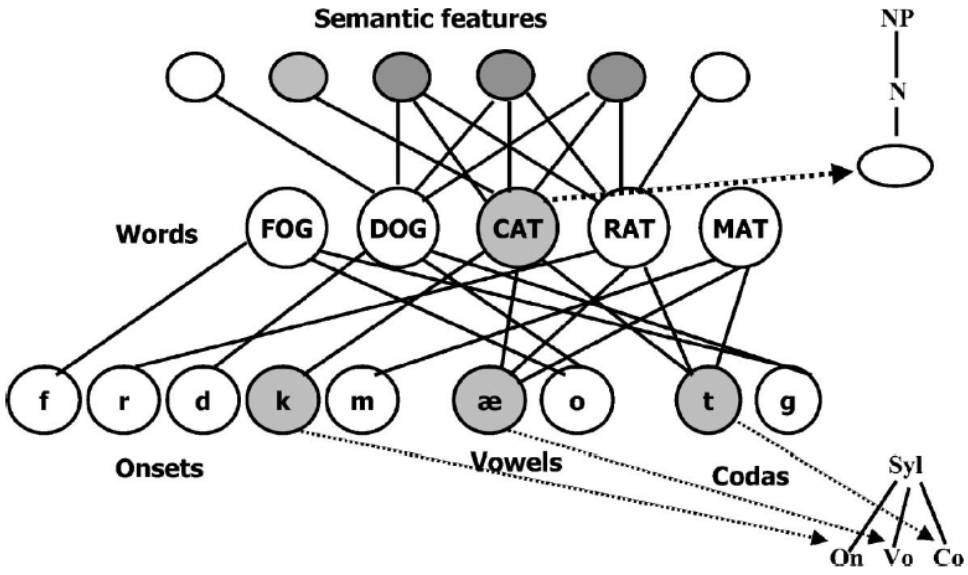


Figure 2.2: A model of language production, illustrating the conceptual, lexical, and phonological levels of representation. Reproduced from Schwartz et al. (2006), Figure 1.

2.3 The interactive spreading activation model of speech production

Models of language production generally agree that when speakers produce an utterance, information must flow through three primary levels of linguistic representation: the conceptual, lexical, and phonological levels. Speakers first conceptualize an idea, which activates the lexical units that are associated with that idea and that fit the syntactic context of the sentence. Following the activation of the appropriate lexical units, the associated phonological representations become active, allowing the speaker to plan the articulatory gestures needed to actually produce speech. A schematization of such a model, reproduced from Schwartz, Dell, Martin, Gahl, and Sobel (2006), is provided in Figure 2.2.

This is where the widespread agreement ends. Much of the discussion in the study of language production has centered on *how* information flows through the representations needed for speech. Two major remaining questions are whether the flow of information is *interactive* or strictly *feedforward*, and whether it is *modular* or *cascading*.

Interactive versus feedforward flow of activation

In *feedforward* models of production planning, such as WEAVER++ (Levelt, Roelofs, & Meyer, 1999), information is hypothesized to flow in one direction only. This means that lexical representations can only receive activation from the conceptual level, and they can

only send activation to the phonological level. Proponents of feedforward models generally argue that they are preferable for their simplicity, since they can account for a large amount of language production data without the assumption of bi-directional information flow.

In *interactive* models of production planning, such as Dell's (1986, 1988) interactive spreading activation model, information can flow in both directions. In other words, once phonological representations become active, they can send activation back up to the lexical level of representation. In general, this interactivity is hypothesized to be reinforcing, since it acts to increase the activation levels of the target representations. When things go awry, however, non-target representations are thought to out-compete the target representations. Where proponents of feedforward models argue that positing bi-directional information flow complicates the model unnecessarily, proponents of interactive models point out that interactivity would seem to be necessary to account for the range of competition effects and speech error data that have been reported in the literature (Dell, Martin, & Schwartz, 2007; Goldrick, Folk, & Rapp, 2010; Schwartz et al., 2006; Vitevitch, Armbruster, & Chu, 2004).

As Vitevitch et al. (2004) point out, it is difficult to imagine how a feedforward model of speech production could at once account for the effects of phonotactic probability, neighborhood density, neighborhood frequency, and onset density that have been demonstrated. If information about phonotactic probability is somehow associated with the phonological representations themselves (via lower resting activation levels for more frequent segments or sequences, for example), then a feedforward model can account for the facilitatory effects of higher phonotactic probability. The same mechanism would also generally be consistent with the facilitatory effects of higher neighborhood density, since phonotactic probability and neighborhood density are directly correlated.

However, such a mechanism would not be able to account for the fact that higher neighborhood density is facilitatory for production even when phonotactic probability is controlled, as in Vitevitch et al. (2004), and it is difficult to see how it could account for the facilitatory effects of high neighborhood frequency and sparse onset density. By definition, the latter variables take into account both phonological and lexical information (neighborhood frequency being the summed lexical frequency of a word's phonological neighbors, and onset density being the number of neighbors that overlap with the target word in their initial segment), and as such, both types of information must be available to the system at the same time.

In contrast, an interactive model of speech production accounts for the observed data using one and the same mechanism: bi-directional flow of activation between lexical and phonological representations. Under this account, increased activation at either level can facilitate production, and can also reinforce activation at the other level. This means that high phonotactic probability is in part facilitatory because more frequent segments have higher resting activation levels, but also because they are connected to more words in the lexicon. Words and segments in dense neighborhoods generally receive greater amounts of activation due to the bi-directional flow of information, and this explains how high neighborhood density can be facilitatory even when phonotactic probability is controlled: words in dense neighborhoods are produced more easily because they ultimately receive reinforcing

activation from their neighbors, via connections at the phonological level.

An explanation of the facilitatory effect of high neighborhood frequency thus follows naturally. When phonotactic probability and the number of neighbors are held constant, higher neighborhood frequency can still facilitate production because high frequency neighbors pass on more activation to the target representation than low frequency neighbors (Vitevitch & Sommers, 2003). The same mechanism is at work, in that interaction between the lexical and phonological levels explains how target representations can receive reinforcing activation from their phonologically related neighbors.

To account for Vitevitch et al. (2004)'s finding that words with sparse onsets were produced with shorter latencies than words with dense onsets (that is, articulation was initiated more quickly for words with a smaller proportion of neighbors overlapping in their onset consonant, when the total number of neighbors was held constant), one additional specification is necessary. Vitevitch et al. (2004) argue that competition between segments vying for the initial slot in the prosodic frame caused words with more initially overlapping neighbors to be produced with longer latencies. With reference to the toy lexicon in Figure 2.2, this means that words such as *cat*, with two neighbors that "disagree" in their initial segment, were initiated more slowly than words such as *dog*, with only one neighbor that disagrees in the initial segment[1].

This line of reasoning is consistent with other findings in the literature (Sevald & Dell, 1994; Yaniv, Meyer, Gordon, Huff, & Sevald, 1990) and will be taken up in greater depth below. The important point for the present discussion is simply that *any* influence of lexical neighbors that are defined by their phonological relationships to the target word is difficult to account for without the assumption of interactive flow of activation. For this reason, all subsequent discussions will assume a model of production that incorporates such interactivity.

Modular versus cascading flow of activation

A second remaining point of contention in models of speech production is whether the flow of information is *modular* versus *cascading*. *Modular* accounts posit that activation can only flow to the next level once a selection has been made at the current level, while in *cascading* accounts, activation can flow to the next level whether a selection has been made or not.

The following concrete example again references the toy lexicon in Figure 2.2. In this simplified lexicon, the semantic features associated with the word *cat* are active to various degrees, the word *cat* is active, and the phonological segments associated with *cat* are also active. In a modular model, the phonological segments can only become active once the lexical representation for *cat* has passed some threshold level of activation; in other words, phonological planning can only begin once the system is completely "sure" of which word it will produce.

[1]This example is not perfect, since the dense onset and sparse onset words in Vitevitch et al. (2004)'s study were equated for the total number of neighbors. If the toy lexicon in Figure 2.2 also contained *dig* as a neighbor for *dog*, then the example would be more comparable to that study.

In a cascading model, by contrast, lexical representations behave like a sieve; activation can flow straight through them, reaching the phonological level even before the system is sure what the target word is. The result is that phonological planning can begin as soon as phonological segments begin to reach some threshold level of activation. This may happen before or after lexical selection has taken place.

Because the primary variable under investigation in this dissertation is articulatory duration, it must be pointed out that a cascaded model provides a more straightforward account of how higher level planning processes could affect the speed of articulation. In a modular model, it would be possible for certain lexical items to send more activation to the phonological level than others, and this overall greater activation level could in principle result in faster articulation. However, this account would predict that more frequent (or generally more accessible) words would always be produced with shorter durations, and this is not the case.

The alternative account is that differences in articulatory duration are the result of phonological segments becoming available at varying rates. This is the account that will be pursued here, and it will be elaborated in more detail below.

Summary of activation flow and assumptions going forward

The model that is assumed in the remainder of the dissertation incorporates both interactivity and cascading flow of information. It will be argued that these two assumptions are necessary to account for results in the literature, and for the results of the present experiments.

The sections that follow present a summary of what is known regarding articulatory duration and latency in the laboratory, as well as in spontaneous, connected speech.

2.4 Articulatory duration and latency in the laboratory

Relatively few studies have found effects of higher level influences, such as word frequency and phonological neighborhood density, on articulatory duration in laboratory speech. The term "laboratory speech" here refers to short (typically one- to two-word) utterances produced in somewhat artificial speaking situations, where the speaker performs a simple but repetitive and non-communicative task such as reading words from a screen, or naming pictures one after the other.

Several studies that explicitly sought to relate lexical level factors to articulatory duration will be presented first – a few that identified significant differences, and a few that did not. Following the discussion of duration proper, the focus will be shifted to studies examining articulatory latency, since latency is a complementary and highly relevant aspect of articulatory timing. It should be noted that the strict division of this discussion into two separate sections is not entirely possible; first, because several studies investigated both aspects of

timing simultaneously, and second, because the two issues are not entirely separable from one another to begin with.

Articulatory duration in the laboratory

In a widely cited study examining the relation between high level and low level planning processes, Balota, Boland, and Shields (1989) primed speakers with either a semantically related, unrelated, or neutral prime, as they prepared to produce single word utterances. Primes and targets were presented orthographically, and the stimulus onset asynchrony (SOA) was varied across three experiments. In general, priming had a small or nonexistent effect on production onset latencies, but at SOAs of 400 and 650 ms (Experiment 1), when the prime and target were presented at the same time (Experiment 2), and when speakers produced a triad of words ending with the prime-target pair (Experiment 3), the duration of words in the related prime condition was approximately 10 to 15 ms shorter than for the same words produced in the unrelated condition.

Balota et al. (1989) interpreted their results with respect to Dell's interactive activation model, and suggested that if activation cascades from the lexical to the phonological level, it could be stipulated that the articulation of each phoneme is initiated as it reaches a threshold level of activation. If the effect of semantically related primes is to boost the activation of the target lexical representation at the moment when phonological encoding is taking place, this could potentially explain the shorter duration of words in the related prime condition.

While Balota et al. (1989)'s study is cited frequently, the effect that they report has proven rather difficult to replicate. The logic behind related subsequent work has been that if semantic priming can affect articulatory duration, the locus of such an effect must reside somewhere in the transfer of activation from lexical representations to articulatory representations. Several studies have therefore asked whether explicitly manipulating phonological encoding processes can have a demonstrable effect on articulatory duration.

Damian (2003) examined single word durations while manipulating planning processes using three different tasks: picture–word interference, picture naming for words blocked by semantic and phonological similarity, and a Stroop interference task intended to disrupt "central planning processes". In all three cases, significant differences were found for onset latencies but not for word durations. In the picture–word interference task, semantically related distractor words gave rise to faster onset latencies when they appeared just before the target picture (SOA = -200 ms), and phonologically related distractors were associated with slower onset latencies when they appeared just after the target picture (SOA = +200 ms). In the blocked naming task, onset latencies for words in semantically related blocks were significantly slower than for words in mixed blocks, while latencies in phonologically related blocks were either significantly faster or not significantly different from mixed blocks, depending on whether a speaking deadline was imposed or not (respectively).

Damian's main purpose in using three different production tasks, and in having subjects speak both at a normal pace and while under a response deadline, was to evaluate Kello, Plaut, and MacWhinney (2000)'s claim that speech production varies flexibly between mod-

ular and cascading activation flow depending on task demands. In Kello et al. (2000), participants completed a Stroop color naming task, in which the names of colors are printed in either congruent or incongruent colored ink, and participants must name the ink color (e.g. the word *blue* is printed in red, and the correct response is "red"). At a normal pace, naming latencies but not articulatory durations were affected by color congruency. When a response deadline was imposed, however, both naming latencies and articulatory durations were affected; under time pressure, participants were slower to initiate *and execute* the articulation of incongruent color words, leading Kello et al. (2000) to argue that task demands (or "central planning processes") affect the flow of information in production planning.

Damian's finding that articulatory duration was unaffected by higher level planning processes, however, is consistent with the conclusion that articulatory planning is not cascaded; Damian concluded that once articulation was initiated, it could not be affected by earlier processing stages.

Another interesting implication of Damian's study is that encoding processes are sensitive to both the strength and timing of competitive processes. Recall that semantic competitors were associated with facilitation at very short SOAs, but with competition in blocked picture naming. This suggests that activating the lexical representations of related words may help facilitate encoding on a very local time scale, but it becomes detrimental on a more global scale. On the other hand, phonological competitors were associated with competition on a local scale (picture–word interference), but with facilitation on a more global scale (blocked naming). One interpretation of these findings is that when competing phonological representations are strongly activated during the process of phonological encoding (in this case, at an SOA of +200 ms), the process of phonological encoding is disrupted, leading to longer onset latencies. On a more global scale, though, it may be that recently activated phonological units are temporarily easier to access, leading to shorter onset latencies in blocked picture naming.

The main problem with this interpretation is that it runs counter to the findings of later studies. Damian and Dumay (2009) also examined the relationship between phonological encoding and articulation, expanding on the blocking manipulation used in Damian (2003). Significant effects were again obtained for onset latencies, but not for word durations. This time, when participants named just one word at a time, trial to trial phonological overlap was inhibitory; for example, speakers were slower to initiate the word *goat* when they had produced the word *green* on the previous trial (as compared to the naming latency for *goat* when it was preceded by *blue* on the previous trial). On the other hand, phonological overlap was facilitatory when the two words were produced in the same phrase; speakers initiated the phrase *green goat* faster than the phrase *blue goat*.

There were no significant differences in articulatory duration in any condition, and the differences in naming latency were not present in a delayed naming task. The authors argued that the effect of phonological overlap on naming latencies must be due to higher level planning processes, and not to articulatory factors, and they suggested two potential accounts of the competition versus facilitation effects. Under one account, competition versus facilitation could be due to whether the overlapping segments compete for the same position

in the prosodic frame. Under the second account, it could be that segmental representations undergo faster decay than lexical representations. The discussion of this study will be taken up again below, in the section on articulatory latency.

The studies reviewed thus far have all examined the relationship between planning processes and the duration of whole words. In a separate but related line of inquiry, several studies have examined the relationship between phonological neighborhood density and the duration of individual segments. By the most commonly used metric, phonological neighborhood density refers to the number of words in the lexicon that differ from the target by a one segment addition, deletion, or substitution. Recall that under the interactive activation model of speech production, neighborhood density is argued to be facilitatory for speech production due to interaction between the lexical and phonological levels of representation; words in dense neighborhoods are thought to be accessed more easily during production because they receive reinforcing activation from their many phonologically related neighbors.

It is possible that examining the duration of individual segments as a function of their neighborhood characteristics could potentially provide similar information as examining the relationship between word frequency and whole word duration. Both approaches seem capable of revealing an effect of increased lexical activation on articulatory speed, if such an effect exists. The primary difference would seem to be that zeroing in on individual segments may provide more precise information regarding the relationship between the timing of phonological encoding processes and articulatory speed; because variation in duration at the word level is subject to so many factors (as described above; see Figure 2.1) it may be that examining a single segment or type of segment makes controlling and understanding such variation simpler than examining a string of segments.

Generally, however, this has not been the primary goal of studies that have examined segment duration as a function of neighborhood characteristics. Such studies have approached the issue with a range of motivations; because variation in duration is correlated with so many separate phenomena, it provides information about many distinct processes. It is important to realize that the significance of neighborhood density for spoken language processing was originally observed in the context of speech perception. Models such as TRACE (McClelland & Elman, 1986), the Cohort Model (Marslen-Wilson, 1989), and the Neighborhood Activation Model (Luce & Pisoni, 1998) are all intended to capture the idea that similar sounding words compete with one another during recognition. Several authors have therefore studied segmental duration in hopes of better understanding whether talkers' knowledge of competition between neighbors in perception can affect the phonetic details of their speech.

Goldinger and Summers' (1989) investigation, for example, was based on the hypothesis that the importance of neighborhood density for speech perception would exert a variability-reducing force on talkers' articulation. Goldinger and Summers hypothesized that articulations would be slower and/or less variable for words in dense neighborhoods, due to their "lower margin of error", the assumption being that talkers' implicit knowledge of the difficulty of recognizing words in dense neighborhoods would lead to more careful pronunciations.

In this study, participants saw two minimal pair words differing only in the voicing of the initial stop consonant (e.g. *dutch – touch*) displayed on the screen and were simply asked to

say them aloud. The dependent measure of interest was the difference in voice onset time (VOT) between the two words in the pair. Each pair was presented 16 times over the course of the experiment. There was a significant interaction between neighborhood density and trial number; the first eight productions of pairs showed no significant difference in VOT, but the last eight productions revealed that pairs in dense neighborhoods were produced with a significantly greater VOT difference than pairs in sparse neighborhoods. For pairs in dense neighborhoods, the VOT difference between the two words increased from 73 to 78 ms over the course of the experiment, while the VOT difference for pairs in sparse neighborhoods decreased, from 74 to 68 ms. Contrary to predictions, variability did not differ significantly as a function of neighborhood density. Unexpectedly, however, pause duration did; the pause duration between words in dense neighborhoods was on average 53 ms longer than the pause duration between words in sparse neighborhoods.

A study by Kilanski (2009) also examined VOT as a function of neighborhood density, as part of a larger investigation concerning the durations of several different types of phonological segments, in different positions in the word. Stimuli varied orthogonally with respect to neighborhood density and lexical frequency, and the hypothesis was that segments would be shorter in more easily perceived words (those with high frequency and low density). Contrary to predictions, however, duration measures revealed a main effect of lexical frequency in the opposite direction, such that high frequency words were produced with longer word-initial consonants (fricative durations and stop VOTs) than low frequency words. Additionally, a significant interaction in the VOT data indicated that for high frequency words, higher density was associated with longer word-initial VOT, but for low frequency words, higher density was associated with shorter VOT. The vowel duration data also ran contrary to predictions, in that vowels in high density words were produced with shorter durations than vowels in low density words.

For the word final consonants and measures of total word duration, Kilanski found shorter word final consonants in high frequency and high density words, and shorter overall word durations for high frequency and high density words. The conclusion drawn was that vowels and word final consonants seemed to have patterned together, while the onsets seemed to pattern "conversely with overall word duration", but no speculation was offered as to why this might have been the case.

The results from both Goldinger and Summers (1989) and Kilanski (2009) suggest that perceptual factors may not provide a clear explanation for durational phenomena in laboratory speech (or at the very least, that the predictions for a listener-centric account of neighborhood density effects on segmental duration are not entirely straightforward). These results will be taken up again in the General Discussion.

The final study to be reviewed in this section is a single-word production experiment specifically designed to pit three accounts of lexically conditioned phonetic variation against one another. Similarly to Goldinger and Summers, Baese-Berk and Goldrick (2009) also examined word-initial VOT for minimal pair words. The hypothesis was that words with voicing-initial minimal pairs (e.g. *cod*, which has a minimal pair word *god*) would be produced with slightly longer VOT than words without such a minimal pair (e.g. *cog*, which has no

minimal pair *gog*), but that this difference would follow one of three patterns.

The *perceptual restructuring account* (based largely on Pierrehumbert, 2002), states that exemplars that are more easily perceived are more likely to be stored in memory, and over time, this leads the memory representation for words that have a voicing-initial minimal pair to have slightly more distinct VOT than the representation for words without such a minimal pair. Under this account, then, the difference between "minimal pair words" and "non minimal pair words" should be consistent, no matter what the speaking context.

The *perceptual monitoring account* states that speakers monitor their speech during production, and it borrows elements from both Levelt et al. (1999), who have proposed a monitoring mechanism to account for the tendency of speech errors to produce real words, and Lindblom (1990)'s H & H Theory, which posits that speakers try to balance the demands on speech production with the listener's needs in perception. Under Baese-Berk and Goldrick's version of the perceptual monitoring account, minimal pair words should only be produced with longer VOT than non minimal pair words when their competitor is present, and in the presence of an actual listener.

Finally, the *speaker internal account* (which assumes an interactive activation model of production such as Dell, 1986, 1988) states that the co-activation of minimal pair neighbors can lead to hyper articulation. Under this account, a minimal pair word's neighbor becomes active due to feedback during the production process, and the activation boost needed to out-compete the neighbor causes minimal pair words to have longer VOT than non minimal pair words.

In Baese-Berk and Goldrick's first experiment, minimal pair (*cod*-type) words and non minimal pair (*cog*-type) words were embedded in a list comprised mostly of fillers, and participants were instructed to simply read each word aloud from the computer screen. In the second experiment, participants were placed in an actual communicative situation, and were tasked with instructing their listener to "click on the (target word)". The target word was displayed on the screen along with two distractor words, and on half of all critical trials, one of the distractors was the target's minimal pair word.

The results were consistent with the speaker internal account. Minimal pair words had very slightly but significantly longer VOT than non-minimal pair words, and the effect was slightly greater when the minimal pair competitor was present (an approximately 4 ms difference when the competitor was not present, and a 10 ms difference when it was). Baese-Berk and Goldrick therefore argued that speaker-internal feedback and competition processes between minimal pair neighbors were behind the VOT effect, with increased competition leading to increased hyperarticulation.

Two important points from this experiment bear further discussion. First, the VOT difference obtained between the two types of words was very unlikely to be perceptible to listeners; Klatt (1976) estimates the just-noticeable-difference for segmental durations in normal listening conditions to be on the order of 25 ms, with an absolute minimum of 10 ms in carefully controlled, single-word perception studies. This suggests that the 4 to 10 ms difference observed by Baese-Berk and Goldrick was not effected by speakers for perceptual reasons, since listeners would have been unable to detect it.

Second, while the results provide support for a speaker-internal account of phonetic variation, the link between co-activation of minimal pair neighbors and hyperarticulated VOT is not entirely clear. Since greater phonological neighborhood density is generally facilitative for production, it is curious that a particular type of neighbor – one differing in the voicing of the initial consonant segment – would act as a competitor for the target word.

This raises several questions. Is it possible that in single word productions, greater neighborhood density is associated with shorter onset latencies, but longer articulatory durations? If so, what mechanism can explain such a difference? Is the particular phonological contrast relevant to the hyperarticulation effect – do different types of minimal pair neighbors have different effects on articulation? For example, would the same effect arise if the two minimal pair neighbors were *cod* and *pod*, rather than *cod* and *god*? What about *nod*? Does the position of the contrast matter? (Would *cod* be equally hyperarticulated in the context of *kid*?) And finally, Baese-Berk and Goldrick only report data on VOT, but their interpretation relies on competition at the lexical level of representation. If lexical activation and competition are driving the hyperarticulation effect, does that mean that all aspects of the words were hyperarticulated? Were the vowels longer than would otherwise be expected?

Because the literature on single word production durations is relatively sparse, it is difficult to make predictions for these questions based on existing data. The literature on single word production latencies is considerably more rich, and may help to address these questions. The following section therefore summarizes and discusses a selection of studies that have investigated the effects of higher level planning factors on articulatory latencies in laboratory speech.

Articulatory latency in the laboratory

Lexical frequency has consistently been shown to influence production latencies in the laboratory, but the locus of the word frequency effect has been debated to a certain extent. An early study by Balota and Chumbley (1985) used a simple delayed production task in order to separate the effects of lexical access and phonological encoding time. The reasoning was that if subjects were given enough time to retrieve a word's lexical representation, then any remaining effect of frequency on articulatory latency should be due to processes below the lexical level. Participants saw a single word printed on the screen and were given a cue to produce the word at delays ranging from 0 to 2900 ms. There was a clear effect of lexical frequency on onset latency, even at delays much longer than the estimated time needed for lexical access. The authors concluded that onset latencies were likely to be influenced by some post-recognition process or processes, since the difference in the time needed to initiate articulation for high versus low frequency words was relatively constant across delays.

In contrast, Jescheniak and Levelt (1994) argued that the lexical frequency effect was localized at a level of representation containing a word's syntactic and semantic information (termed the lemma), but not phonological or articulatory information (the lexeme). Seven experiments were conducted to determine the locus of the word frequency effect, and robust frequency effects were obtained for picture naming and translation tasks. A more

"ephemeral" effect of word frequency was found for noun gender decisions, and no significant effect of frequency was found for latencies in either an object recognition task or a delayed naming task. As a potential explanation for the discrepancy with Balota and Chumbley's findings, Jescheniak and Levelt point out that their materials consisted exclusively of monosyllabic items, while Balota and Chumbley's were primarily multisyllabic.

Taking these two studies together, it seems likely that processes at both the lexical and phonological levels of representation contribute to production latencies to a certain extent. It is possible that phonological processing is rather easily overshadowed by lexical processing, especially in tasks where retrieval of the word form requires processing the word's meaning, and perhaps especially for short words. Many subsequent studies have zeroed in on the phonological processing component by carefully manipulating the phonological properties of the stimuli to be produced, and by removing or minimizing the need for lexical retrieval in performing the task.

In Yaniv et al. (1990), for example, speakers were instructed to prepare pairs of CVC syllables, the order of which was given at the last second. This sort of syllable preparation paradigm is more similar to the task in Balota and Chumbley (1985) than to Jescheniak and Levelt (1994), and indeed, the phonological manipulations in Yaniv et al. (1990) had a significant effect on speakers' production latencies. When the syllables in the pair had the same or similar vowels (e.g. /i/ and /ɪ/), response latencies were significantly longer than when vowels were dissimilar. This was true whether the initial consonants in the pair were the same or different, but similarity of the final consonants had an effect on production latencies as well; the inhibitory effect of vowel similarity was significantly reduced when the final consonants were different, suggesting that the planning of two-word utterances is rendered more difficult by greater amounts of word final (or perhaps rime) overlap.

In a similar task, Sevald and Dell (1994) asked speakers to produce as many alternating CVC syllables as possible in eight seconds (for example, *pick, tick, pick, tick*, etc.). Seemingly in contrast with Yaniv et al. (1990)'s findings, however, participants' productions were facilitated when vowels and final consonants were overlapping. Speakers produced significantly more syllables for pairs such as *pick–tick* than for pairs such as *pick–pin*. Sevald and Dell therefore argued for a "sequential cuing effect" – a stipulation that segments are encoded from left to right – within the context of an interactive model of speech production. If phonological encoding is sequential, and if activation can spread from the phonological level back up to the lexical level, then competition for the final consonant slot in pairs such as *pick–pin* could accumulate as the two competing words are continuously re-activated, explaining why such pairs are more difficult to produce in alternating succession.

In general, studies manipulating the phonological similarity between the items to be produced seem to agree that segmental overlap between concurrently active representations is important for determining articulatory latencies. Whether such overlap is inhibitory or facilitatory, however, apparently depends on a variety of complex factors. In Yaniv et al. (1990)'s study, overlapping vowels were inhibitory, and inhibition was even greater when the final consonants overlapped as well. But in Sevald and Dell (1994), overlap between segments in the rime was facilitatory.

One factor certain to be relevant in reconciling these (and other) results is the nature of the task being performed. Recall that Yaniv et al. (1990)'s participants prepared a single two-syllable pair in advance, while speakers in Sevald and Dell (1994) were required to produce as many alternating syllables as possible in eight seconds. It could be the case, for example, that producing a single two-word phrase places significantly different demands on the production system than producing a series of phonologically related syllables.

Damian and Dumay's (2009) study, reviewed in the section on articulatory duration, is also highly relevant to the present discussion: given two words overlapping only in their initial consonant (e.g. *green* and *goat*), production latencies were inhibited in a single word naming task, but facilitated when participants planned a two-word phrase. That is, "goat" was initiated more slowly on a trial following the word "green", but the phrase "green goat" was initiated significantly faster than "blue goat". Again it would seem that the planning of two-word phrases is different in some crucial way from the planning of single words, with overlap in initial segments leading to faster latencies only when two words are planned at once.

Meyer (1990, 1991) developed a procedure specifically designed for examining effects of phonological overlap in two-word utterances. In the form preparation paradigm, participants memorize pairs of words varying in their phonological similarity and are asked to produce the second word when prompted with the first. Where Yaniv et al. (1990) found no significant facilitation for word-initial overlap, Meyer did: participants were faster to initiate responses when the initial segment or segments of the words were identical (e.g. *melon–metal*) as compared to when they overlapped in their final segments (*pocket–ticket*) or were non-overlapping. Moreover, the magnitude of the facilitation effect increased with the amount of initial overlap, such that words overlapping in their initial two segments were produced faster than words overlapping in only their first segment.

Roelofs (1999) replicated Meyer's results using the same paradigm, and also extended the original findings, confirming that segmental (not featural) overlap was driving the effect. Roelofs' participants exhibited significant facilitation for *tennis–terrace* pairs, but not for *tennis–devil* pairs.

Damian and Dumay (2007) investigated the effects of phonological overlap using two different tasks: picture–word interference, and colored picture naming. In Experiments 1 and 2, participants saw a picture and a printed word distractor displayed at the same time (SOA = 0 ms) and were asked to name the picture. In Experiment 1, picture names were single nouns (e.g. *barrel*), and in Experiment 2, speakers were instructed to produce a determiner–adjective–noun phrase corresponding to the ink color of the picture (*the green barrel*). Distractor words in both experiments overlapped with the target picture either in their initial or final consonants (e.g. *basil* or *squirrel*). In both experiments, in all cases, phonological overlap was associated with shorter production latencies.

In a third experiment, participants again named colored pictures, but this time phonological overlap between the adjective and the noun was manipulated. Consistent with the other two-word production studies reviewed above, initial overlap between words resulted in faster latencies; *blue bed*, for example, was initiated more quickly than *green bed*.

The studies reviewed thus far have all examined phonological encoding by explicitly co-activating either overlapping or non overlapping phonological representations, through external manipulations. According to the interactive activation model of speech production, though, such co-activation occurs constantly in the normal course of speech production. Whenever a speaker intends to produce a word, both the target word and its phonologically related neighbors are thought to become active through feedback processes.

As described above, several studies have indicated that greater phonological neighborhood density is generally associated with facilitated production. This would seem to be consistent with the results of production studies that have used picture–word interference, and those that have examined the planning of two-word utterances; it is possible that the same mechanism is responsible for those findings and the findings from studies investigating neighborhood density effects. In Vitevitch and Sommers (2003), for example, words in dense neighborhoods were named more quickly and resulted in fewer tip-of-the-tongue ("TOT") states than words in sparse neighborhoods, suggesting that the co-activation of shared segments by phonologically related lexical representations is generally facilitatory when it occurs on a short time scale. In an experiment with older adults, neighborhood frequency (the summed lexical frequency of a word's phonological neighbors) was shown to have a similar effect; older adults named pictures faster and produced fewer TOTs for words with higher than lower neighborhood frequency.

Because neighborhood density and phonotactic probability are generally highly correlated, Vitevitch et al. (2004) investigated these two variables separately to determine whether they did in fact have independent effects. In a set of simple picture naming studies, pictures whose names had higher phonotactic probability were named faster than those with phonotactically less probable names, and this was true when neighborhood density was held constant, as well as when the two factors varied orthogonally. In an additional experiment that controlled for phonotactic probability, onset density (the proportion of neighbors overlapping with the target in the initial segment only) was also shown to have a significant effect on naming latencies. However, contrary to what might be expected, words with a greater proportion of neighbors overlapping in the onset were named more slowly than words with fewer neighbors overlapping in the onset (866 vs. 819 ms, respectively). Again, it is therefore not entirely clear how factors of planning and positional segment overlap combine to sometimes yield facilitation, and sometimes inhibition.

Recall that Damian and Dumay (2009) found that initially overlapping segments had different effects depending on the number of words planned. (That is, *goat* was initiated more slowly following *green* when each trial required a one word response, but *green goat* was initiated more quickly than *blue goat* when a two-word utterance was required.) The authors offer two possible explanations for this apparent discrepancy. One possibility is that production planning takes into account a segment's position within the prosodic frame. Under this explanation, *goat* would be more difficult to produce on a trial following *green* due to the sequential cuing effect; the overlapping initial segments cause the two words to remain active, causing their segments to compete for later slots in the word. For the two-word phrase *green goat*, however, segmental overlap between the words does not incur

competition because the words are specified for different positions in the phrase. Because the production system "knows" they are not vying for the same temporal position, co-activation of the same segmental representation is facilitatory rather than inhibitive.

The second possible explanation offered by Damian and Dumay (2009) involves the temporal proximity of segments to one another. They propose that if segmental activation dissipates very quickly, then the temporal proximity of segments to one another in a two-word phrase could cause positional segmental overlap to be facilitatory, while segments produced farther apart in time would not benefit from close temporal proximity, but rather would suffer from positional competition via the segmental cuing effect.

Summary of duration and latency in laboratory speech

Nearly 30 years of research employing a variety of experimental tasks has indicated that the phonological encoding process is sensitive to the structure, timing, and similarity of the target sounds to any co-activated or potentially competing sounds. In general, greater activation of segmental representations appears to be associated with shorter onset latencies for the target word, and a limited amount of evidence also indicates that greater lexical activation may be associated with shorter articulatory durations under certain circumstances.

More specifically, positional segmental overlap between co-activated lexical representations is generally facilitatory when segments occur in word initial position and are temporally close to the target; picture-word interference, two-word naming tasks, and manipulations of phonological neighborhood density have all indicated that the simultaneous activation of phonologically overlapping lexical representations tends to speed production latencies for the target word.

The picture is slightly more complicated than that, however, since some evidence indicates that the activation of overlapping representations may be inhibitory at longer temporal spans, or perhaps in the context of competition for the same slot in a prosodic frame. Finally, while neighborhood density is generally facilitatory for production latencies, onset density in particular seems to have the opposite effect on latency; words with many neighbors tend to be initiated more quickly, but when the majority of those neighbors overlap in their initial consonant, latencies are slower.

We now turn our attention to a difference source of evidence regarding the activation and planning of words and segments. Within the past decade especially, researchers have begun to ask questions about how processes of lexical access and encoding that have been examined in the laboratory scale up to connected speech. In the next section, studies examining articulatory duration in conversational speech are summarized and discussed.

2.5 Duration in conversational speech

One of the earlier studies examining word durations in connected speech was Fowler and Housum (1987), who looked at the durations of content words as a function of "newness"

in discourse. Not all of the speech examined was entirely spontaneous (most of the paper focused on a short story read in a casual, naturalistic way), but the study provided some foundational data on the extent to which duration varies depending on whether a word has been mentioned before, and on whether listeners are sensitive to such variation.

Experiment 1 compared first and second mentions of words in connected speech and found significant effects on duration and amplitude, but not pitch; first mentions were both longer and louder than subsequent mentions. A set of follow-up experiments found that the phonetically reduced "old" words were more difficult for listeners to recognize in isolation, but listeners were able to use other cues (such as contextual predictability) to combat this difference in intelligibility. Moreover, listeners appeared to use their knowledge that a word had been produced before to help recall the earlier context; in a priming task comparing the efficacy of old versus new primes, old primes facilitated the recall of targets more than new primes.

Fowler and Housum made no strong claims regarding the role of the speaker versus the listener with respect to the relationship between duration and previous mention. Rather, they characterized the relationship between talkers and listeners as symbiotic, concluding that whether or not speakers' reduction of previously mentioned words was effected for their listeners or not, such reduction did not appear to hamper listeners' comprehension of words in context, and in some respects it even seemed to help.

Several studies by Bell and colleagues examined whether phonetic reduction affected content versus function words differently. In their 2003 study, Bell et al. looked at the ten most frequent function words in English and found that they were longer in duration, more likely to contain full vowels, and more likely to contain pronounced coda obstruents (in the case of *it*, *that*, *and*, and *of*) when they occurred in the context of a disfluency, in a less predictable context, or in utterance initial or final position.

The authors point out that their findings with respect to vowel fullness extend Fox Tree and Clark's (1997) study of the function word *the*; in both studies, *the* was more likely to be produced with a full vowel (rather than a schwa) in the context of a disfluency. Bell et al. (2003)'s findings, however, imply that the option to produce full vowels in function words can now be understood as being lexically specific, since in their study the vowel fullness findings only held for *the*, *and*, and *it*.

On the other hand, durational lengthening occurred across the board, for all ten of the function words studied. The authors argue that this finding can best be understood in the context of a spreading activation model of speech production that allows cascaded information flow. Bell et al. (2003) hypothesize that words that are more probable in their context have higher levels of lexical activation, and that cascading activation from the lexical level to the phonological level can explain the gradient effect of contextual probability on articulatory duration. Note that these findings and this argument are consistent with Balota et al. (1989)'s study of the effects of semantic priming on articulatory duration in single word productions. In both cases, a boost in activation at the lexical level is supposed to cascade down to the phonological level, with the relatively faster encoding of phonological segments leading to faster execution of articulatory gestures.

Bell et al. (2009) largely replicated their earlier findings for English function words using a different corpus of spontaneous speech, and also provided a comparison of contextual predictability effects on function words versus content words. Independent effects of contextual predictability and word frequency were found. The duration of content words was most strongly related to their frequency and to the conditional probability given the following word. There was also a highly significant interaction between these predictors; the shortening effect of high contextual predictability was strongest for high frequency words. Conditional probability given the previous word, however, was not a significant predictor of content word duration.

For function words, the effect of word frequency was not as clear cut. While there was a binary split between the highest frequency words versus the mid and low frequency words, with frequency in the highest range being associated with durational shortening, this split was argued to be due to differences in "combinatorial behavior and form variation ... rather than a difference in behavior attributable directly to their frequency of occurrence." The effect of contextual probability, however, was clearly significant. Higher conditional probability given both the preceding and following words was associated with shorter duration in function words.

In contrast to their 2003 paper, in which they proposed cascading activation as a potential explanation for predictability effects on word durations, Bell et al. (2009) proposed an explicit mechanism whose purpose is to coordinate the speed of lexical access with the speed of articulation. The authors argue that such a mechanism is consistent with their finding that lexical frequency interacted significantly with conditional probability, such that lower frequency words showed *weaker* effects of conditional probability than higher frequency words. The logic was that such an interaction is consistent with a ceiling effect on lengthening, rather than a floor effect on reduction, suggesting that for the least accessible words, the local speech rate can only become so slow before an actual disfluency is triggered.

In these papers and in others not discussed here, Bell, Jurafsky, and colleagues have thus taken a probability-driven approach to understanding phonetic reduction, with the logic that the speed of lexical access is likely to be causally related in some way to the speed of articulation (c.f. Jurafsky et al., 2001; Jurafsky, Bell, & Girand, 2002; Jurafsky, 2003). An approach that is similar in some respects can also be seen in work by Gahl and colleagues, which also suggests that probabilistic information about a word's usage affects the speed with which speakers articulate – and presumably encode – words in connected speech.

In a study comparing the duration of homophones in spontaneous speech, Gahl (2008) found that when other factors were brought under statistical control, the member of the homophone pair with higher lemma frequency was produced with shorter duration than its lower frequency twin. The significance of this finding for the present discussion is that it is consistent with the idea that faster articulation is at least in part dependent on higher level factors. Because homophones have the same segmental representation, the finding that lemma frequency (and not just phonological form frequency) affects word duration in conversational speech again suggests that greater activation at higher levels of representation is somehow translated into faster articulatory speed.

Work by Gahl and Garnsey (2004) additionally suggests that "higher level activation" is not limited to effects of lexical frequency and local conditional probability. Rather, speakers' knowledge of a word's usage appears to extend to more abstract information concerning the probability that a word will occur in a particular syntactic frame. In this study, the duration of verbs in read speech depended on their bias towards a given usage. When verbs were embedded in sentences that were consistent with their most frequent syntactic usage, they were produced with shorter duration than when they were embedded in less probable syntactic frames. This finding suggests that information about a word's usage extends to the level of syntactic probability, and that such abstract grammatical factors may also be relevant for understanding the accessibility of a word in a given context.

Finally, consistent with work examining the effect of phonological neighborhood density on lexical accessibility in single word productions, Gahl et al. (2012) recently demonstrated that neighborhood density was also positively correlated with phonetic reduction in spontaneous speech. In that study, measures of both word duration and vowel quality indicated that greater neighborhood density was associated with phonetic reduction. Words with more neighbors had shorter total durations, and more centralized vowels, supporting the view that higher levels of lexical activation may be associated with both shorter production latencies (in laboratory studies) and shorter durations (in connected speech).

The studies reviewed thus far in this section all speak to a relationship between the probability of a word in a given context and the duration of that word in connected speech. One interpretation of this relationship (the one that has been pursued thus far) implicates lexical accessibility or activation as the driving force behind phonetic reduction processes. Under this interpretation, the speed with which a speaker can access a word in memory is directly tied to the speed with which he or she can articulate that word – either by way of activation cascading to the level of phonological representations, or perhaps by an explicit mechanism coordinating the speed of lexical access with the speed of articulation.

However, a different class of explanation posits that local adjustments in speaking rate fall out naturally from a principle of maximal efficiency. According to the *smooth signal redundancy hypothesis* (Aylett & Turk, 2004, 2006), the total amount of information present in the speech signal at any given time is approximately the same – that is, the redundancy in the signal is "smooth". Such an account predicts that when more information is available at higher levels of linguistic structure (such as when a word, syllable, or segment is highly predictable given its particular conversational, syntactic, or phonotactic context), the amount of information conveyed by the acoustic signal can be less. In other words, if a given sound is highly predictable for contextual reasons, then less acoustic evidence is necessary it order for it to be recognized.

Aylett and Turk have argued that the primary way in which the smooth signal redundancy hypothesis is implemented is through a language's prosody. The idea is that speakers' compensation for reduced levels of information in less predictable stretches of speech has largely been codified in the language's grammar through prosodic prominence. This hypothesis has been tested with respect to both syllable durations and vowel quality in connected speech.

Aylett and Turk (2004) examined syllable durations as a function of both prosodic prominence and several measures of language redundancy. Redundancy was quantified using measures at three levels of linguistic representation: syllable trigram probability, word frequency, and the number of previous mentions of the word in the discourse. Multiple regressions showed that several prosodic variables and all three measures of linguistic redundancy were correlated with syllable durations, once other factors were brought under statistical control.

Aylett and Turk (2006) extended this work, showing that vowels in syllables with high language redundancy were produced with both shorter duration and qualitatively more reduced (schwa-like) pronunciations. In both studies, language redundancy and prosodic prominence were found to be highly correlated, and the authors argued that because the variance accounted for by prosodic versus redundancy factors was largely overlapping, prosody (including durational lengthening) should be interpreted as compensation for unevenness in the distribution of linguistic information.

Two major points should be noted with respect to the smooth signal redundancy hypothesis. The first is that it generally makes the same predictions as accessibility-based accounts of phonetic reduction; in both cases, higher frequency and contextual predictability are predicted to be associated with shorter segmental durations. This underscores the difficulty of determining when phonetic duration should be attributed to speaker-oriented versus listener-oriented factors, as described above in the section enumerating the many sources of durational variation in language.

The second point is that because experimental studies have made considerable progress in determining the architecture and dynamics of the speech production system as regards short utterances produced in the laboratory, it is now possible to begin asking whether activation based accounts can fruitfully be extended to conversational speech. Again, this does not discount the fact that speakers are able to and likely do adjust their pronunciations in consideration of their listeners in certain contexts, and it does not discount the idea that certain aspects of such adjustments may be incorporated into a language's grammar over time. Rather, it begs the question of how far accessibility-based accounts can go in explaining phonetic variation in natural speech, and this is the overarching motivation for the studies presented in the dissertation.

2.6 Duration and motor control

The preceding sections have reviewed the literature on lexical access and phonological encoding in order to lay the groundwork for an account of phonetic duration that is consistent with what is known about such higher level planning processes. However, the only way for processes of linguistic planning to interact with phonetic duration is through the motor system. This section provides a brief summary of current ideas in motor planning, in order to provide a basic framework for understanding the link between phonological planning and the movements of the articulators.

Most current theories of motor planning appeal to the notion of *optimal control*, the idea that "the motor system aims to minimize a cost function that reflects some combination of effort, variability, or the satisfaction of task goals." (Haith & Krakauer, 2013, and references therein). To better understand the nature of the optimizing function, it will first be necessary to describe the concept of *signal dependent noise*. Harris and Wolpert (1998) first introduced the idea of signal dependent noise in an influential paper modeling the trajectory and duration of eye saccades and arm movements. They begin from the observation that noise is inherent in the firing of motor neurons, and they reason that the longer such noise is permitted to accumulate, the greater the potential deviation from the target location becomes. To minimize deviations from the target, then, one would predict that motor movements would be executed as quickly as possible.

However, this is not consistent with the observation that more precise movements tend to be longer in duration. To reconcile these observations, Harris and Wolpert proposed that the amount of signal-dependent noise is proportional to both the duration of movement and the mean level of the signal. The result is that, "in the presence of such signal-dependent noise, moving as rapidly as possible requires large control signals, which would increase the variability in the final position. As the resulting inaccuracy of the movement may lead to task failure or require further corrective movements, moving very fast becomes counterproductive."

The significance of the concept of signal-dependent noise is that it helps explain the precise nature of the speed–accuracy trade-off in motor planning. The existence of noise in the system dictates that faster movements are generally better, but the proposed relationship between the amount of noise and the magnitude of the signal imposes an upper limit on speed; both very slow and very fast movements are dispreferred. Harris and Wolpert therefore propose that movements are executed with the minimum duration possible, given the precision required by the task: "the temporal profile of the neural command is selected . . . to minimize the movement duration for a specified final positional variance determined by the task."

The implication for speech planning is that all else being equal, the movements of the articulators should operate following a similar principle of optimal control. The precision of the movements required for speech articulation imposes a relatively constant tolerance for positional variance, and the degree of tolerance plus the signal-dependent noise should determine the duration of articulatory movements, all else being equal. Assuming that speech articulation behaves similarly to other types of motor control, articulatory movements should be executed as quickly as possible given the task demands.

One question that arises is how the articulatory system "knows" what the tolerance level is for the positional variance of the articulators. It stands to reason that a certain amount of motor learning would be necessary in order to achieve sufficient knowledge of the "relatively constant tolerance for positional variance" described above. Data that speak directly to this question come from the speech acquisition literature, from studies that have examined the durational changes in children's articulations over time.

In general, evidence from the developmental speech literature indicates that children's

articulatory movements become faster and less variable through a protracted learning process. Nittrouer (1993) compared the durational properties of stops and vowels produced by adults and children aged three, five, and seven years old. The durations of the schwa, stop closure, voice onset time, and full vowel were measured in the sequence /əˈCV/ (e.g. *a key*, *a two*) for words in the carrier phrase, "It's a ... Bob." In general, durations in children's speech were both longer and more variable than adults'. The between group differences in VOT and in full vowel duration, however, did not reach statistical significance.

More recent studies using kinematic tracking procedures have generally corroborated Nittrouer's acoustic data. Grigos, Saxman, and Gordon (2005) tracked the kinematic movements of the lips and jaw in conjunction with VOT over a period of time in which English speaking children acquired a perceptible voicing contrast (ages ranged from 19 to 21 months). The transition from no contrast to contrast was marked by a sudden jump in VOT values for the voiceless stops; children followed a pattern of initial overshoot of adult long-lag VOT values, followed by gradual reduction and refinement over time of the gestures associated with stop aspiration. There were no significant changes in the duration of articulatory movements over time, but the displacement and peak velocities of the articulators did exhibit significant change. The maximum displacement of the jaw and maximum speed of jaw and lower lip movements increased as children acquired the voicing contrast.

Grigos (2009) expanded on these findings, reporting reduced variation in jaw movement as children acquired the voicing contrast. It was argued that the observed reduction in movement variability reflected improved coordination between the oral and laryngeal articulators following acquisition of the voicing contrast.

Beyond gross differences in speed and variability between children and adults, some evidence has indicated that pattern frequency also affects motor control, again suggesting that practice plays an important role in improving gestural coordination. Munson (2001) measured the duration of consonant clusters in children's and adults' production of nonwords varying in phonotactic probability. For both children and adults, less frequent sequences were produced with greater variability in duration, and for children (but not adults), durations were also longer overall for less frequent sequences.

One interpretation of these results is that the optimization of articulatory effort and variability in motor control follows a very protracted development, since effects of phonotactic probability on variability were present even for the adults in Munson's study. However, a second interpretation is that both motor control and phonological encoding played a role in determining articulatory speed. It is possible that less frequent sequences are encoded more slowly at the level of phonological representation, and that this slower encoding process gives rise to longer articulation times. Without further data, it is difficult to determine the extent to which practice encoding phonological segments versus practice executing motor control sequences contributed to Munson's findings. It seems likely that, to a certain extent, both factors played a role.

Complicating the picture even further, Roelofs cites evidence that difficulty in higher level encoding is directly associated with the time needed to initiate movements; Rosenbaum (1980, cited in Roelofs, 1999) found that participants were slower to initiate arm movements

when they were less certain of the direction and extent of movement that would be required. It is therefore possible that an interaction between difficulty in phonological encoding and articulatory planning also plays a role.

In sum, the principle of optimal control provides a starting point for interpreting differences in articulatory duration due to difficulty in motor planning. Children's articulations in particular are expected to be longer and more variable than adults' because the precise relationships between motor commands and positional targets have yet to be fully learned; the complexity of the coordination problem dictates that until sufficient learning has occurred, more control can be maintained over slower movements. Finally, data relating articulatory duration and phonotactic probability suggest that effects of phonotactic pattern frequency may arise from multiple sources. The effects of phonological encoding time, encoding difficulty, and previous motor practice are all likely to contribute to the speed with which articulations are executed.

2.7 Research questions

The purpose of the preceding literature review has been to provide a broad overview of the many factors known to affect the duration of words and segments in speech production. Because the sources of durational variation are so wide-ranging, the goal of the dissertation is to examine one particular set of durational measures in a range of contexts, the intent being to shed light on how the demands of the speaking situation interact with the factors under consideration.

More concretely speaking, the durational measures to be examined in the chapters that follow are word-initial voice onset time, rime duration, and the total word duration of monosyllabic English words. These measures were chosen because they address several questions that have arisen in the review of the existing literature. Specifically, while some studies have suggested (directly or indirectly) that the speed of phonological encoding may provide a link between increased lexical activation and faster articulatory duration, thus far the evidence for this claim comes almost exclusively from measurements of total word durations in connected speech (and, to a limited extent, in single word productions, e.g. Balota et al., 1989).

The dissertation therefore examines the duration of articulatory movements associated with units smaller than the whole word, with the logic that looking at smaller units of articulatory organization can potentially provide more precise information with respect to phonological encoding. Voice onset time, rime duration, and total word duration will be examined in the context of three experiments. The hope is that in considering word duration in conjunction with voice onset time and rime duration, a clearer picture will emerge with respect to the relationship between activation at the lexical level, and encoding processes at the phonological level.

The specific research questions to be addressed in the dissertation are as follows:

1. To what extent does the existence of specific minimal pair neighbors in the lexicon contribute to variability in phonetic duration?

2. To what extent does the structure of the lexicon more generally contribute to systematic differences in phonetic duration?

3. Is there evidence that the speed of phonological encoding is related to articulatory duration in connected speech?

The chapter concludes with a discussion of the specific motivations, methodologies, and predictions associated with each of these research questions.

1. To what extent does the existence of specific minimal pair neighbors in the lexicon contribute to variability in phonetic duration?

Baese-Berk and Goldrick (2009) suggested that the existence of specific minimal pair relationships between words in the lexicon is relevant for understanding speaker-internal processes of feedback and competition. Recall that in that study, words with a minimal pair neighbor differing only in the voicing of the initial segment were produced with slightly longer VOT than words with no such neighbor, and this difference persisted even when the other member of the minimal pair was not present.

Baese-Berk and Goldrick's explanation of this finding was that when a minimal pair neighbor becomes active, the target word requires a boost in lexical activation to out-compete it, resulting in the slight hyperarticulation of the target word relative to words without such minimal pair neighbors. As discussed above, however, this explanation raises many questions regarding the link between minimal pair neighbors and hyperarticulation. Is the specific phonological contrast of any importance? Does the contrast between neighbors lead to hyperarticulation of particular phonological segments, or does the proposed boost in lexical activation affect all aspects of the target word equally?

To address these questions, a novel word learning study was conducted. In Experiment 1, preschool-aged children were taught two novel words that created specific minimal pair relationships with existing words in the lexicon. Over the course of three days, children learned the words *tog* and *keet* (contrasting with the known words *dog* and *cat*, among others), and longitudinal measurements were taken of participants' voice onset time, rime duration, and total word duration as the novel words were incorporated into the lexicon.

The specific research question was whether the articulatory durations for either the novel words or the known words would change as the new lexical representations became more established. The prediction was that if hyperarticulation is truly due to a boost of activation at the lexical level of representation, then the total word duration for both the novel words and the known words should increase over time, with voice onset time and rime duration increasing proportionally with total word duration.

If the particular phonological relationship between minimal pair words was important in some way, however, then the *tog–dog* and *keet–cat* pairs should behave differently. The voice onset time for *tog* would be expected to increase disproportionately to its rime duration and total word duration, while the voice onset time for *keet* should remain relatively unchanged. Such a finding would indicate that the importance of minimal pair neighbors for speaker-internal processes of speech production is primarily situated at the phonological, rather than lexical, level of representation.

2. To what extent does the structure of the lexicon more generally contribute to systematic differences in phonetic duration?

Numerous studies in the literature have indicated that feedback between lexical and phonological representations is likely to have important implications for production planning. In particular, activation and competition processes between similar words have been shown to affect the speed with which target words and their phonological representations can be accessed and encoded. Studies of single- and two-word production in the laboratory have indicated that particular phonological relationships between co-activated lexical representations can either be facilitatory or inhibitory, although the nature of this relationship is not yet well understood.

Experiment 2 therefore pursues a reanalysis of Baese-Berk and Goldrick's production data for minimal pair words produced in isolation[2]. While the original analysis examined VOT as a function of minimal pair status, the reanalysis presented in Chapter 5 examines VOT, rime duration, and total word duration as a function of the phonological neighborhood characteristics of the target words.

The reanalysis takes as its starting point the observation that while the "minimal pair" and "no minimal pair" words examined in the study were carefully matched with respect to their lexical and phonotactic frequencies, words in the minimal pair group had, on average, significantly more total neighbors than words in the no minimal pair group. In addition, because most phonological neighbors in English tend to differ by their onset consonant, the set of minimal pair words also had significantly more neighbors differing in their onset consonant than the set of no minimal pair words.

The specific research questions addressed in Experiment 2 were therefore whether VOT, rime duration, and total word duration differed significantly as a function of either total neighborhood density or position-specific density, and the prediction was that they would. The significance of this study is that such a finding would provide a more straightforward link between the results reported in Baese-Berk and Goldrick's original paper and studies of laboratory speech that have reported effects of phonological overlap on production planning.

[2]I gratefully acknowledge Melissa Baese-Berk and Matt Goldrick for sharing their data for the purposes of this reanalysis.

3. Is there evidence that the speed of phonological encoding is related to articulatory duration in connected speech?

Experiment 3 builds on the findings of Experiments 1 and 2, asking whether there is any evidence that the structure of the lexicon affects phonetic duration in connected speech. The specific research question in this chapter is whether the density and phonological structure of a word's neighborhood have any effect on the word's VOT, rime duration, and total duration in spontaneous, conversational speech. The logic is similar to that of the previous experiments; if VOT and rime duration are affected in some way by the number and/or type of a word's phonological neighbors, and if such effects are disproportionate to any effects on total word duration, this suggests that feedback processes and fluctuations in the speed of phonological encoding may be responsible for the differences in phonetic duration that have been observed in previous studies.

Gahl et al. (2012) recently demonstrated that greater phonological neighborhood density is associated with shorter total word durations in spontaneous speech, suggesting that phonological overlap between a target word and its neighbors is in this case facilitatory. The prediction with respect to total neighborhood density and total word duration is therefore straightforward: words with more neighbors are predicted to have shorter total durations, all else being equal.

Such an effect may actually be reflective of a more complicated relationship, however. Since the total duration of a monosyllabic word is comprised primarily of the duration of its vowel, Gahl et al. (2012)'s findings may actually point to an effect of vowel overlap on vowel duration; the fact that most English phonological neighbors differ by their onset consonant entails that many English neighbors share the same vowel. For this reason, it is additionally predicted that the type of neighbors that overlap with the target word will affect articulatory duration in a position-specific way. Words with a disproportionate number of neighbors overlapping with the target in their initial segment are predicted to have relatively short VOT, and words with a disproportionate number of neighbors overlapping with the target in their final segments are predicted to have relatively short rime durations.

If it is the case that phonetic durations in spontaneous speech are affected by the number and segmental makeup of phonologically similar words in the lexicon, then it stands to reason that the speed of phonological encoding provides the link (or at least *a* link) between higher level planning processes and articulatory duration.

Chapter 3

Experiment 1: Phonetic duration in word learning

3.1 Background to Experiment 1

Previous findings

Two lines of research are directly relevant to the word learning study pursued in this chapter. The first is concerned with the developmental changes in children's motor control that affect the coordination and duration of their articulatory gestures, and the second is concerned with the relationship between the lexicon and children's phonological knowledge. Several studies reviewing developmental changes in children's articulatory control are reviewed in Chapter 2, but will also be summarized here. The connection between the lexicon and the development of phonological representations will then be considered in more depth below.

Summary of developmental motor control and articulatory duration

As discussed in the section on articulatory duration and motor control, the principles of optimal control and signal-dependent noise dictate that in general, articulatory movements will be executed as quickly as possible while maintaining an acceptable, relatively constant level of precision. In other words, there is a tradeoff between articulatory speed and precision, and the job of the motor control system is to learn the most optimal, efficient set of control parameters, "to minimize a cost function that reflects some combination of effort, variability, or the satisfaction of task goals." (Haith & Krakauer, 2013)

Studies of developmental changes in articulatory speed have been consistent with the idea that articulatory movements speed up as the motor system becomes more efficient. Nittrouer (1993), for example, measured the durations of a small set of consonant and vowel gestures produced by three-, five-, and seven-year-olds, and found that in all cases, children's movements were longer and more variable than adults' (although the differences in VOT and full vowel duration did not reach statistical significance).

In a longitudinal study of children's acquisition of the English stop voicing contrast, Grigos et al. (2005) and Grigos (2009) recorded both acoustic and kinematic tracking data of the lips and jaw as children acquired a reliable contrast between /b/ and /p/. The kinematic data revealed that the maximum speed of the articulators increased over time, and the variability in jaw movement decreased over time. The acoustic data indicated that when children began to produce a reliable voicing contrast (around 20 months of age, in this study), they initially overshot the target VOT values, producing a longer duration of aspiration noise than adults. A period of gradual refinement then followed, where children presumably gained more control over the coordination of the oral and laryngeal gestures, and VOT gradually shortened to more adult-like values.

The finding that young children's long-lag VOT follows a pattern of durational overshoot followed by gradual shortening over time is consistent with an earlier study conducted by Macken and Barton (1980), and Grigos et al. (2005) argue that it is also consistent with the physiological demands placed on the articulators; Lofqvist (1980; 1980, cited in Grigos et al., 2005) has posited that the laryngeal gesture required to produce long-lag VOT is ballistic in nature, which helps explain why the acquisition of voiceless plosives in English follows the observed trajectory. Learning to produce adult-like long-lag VOT values consists of first learning to abduct the vocal folds sufficiently, and later learning to control the variability in the abduction gesture itself, as well as the timing between the oral release of the stop, vocal fold abduction, and the resumption of modal voicing.

Developing efficient motor control of the articulators is not the only challenge in achieving adult-like articulatory durations, however. Speaking – transforming ideas into articulations – requires several levels of abstract linguistic representations, with the phonological level providing the interface between words and articulations. The next section provides a brief overview of what is known about the development of adult-like phonological representations, focusing on the development of adult-like durational patterns in speech production.

The lexicon, phonological knowledge, and articulatory duration

There is good reason to believe that children's phonological knowledge changes as they learn more words. Corpus studies of the growth of the child lexicon have suggested that when the lexicon is small and sparse, children may be able to store and access words using less structured and/or less phonologically specified representations (Charles-Luce & Luce, 1990, 1995), and numerous experimental studies have provided support for this idea.

Metsala (1999), for example, showed that three- and four-year olds performed better on tasks of phonological awareness for words in phonologically dense neighborhoods of the lexicon, suggesting that the need to distinguish similar neighbors in memory may encourage the development of more phonologically specified representations. Similarly, using a likeness categorization task, Storkel (2002) provided evidence that children the same age based their similarity judgments on more segmental representations for words in dense neighborhoods, but on more holistic, acoustically based representations for words in sparse neighborhoods.

The idea underlying this line of research on developmental speech perception is that phonological representations become more abstract as children learn more (and more varied) examples of words, a hypothesis sometimes referred to as lexical restructuring. A similar proposal has also been suggested with respect to developmental speech production, with some researchers arguing that the duration of children's articulatory movements partially reflect the extent to which they have developed abstract, recombinable motor control structures.

The logic of studies in this vein has typically been that nonword repetition tasks can be used to tap into the link between lexical and phonological representations. Munson, Edwards, and Beckman (among others) have often argued for a model of phonological acquisition in which phonological representations are gradually abstracted over the growing lexicon throughout development. In such a model, phonological representations are initially quite context-dependent, such that knowledge of a given sound is closely tied to the words in which it appears in the lexicon. Known words therefore provide crucial data on the types of phonological combinations that are possible, and they also provide the occasion to practice sequence-specific motor schemas.

Nonword repetition is seen as testing the state of phonological abstraction at a given point in time; speakers are more likely to achieve context-independent motor and phonological representations earlier for segments that occur frequently and in many different contexts, whereas infrequent segments with more restricted distributions in the lexicon may lag behind in developing robust, independent representations until a sufficient number of exemplars in a variety of word types are learned.

In a nonword repetition paradigm, the duration of biphone sequences in particular is predicted to reflect the frequency with which the sounds in the pair occur next to each other in the lexicon. Infrequent sequences are predicted to be produced with longer and more variable durations because the planning structures necessary to articulate them are less practiced, and less automatized. Following this logic, Munson (2001) and Edwards, Beckman, and Munson (2004) examined the effect of phonotactic probability on children's production of biphone sequences, and Munson, Swenson, and Manthei (2005) looked at the relationship between phonotactic probability, phonological neighborhood density, and vowel duration.

In Munson (2001), adults and children heard and repeated back nonwords containing biphone sequences with high versus low transitional probabilities. The dependent measures of interest were gross repetition accuracy (operationalized as the number of segmental substitutions made), the absolute duration of the biphone sequences, and variability in biphone sequence duration. The results showed that children's productions of low probability sequences were overall less accurate, and had longer and more variable durations, as predicted. Adults were also found to produce low frequency sequences with greater durational variability, but there was no significant difference in absolute duration between high and low frequency sequences.

Edwards et al. (2004) further extended Munson (2001), replicating the earlier findings using a larger, more varied stimulus set and a much larger group of children. The difference in absolute duration between high and low frequency sequences was greatest for the

youngest group of children, and decreased with age. Importantly, Edwards et al. (2004) additionally identified a significant correlation between vocabulary size and gross repetition accuracy; children with smaller vocabularies made more segmental substitutions (e.g. replaced the sequence /fk/ with the sequence /ft/) when repeating low frequency sequences than did children with larger vocabularies. The regression analyses also showed that expressive vocabulary size was a better predictor of gross accuracy than was receptive vocabulary, perhaps suggesting that production-specific knowledge is a better predictor of performance in speech production tasks.

Munson et al. (2005) also investigated the durational properties of children's productions of nonwords, but they extended the investigation to include real words as well. In this study, children in two age groups (four year-olds and seven year-olds) repeated nonwords varying in phonotactic probability as well as real words varying in neighborhood density. Two dependent measures were examined: repetition latency and vowel duration. The high and low density real word stimuli were matched for phonotactic probability, segmental content, age of acquisition, and frequency, such that any observed differences in latency or duration could be attributed to neighborhood density, and not to a confounding variable.

There were main effects of lexical status and age group: repetition latency was longer for nonwords than for real words, and longer for younger than for older children. The effect of phonotactic probability was significant for both groups of children, such that both younger and older children initiated articulation more quickly for high probability nonwords than for low probability nonwords. The effect of neighborhood density on repetition latency was significant only for the older group of children, who began articulating real words in high density neighborhoods more slowly than words in low density neighborhoods. These findings are all consistent with previous work with adults, indicating effects of lexical competition for real words and phonotactic facilitation for nonwords in a repetition task (Vitevitch & Luce, 1999, 2005).

With respect to vowel duration, the only significant effects were of lexicality and phonotactic probability. Nonwords were produced with shorter vowels than real words, and nonwords with high phonotactic probability were produced with shorter vowels than nonwords with low phonotactic probability. There was no difference in duration between vowels in high density versus low density words, for either group of children. The findings regarding vowel duration in nonwords are consistent with Munson (2001) and Edwards et al. (2004) in demonstrating that segments in low probability sequences are produced with longer durations. It is interesting, however, that vowels in nonwords were produced with significantly *shorter* duration than vowels in real words, and also that no differences with respect to neighborhood density emerged in the analysis of vowel duration in real words. If phonotactic familiarity were the only factor contributing to the duration of individual segments, then known words should always be produced with shorter duration than nonwords, since the exact phonotactic sequences present in known words are by definition already known.

Munson et al. (2005) do not speculate as to the reason for the lexicality effect, but one possible explanation is that the existence of a lexical representation fundamentally changes the dynamics of production by creating the possibility of lexical-phonological feedback. This

could potentially explain why real words were produced with longer vowel durations than nonwords, but why vowel durations in nonwords were shorter for more frequent sequences. Under this account, representational independence and articulatory practice at the segmental level facilitates the production of phonotactic sequences, and this effect is most easily observed in nonwords, for which the production routines must be assembled on the fly. In real words, however, it could be that the existence of a lexical-level representation overshadows or perhaps wipes out any effects of phonotactic facilitation.

A possible mechanism for such an effect could follow the logic of Baese-Berk and Goldrick (2009)'s study of VOT in minimal pair words; it could be that in single-word productions, the target lexical representations need to overcome competition from non-target representations, and this competition leads to global hyperarticulation for all segments in the target word. This would be consistent with Munson et al. (2005)'s findings: the need to resolve lexical competition in perception before selecting a word and initiating articulation would lead to longer latencies in repeating words with more competitors, and higher activation resulting from lexical-phonological feedback could lead to longer word durations.

If we are to adopt this tentative hypothesis, several issues arise. For one, even if we accept that higher lexical activation leads to longer word duration in single-word productions, the reason for such a relationship is not immediately clear. As reviewed in Chapter 2, Gahl et al. (2012) demonstrated that words with more phonological neighbors were produced with shorter, more reduced pronunciations in conversational speech, and these authors speculated that the effect originated in the speed of lexical access; words with more phonological neighbors receive more feedback from phonological representations, presumably allowing them to be selected and planned more quickly. It is possible that in conversational speech, competition between phonologically similar neighbors is reduced as compared to single-word productions, since many other factors aid in boosting activation of the target word during running speech. *Cat* and *cab* have far less reason to compete for selection in a conversation about furry animals than in a task involving word recognition and repetition.

Single-word productions, on the other hand, do not benefit from the syntactic and contextual information relevant in conversational speech. Rather, one hypothesis is that absent contextual facilitation, lexical-phonological feedback results in globally longer durations in single-word productions. A possible way to tie together both Gahl et al. (2012)'s and Munson et al. (2005)'s findings could be that the effect of lexical-phonological feedback differs according to the speaking task. In Gahl et al. (2012), and in conversational speech in general, perhaps phonological neighbors do not compete for selection so much as provide support for fast lexical access and phonological encoding. In Munson et al. (2005), and in tasks involving perception and production of single words more generally, the number of lexical representations competing for selection is likely greater, as is the amount of time that elapses between initial activation of a word and its eventual articulatory realization. A reasonable hypothesis would therefore be that there is something about relatively prolonged lexical-phonological feedback that leads to longer word durations.

The primary hypothesis to be pursued in the present study is that the existence of phonologically related neighbors in the lexicon gives rise to increased articulatory duration in

single-word productions. This hypothesis assumes the interactive spreading activation model of speech production described in the Background to the dissertation, and it asks whether the need to out-compete a minimal pair neighbor leads to some form of hyperarticulation, as posited in Baese-Berk and Goldrick (2009), and as would be consistent with Munson et al. (2005).

The present experiment addresses this hypothesis by way of a word-learning experiment with preschool-aged children. This experimental paradigm was chosen because it simultaneously addresses multiple questions regarding the nature of the connections between phonological representations and the lexicon. Munson et al. (2005) suggested that general effects of lexical competition on articulatory duration may not be obvious in preschool-aged children, due to the relative sparseness of their lexicons. However, it has been repeatedly hypothesized that increasing lexical competition in preschoolers contributes to the restructuring of phonological representations (Charles-Luce & Luce, 1990, 1995; Metsala, 1999; Storkel, 2002). Taking these two hypotheses together, it may be the case that global neighborhood density has no measurable effect on children's articulations, but that the introduction of specific minimal pair neighbors into the lexicon *would* produce observable competition, when words with particular phonological relationships to one another are examined more closely. The present experiment directly tests this possibility.

A second question addressed is whether the type of phonological relationship between two neighbors affects the extent to which they interact with one another. It is reasonable to hypothesize, for example, that words that differ by their onset consonant (e.g. *cat–pat*) might have a closer relationship than words that differ by their vowel (e.g.*cat–kit*). This would predict stronger hyperarticulation effects when pairs such as *cat–pat* become co-activated, and weaker or perhaps nonexistent effects with *cat–kit* pairs.

Finally, the preschool word learning paradigm has several potential practical benefits. Because preschoolers are generally more actively engaged in word learning than adults, it is possible that it will simply be easier to establish full-fledged lexical representations more quickly and easily with preschoolers than with adults; adult word learning experiments have typically yielded mixed results, and have suggested that creating new lexical representations with adults in the laboratory is actually somewhat challenging (Gaskell & Dumay, 2003; Leach & Samuel, 2007). Additionally, precisely because the child lexicon is relatively sparser than the adult lexicon, adding a single word may have a relatively larger, more observable effect on children's productions than on adults'.

Research questions

1. To what extent do the durations of individual segments and whole words vary systematically as a function of their general neighborhood density and/or phonotactic probability in children's productions of already known words?

2. Does the production of a novel word change as it becomes integrated into the lexicon?

3. Does the increased lexical-phonological feedback created by the acquisition of a new word affect the production of already known words? If so, does the particular phonological relationship between words have an effect?

3.2 Methods

Participants

Participants were 23 children with an average age of 4 years;2.8 months (range = 3;4 to 5;2, SD = 0;6) on their first day of testing. All children attended the same preschool, exhibited no signs of speech, language, or hearing disorder, and were progressing through school normally, as reported by parents and teachers. As a prerequisite to participation in the study, all participants' parents completed a consent form and language background questionnaire. Many parents reported significant exposure to languages other than English; this information is summarized briefly in Table 3.1. For some children, languages other than English were spoken by one or more family members in the home. Some children had attended Spanish immersion daycare or had non-English speaking caregivers for some period of time. However, all children attended day-long preschool in English, spoke English fluently, and scored within the normal range on a standardized expressive vocabulary test (the Expressive One Word Picture Vocabulary Test, Brownell, 2000). Due to the relative difficulty of finding strictly monolingual children in the community where the study was conducted, the large amount of variation in language background making it difficult to create clear-cut groupings of participants, and in light of the children's demonstrated proficiency in English, any effects of multilingual language background are not considered here.

Stimuli

The stimuli for Experiment 1 can be grouped into three global types: baseline words, novel words, and minimal pair words. Over the course of three days, children provided baseline pronunciations for twenty already known, familiar baseline words. They were then taught two novel words, "keet" ([kit]) and "tog" ([tɔg]). On the final day of the experiment, they again produced a subset of the baseline words, along with the novel words and the already known words that now formed minimal pairs with the novel words (*cat*, *coat*, and *feet* all formed minimal pairs with the novel word *keet*, and *dog* formed a minimal pair with the novel word *tog*).

The experimental stimuli are shown in Table 3.2. The purpose of the baseline stimuli is two-fold. First, they allow us to ask which phonological and lexical factors, if any, affect children's voice onset time and word durations for already known, familiar words. These analyses are presented in Section 3.3. Second, the production of the baseline words on Days 2 and 3 provides experimental control; since the baseline words are phonologically dissimilar to the novel and minimal pair words, any changes in word duration or VOT that occur

participant	sex	age	EOWPVT	other language(s)
01	M	4;1	114	Spanish
02	M	3;10	75	Spanish
03	F	4;7	122	Spanish
04	M	4;9	117	–
05	F	4;8	109	Swedish
06	F	5;0	111	Urdu, Arabic
07	F	4;0	–	Spanish
08	F	4;10	122	–
09	F	4;4	120	Chinese
10	M	4;8	96	Chinese
11	M	4;4	109	Spanish, Japanese
12	F	5;2	125	–
13	F	4;10	107	Spanish
14	F	3;10	145	Spanish, Korean
15	M	4;1	122	Spanish
16	F	4;3	–	–
17	F	4;0	86	Chinese
18	F	3;4	138	Hindi
19	F	4;2	145	German
20	F	3;5	94	Hebrew
21	M	4;2	–	–
22	F	3;10	126	–
23	F	3;7	137	–
summary	**7 M**	**4;2.8**	**116**	**7 monolinguals**

Table 3.1: Participants in Experiment 1. Ages given are for the first date of testing (with the average time elapsed between Day 1 and Day 3 of the study being 11.6 days). "EOWPVT" is the child's score on the Expressive One Word Picture Vocabulary Test (Brownell, 2000), standardized for age, with "–" indicating the child did not complete the test. "Other language(s)" are based on parental report from the language background questionnaire.

over time in the baseline words must be due to the repetitive nature of the experimental paradigm, and not to the exposure to the novel words.

Because one of the primary variables of interest was voice onset time (VOT), care was taken to match the stimuli for place of articulation of the stop consonant and for the height of the following vowel, since these factors are known to affect VOT for universal, physiological reasons (Cho & Ladefoged, 1999). Stimuli were also matched as closely as possible for syllable structure. All words are monosyllabic, and the majority are of the form consonant-vowel-consonant (CVC). However, a further significant constraint on stimulus selection was that all words needed to be highly imageable and familiar to a preschool-aged child. For this reason, several words with complex codas (e.g. *pants*, *toast*) were included.

Table 3.3 provides summary statistics for the experimental stimuli, including the results of a t-test comparing the high and low density words on each measure. The mean values for number of phonological neighbors, average segmental frequency, average biphone frequency,

ND	vowel height	onset	word	type	days	syll structure	# of neighbors
HD	hi V	p	peach	baseline	1	CVC	16
HD	hi V	t	tail	baseline	1	CVC	18
HD	hi V	k	coat	minpair	1, 3	CVC	19
HD	hi V	k	cake	baseline	1	CVC	21
HD	(hi V)	–	feet	minpair	1, 3	CVC	17
HD	lo V	p	pen	baseline	1, 2	CVC	16
HD	lo V	t	top	baseline	1	CVC	14
HD	lo V	t	tie	baseline	1	CV	22
HD	lo V	k	cat	minpair	1, 3	CVC	27
LD	hi V	p	pig	baseline	1	CVC	12
LD	hi V	t	toast	baseline	1	CVCC	8
LD	hi V	k	keys	baseline	1	CVC	8
LD	hi V	k	comb	baseline	1, 2, 3	CVC	10
LD	(hi V)	–	face	baseline	1	CVC	6
LD	lo V	p	pants	baseline	1, 2	CVCCC	3
LD	lo V	p	pipe	baseline	1	CVC	10
LD	lo V	t	tongue	baseline	1, 2, 3	CVC	10
LD	lo V	t	tent	baseline	1, 2, 3	CVCC	10
LD	lo V	k	cow	baseline	1	CVC	9
LD	lo V	–	dog	minpair	1, 3	CVC	7
HD	hi V	k	keet	novel	1, 2, 3	CVC	21
LD	lo V	t	tog	novel	1, 2, 3	CVC	10

Table 3.2: Stimuli for Experiment 1. "ND" is neighborhood density classification, determined by a median split. The number of neighbors are based on the Child Mental Lexicon (Storkel & Hoover, 2010). "Days" are the days of the study on which the word was produced.

age of acquisition, and log word frequency are given first; all of these measures are drawn from the Child Mental Lexicon (Storkel & Hoover, 2010; http://www.bncdnet.ku.edu/cgi-bin/DEEC/post_ccc.vi). The CML is the result of combining several corpora of spontaneous and elicited speech produced by kindergarten and first grade children. Because the current study is concerned with the speech production of children of roughly the same age, this corpus was considered to provide the most relevant lexical and phonological statistics. However, since it is conceivable that the frequency with which a child hears a word spoken could also impact children's lexical representations, a measure of word frequency in conversational adult speech is also included for comparison: frequency (in words per million) in the SUBTLEX-US database of American English movie subtitles (Brysbaert, New, & Keuleers, 2012).

As seen in Table 3.3, the high and low density stimuli differed significantly only with respect to the number of neighbors in the child lexicon. Mean segment frequency, mean biphone frequency, age of acquisition, log word frequency in child speech, and word frequency in adult speech were all statistically equivalent for high versus low density stimuli. The novel word stimuli were also within the range of phonotactic probability for the known word stimuli; the mean segment frequencies for *keet* versus *tog* were 0.064 and 0.034, and the mean biphone frequencies were 0.003 and 0.001, respectively. This places *tog* on the low end

	# of neighbors	segment frequency	biphone frequency	age of acquisition	log word frequency (CML)	frequency (SUBTLEX)
HD	18.8	0.060	0.0049	3.96	3.21	50.14
LD	8.5	0.057	0.0047	4.46	2.40	72.71
t test	$p < 0.01$	$p = 0.6$	$p = 0.9$	$p = 0.3$	$p = 0.08$	$p = 0.5$

Table 3.3: Summary statistics (mean values) for stimuli in Experiment 1. Age of acquisition ratings come from Kuperman et al. (2012), and "frequency (SUBTLEX)" refers to word frequency (per million words) in the SUBTLEX-US database of American English movie subtitles (Brysbaert et al., 2012). All other measures come from the Child Mental Lexicon (Storkel & Hoover, 2010).

of phonotactic probability as compared to the other words in this study. However, the fact that *tog* has ten phonological neighbors in the child lexicon suggests that this should not be a cause for concern. The phonotactic probability for *keet* places it squarely in the center of the distribution of the known words.

Procedure

For all experimental tasks, children were tested in a quiet room at their preschool, and were seated at a small table across from the experimenter (the author). The experiments were presented as games involving picture naming and guessing, and all sessions were recorded in their entirety[1]. Children wore an Audio Technica Pro 70 lapel microphone and their speech was recorded onto a Marantz PMD661 solid state recorder, with 16-bit quantization and a sampling rate of 44 kHz. Recordings were later downsampled to 22 kHz before acoustic analysis.

Day 1

Two experimental tasks were administered on Day 1: picture naming for known words, and the first round of word learning.

Baseline: picture naming. For the picture naming task, children were presented with colorful drawings on laminated 4" x 6" cards and asked to name the objects depicted. Twenty cards depicting the twenty baseline words were shuffled and presented in a random order. After all twenty had been named, the cards were re-shuffled and named a second time in a different random order. The pictures used to elicit the twenty baseline words are shown in Figure 3.1.

Most of the children immediately recognized most of the words. In some cases, the child gave a different name for the picture or was unable to come up with a name. In these cases,

[1]Due to equipment failure, several of the recordings were lost. Fortunately, the majority of these were for Day 2 of the experiment, and therefore did not affect the baseline or post-test measures. For the one baseline recording and one post-test recording that were lost, these sessions were re-run in their entirety.

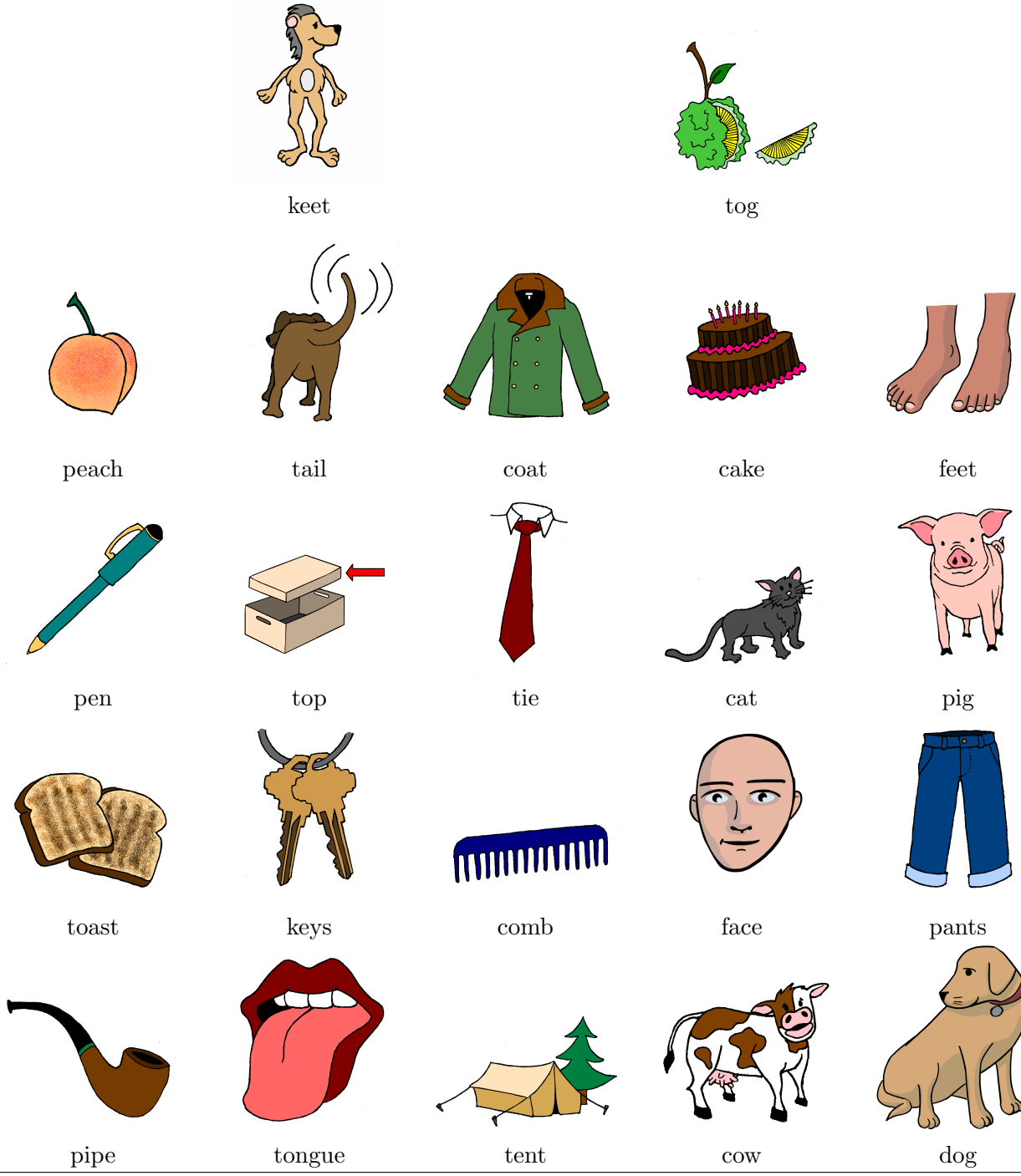


Figure 3.1: Pictures used to elicit words in Experiment 1.

	Day 1	Day 2	Day 3	TOTAL
keet	67/108 (62.0)	130/146 (89.0)	124/132 (93.9)	321/386 (83.2)
tog	72/123 (58.5)	141/163 (86.5)	114/120 (95.0)	327/406 (80.5)
peach	30/41 (73.2)	–	–	30/41 (73.2)
tail	41/46 (89.1)	–	–	41/46 (89.1)
coat	28/43 (65.1)	–	76/83 (91.6)	104/126 (82.5)
cake	45/48 (93.8)	115/116 (99.1)	108/108 (100.0)	268/272 (98.5)
feet	38/39 (97.4)	–	107/111 (96.4)	145/150 (96.7)
pen	35/43 (81.4)	83/87 (95.4)	–	118/130 (90.8)
top	33/44 (75.0)	–	–	33/44 (75.0)
tie	42/49 (85.7)	–	–	42/49 (85.7)
cat	47/47 (100.0)	–	95/95 (100.0)	142/142 (100.0)
pig	103/105 (98.1)	–	–	103/105 (98.1)
toast	34/42 (81.0)	–	–	34/42 (81.0)
keys	33/35 (94.3)	–	–	33/35 (94.3)
comb	37/45 (82.2)	97/110 (88.2)	81/88 (92.0)	215/243 (88.5)
face	38/39 (97.4)	–	–	38/39 (97.4)
pants	43/46 (93.5)	108/111 (97.3)	–	151/157 (96.2)
pipe	25/44 (56.8)	–	–	25/44 (56.8)
tongue	43/46 (93.5)	95/98 (96.9)	96/96 (100.0)	234/240 (97.5)
tent	34/45 (75.6)	98/103 (95.1)	91/93 (97.8)	223/241 (92.5)
cow	41/44 (93.2)	–	–	41/44 (93.2)
dog	58/58 (100.0)	–	114/114 (100.0)	172/172 (100.0)

Table 3.4: Number of spontaneous productions as a proportion of the total productions of each word, for each day. Words produced in larger phrases (e.g. "a birthday cake" instead of "a cake") are included as correct productions in this table of counts, but were excluded from the durational analyses. The percentages given in parentheses can be considered a rough index of identification accuracy.

the experimenter supplied the target word, and the child was asked to repeat it back. All repetitions (including those occurring on subsequent days of testing) were coded as such and excluded from the analyses; only tokens where children spontaneously produced the target word are analyzed here. Words preceded by a determiner (e.g., "a dog") were included, but words in longer phrases (e.g. "a birthday cake" instead of "a cake") were excluded.

Table 3.4 gives the number of spontaneous productions for each word produced on each day, as a proportion of the total number of tokens produced. The percentage of spontaneous productions (given in parentheses) can be taken as a rough index of familiarity with the target words. Note that these percentages are used as an approximation of identification accuracy and entered as a potential predictor in the models of word duration and VOT described in Section 3.3.

Word learning: story-telling. After providing baseline pronunciations, children were introduced to the novel words. Participants were told they would play a story-telling game with the experimenter; the experimenter would read them a story, and it was their job to act out the story using the characters provided. The "characters" were drawn in the same style

as the pictures depicting the baseline words, and were printed on small laminated cards, cut out so that they could be used as figurines. The experimenter first introduced the characters, saying, "These are the characters in our story. This one is called a *keet*, and this one is called a *tog*." The children were asked to repeat the names back before the story began. As the experimenter read the story, the children were encouraged to pick up the figurines and move them around on the backdrop to act it out. The text of the story was as follows:

> This is a story about a *keet* and a *tog*. Which one of those is the keet? And which one of those is the tog?
>
> One day, the keet is walking through the forest when he sees his friend, the pig. The pig is standing next to a tree. The keet walks up to the pig. "Hello, pig!" says the keet. "How are you today?"
>
> "Hello, keet!" says the pig. "I am *so* hungry! I didn't have breakfast this morning. And what I would like is a juicy, delicious piece of fruit." Do you see a juicy, delicious piece of fruit?
>
> "Well, we are in the forest," says the keet, "and sometimes fruit grows on trees." Can you put the tog in a tree? "Maybe if we go looking, we can find some fruit for you to eat."
>
> So the keet and the pig go for a walk in the forest. They search and search, and they look and look, when suddenly the pig says, "Look! Up in that tree! A *tog*!"
>
> "A tog?" says the keet. "What's a tog?"
>
> "A tog is my favorite kind of fruit!" says the pig. "It tastes delicious! How can we get it down?" How do you think they're going to get it down?
>
> Well, the keet decides to climb the tree. The keet climbs up in the tree, and he gets the tog, and he brings it down to his friend, the pig. "Here you go, pig!" says the keet. "A delicious tog for you to eat!"
>
> "Thank you, keet!" says the pig. And he gobbles up the tog: *om nom nom.* Can you make the pig gobble up the tog?
>
> The end.

At the end of the story, participants were told it was their turn to tell the story, and the experimenter would act it out. Most children were too shy to re-tell the entire story. To ensure that all participants had heard and understood the story, and to encourage them to produce the novel words several times, the experimenter asked leading questions, such as "Who was walking through the forest?" and "What did the pig say?". After completing the re-telling task, the children were taken back to the classroom. The entire experimental session on Day 1 typically lasted eight to twelve minutes.

I'm thinking of something . . .
pants . . . that you can wear. . . . that you put on your legs.
pen . . . that you can write with. . . . that you might use at school.
cake . . . that's made of chocolate. . . . that has candles on it.
comb . . . that you use to fix your hair. . . . that's blue.
tongue . . . that's part of your body. . . . that's in your mouth.
tent . . . that you use when you go camping. . . . that you can sleep inside.
keet . . . that's a kind of animal. . . . that lives in the forest. . . . that's brown. . . . that's friends with a pig.
tog . . . that grows in a tree. . . . that's a kind of fruit. . . . that's green and bumpy. . . . that the pig likes to eat.

Table 3.5: Example hints used for the 2AFC word learning task in Experiment 1.

Day 2

Word learning: 2AFC. On Day 2, the experimental session focused on reinforcing the novel words *keet* and *tog*. Children played a board game with the experimenter, the goal of which was to help some "children" game pieces reach the end of the path, where they would reach a birthday party for one of their friends. On each trial, participants were presented with a random pair of the printed cards from Day 1, and were given a "hint" which one the experimenter was thinking of. If they "guessed" correctly, they could move the game piece ahead one space. The hints focused on physical and functional attributes of the target words, reinforcing the semantic properties of the novel words. Examples of the hints used in this two-alternate forced choice (2AFC) task are given in Table 3.5.

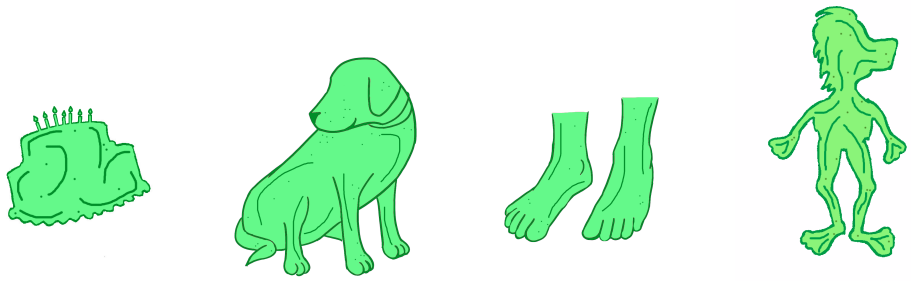
Only eight target words were elicited in the 2AFC task: two unrelated control words (*pants* and *pen*), the two novel words (*keet* and *tog*), and four phonetic control words starting with comparable onsets as the novel words (*cake*, *comb*, *tongue*, and *tent*). This reduced subset of target words ensured that children received ample practice with the novel words, and that they produced multiple tokens of each target word. There were 24 spaces from the beginning to the end of the game board, ensuring that all participants provided at least three repetitions of each target word (the number needed to move one game piece from the start line to the finish line). The picture cards were shuffled after each run through the stack, such that each pairing of pictures was random. Some children enjoyed the game and wanted to move multiple game pieces to the finish line, causing the total number of repetitions per child to vary considerably. The Day 2 experimental session lasted approximately eight to fifteen minutes.

Day 3

Post-test: picture naming. On Day 3, participants completed a second picture naming task, this time targeting ten words: the novel words *keet* and *tog*, the phonetic control words *cake*, *comb*, *tongue*, and *tent*, and the minimal pair words *cat*, *coat*, *feet*, and *dog*. The words *cat* and *coat* differ from the novel word *keet* by a vowel change, while *feet* and *dog* differ from the novel words only by their onset consonant. These minimal pair words were chosen because they are highly imageable and familiar to preschool children, and they represent two types of phonological neighbors that have traditionally been considered equivalent in models of the lexicon: they all differ from the novel words by a one phoneme substitution.

The Day 3 picture naming task was again presented as a board game played by the child and the experimenter. Participants were presented with three frog game pieces attempting to cross a pond. Each frog had a colored bag associated with it, and each bag contained "lily pads" in the shape of the target objects. The lily pads were versions of the previously used pictures (which were by now very familiar to the children) that were colored entirely green. Examples of some of the "lily pad" game pieces are provided in Figure 3.6. Participants drew one lily pad at a time out of one of the colored bags, and after "guessing" the name of the picture, they placed the lily pad in the pond. Once the lily pads for one frog were in place, the child made the frog hop across the pond to eat the fly that was waiting for it on the other side.

The frog game board had fifteen spaces, such that moving three frogs across the board resulted in 45 total productions; the target words were thus produced either four or five times each. Children selected the order of the colored bags to draw from and drew pictures out of each bag in a random order. Table 3.7 is a simplified version of Table 3.4 summarizing the number of analyzable tokens of the Day 3 target words produced on each of the three days.



cake dog feet keet

Table 3.6: Example "lily pad" versions of target pictures used for picture naming on Day 3 of Experiment 1.

Word	Stimulus Type	Day 1	Day 2	Day 3	TOTAL
keet	novel	67	130	124	321
tog	novel	72	141	114	327
cake	control	36	115	76	227
comb	control	37	97	81	215
tongue	control	43	95	96	234
tent	control	34	98	91	223
cat	CC minimal pair	47	–	95	142
coat	CC minimal pair	28	–	76	104
feet	VC minimal pair	38	–	107	145
dog	VC minimal pair	51	–	103	154

Table 3.7: Number of tokens included in the longitudinal analyses. Only words that were elicited on Day 3 are shown here. See Table 3.4 for the total number of tokens produced for all words.

Day 4

Expressive vocabulary test. On the final day of testing, children completed the Expressive One-Word Picture Vocabulary Test (Brownell, 2000), a standardized test of expressive vocabulary that has been extensively normed for use with a wide age range. Participants are presented with a series of colored illustrations depicting a variety of objects and actions, and they continue naming pictures until supplying six incorrect or unknown responses in a row. Three participants were unavailable or did not elect to participate on Day 4. These children all spoke exclusively (S16, S21) or predominantly (S07) English at home, and showed no signs of developmental or linguistic delay, so their data have not been excluded. All children who completed testing on Day 4 scored within the normal range for their age; their scores are provided above in Table 3.1.

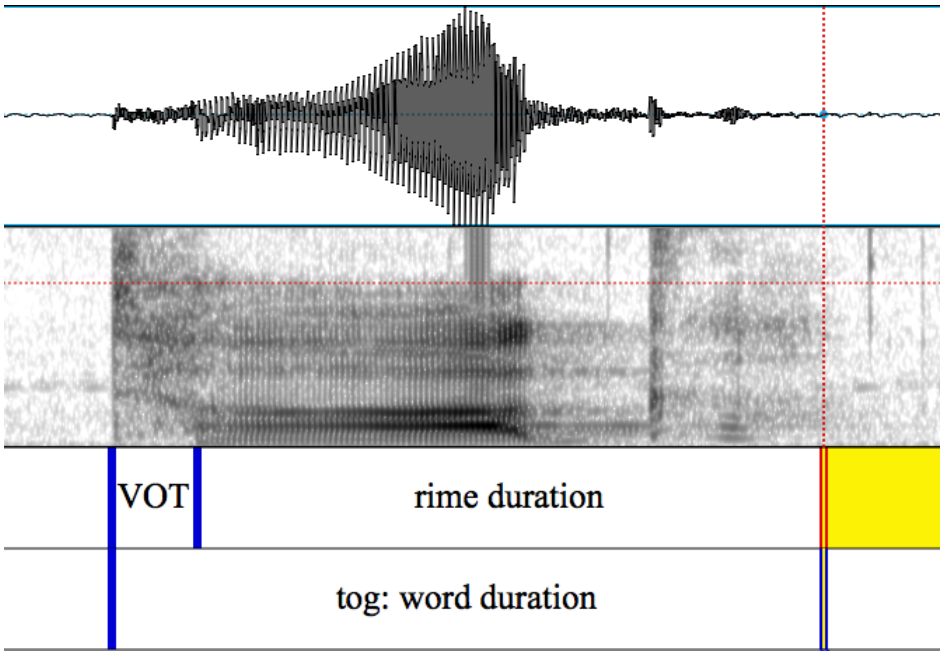


Figure 3.2: Spectrogram and waveform showing example segmentation boundaries for *tog*. Total word duration is marked on the lowest tier. VOT and rime duration are marked on the higher tier.

Acoustic analysis

The primary dependent measures under investigation are total word duration, voice onset time (VOT), and rime duration. All measurements were made by hand by the author using the Praat software for phonetic analysis (Boersma & Weenink, 2012) and were defined as follows. Word duration was the time at which any acoustic information associated with the speaker's articulation could no longer be detected, minus the time of the initial consonant release burst. For *face* and *feet*, the onset of the word was the point at which high frequency frication noise became visible on the spectrogram. Tokens of *dog* where the initial consonant was prevoiced were excluded from the durational analyses, but will be discussed briefly in Section 3.3. VOT was defined as the time of vowel onset (the first zero crossing at the beginning of the periodic vibration associated with the vowel) minus the time of the consonant release burst. Rime duration was defined as the offset of noise at the end of the word minus the vowel onset. Figure 3.2 provides example segmentation boundaries for the novel word *tog*.

3.3 Results

Day 1

Baseline word durations

Known words. The first question of interest was whether any of the lexical and phonological factors known to affect word duration in adult speech would affect children's productions of already known words. A mixed effects regression model was fit to the word duration data from Day 1. Subject and word were first entered as random effects, followed by the control parameters number of phonological segments and trial number in the experiment. Words with more segments unsurprisingly had longer durations. Increasing trial number was associated with a slight decrease in word duration, perhaps suggesting a practice or recency effect, but this predictor missed statistical significance (pMCMC = 0.2) and was dropped from the model.

The following variables were then investigated as potential predictors of word duration: frequency in the Child Mental Lexicon (CML; Storkel & Hoover, 2010); frequency in the SUBTLEX database of English movie subtitles (Brysbaert et al., 2012); number of phonological neighbors in the CML; age of acquisition rating in Kuperman et al. (2012); mean identification accuracy on Day 1 of the experiment (see Table 3.4); mean positional frequency of the phonological segments comprising each word, based on the CML; mean frequency of the biphone sequences comprising each word, based on the CML; and existence of a minimal pair based on the voicing of the first segment. The novel words *keet* and *tog* were excluded from this analysis. Predictors were added in a stepwise fashion in the order given above, and non significant predictors were dropped from the model.

Higher frequency in the CML initially appeared to be associated with shorter word durations, but closer inspection revealed this effect to be driven by the words *pants* and *keys*, which happen to have intrinsically longer durations due to their segmental content and also have a frequency of occurrence of 0 in the CML. When these words were excluded from the analysis, there was no trend suggesting an effect of CML word frequency. Furthermore, despite the fact that word frequency in the CML and word frequency in SUBTLEX are significantly correlated (Kendall's rank correlation tau = 0.476, z = 2.92, $p < 0.01$), there was also no trend suggesting a relationship between SUBTLEX frequency and word duration. No frequency measure was therefore retained in the final model.

Two predictors were initially found to account for a significant proportion of variance: the control parameter number of segments and the number of phonological neighbors in the CML. Words with more neighbors were produced with shorter word durations (pMCMC < 0.05). Because the high versus low density stimulus words were carefully matched for frequency and phonotactic probability, there was no significant correlation between number of phonological neighbors and either of the measures of phonotactic probability (z = 1.02, $p = 0.31$ for mean positional segment frequency; z = 1.37, $p = 0.17$ for mean biphone frequency), or between number of neighbors and either of the word frequency measures (z =

predictor	β	t	pMCMC
intercept	0.3012	4.65	0.0001
baseline word duration	0.9836	4.95	0.0001

Table 3.8: Fitted model predicting word duration for already known words on Day 1. This model includes random effects for subject and word. No other factors were found to be significant once baseline word durations were taken into account.

1.67, $p = 0.10$ for frequency in the CML; z = -0.46, $p = 0.65$ for frequency in SUBTLEX). There was also no correlation between number of neighbors and age of acquisition (z = -0.10, $p = 0.92$)[2]. However, despite the fact that the stimulus words were balanced as closely as possible for segmental content, the possibility remained that the intrinsic duration of the segments differed between high and low density words.

To investigate this possibility, average durations were first calculated for all segments in the Buckeye corpus of conversational speech. This is not the ideal source of information on average segment duration, since the Buckeye corpus is running, spontaneous speech produced by adults, and not citation pronunciations produced by children. However, the Buckeye data provides a single, very large sample of speech, which was thought to be preferable to compiling data on intrinsic segment durations from a variety of published studies using different subjects and methodologies, which in any case would still be based on adults rather than children. Average segment durations were therefore determined from the Buckeye corpus, and a baseline word duration was calculated for each stimulus word in the present set. The baseline durations of high density words were on average 29 ms shorter than low density words. A two-tailed t test comparing baseline durations for the two groups was not significant ($t(14.1) = -1.14$, $p = 0.27$), but because the difference is in the same direction as that found in the initial regression model, it was important to determine whether this confound in intrinsic duration had given rise to the apparent effect of neighborhood density on word duration.

A second mixed effect regression was therefore fit, this time including baseline word duration as a possible predictor. Baseline duration accounted for a significant proportion of variance, and indeed, when it was included as a control variable, no other predictors were found to be significant. That is, despite careful matching of the stimulus words, the apparent effect of neighborhood density on word duration was found to be an artifact of the stimulus set. The final model predicting total word duration is summarized in Table 3.8.

Novel words. To test whether the novel words were produced with significantly different total duration from the already known words, a mixed effects regression was fit to the Day 1 data, this time including the novel words *keet* and *tog*. Subject and word were again entered as random effects, and baseline word duration was again entered as a fixed effect; it remained a significant predictor when the novel words were included. Lexicality (known

[2]Kendall's rank correlation was used for all correlation tests, because it is more appropriate for small sample sizes, such as the 20 word stimulus set.

vs. novel word) was then added to the model. The novel words were produced with slightly longer duration on average, but this effect was not significant ($\beta = 0.0331$, $t = 0.82$, $p = 0.4$).

Baseline VOT

Known words. The same modeling procedure was followed for the analysis of VOT as for the analysis of total word duration: random effects were added first, followed by control parameters, and lastly by the lexical and phonological variables of interest. The dataset for the VOT model was slightly smaller than that used for the word duration model, because only words beginning with voiceless stops were analyzed (i.e. *face*, *feet*, and *dog* were excluded). A model predicting VOT of the known words only was fit first.

Because VOT is subject to variation from different sources than word duration, the control parameters entered into the VOT model additionally included consonant place of articulation, vowel height, and rime duration. The rationale for including rime duration as a predictor was as follows. VOT can be expected to expand or contract roughly in proportion to the rest of the word; words that are articulated quickly will have shorter VOT, and slowly articulated words will have longer VOT (Boucher, 2002; Smiljanić & Bradlow, 2008a, 2008b). All else being equal, VOT is expected to increase or decrease in proportion to rime duration.

A primary question of interest in the present investigation is whether the existence of minimal pairs with particular phonological properties have an effect on the phonetic realization of words above and beyond any lexical level effects; the VOT analysis is concerned with whether words such as *peach* (with its minimal pair *beach*) are produced with disproportionately longer (or shorter) VOT. That is, the question is whether minimal pair status has an effect on the VOT of a word-initial consonant in addition to any effects of phonological neighborhood on the total duration of the word. However, to include total word duration (or baseline word duration, as calculated above) as a predictor for VOT is partially redundant, since VOT is by definition included in the total duration of the word. For this reason, the duration of the word excluding the VOT (i.e. the rime duration) was entered as a control parameter.

Control variables were entered into the model in the order in which they were considered likely to yield effects: rime duration, onset consonant, vowel height, trial number, and number of phonological segments. The only control variables that predicted a significant amount of variance were rime duration and onset consonant, so they were retained. The same lexical and phonological variables investigated in the word duration analysis were then entered into the model in a stepwise fashion. The mean biphone frequency for the word was found to be a significant predictor of VOT. The frequency of the initial biphone alone was also found to be significant. However, because mean biphone frequencies for the stimuli were more evenly distributed than the initial biphone frequencies, and because the model including mean biphone frequencies accounted for more variance than the model including initial biphone frequency, the mean biphone frequency was taken to be the better predictor. No other

predictor	β	t	pMCMC
intercept	0.0653	8.95	0.0001
rime duration	0.0454	5.22	0.0001
onset = /t/	0.0057	1.66	0.1594
onset = /k/	0.0129	3.62	0.0036
mean biphone freq	-1.7419	-3.60	0.0046

Table 3.9: Significant predictors of VOT for already known words on Day 1. This model includes random effects for subject and word.

predictors reached statistical significance. The model predicting VOT for known words is summarized in Table 3.9.

Novel words. After fitting the model predicting VOT of the already known words, the novel words *keet* and *tog* were added to the dataset, and a second model was fit. All predictors from the model based only on known words remained significant. The only additional predictor that contributed predictive power for the expanded dataset was lexicality; novel words were produced with longer VOT than already known words ($\beta = 0.0137$, $t = 2.09$, $p = 0.0576$), and a chi-squared test indicated that the model including lexicality accounted for significantly more variance than the model without lexicality ($\chi^2 = 4.66$, $p = 0.03$).

Summary of Day 1 results

The only significant predictor of total word duration on Day 1 was baseline word duration. Once the baseline for the segmental content of the word was taken into account, no other lexical or phonological variables were related to total word duration. In general, VOT was highly predictable from word duration. However, several additional predictors emerged as significant in the VOT analysis: VOT varied by onset consonant, by mean biphone frequency of the word, and by lexicality. Words beginning with /k/ had significantly longer word-initial VOT than words beginning with /p/ or /t/; words with higher phonotactic probability, as measured by their mean biphone frequency, were produced with shorter VOT; and the novel words *keet* and *tog* were produced with significantly longer VOT than the already known words.

Changes in duration over time

Changes in total word duration

The words available for longitudinal analysis are a subset of the words examined in the baseline analysis. Twelve of the words produced on Day 1 were also produced on subsequent days. As described above, the two novel words (*keet* and *tog*) and a set of four phonetic control words (*cake*, *comb*, *tongue*, and *tent*) were produced on all three days of the experiment. In addition, two control words that were phonetically dissimilar from the novel words (*pen* and *pants*) were also produced on Day 2, and four minimal pair words differing from the

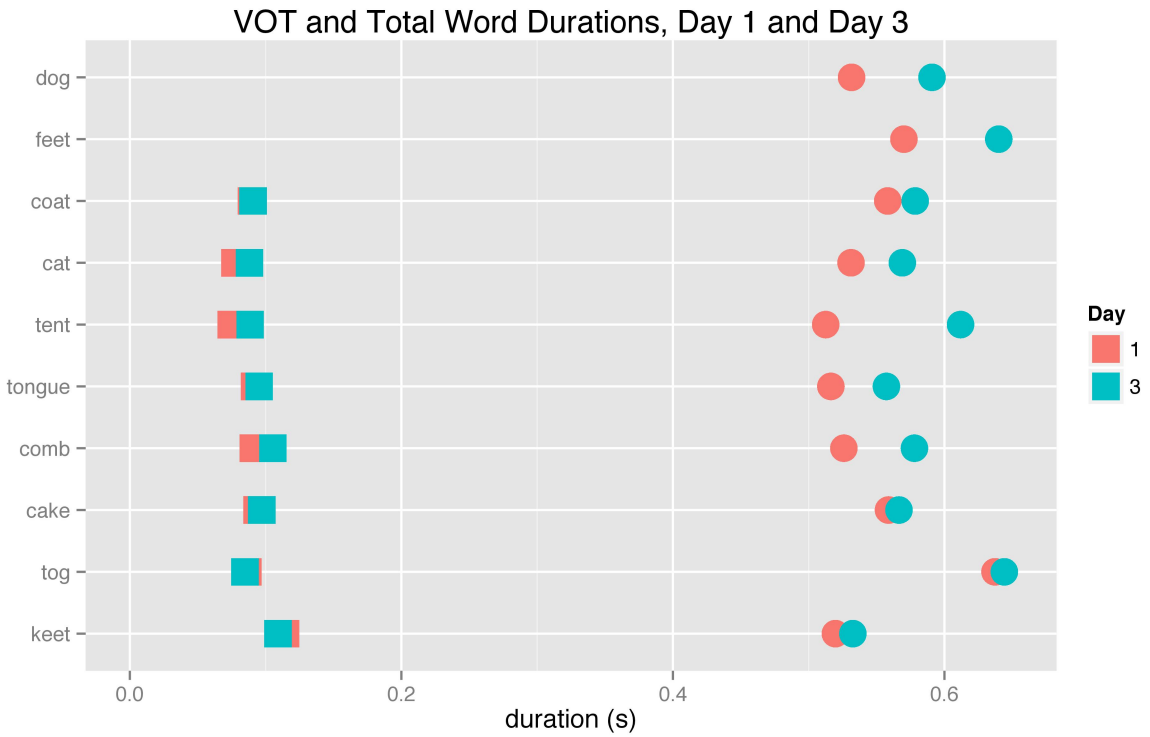


Figure 3.3: Raw mean durations for the ten words produced on Days 1 and 3. Squares represent voice onset time, circles represent total word duration, and the two colors represent Days 1 and 3.

novel words in either their first (*feet*, *dog*) or second (*cat*, *coat*) phonological segment were produced on Day 3.

Figure 3.3 shows the raw mean durations for the ten words that were produced on Days 1 and 3. The figure is oriented to resemble a spectrogram for ease of interpretation; time = 0 is equivalent to the location of the stop burst, the colored squares represent the onset of the vowel on Days 1 and 3 (red and blue squares, respectively), and the colored circles represent the offset of the word on Days 1 and 3. (A word's mean rime duration is thus the difference between the circle and square of the same color.) Words are ordered from the bottom as follows: novel words, phonetic controls, so-called "CC minimal pair" words, and "VC minimal pair" words.

Perhaps the most striking aspect of Figure 3.3 is the clear main effect of Day on total word duration: on average, all words had longer duration on Day 3 as compared to Day 1. Figure 3.4 provides a more detailed depiction of this overall increase in total duration, showing just the change in word duration from Day 1 to Day 3.

To determine whether the increase in word durations over time was statistically significant, two mixed effects models were fit to the total word duration data. For both models,

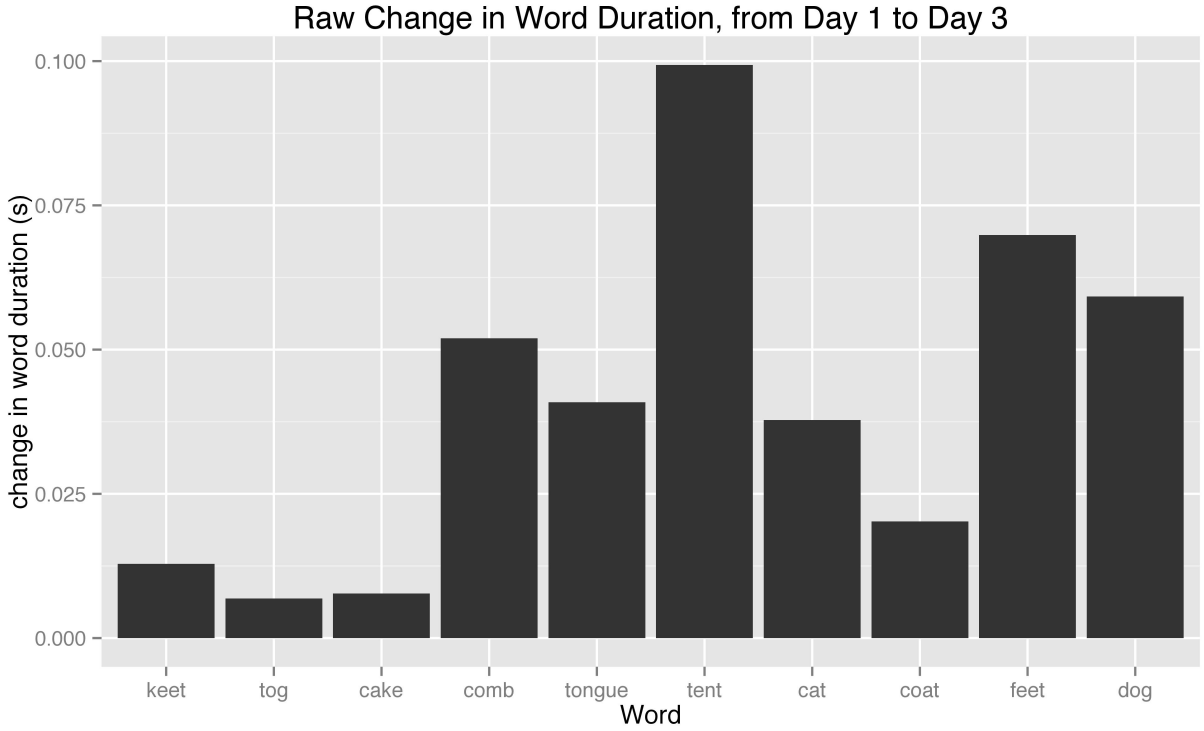


Figure 3.4: Raw change in word duration for the ten words produced on Days 1 and 3.

subject was first entered as a random effect. Due to the relatively small size of the data set, two alternate strategies were then explored to account for the grouping of observations by word. In the first model, word was entered as a fixed effect, followed by day, followed by their interaction. Both word and day returned significant coefficients, but their interaction did not. In particular, the words *feet*, *keet*, and *tog* were produced with significantly longer duration than the other words, and the word *tongue* was produced with significantly shorter duration. The main effect of day indicated that all words were produced with longer duration on Day 3 as compared to Day 1. This model is summarized in Table 3.10.

The second modeling strategy was based on examination of the plots in Figures 3.3 and 3.4. The words in each *a priori* category seem to have behaved similarly, with two clear exceptions; the novel words *keet* and *tog* showed very little increase in average total duration, the phonetic control words *cake* and *tongue* and the CC minimal pair words *cat* and *coat* exhibited similarly small increases, and the VC minimal pair words *dog* and *feet* exhibited the largest numerical increase as a group. The two seeming outliers are the words *comb* and *tent*, which were intended to serve as phonetic control words, but which increased considerably in average duration from Day 1 to Day 3.

Recall the data on identification accuracy (or more precisely, the proportion of spontaneously produced target words) presented in Table 3.4. While most children recognized

predictor	β	t	p**MCMC**
intercept	0.553	23.57	< 0.0001
word = cat	-0.021	-0.95	0.3392
word = coat	-0.008	-0.32	0.7424
word = comb	0.000	0.02	0.9882
word = dog	0.001	0.03	0.9762
word = feet	0.041	1.86	0.0540
word = keet	-0.038	-2.16	0.0348
word = tent	0.001	0.05	0.9674
word = tog	0.055	3.12	0.0014
word = tongue	-0.038	-2.03	0.0372
Day 2	0.018	1.37	0.1664
Day 3	0.040	3.38	0.0004

Table 3.10: Significant predictors of total word duration for the ten words produced on Days 1 and 3. This model includes a random effect for subject.

the pictures of the comb and the tent, some of them were initially unable to spontaneously supply the intended label for the object. Figure 3.5 plots the proportion of spontaneously produced labels against the age of acquisition rating in Kuperman et al. (2012). Kendall's rank correlation returns a highly significant relationship between these two variables (z = -2.60, $p = 0.009$), suggesting that mean naming accuracy on Day 1 was related to children's familiarity with the target words. Perhaps more importantly for the analysis of word duration, *tent* and *comb* are clear outliers in both mean accuracy and age of acquisition rating, especially as compared to the rest of the words elicited on Day 3.[3]. It is unclear at present why lower initial familiarity with a word would result in a greater effect of recent exposure on word duration in a picture naming task. This question must be left up to future research, and at present this finding is taken simply as evidence that *comb* and *tent* should not be grouped with the other phonetic control words in the durational analyses.

The second model examining longitudinal changes in word duration therefore asked whether there were significant changes in duration for the following groups of words: novel words (*keet* and *tog*) vs. control words (*cake* and *tongue*) vs. CC minimal pair words (*cat* and *coat*) vs. VC minimal pair words (*feet* and *dog*). Subject was entered as a random effect, followed by a fixed effect of baseline word duration based on segment durations in the Buckeye corpus, which was found to be a highly significant predictor in the analysis of the 20 words produced on Day 1. Fixed effects of word type, day, and their interaction were then entered. Unsurprisingly, baseline word duration was again a highly significant predictor. Beyond the effect of baseline duration, however, novel words and the VC minimal pair words had significantly longer overall duration than the control words, and all words were produced with significantly longer duration on Day 3. This model is summarized in Table 3.11

[3]The other obvious outlier, *coat* can be explained by the fact that children overwhelmingly preferred the term *jacket*. This is likely a natural consequence of the climate where the study was conducted.

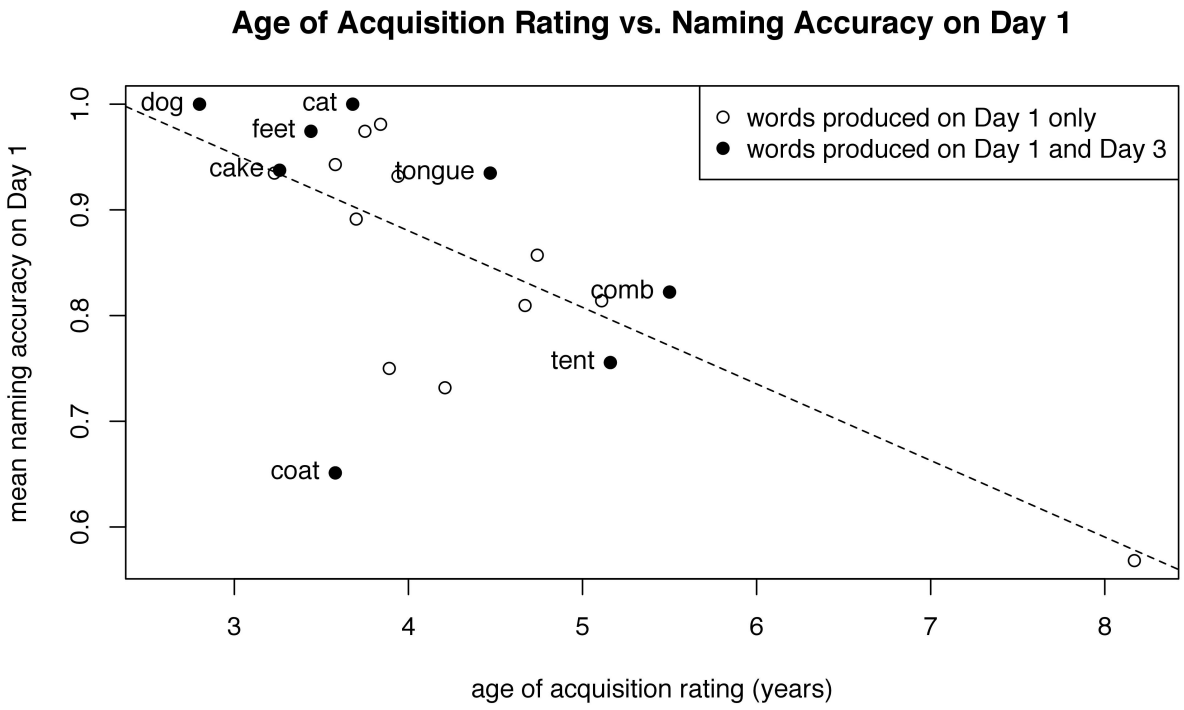


Figure 3.5: The relationship between mean naming accuracy on Day 1 (as indexed by the proportion of spontaneously produced tokens) and age of acquisition rating for stimuli in the present experiment. See text for details.

predictor	β	t	p**MCMC**
intercept	0.389	5.91	0.0001
baseline word duration	0.565	2.37	0.0204
Word Type = novel	0.030	2.57	0.0106
Word Type = CC minimal pair	-0.014	-0.76	0.4486
Word Type = VC minimal pair	0.048	3.16	0.0018
Day 2	0.018	1.27	0.1990
Day 3	0.032	2.61	0.0090

Table 3.11: Significant predictors of total word duration for word types produced on Days 1 and 3. This model includes a random effect for subject.

	dec	inc
cake	4	11
cat	6	15
coat	6	12
comb	6	12
feet	6	12
keet	7	14
tent	3	14
tog	8	11
tongue	7	13

Table 3.12: Number of participants producing increases vs. decreases in mean word duration on Day 3 (as compared to Day 1) for words in Experiment 1.

As a group, then, children reliably produced all words with longer duration on Day 3, as compared to Days 1 and 2. To explore the possibility that some participants may have produced a greater increase in word duration than others, random slopes for the effect of day were added to the models described above. In both cases, the random slopes contributed significant predictive power to the models, and in both cases, the t statistic for the main effect of day dropped below the level of statistical significance ($t = 1.80$ for the model with fixed effects for word; $t = 1.45$ for the model with word type as a predictor, excluding *comb* and *tent*). This suggests that the main effect of day reported above was driven by a subset of the participants, and that not all children produced longer durations on Day 3.

To better understand this finding, by-word mean durations were calculated for each participant, and participants were categorized as either "increasers" or "non-increasers" according to whether they produced a majority of the ten target words with longer average duration on Day 3. Sixteen out of the 23 total participants produced a majority of the ten words with longer duration on Day 3. Two participants did not provide any data for Day 3, two participants produced exactly half of the words with longer duration, and three produced more words with shorter duration on Day 3.

The by-word means also suggest that all ten words were equally likely to be produced with longer duration on Day 3. Table 3.12 gives the number of children who produced each word with relatively shorter vs. longer duration on Day 3 of the experiment. (The rows sum to different totals because not all participants provided usable data on both days.)

In the aggregate, then, most participants produced most words with longer durations on Day 3 as compared to Day 1. This finding, combined with the statistical analyses indicating no significant interaction term between word type and day, indicates that all words were equally likely to increase in duration over the course of the experiment, and that the magnitude of the observed increase in word duration was consistent across words. There is no statistical evidence that the words' phonological properties had an effect on their overall increases in word duration over the course of the experiment.

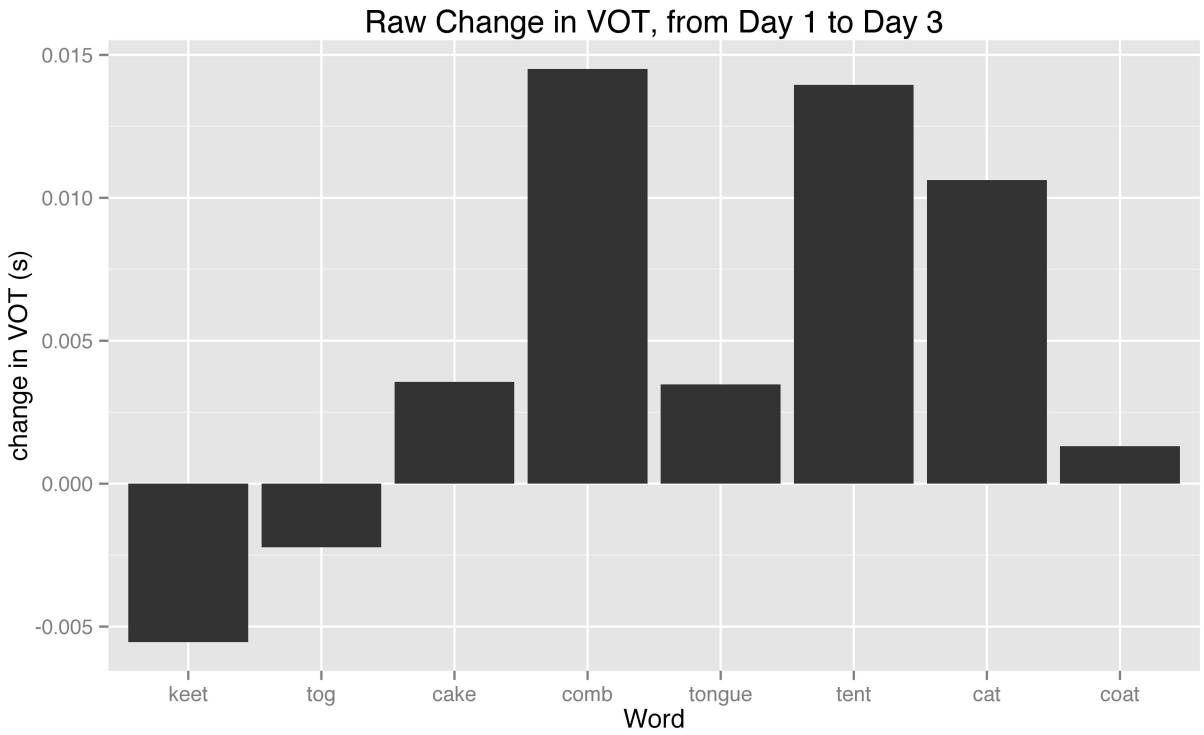


Figure 3.6: Raw change in VOT from Day 1 to Day 3 for words beginning with voiceless stops.

Changes in VOT

The raw means for VOT are shown above in Figure 3.3. Figure 3.6 provides a more detailed depiction of the raw differences in VOT between Day 1 and Day 3. For already known words, the changes in VOT track well with the changes in total word duration; *tent* and *comb* show the greatest increases in VOT, and all other known words exhibit smaller increases. (Note that *feet* and *dog* are not plotted here, since they do not begin with voiceless stops.) Strikingly, however both novel words exhibit overall *decreases* in VOT from Day 1 to Day 3. This is in contrast to the findings for total word duration; the novel words were qualitatively identical to the known words with respect to word duration, but were qualitatively different with respect to VOT.

To determine whether any longitudinal changes in VOT were statistically significant, the same modeling strategy was applied in the analysis of VOT as for the analysis of total word duration. Two models were fit, the first of which compared all words beginning with voiceless stops to all other voiceless stop words, and the second model compared groups of words to one another according to their *a priori* word type.

Random intercepts for the effect of subject were first added. The control variables exam-

predictor	β	t	p**MCMC**
intercept	0.0712	12.36	< 0.0001
rime duration	0.046	8.87	< 0.0001
word = cat	-0.0091	-1.91	0.06
word = coat	-0.0004	-0.07	0.94
word = comb	0.0059	1.42	0.16
word = keet	0.0246	6.51	< 0.001
word = tent	-0.0099	-2.43	0.01
word = tog	-0.0098	-2.61	0.01
word = tongue	-0.0026	-0.64	0.53
Day 2	-0.0083	-2.82	0.004
Day 3	-0.0006	0.21	0.83

Table 3.13: Significant predictors of VOT for the eight words beginning with voiceless stops produced on Days 1 and 3. This model includes a random effect for subject.

ined were the same as in the VOT model fit for the baseline words on Day 1. Rime duration was again highly predictive of VOT, but onset did not contribute significant predictive power to the model when the fixed effect of word was also included. The fixed effect of day revealed that VOT was systematically shorter on Day 2 as compared to the other two days, but there was no overall increase or decrease in VOT when comparing Day 1 to Day 3 (beyond the across-the-board increase in word duration, which is captured by the fixed effect of rime duration).

One possible reason for the overall lower VOT values observed on Day 2 is that the task performed on that day was qualitatively different. Recall that Days 1 and 3 were both picture naming tasks, while Day 2 was a two alternative forced choice task intended to focus children's attention on the semantic properties of the words. This point will be considered in more depth in the discussion.

The second model fit to the VOT data used word type as a predictor, rather than modeling all words separately. This model did reveal a main effect of onset consonant, as well as significant interactions between onset consonant and word type, and between word type and day. Words beginning with /k/ had significantly longer word-initial VOT overall, and this effect was especially pronounced for the novel words. However, novel words were produced with significantly shorter VOT on Day 3 as compared to Day 1. Note that in order to examine the interaction between Day and Word Type, it was not possible to include the data from Day 2 in the model, since the CC minimal pair words were not produced on Day 2.

Since it is not clear whether the VOT for phonologically voiced stops should be expected to behave similarly to VOT for phonologically voiceless stops, the word *dog* was not included in the statistical analyses of VOT presented above. While /d/ is normally considered to be a voiced stop in phonological descriptions of English, VOT for English /d/ typically falls in a bimodal distribution; while some proportion of tokens are truly fully voiced, most tokens are realized with short-lag VOT on the order of 20 ms. One of the questions of interest in the

predictor	β	t	p**MCMC**
intercept	0.0412	5.66	0.0001
rime duration	0.0962	10.42	0.0001
onset = /k/	0.0083	2.07	0.0372
Word Type = novel	0.0003	0.043	0.9866
Word Type = CC minimal pair	-0.0392	-0.88	0.3970
Day 3	0.0048	1.08	0.2840
onset = /k/ : Word Type = novel	0.0298	4.84	0.0001
onset = /k/ : Word Type = CC minimal pair	0.0297	0.67	0.5238
Word Type = novel : Day 3	-0.0138	-2.15	0.0378
Word Type = CC minimal pair : Day 3	0.00004	0.01	0.9972

Table 3.14: Significant predictors of VOT for word types produced on Days 1 and 3. This model includes a random effect for subject.

	Day 1	**Day 3**
pre-voiced	7	11
short-lag	51	103

Table 3.15: Number of tokens of *dog* beginning with pre-voiced vs. short-lag stops in Experiment 1. There was no significant difference between days (Fisher's exact test $p = 0.6088$).

current investigation is whether adding new words to the lexicon affects the pronunciation of already known words. If speakers change their pronunciation of already known words so as to maximize contrast with newly learned words, a possible manifestation of such a change for the word *dog* would be a greater proportion of pre-voiced tokens of word-initial /d/, since pre-voicing would maximize the phonological contrast between *dog* and the novel word *tog*.

To test this hypothesis, all tokens of *dog* were coded for whether they were produced with pre-voiced or short-lag word-initial /d/. Table 3.15 provides the number of tokens falling into each category, broken down by day. Seven of 51 total /d/'s (13.7%) were produced with pre-voicing on Day 1, versus 11 of 103 tokens on Day 3 (10.7%). Fisher's exact test returns a p-value of 0.6088 for this table of counts; there was no difference in the number of tokens of *dog* produced with pre-voicing on Day 1 versus Day 3.

Finally, it is possible that the VOT for tokens of *dog* produced with short-lag stops could decrease in response to learning *tog*. To test this possibility, a random effect for subject and fixed effects for rime duration and day were entered into a model predicting VOT for short-lag tokens of *dog*. The effect of day did not reach even marginal statistical significance (β = -0.0012, t = -0.76, pMCMC = 0.4449). There is no evidence that learning the word *tog* affected children's VOT for the word-initial stop in *dog*.

Changes in rime duration

The changes in rime duration that occurred over the course of the experiment can largely be deduced from the changes in total word duration and VOT, since the rime is simply

the portion of the word that is not the VOT. However, one question does merit further investigation: since VOT was found to be significantly shorter on Day 2 than on the other two days, and since total word duration was found to increase numerically but not significantly from Day 1 to Day 2, it is as yet unclear whether rime duration was significantly greater on Day 2.

To answer this question, two mixed effects models were fit to the rime duration data, following the same strategy outlined in the previous analyses. Both models indicated significantly longer rime durations on Days 2 and 3, as compared to Day 1. (For the model treating word as a fixed effect: $\beta = 0.0258$, $t = 2.09$, pMCMC $= 0.0346$ for Day 2; $\beta = 0.0353$, $t = 3.24$, pMCMC $= 0.0004$ for Day 3. For the model examining word types: $\beta = 0.0257$, $t = 1.92$, pMCMC $= 0.0588$ for Day 2; $\beta = 0.0260$, $t = 2.30$, pMCMC $= 0.0240$ for Day 3). Neither model returned a significant interaction between word type and day, indicating that rime duration increased over time irrespective of the phonological content of the word. However, when Day 2 was taken to be the baseline for comparison, no significant difference was found between rime durations on Day 2 versus Day 3, meaning that rime duration increased significantly from Day 1 to Day 2, but did not continue to increase significantly from Day 2 to Day 3.

Summary of changes over time

Two modeling strategies were undertaken in order to account for the fact that observations were statistically grouped by word: for each of the variables under investigation (total word duration, VOT, and rime duration), one mixed effects model treated word as a fixed effect, and a second model grouped words into one of four word types: control words, novel words, and two different types of minimal pair words. In the analyses of the grouped words, a fixed effect of baseline word duration was included to account for the different segmental content of each word, and the intended control words *comb* and *tent* were excluded because they were found to behave differently from the other control words. It was speculated that this difference might be due to the fact that *comb* and *tent* were outliers with respect to both age of acquisition rating and identification accuracy on Day 1 of the experiment.

Total word duration increased significantly from Day 1 to Day 3 of the experiment, and this occurred irrespective of word type; control words, novel words, and minimal pair words were all significantly longer on Day 3. The analysis of VOT, however, revealed an effect of lexicality: while VOT for known words increased in proportion to the rest of the word, VOT for novel words decreased over the course of the experiment. A closer look at the rime duration data indicated that, as might be expected from the analysis of total word duration, rime duration was significantly longer on Days 2 and 3 (as compared to Day 1) of the experiment.

3.4 Discussion

The goal of the present experiment was to investigate in greater detail the effects of lexical-phonological feedback on the durational properties of children's production of known and novel words. On the first day of the experiment, children completed a picture naming task, naming pictures of familiar objects with monosyllabic names of varying phonological neighborhood density and beginning predominantly with voiceless stops. Participants were then taught two novel words, *keet* and *tog*, thereby creating minimal pair relationships with several known words. Post word learning productions of *cat* and *coat* (CC minimal pair words) as well as *feet* and *dog* (VC minimal pair words) and a set of phonologically dissimilar control words were elicited in a follow-up picture naming task on Day 3.

The analyses of word duration, voice onset time, and rime duration suggested that the existence of phonologically related words did not have a significant influence on any of the durational properties under investigation. In the analysis of baseline productions, neighborhood density was not found to be related to any of the durational measures; neither total word duration, nor VOT, nor rime duration were affected by the total number of phonological neighbors in the lexicon. Moreover, the learning of the novel words did not seem to affect the durational properties of the known words or of their component parts. Rather, all words, including words that were phonologically unrelated to the novel words, increased equally in duration from Day 1 to Day 3 of the experiment. Although the difference in magnitude was not found to be statistically reliable, this effect was particularly striking for the control words *comb* and *tent*, which were outliers with respect to both age of acquisition rating and mean naming accuracy on Day 1.

While the results for word duration were qualitatively identical across word types, the findings for VOT were more nuanced. The baseline data indicated that stops in words with low phonotactic probability were produced with longer VOT, when the total duration of the word was controlled for. That is, even in already familiar words, the phonotactic probability of the word as a whole had a measurable effect on the timing of the gestures associated with the initial consonant. The effect of phonotactic familiarity was especially pronounced for the novel words, which were produced with significantly longer VOT than the known words on Day 1, and which were the only group of words to demonstrate a significant *decrease* in VOT from Day 1 to Day 3 of the experiment. (All other words exhibited an increase in VOT from Day 1 to Day 3 commensurate with the increase in total word duration observed from Day 1 to Day 3.) Taken together, these findings strongly suggest that children's coordination of articulatory gestures is highly dependent on the amount of practice accrued for the sequences being produced, and that at least for children of this age, phonotactic probability continues to be a significant determiner of articulatory duration even for familiar words.

What to make of the overall increase in word duration that occurred from Day 1 to Day 3 of the experiment? It should be noted that words did not increase (or decrease) in duration over the course of the task being administered on any given day; in none of the analyses was trial number a significant predictor (or even marginally related) to any of the durational measures. The increase in word duration seems to have occurred "wholesale" – in one fell

swoop – as a result of leaving the experimental setting and returning at a later date.

A possible explanation for this finding is that children approached the tasks on Days 2 and 3 with considerably different expectations than on Day 1. On the first day of the experiment, children had no *a priori* expectations for which words would be produced (or indeed, what their task would be at all). On Day 2, however, upon returning to the same experimental setting and seeing the same colored, laminated cards they had seen for the first time on Day 1, participants would have immediately adjusted their expectations for which words they would be expected to produce. This suggests two possible explanations for the observed increases in total word duration that occurred over the course of the experiment.

One possible explanation invokes the idea of increased lexical competition between words belonging to the same semantic (or in this case, situational) set. On Days 2 and 3 of the experiment, the set of words that had been produced on Day 1 were *all* much more viable competitors for selection than they had been on Day 1. The fact that children were expected to produce one of a small set of previously experienced words on any given trial may have increased the amount of competition equally for all of the words in the set, thereby increasing the duration of each word produced by a comparable amount. This explanation may be consistent with the numerically greater increase in word duration exhibited by the low familiarity words *tent* and *comb*, which would have had to "work" much harder to be selected than the other, more familiar words; if lexical competition is driving the increase in duration, then the representations for low familiarity words would need a particularly large boost in activation to overcome their competitors, which could cause them to have particularly large increases in duration.

Recall, however, that VOT was found to be significantly shorter for the words produced on Day 2 than for the same words produced on Days 1 and 3. Since VOT was repeatedly found to be sensitive to frequency effects in the present experiment, the decrease in VOT on Day 2 may have been the result of practicing the target sequences on Day 1. If this is true, then the subsequent increase in VOT on Day 3 could be attributed to lexical competition in the same way as the increase in rime durations on Days 2 and 3. This is perhaps not entirely outlandish, bearing in mind that in Munson et al. (2005), vowels in real words were produced with longer duration than vowels in nonwords. The same principle could be underlying both findings; namely, that once a word form hits some threshold level of phonotactic familiarity, lexical level influences on whole word duration can overshadow the influence of phonotactic familiarity on segment durations.

The second possible explanation also has to do with the small set of repeatedly produced words, but does not rely on production-internal factors. Rather, it is possible that given a known set of potential targets on Days 2 and 3 of the experiment, children may have produced each individual word with something akin to contrastive focus[4] Katz and Selkirk (2011) found that in read sentences, adult English speakers produced contrastively focused words with longer duration, higher pitch, and greater pitch movement. Thus if children

[4]I thank Jonah Katz for helpful discussion related to this interpretation. Any misunderstandings are, of course, my own.

	Day 1	**Day 2**	**Day 3**	β	t	p**MCMC**	**corr**	t	p
mean pitch (Hz)	267.9	291.0	277.9	10.67	3.8	< 0.0004	0.09	4.4	< 0.0001
peak pitch (Hz)	323.0	346.5	337.2	16.35	3.9	< 0.0001	0.14	6.9	< 0.0001
change in pitch (Hz)	98.3	110.1	109.5	13.15	3.25	< 0.0026	0.22	11.2	< 0.0001

Table 3.16: Post-hoc analysis of pitch in Experiment 1. Beta coefficients given are for a mixed model comparing pitch on Day 1 to Day 3, with random intercepts for participant and word. Pearson's product-moment correlations between pitch and duration are also given.

realized each word as an element contrasting with other possible elements in a fixed set of alternatives, the increases in total word duration observed in the present experiment should be correlated with higher pitch and greater pitch movement.

To investigate this possibility, a post-hoc analysis of three pitch-related variables was conducted: the mean pitch, peak pitch, and change in pitch was calculated for all words in Experiment 1. An automatic script recorded the average pitch (in Hertz) for the rime of each word, the highest pitch value detected within the rime, and the difference between the highest pitch and the lowest pitch within the rime. Table 3.16 gives the average of each of these values for each day of the experiment, as well as the beta coefficient for a mixed model comparing pitch on Day 3 to Day 1 (with random intercepts for participant and word), and the Pearson's product-moment correlation between pitch and total word duration.

As can be seen in Table 3.16, all three pitch measures increased significantly over the course of the experiment; average pitch, peak pitch, and change in pitch were all greater on Day 3 as compared to Day 1. In addition, all three were significantly correlated with the total word duration. These data are consistent with the second interpretation offered for the overall increases in duration; children appear to have produced words on Days 2 and 3 with something akin to contrastive focus. Note, however, that while the correlations between pitch and word duration are highly reliable, they are also rather weak, suggesting that contrastive focus is likely not the whole story.

A final point that bears discussion is that if lexical competition did contribute to the increases in word duration observed in the present experiment, it remains somewhat unclear why greater lexical competition would be associated with longer word durations in citation speech. This would seem to run counter to Gahl et al. (2012)'s findings for conversational speech, which demonstrated that words with more phonological neighbors were produced with shorter, more reduced articulations, and counter to previous findings that words with more neighbors are produced with shorter onset latencies in picture naming experiments with adults (Vitevitch & Sommers, 2003). A crucial missing piece of the puzzle is whether real words with more phonological neighbors are produced with longer durations in adults' citation speech. Baese-Berk and Goldrick (2009) reported longer VOT for words with more neighbors, but these authors did not report total word duration; it is quite possible that the VOT effect reported was actually an artifact of increased word duration. This will be the subject of Chapter 4.

If words in denser phonological neighborhoods are found to be produced with longer

durations in adult citation speech, but shorter durations in conversational speech, then the relationship between feedback and articulatory realization must be dependent on the nature of the task. One difference between conversational speech and single-word productions is the time that elapses between lexical selection and articulation. In conversational speech, this time is much shorter due to the strict time constraints imposed by running speech, and to the considerable facilitation provided by the syntactic and semantic context. In citation speech, there is no such time constraint, and also no such contextual facilitation; for single-word production, the time from lexical selection to articulation, and perhaps the influence of lexical-phonological feedback, may be crucially greater, and this may cause the number of neighbors in the lexicon to have a qualitatively different effect on phonetic duration.

3.5 Conclusion

In this chapter, the durational properties of preschoolers' single-word productions were shown to vary according to phonological and lexical factors. VOT was inversely correlated with phonotactic familiarity, while overall word duration lengthened over the course of the experiment. Two potential explanations were advanced for the increase in word duration over the course of the experiment, both of which centered on the fact that the same small set of words was presented on multiple days. There was no evidence that phonological neighborhood density had any effect on segment or word duration in this experiment. This is consistent with previous work demonstrating that processing differences between words in high versus low density neighborhoods seem to emerge later in development (Munson et al., 2005). The main question that emerges from this investigation is therefore whether increased lexical-phonological feedback for words in high density neighborhoods leads to increased durations in adult single-word productions, or whether it is simply not the case that feedback is reliably associated with longer durations. This will be the topic of investigation of Experiment 2.

Chapter 4

Experiment 2: Phonetic duration in single word productions

4.1 Background to Experiment 2

Activation and articulatory duration

Higher frequency is generally associated with shorter articulatory duration, and this has been demonstrated for representations at the lexical level (e.g. Bell et al., 2009) as well as representations at the sublexical level (e.g. Munson, 2001). In part, shorter durations for articulatory gestures associated with more frequent phonetic sequences can be attributed to increased familiarity or automaticity; as reviewed in Chapter 2, movements tend to reduce in duration and become more efficient as practice with a given motor schema increases (Haith & Krakauer, 2013; Harris & Wolpert, 1998).

However, this cannot be the whole story. For one, if practice or automaticity were the only reason for reduced articulatory duration, then different words made up of identical sequences (that is, homophones) should be produced with identical durations, all else being equal, but Gahl (2008) showed that in connected speech, higher frequency homophones are produced with shorter durations than their lower frequency counterparts. This suggests that the effect of lexical frequency extends beyond its association with motor practice; the fact that higher frequency words are generally more "accessible" may help to explain why they tend to be produced with shorter duration in connected speech.

The idea that words are produced with reduced articulations when they are more accessible is also consistent with the many studies demonstrating a relationship between duration and contextual predictability. The Probabilistic Reduction Hypothesis (Jurafsky et al., 2001; Bell et al., 2003) captures this observation, positing that more probable words tend to be realized with reduced productions in connected speech because of their relatively higher accessibility. It is important to note that most studies implicating a relationship between accessibility and word duration have focused on connected speech. While more accessible words, as indexed by their unigram frequency, are typically produced with shorter onset

latencies in single word production tasks, lexical frequency has not proven to be reliably related to the duration of isolated words (Balota & Chumbley, 1985; Levelt et al., 1999). It may well be that absent the need to string multiple words together, effects of lexical accessibility are primarily evident in the time it takes to initiate an articulatory plan, and not in the time it takes to execute that plan.

In spreading activation models of speech production, a unit that is relatively more accessible is considered to have a higher activation level (relative to its competitors, or perhaps relative to its own activation level in another context, for example). As laid out in Chapter 2, the dissertation assumes a model of speech production in which lexical and phonological representations interact. Such interaction is said to increase the activation levels of candidate representational units, giving rise to complex patterns of competition and facilitation. High frequency words can be modeled as having high resting activation levels, corresponding to their relatively higher accessibility when compared to low frequency words. Some researchers have hypothesized that higher lexical activation may extend to higher activation for the target word's phonemes; more accessible words may be easier to plan, and this may be reflected in their articulatory durations.

This relationship is not entirely straightforward, however; the seemingly conflicting findings reviewed in Chapter 2 indicate that the dynamics of lexical selection and phonological encoding are not well understood. The goal of the present chapter is therefore to better understand whether feedback between lexical and phonological representations can affect the durational properties of words produced in isolation. It is hypothesized that changes in the relative accessibility of phonological segments during production planning may give rise to differences in articulatory duration, and that such changes in accessibility may be related to a given word's neighborhood structure. In particular, the hypothesis pursued in this chapter is that a target word's phonological relationship to its neighbors can have effects on the duration of its segments.

In the next section, studies investigating the effect of phonological neighborhood density on articulatory duration in single word productions are reviewed briefly, motivating the analyses presented in the remainder of the chapter.

Previous findings on neighborhood density and duration in isolated words

Interactive spreading activation models of speech production (e.g. Dell, 1986, 1988) predict that words with many phonological neighbors should be more accessible due to lexical-phonological feedback. Feedforward activation is predicted to spread from the target word to its constituent phones, at which point feedback from the phones results in the activation of the target word's phonological neighbors. Reverberating activation between these neighbors and their constituent phones is then thought to improve lexical access for words in dense neighborhoods. Words with many neighbors receive more reinforcement, allowing them to be accessed more quickly and encoded more accurately: evidence from naming latencies

and speech errors suggests that increased feedback facilitates lexical access and phonological encoding (Vitevitch, 2002; Vitevitch & Sommers, 2003).

However, studies exploring the relationship between neighborhood density and speech articulation have suggested a more complex picture. One of the earlier studies examining neighborhood density and articulation was Wright (2004). Wright was not interested in articulatory duration, but rather in whether speakers adjust their pronunciation for words that are relatively difficult to recognize in order to make them more perceptible to listeners. Wright compared the resonant frequencies of vowels in "lexically easy" versus "lexically difficult" words, operationalized as high frequency/low density words and low frequency/high density words, respectively; low frequency words whose sound sequences overlap with many other words are potentially highly confusable, while high frequency words whose sound sequences render them relatively phonologically unique should be easier to recognize. The hypothesis was that speakers may use information regarding perceptual difficulty to make their speech more intelligible to listeners, and Wright's vowel formant frequency measurements supported this hypothesis. Vowels in lexically easy words were found to be relatively reduced, as measured by their distance from the center of the vowel space, while vowels in lexically difficult words were found to be relatively expanded, with formant frequencies closer to the periphery of the vowel space.

Munson and Solomon (2004) followed up on Wright (2004), qualitatively replicating the results from that study using a subset of Wright's stimuli. Munson and Solomon also reported vowel space expansion for "lexically difficult" words, again referring to low frequency words with many neighbors. In addition to measuring vowel space expansion, these authors also measured vowel duration. They reported that low frequency/high density words were produced with significantly shorter vowel durations, indicating that the observed vowel space expansion cannot have been due to a simple association between vowel duration and articulatory exaggeration.

Munson and Solomon also observed that the variables of interest in Wright's study were partially confounded in several respects. For one, the words with higher neighborhood density in his study and in their Experiment 1 also had higher phonotactic probability than words with low neighborhood density, and the segmental makeup of the two lists was also not balanced. More importantly, though, they note that the conflation of word frequency and neighborhood density makes it impossible to determine the independent effect of each. Their Experiment 2 therefore orthogonalized the two variables. The results indicated that vowels in high frequency words were produced with shorter durations and more reduced vowel targets than vowels in low frequency words, but vowels in high density words were only produced with more reduced vowel targets as compared to low density words; no independent effect of neighborhood density on vowel duration was found.

To better understand the effects of frequency versus neighborhood density on vowel articulation, Munson (2007) expanded on the single word reading task by including a delayed production condition. In the immediate condition, subjects were told to read the word on the screen as quickly as possible, and in the delayed condition, a 1000 ms delay was enforced between the stimulus onset and the signal to produce the word. Vowels in high frequency

words were again produced with shorter duration than vowels in low frequency words. There was also a main effect of condition: vowels in the delay condition were produced with very slightly but significantly longer duration than vowels in the immediate condition. No effect of neighborhood density was found on vowel duration, however.

Munson (2007) also analyzed reaction times, finding that high frequency, low density words were produced with slightly faster response latencies in the immediate response condition, but with slightly slower latencies in the delayed condition (although the latter difference was not significant in post-hoc testing). One interpretation of this result is that reaction times in the immediate response condition primarily reflected the speed of recognition and/or lexical access, and once the speed of lexical access was taken out of the picture, as in the delayed response condition, phonological encoding was slightly faster for words with many neighbors.

Kilanski (2009) explored the duration of several types of phonological segments in stimuli orthogonalized for neighborhood density and lexical frequency. There was a main effect of neighborhood density on vowel duration; vowels in high density words were produced with shorter duration than vowels in low density words. Kilanski also examined voice onset time (VOT) and fricative duration for word-initial consonants. With respect to VOT, there was a main effect of frequency such that high frequency words were produced with longer VOT than low frequency words overall. There was also a significant interaction between lexical frequency and neighborhood density: in high frequency words, higher density was associated with longer VOT, but for low frequency words, higher density was associated with shorter VOT. The same frequency/density interaction was found in the data on fricative duration, numerically speaking, but only the main effect of frequency was significant; words with higher frequency were also found to have longer word initial fricatives.

Kilanski also examined word final consonants and total word durations, finding shorter word final consonants in high frequency and high density words, as well as shorter overall word durations for high frequency and high density words. The conclusion drawn was that vowels and word final consonants seemed to have patterned together, while the onsets seemed to pattern "conversely with overall word duration", but no speculation was offered as to why this might have been the case.

Several studies by Rebecca Scarborough have investigated the effects of neighborhood density on vowel articulation, but Scarborough's work has primarily focused on coarticulatory vowel nasalization, and so will not be reviewed in detail here. Scarborough (2004) and Scarborough (2012) both report greater vowel nasalization in the context of a nasal consonant for words with many phonological neighbors. Scarborough (2012) also reports data on vowel duration, finding that vowels in high density words and nonwords were produced with significantly longer duration. This would seem to run counter to the other studies reviewed here, since high neighborhood density generally seems to have been associated with shorter vowel durations, if anything. Given that the stimuli used in Scarborough (2012) were not provided, however, it is not possible to determine whether the difference in vowel duration between the two stimulus sets could have been due to some confound in the experimental materials.

The final study to be reviewed here examined the effect of minimal pair status on word-initial VOT. Baese-Berk and Goldrick (2009) compared VOT for word-initial voiceless stops in words with a minimal pair neighbor based on the voicing of the first consonant (e.g. *cod*, which forms a minimal pair with *god*) to VOT in words with no such neighbor (e.g. *cog*, which has no counterpart *gog*). Words for which a voiced neighbor exists (henceforth "minimal pair words") were found to have significantly longer VOT than non-minimal pair words. Baese-Berk and Goldrick argued that feedback would lead to competition between the minimal pair neighbors, requiring a boost in the lexical activation of the target word in order to overcome its competitor. This boost in lexical activation was hypothesized to result in longer voice onset time for the word-initial consonant.

It is curious that the existence of a minimal pair neighbor could lead to longer voice onset time for several reasons. For one, as reviewed above, when studies have found a possible association between neighborhood density and duration, segments have typically been found to be shorter in words with more neighbors, not longer. However, it is important to note that for the most part, these studies have examined word-medial vowel duration, and not word-initial consonant duration; the results from Kilanski (2009) (and from Goldinger & Summers, 1989, reviewed in Chapter 2) could be seen as supporting evidence for Baese-Berk and Goldrick's suggestion that increased activation, due to feedback from phonological neighbors, is associated with longer word-initial consonants. Moreover, it should be noted that previous studies have focused on the effects of total neighborhood density, and not on the specific minimal pair relationship examined in Baese-Berk and Goldrick (2009). It could be the case that the nature of the phonological similarity between two minimal pair words can affect their pronunciation, perhaps in a different way from general neighborhood density.

Still, as argued previously, if higher lexical activation due to competition between minimal pair neighbors is responsible for longer VOT, it is not clear how higher lexical activation due to feedback could also be responsible for shorter total word durations, and for shorter vowel durations, as would seem to be the case in the studies reviewed above.

To better understanding the relationship between feedback processes and articulatory duration, the remainder of the chapter is dedicated to a reanalysis of the data reported in Baese-Berk and Goldrick (2009)[1]. The primary research questions to be addressed are as follows:

Research questions

1. Is longer VOT in minimal pair words related to minimal pair status *per se*, or is it a product of greater total neighborhood density?

2. More generally, does the makeup of a word's phonological neighborhood provide explanatory power with respect to the duration of its constituent segments?

[1]I thank Melissa Baese-Berk and Matt Goldrick for graciously sharing their data with me for the purposes of this reanalysis.

3. Can the longer VOT reported for minimal pair words be attributed to an overall expansion in total word duration (either for words with particular minimal pair neighbors, or for words with more neighbors overall), or is it the case that onset consonants behave differently from syllable rimes?

4.2 Methods

The raw data for Experiment 2 come from Experiments 1a and 1b of Baese-Berk and Goldrick (2009). Specific details regarding the original rationale and collection of the data are available in that paper, but aspects of the data collection that are relevant to the present study are summarized here.

Participants

Participants were 25 Northwestern University undergraduates. Data from 22 participants was reported in Baese-Berk and Goldrick (2009), and data from three additional participants (also collected by those authors) is included here. Baese-Berk and Goldrick's Experiment 1 was split into two parts; each part followed the same procedure, but their Experiment 1a used a set of (12) participants to examine VOT in words beginning with /p/, while their Experiment 1b used a different set of (10) participants to examine VOT in words beginning with /t/ and /k/. The data set provided for the current investigation included recordings of 13 participants reading /p/ words, and 12 participants reading /t/ and /k/ words. Data from all 25 of these participants is analyzed here.

Stimuli

Stimuli were 47 pairs of matched words (16 /p/ word pairs, 19 /t/ word pairs, and 13 /k/ word pairs), for a total of 94 words beginning with /p/, /t/, or /k/. Words in each pair were matched for their initial consonant and vowel, as well as their sum segmental probability, mean biphone probability, and phoneme length. Measures of phonotactic probability were drawn from the Phonotactic Probability Calculator (Vitevitch & Luce, 2004). In addition, all words were low frequency, with less than 20 occurrences per million in the CELEX database (Baayen, Piepenbrock, & Van Rijn, 1995). Crucially, for each word pair, one word had a minimal pair neighbor based on the voicing of the first segment (e.g. *pie* has a neighbor *buy*), while the other did not (e.g. *pipe* has no neighbor *bipe*). Note that an important aspect of Baese-Berk and Goldrick's design, which is also important for the current investigation, is that the voiced neighbors were never presented during the experiment, and the target words were also embedded in a list comprised of two-thirds filler words. Any differences arising between the two stimulus types (minimal pair words vs. non-minimal pair words) can therefore be attributed to some aspect of the stimuli themselves, and not to a conscious awareness of the minimal pair manipulation.

/p/-initial stimuli		/t/-initial stimuli		/k/-initial stimuli	
minimal pair	*no minimal pair*	*minimal pair*	*no minimal pair*	*minimal pair*	*no minimal pair*
pie	pipe	tab	tat	cob	cog
peek	peel	tan	tag	cod	cop
palm	pomp	tank	tap	kilt	kin
pore	pork	teal	teat	kit	kiln
punk	pulp	teem	teethe	core	corn
punch	pulse	tick	tiff	cuss	cub
pun	pup	tuck	tuft	cuff	cud
pad	pal	ted	tempt	curl	curb
pall	paunch	tense	tenth	coo	coot
peat	peal	tart	tar	cab	cad
pare	pep	taunt	torch	cape	cake
pill	pinch	tore	taut	code	comb
pig	pith	torque	torn		
poll	poach	tomb	tooth		
pox	posh	tame	taint		
putt	pub	tyke	tithe		
		tile	tights		
		toe	toast		
		tote	toad		

Table 4.1: Stimuli from Experiments 1a and 1b of Baese-Berk and Goldrick (2009), reanalyzed in the present Experiment 2.

The 47 word pairs are provided in Table 4.1, and a summary of the matching statistics is given in Table 4.2. For Baese-Berk and Goldrick, the crucial manipulation was whether words had a minimal pair neighbor based on the voicing of the first consonant, and for this reason, they did not consider the neighborhood characteristics of the target words more generally. However, the current investigation pursues the hypothesis that the general neighborhood characteristics of the words was responsible for the effect observed in the original study, and so the mean number of phonological neighbors based on the Hoosier Mental Lexicon (Nusbaum, Pisoni, & Davis, 1984) has been included in Table 4.2 for reference.

Procedure

Participants read each stimulus word from a computer screen in a self-paced reading task, pressing a button on a button box to advance to the next trial. All words were presented three times each; the entire word list was presented in three different, randomized orders, with a short break in between lists. Audio was recorded at 22.05 kHz using a flash memory recorder and a headset microphone in a sound-attenuated booth.

		minimal pair	no minimal pair	t-test	p value
/p/ stimuli	sum segmental probability	1.104	1.199	$t(15) = 0.31$	$p > 0.75$
	mean biphone probability	0.004	0.003	$t(15) = 1.91$	$p > 0.07$
	mean phoneme length	3.125	3.375	$t(15) = -1.73$	$p > 0.10$
	mean word frequency	4.875	4.125	$t(15) = 0.47$	$p > 0.65$
	mean # neighbors	24.25	14.50	$t(15) = 3.22$	$p < 0.01$
/t/ stimuli	sum segmental probability	1.2	1.2	$t(18) = 0.01$	$p > 0.95$
	mean biphone probability	0.005	0.004	$t(18) = 0.61$	$p > 0.50$
	mean phoneme length	3.2	3.4	$t(18) = -1.42$	$p > 0.15$
	mean word frequency	5.76	6.55	$t(18) = -0.51$	$p > 0.60$
	mean # neighbors	22.32	14.63	$t(18) = 2.67$	$p < 0.05$
/k/ stimuli	sum segmental probability	1.2	1.2	$t(11) = -0.33$	$p > 0.75$
	mean biphone probability	0.004	0.004	$t(11) = -0.042$	$p > 0.65$
	mean phoneme length	3.08	3.25	$t(11) = -1.0$	$p > 0.30$
	mean word frequency	8.3	6.2	$t(11) = 1.0$	$p > 0.30$
	mean # neighbors	25.00	20.83	$t(11) = 2.16$	$p = 0.05$

Table 4.2: Summary statistics for stimuli in Baese-Berk and Goldrick's Experiment 1, re-analyzed in the present Experiment 2. Phonotactic measures come from the Phonotactic Probability Calculator (Vitevitch & Luce, 2004), word frequency from the CELEX database (Baayen et al., 1995), and the number of phonological neighbors from the Hoosier Mental Lexicon (Nusbaum et al., 1984).

Acoustic analysis

Both Baese-Berk and Goldrick (2009) and the present study used the Praat software for acoustic analysis (Boersma & Weenink, 2012). Baese-Berk and Goldrick report that "voice onset time was measured for each token from the stop burst on the waveform to the first zero crossing after the onset of periodicity." The same criteria were used to measure VOT in the present analysis.

In addition to investigating the more general influence of phonological neighbors on VOT, the present analysis is also concerned with the relation between total word duration and VOT; it is possible that the effect of minimal pair neighbors on VOT reported in Baese-Berk and Goldrick (2009) can more accurately be characterized as an effect of total number of neighbors on total word duration. For this reason, word duration measurements were also taken using Praat, and the end of the word was considered to be the offset of any noise in the waveform display related to the articulation of the stimulus. It should be noted that the onset of the word was taken to be the stop release burst, rather than the beginning of the stop closure, since in isolated productions it is not possible to reliably identify the beginning of the stop closure based on acoustic data alone. An example of segmentation boundaries for the word *pad* is provided in Figure 4.1.

Tokens that were mispronounced or disfluent were excluded from all analyses. This resulted in a reduction of the dataset by 1.94%. This is slightly higher than Baese-Berk and Goldrick's figure of 1.5%, which is likely due to the inclusion of three additional subjects in

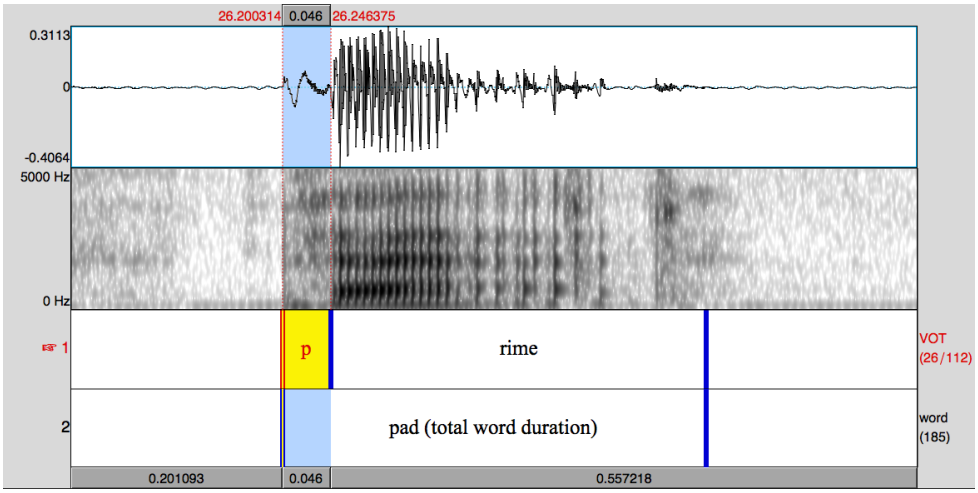


Figure 4.1: Example segmentation boundaries for the word *pad* in Experiment 2.

the present analysis.

4.3 Statistical analysis and results

Because the word duration data did not figure in to the original investigation reported in Baese-Berk and Goldrick (2009), but the current study is specifically interested in the relation between total word duration and voice onset time, it was important for the same experimenter to take both measurements using the same criteria for boundary location, rather than matching Baese-Berk and Goldrick's original VOT measurements with a new set of word duration measurements. Total word duration and VOT were therefore measured by the author, as described in the section on acoustic analysis.

The structure of this section is as follows. The present set of VOT data are first analyzed following the same statistical method outlined in Baese-Berk and Goldrick (2009), and then in a complementary analysis by regression modeling. The purpose of this initial analysis is to ascertain whether the original measurements and the present measurements yield comparable results. Next, the relationship between VOT, total word duration, and phonological neighborhood density is explored in a series of linear mixed effects regressions. To preview the results, no significant effect of minimal pair status is found *per se*. However, when VOT is examined in conjunction with total word duration, the number and type of phonological neighbors in the lexicon are revealed to have significant effects on the durational properties of the words.

		minimal pair	no minimal pair	V	p value
/p/ stimuli	by participants	0.0766	0.0736	3	0.001
	by items	0.0767	0.0736	42	0.19
/t/ stimuli	by participants	0.0879	0.0876	36	0.85
	by items	0.0879	0.0869	79	0.54
/k/ stimuli	by participants	0.0923	0.0937	50	0.42
	by items	0.0923	0.0936	47	0.57
all combined	by participants	0.0854	0.0846	268	0.21
	by items	0.0852	0.0841	472	0.34

Table 4.3: Results of the analysis of VOT measurements using Wilcoxon matched-pairs signed rank tests. Only the by-participants analysis for /p/-initial stimuli reaches statistical significance.

Analysis following Baese-Berk and Goldrick

Baese-Berk and Goldrick first calculated the mean VOT of each participant's three productions of each word, resulting in a set of by-participant, by-word means. This strategy was used to reduce the variability inherent in speech production data. These mean VOT values were then submitted to by-participant and by-items analyses using Wilcoxon matched-pairs sign rank tests, owing to the non-normal distribution of the data.

When this procedure is used to analyze the present set of VOT measurements, a significant difference between words with vs. without minimal pairs is only found in the by-participants analysis for /p/-initial words; all other Wilcoxon test statistics return non-significant values. These results are presented in Table 4.3.

With the exception of the /k/-initial stimuli, the by-participant and by-item means are qualitatively in line with the reported findings; on average, words with minimal pair neighbors are produced with 1 ms longer VOT than words without minimal pair neighbors. While this difference does not reach statistical significance, it is worth noting that the magnitude of the effect reported in Baese-Berk and Goldrick was on the order of 3 ms. Thus while the measurements made for the purposes of the present experiment do not result in a perfect replication of Baese-Berk and Goldrick's findings, it seems likely that on average, the present VOT measurements differed from the original measurements by only a few milliseconds. It is perhaps unsurprising that it would be somewhat difficult to replicate an effect size of 3 ms, especially given that different experimenters made the measurements, likely using slightly different criteria for marking segmental boundaries, and perhaps slightly different settings for displaying the acoustic data on the computer screen. In addition, recall that the present analysis includes data from an additional three subjects, which could also have caused the means to shift slightly.

The following section is comprised of three analyses. First, an attempt is made at replicating the original findings for VOT using a mixed effects regression model. Second, the total word durations are modeled in a similar fashion, this time asking whether the total number of phonological neighbors has an effect on total word duration, as predicted. Finally, VOT is

predictor	β	t	p**MCMC**	**AIC**	**loglik**	χ^2	$p(\chi^2)$
intercept	-2.479	-40.71	0.0001	-534.1	271.0		
onset consonant				-544.4	278.2	14.4	0.0008
consonant = /p/	-0.165	-1.97	0.0106				
consonant = /k/	0.067	3.09	0.0016				
minimal pair = yes	0.011	0.64	0.5034	-542.8	278.4	0.4	0.5176

Table 4.4: Mixed effects regression model predicting log-transformed voice onset time using fixed effects of consonant (/p/ vs. /t/ vs. /k/) and minimal pair status. This model includes random effects for participant and word. The interaction between consonant and minimal pair status was also non-significant.

considered in conjunction with total word duration; once any increases or decreases in total word duration are taken into account, is there any evidence that VOT is disproportionately longer or shorter depending on a word's phonological neighborhood?

Analysis by mixed effects regression

Regression analysis 1: Does minimal pair status affect voice onset time?

The purpose of the first regression analysis is to see whether dealing with the variation in VOT productions by using a mixed effects regression model, rather than by analyzing the grand means, will result in a statistically significant replication of Baese-Berk and Goldrick's original findings. Because the VOT measurements followed a slightly skewed distribution, they were first log-transformed for the purposes of modeling. Random intercepts were included to account for the grouping of observations by participant and by word, and the model included fixed effects in accordance with the experimental manipulations in Baese-Berk and Goldrick (2009): consonant and minimal pair status.

Table 4.4 gives the estimated coefficients for these predictors, as well as their associated t and p values, as estimated by MCMC simulations. The table also includes the same criteria used for model evaluation in the analyses in the rest of the chapter: the Akaike Information Criterion, model log likelihood, and the results of a chi-squared test comparing nested model with one another. This relatively simple model indicates that consonant was a significant predictor, reflecting the fact that VOT was shortest for /p/ and longest for /k/, but minimal pair status was not a significant predictor. The interaction between consonant and minimal pair status was also examined, but it was also non-significant (β = 0.046, t = 1.17, pMCMC = 0.22 for /p/ words with minimal pairs; β = -0.012, t = -0.28, pMCMC = 0.77 for /k/ words with minimal pairs).

The results of the analysis by mixed effects regression model are consistent with the analysis by Wilcoxon matched pairs sign rank tests. Both analyses indicate that minimal pair status did not have a significant effect on VOT, but again, this is not particularly surprising for the reasons discussed above: any one of several factors could explain why the difference in VOT did not reach statistical significance in this reanalysis.

random effect	SD	MCMC median	HPD95lower	HPD95upper
word (intercept)	0.075	0.069	0.058	0.081
speaker (intercept)	0.204	0.149	0.119	0.183
residual	0.214	0.215	0.210	0.220

Table 4.5: Random effects in the simplest model predicting log-transformed VOT, using fixed effects of consonant and minimal pair status.

An additional reason why the previously reported difference in VOT might be somewhat fragile is that the effect may be better characterized as an influence of phonological neighborhood density more generally on total word duration; because the set of words with minimal pair neighbors also had more total phonological neighbors, it is possible that higher neighborhood density was associated with longer word duration, and that this difference was reflected in slightly longer VOT for words with minimal pair neighbors. The next section is devoted to exploring this possibility.

Regression analysis 2: Does number of phonological neighbors affect total word duration?

Making generalizations about the effects of psycholinguistic variables on word durations is complicated by the fact that many factors are known to affect word duration. In addition, these factors are often correlated with one another in complicated ways. The procedure for bringing these factors under statistical control will first be described, taking into consideration all known potential predictors of total word duration and their interrelationships, followed by a description of the fitted model.

Random variation. Due to the experimental design, all observations can be grouped by speaker and by word: some speakers will tend to speak faster than others, and all words will likely have some idiosyncratic properties that are not perfectly captured by the fixed effects of the model. The model therefore includes random intercepts for speaker and word.

Segmental makeup. Because the original experimental design required careful matching of the stimuli in the two groups (minimal pair words vs. non minimal pair words), the 94 words under consideration are all quite similar with respect to their segmental and phonotactic properties: all words begin with a voiceless stop consonant followed by a vowel, and all have relatively restricted values of lexical frequency and phonotactic probability. However, the stimulus words do vary somewhat in their vowels and coda consonants, and these differences in segmental makeup will have an important effect on total word duration. Both the number and identity of the phonological segments that make up each word need to be taken into account before it is possible to determine whether any psycholinguistic factors had a systematic effect on word duration.

In order to control for segmental content, a baseline word duration was calculated for each word based on the average phone durations in the Buckeye Corpus of Conversational Speech, a word- and phone-aligned corpus of approximately 300,000 words produced in

	number of words	*examples*
CV	3	pie, toe
CVC	26	tick, cake
CVCC	4	pox, tights
CVG	40	pad, comb
CVGC	17	pinch, tart
CVGCC	1	tempt
CVGG	3	torn, kiln

Table 4.6: Syllable structures for stimuli in Experiment 2. "C" stands for a voiceless consonant, "G" stands for a voiced consonant, and "V" stands for a vowel.

natural conversational interactions by 40 native English speakers from Columbus, Ohio (Pitt et al., 2007). By-speaker, by-phone mean durations were calculated over the entire corpus, and these mean durations were averaged together to produce a grand mean duration for each phone. Mean phone durations were then added together to produce a baseline word duration for each of the words under investigation, and the baseline word durations were log-transformed to improve the normality of their distribution.

This measure is clearly not ideal. For one, the average phone durations do not take into account the position of the phones in the words of the corpus, and it stands to reason that singleton word-initial phones would be produced with longer durations than word-medial phones in consonant clusters, for example. However, this drawback is mitigated by two additional factors. First, the overall phonological similarity of the words under investigation means that any biases present in the baseline word durations should affect all words roughly equally. Second, a fixed effect of syllable structure (to be described below) accounts for any expansions or contractions in segmental duration that occur for phonological reasons. In light of these mitigating factors, the baseline word duration measure should be sufficient to account for variation in total word duration due to the number of segments that make up each word and their intrinsic durations.

Syllable structure. The same phone can be longer or shorter depending on its phonological environment. The vowel in a consonant-vowel-consonant ("CVC") word will be longer if the final consonant is voiced than if it is voiceless, for example. To account for such expansions and contractions due to regular phonological processes, each word was coded for its syllable structure. Words fell into one of seven categories: CV, CVC, CVCC, CVG, CVGC, CVGCC, and CVGG, where "C" stands for a voiceless consonant, and "G" stands for a voiced consonant. Table 4.6 summarizes the number and types of words in each category.

As with the baseline word durations, the syllable structure predictor is not a perfect measure in and of itself. However, in combination with the baseline word durations and the overall phonological similarity of all of the words under investigation, it provides a good baseline for examining the psycholinguistic factors of interest.

Word frequency. All else being equal, frequent words tend to be produced with shorter

durations than infrequent words (although this effect is only well supported in data from connected speech). As described above, Baese-Berk and Goldrick carefully balanced the stimuli in their two groups based on word frequency counts from the CELEX database (Baayen et al., 1995), and they limited their investigation to words with a frequency of occurrence of 20 times per million or less. While these criteria ensured that only low frequency words were chosen as stimuli, and that the words in the two stimulus groups were roughly balanced for frequency, some recent investigations have suggested that the SUBTLEX word frequency counts may serve as better predictors in psycholinguistic investigations of spoken language, since they are based on spoken (rather than written) language (Brysbaert et al., 2012). For this reason, word frequency based on the SUBTLEX-US database of English movie subtitles was entered as a predictor. Word frequencies were log-transformed in order to improve the normality of their distribution.

Phonotactic probability. Just as frequent words tend to be produced with shorter duration than infrequent words, frequent sounds and sequences of sounds also tend to be shorter, and this effect has been observed in isolated word productions Munson (2001). Two measures of phonotactic probability were investigated as predictors of total word duration: the mean positional segmental probability of the segments comprising each word, and the mean biphone probability of the biphones in the word. Both measures were drawn from the online Phonotactic Probability Calculator (Vitevitch & Luce, 2004). Because the distribution of biphone probabilities was heavily skewed, these values were log-transformed to reduce the influence of the few words with relatively high biphone probabilities. The mean segmental probabilities were approximately normally distributed, so they were not transformed.

Trial number. Having recently produced a given word and/or increasing one's speaking rate in order to finish a lengthy word-reading task could cause words produced later in the experiment to have shorter durations than words produced earlier in the experiment. To account for these potential sources of variation, trial number was entered into the model as a potential predictor of word duration.

Number of phonological neighbors. As summarized in the Background to Experiment 2, the relationship between a word's phonological neighborhood density and its duration in isolated productions is not clear. The hypothesis currently under investigation is that having a greater number of phonological neighbors, which for Baese-Berk and Goldrick's stimuli was found to be correlated with having a minimal pair based on initial consonant voicing, may have been associated with longer word durations overall, and this expansion in total word duration may have been reflected in slightly longer voice onset times.

The Child Mental Lexicon (Storkel & Hoover, 2010) was used to calculate the total number of phonological neighbors, as well as the number of each "type" of phonological neighbor, based on the Hoosier Mental Lexicon (HML; Nusbaum et al., 1984). The Child Mental Lexicon provides an option for generating statistics from either a corpus of child speech or from the HML. It was used to obtain the data on phonological neighbors because it provides a simple, web-based interface that includes the option to subdivide the total number of neighbors into neighbor types; in addition to the total number of neighbors in the HML, the number of "rime neighbors" vs. "onset neighbors" (words sharing the same

	baseline dur	word freq	mean seg prob	mean biphone prob	# neighbors
word freq	0.183				
mean seg prob	-0.032	0.138			
mean biphone prob	0.258	0.134	0.600		
# neighbors	-0.603	0.066	0.241	-0.028	
trial #	-0.025	-0.120	-0.176	-0.178	0.062

Table 4.7: Pairwise Spearman correlations among numerical predictors in the model of total word durations.

rime vs. the same onset as the target word, respectively) was also calculated. All three predictors (total number of neighbors, rime neighbors, and onset neighbors) were ultimately investigated as potential predictors, but the initial focus in this analysis will be on the total number of neighbors.

Minimal pair status. Because Baese-Berk and Goldrick found that minimal pair status affected voice onset time, and because it is hypothesized that the reported difference in VOT may have been correlated with a difference in total word duration, minimal pair status was also examined as a potential predictor. It could be the case that minimal pair status is a better predictor of word duration than the total number of neighbors, and it is also possible that minimal pair status could add predictive power over and above that of neighborhood density more generally. Both possibilities were investigated.

Correlations among variables. Table 4.7 gives the pairwise Spearman correlations for numerical predictors in the model of total word durations. Following Baayen (2008), the condition number of the predictors, κ, was calculated to be 60.95, indicating a high level of multicollinearity. As expected, the correlations between the two measures of phonotactic probability, and between the mean segmental probability and number of neighbors were particularly high. Perhaps less expected was the correlation between baseline word duration and number of neighbors; this is likely explained by the fact that shorter words are made up of fewer segments and also tend to have more neighbors.

In order to minimize multicollinearity, residualized measures of several variables were examined as predictors. Two models were fit. In the first model, baseline word duration was entered as a predictor, and residualized measures of word frequency, mean biphone probability, and number of phonological neighbors were entered as predictors, because these variables exhibited the highest levels of correlation with baseline word duration. Simple linear models predicting word frequency, biphone probability, and phonological neighborhood density from baseline word duration were constructed, and the residuals from each of those models were used as predictors. In this way, the variance associated with baseline word duration was effectively factored out of each of those predictors (but still captured by the baseline word duration predictor itself). The non-residualized variables syllable structure, mean segmental probability, trial number, and minimal pair status were also entered into the model.

Predictors were added to the model in the order in which they were expected to contribute

predictor	β	t	p**MCMC**	**AIC**	**loglik**	χ^2	$p(\chi^2)$
intercept	-0.051	-0.33	0.6614	-4330.6	2169.3		
baseline duration	0.499	5.31	0.0001	-4350.5	2180.2	21.9	0.0001
syllable structure				-4379.4	2200.7	40.9	0.0001
structure = CVC	-0.159	-3.15	0.0004				
structure = CVCC	-0.092	-1.16	0.1962				
structure = CVG	-0.043	-0.89	0.3106				
structure = CVGC	-0.171	-2.63	0.0042				
structure = CVGCC	-0.246	-2.32	0.0092				
structure = CVGG	-0.051	-0.72	0.4360				
residualized neighborhood density	-0.002	-1.87	0.0326	-4381.1	2202.6	3.8	0.0528

Table 4.8: First fitted model examining predictors of (log-transformed) total word duration in Experiment 2. This model includes random intercepts for participant and word. No other predictors approached statistical significance.

random effect	**SD**	**MCMC median**	**HPD95lower**	**HPD95upper**
word (intercept)	0.070	0.059	0.051	0.068
speaker (intercept)	0.139	0.091	0.073	0.109
residual	0.119	0.121	0.118	0.124

Table 4.9: Random effects in the first fitted model predicting (log-transformed) total word duration in Experiment 2.

predictive power: baseline word duration, syllable structure, trial number, word frequency, biphone probability, segmental probability, number of phonological neighbors, and minimal pair status. Only predictors whose estimated coefficients were significantly different from zero, and which resulted in a significant improvement in model fit as judged by their effect on the Akaike Information Criterion (AIC) and the model's log-likelihood were retained. Non-significant predictors were dropped before additional predictors were added.

The first fitted model is shown in Table 4.8. The only predictors that accounted for a significant proportion of variance were baseline word duration, syllable structure, and the residualized measure of phonological neighborhood density, all of which had significant pMCMC values at the $\alpha = 0.05$ level. No other predictors even approached the level of statistical significance. The multicollinearity of the predictors in the fitted model was calculated to be $\kappa = 12.39$, which is considered unlikely to be problematic (Baayen, 2008). A chi-squared test indicated that the model including the residualized measure of neighborhood density was a better fit to the data than a model without neighborhood density ($\chi^2 = 3.75$, $p = 0.05$). Figure 5.1 plots the partial effects of the significant predictors on log-transformed total word duration.

As a complementary analysis, a second model was fit to the word duration data using the raw data on phonological neighbors, but residualized measures of baseline word duration and mean segmental probability. The results were qualitatively identical to the first model; only

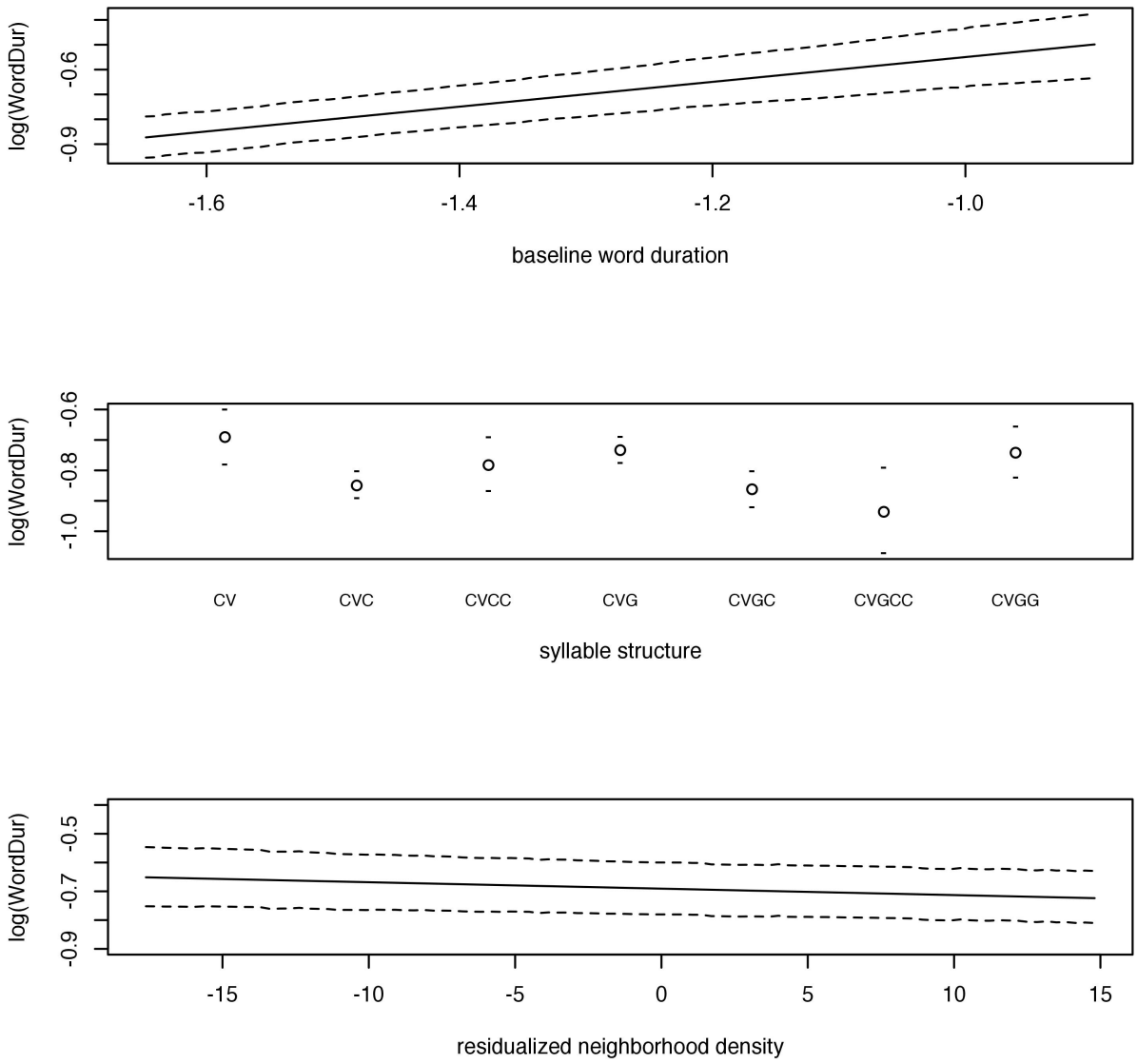


Figure 4.2: Partial effects plot for the first fitted model of total word duration in Experiment 2.

predictor	β	t	p**MCMC**	**AIC**	**loglik**	χ^2	$p(\chi^2)$
intercept	-0.113	-0.78	0.3372	-4330.6	2169.3		
residualized baseline duration	0.414	4.78	0.0001	-4336.4	2173.2	7.8	0.0052
syllable structure				-4357.2	2189.7	33.01	0.0001
structure = CVC	-0.159	-3.15	0.0004				
structure = CVCC	-0.092	-1.16	0.1944				
structure = CVG	-0.043	-0.89	0.3084				
structure = CVGC	-0.171	-2.64	0.0030				
structure = CVGCC	-0.246	-2.32	0.0082				
structure = CVGG	-0.051	-0.72	0.4260				
neighborhood density	-0.024	-5.21	0.0001	-4381.1	2202.6	25.7	0.0001

Table 4.10: Second fitted model examining predictors of (log-transformed) total word duration in Experiment 2. This model includes random intercepts for participant and word. No other predictors approached statistical significance.

random effect	**SD**	**MCMC median**	**HPD95lower**	**HPD95upper**
word (intercept)	0.070	0.059	0.051	0.068
speaker (intercept)	0.139	0.090	0.076	0.107
residual	0.119	0.121	0.118	0.124

Table 4.11: Random effects in the second fitted model predicting log-transformed total word duration in Experiment 2.

residualized baseline word duration, syllable structure, and total number of neighbors were significant predictors of word duration. This time, however, the total number of neighbors was found to be highly statistically significant, with a chi-squared test returning a p value less than 0.0001. The estimated coefficients and their associated statistics are given in Table 4.10.

As an index of goodness-of-fit, the squared correlation between fitted and observed values for both models of total word duration was approximately 0.675. This is a rather high proportion of variance accounted for, which can be explained by the fact that the combination of baseline word duration and syllable structure were quite effective as control variables. In fact, a model with these two predictors alone resulted in a squared correlation of 0.674624; the unique proportion of variance explained by the neighborhood density predictor was only 0.0000025.

To better understand the magnitude of the effect of neighborhood density on total word duration, the observed means can be compared for pairs of words that are predicted to have roughly the same duration according to their segmental content and syllable structure, but differ greatly in their neighborhood density. One such pair is *pill* and *teethe*, whose baseline durations are 224.7 and 226.7 ms, but which have 36 and 11 phonological neighbors, respectively. The observed mean duration for *pill* was 437.2 ms, and the mean for *teethe* was 495.6 ms, a difference of 58.4 ms. To take a second example, *tank* and *torch* had 16 vs.

8 neighbors, baseline durations of 363.8 vs. 366.0 ms, and observed mean durations of 466.0 and 546.6 ms, for a difference of 80.6 ms. Thus while the proportion of variance explained by the residualized neighborhood density predictor is very small, the magnitude of the effect appears to be on the order of milliseconds or tens of milliseconds.

Perhaps more important than the finding that neighborhood density had a reliable effect on total word duration is the fact that this effect was in the opposite direction of what was predicted; all else being equal, words with more neighbors were produced with shorter total durations. This is in line with Gahl et al. (2012)'s findings for conversational speech, but it does nothing to explain why words with minimal pair neighbors were produced with longer VOTs in Baese-Berk and Goldrick (2009). The effect of minimal pair status on VOT cannot have been an artifact of expansion in total word duration, as was hypothesized.

If words with more phonological neighbors have longer VOT but shorter total duration, this suggests that a word's neighbors can have different effects on different aspects of its duration. It is well known that the vocabulary of English (as well as that of other languages) is skewed such that words tend to have more "rime neighbors" than "onset neighbors" (De Cara & Goswami, 2002; Gupta & Dell, 1999). That is, words tend to have more neighbors that contrast in the onset and share a rime (e.g. *cat, rat, mat, sat,* etc.) than words that contrast in the rime and share an onset (e.g. *cat, kit, cot, cab,* etc.). Given this skewing in the type of neighbors in the English lexicon, it is possible that the effect of "minimal pair status" on VOT, as well as the effect of "total number of neighbors" on total word duration can be accounted for with one parsimonious explanation: competition between segments for the onset "slot" may lead to longer VOT, while agreement or support for segments in the rime "slot" may lead to a shorter rime. This predicts that words with many rime neighbors should have longer VOT but shorter rimes, while words with many onset neighbors should have shorter VOT (and perhaps longer rimes).

In light of this hypothesis, the total number of onset versus rime neighbors in the Hoosier Mental Lexicon was calculated for the 94 words under investigation using the Child Mental Lexicon online calculator (Storkel & Hoover, 2010). On average, minimal pair words had more total neighbors than non-minimal pair words (23.6 vs. 16.2 total neighbors, respectively), as well as more onset neighbors and more rime neighbors (9.6 vs. 7.7 onset neighbors, and 6.9 vs. 4.2 rime neighbors, respectively). The fact that words in the current stimulus set had on average more onset than rime neighbors suggests that the current stimuli are perhaps not entirely representative of monosyllabic words in English, but this is not surprising given how carefully they were selected to conform to various matching criteria. The strong correlation between the different neighborhood metrics indicates that care should be taken in attributing any effects on duration to one type of neighbor over another. However, the fact that the numbers of rime versus onset neighbors are not perfectly correlated indicates that it may be possible to determine whether a given type of neighbor is more predictive of VOT versus rime duration. This is the subject of the section that follows.

Regression analysis 3: Are VOT and/or rime duration disproportionately longer or shorter depending on a word's neighborhood characteristics?

Baese-Berk and Goldrick (2009) found that words with minimal pair neighbors based on the voicing of the first segment had longer VOT than words without such minimal pairs, and the current investigation has thus far determined that 1) Baese-Berk and Goldrick's minimal pair words had more phonological neighbors overall, and 2) words with more total neighbors were produced with shorter total durations, all else being equal. To explore the possibility that different types of phonological neighbors can affect word duration differently, the final set of regression analyses pursues the hypothesis that onset versus rime neighbors have different effects on VOT versus rime duration.

In addition to the variables entered into the model for total word duration, several additional predictors were examined in the models of VOT and rime duration. These variables are summarized below, followed by a description of the fitted models.

Place of articulation. As described above, VOT is known to vary according to place of articulation, with labial stops typically having the shortest VOT, followed by alveolar stops, and then velar stops. Place of articulation (/p/ vs. /t/ vs. /k/) was therefore entered into the model of VOT.

Vowel height. Stops produced before high vowels tend to have longer VOT than stops before low vowels. Vowel height (high vs. low) was therefore entered as a control parameter. The vowels /i ɪ e o u/ were considered high, and all other vowels were considered low.

Biphone probability of the first biphone ("B1"). In addition to the word's mean positional segment probability and mean biphone probability, the biphone probability of only the first biphone was also examined as a predictor. It could be the case that the probability of B1 will be a better predictor of VOT than measures of phonotactic probability intended to characterize the whole word.

Observed rime duration. All else being equal, VOT is likely to increase or decrease along with the rest of the word. Since the present question is whether VOT is disproportionately longer or shorter than would otherwise be expected, the observed rime duration was used as a control parameter in the model of VOT. Note that this predictor in conjunction with the place of articulation predictor obviates the need for a measure of baseline word duration. Observed rime durations followed a skewed distribution, and were therefore log-transformed in order to improved their normality.

Baseline rime duration. To provide a more sensible measure of baseline duration, the baseline word durations used in the model of total word duration were converted to baseline rime durations; the average duration of the initial segments in the Buckeye Corpus was subtracted from the baseline word durations, yielding a prediction for each word's baseline rime duration. Baseline rime durations were also log-transformed to yield a more normal distribution. The same syllable structure predictor as was used to model the total word durations was also entered into the model of rime durations, since all words began with a CV sequence; that is, the syllable structure predictor is already a rime structure predictor, since the stimulus words only differ in their rimes.

Modeling procedure. The same modeling procedure was followed as before. Only predictors whose coefficients differed significantly from zero, and which improved model fit as determined by a chi-squared test comparing the log-likelihood of a model with versus without the predictor were retained. Random intercepts for participant and word were first entered (and determined to significantly improve model fit). Fixed effects were then entered in the order in which they were expected to improve model predictions. For the model of VOT, this order was as follows: observed rime duration, consonant, vowel height, trial number, word frequency, mean biphone probability, mean segmental probability, B1 probability, neighborhood density, and minimal pair status. For the model of rime duration, the order of entry was: baseline rime duration, syllable structure, trial number, word frequency, mean biphone probability, mean segmental probability, neighborhood density, and minimal pair status.

Table 4.14 gives the pairwise Spearman correlations for the numerical predictors in the two models. Measures of phonotactic probability and neighborhood density were particularly highly correlated with one another and with the baseline duration measures, and a strategy of residualization was again adopted in order to minimize multicollinearity. The analysis and results of the VOT model will be presented first, followed by the analysis of rime duration.

Regression analysis 3: VOT results

A control model was first fit to the VOT data, including all non-neighborhood related variables that accounted for a significant proportion of variance. Two strategies were then adopted to determine which metrics of neighborhood density were the best predictors of VOT. First, a residualized measure of rime density was added to the control model; the residuals from a simple linear regression predicting rime density from log-transformed rime duration were entered as a predictor, effectively asking whether rime density can account for any additional variance once the predictive power associated with rime duration is removed. This model is summarized in Table 4.12. A residualized measure of onset density (again removing variation associated with rime duration) was then added on top of the rime density predictor. Onset density did not contribute any additional predictive power, but it is shown in Table 4.12 for reference.

When residualized onset density was entered *instead* of rime density, it did account for significant variance, but its coefficient was less likely to be different from zero than the rime density coefficient ($\beta = 0.003$, $t = 1.99$, pMCMC $= 0.04$, for onset density), and the difference in log likelihoods for a model with versus without onset density is much less than the difference between models with versus without rime density (log likelihood = 312.7, $\chi^2 = 4.1$, $p = 0.04$ for onset density; log likelihood = 321.6, $\chi^2 = 21.9$, $p < 0.0001$ for rime density). This combined with the fact that onset density and rime density are highly correlated, and the fact that when the two are entered into the same model, onset density provides no explanatory power, suggests that rime density is the more effective predictor of VOT.

predictor	β	t	pMCMC	AIC	loglik	χ^2	$p(\chi^2)$
intercept	-2.307	-37.27	0.0001	-534.1	271.0		
rime duration	0.154	6.11	0.0001	-569.5	289.8	37.4	0.0001
onset consonant				-581.6	297.8	16.1	0.0003
consonant = /p/	-0.153	-1.98	0.0120				
consonant = /k/	0.0581	3.26	0.0012				
vowel height = hi	-0.054	-3.81	0.0001	-595.7	305.8	16.0	0.0001
trial number	-0.0003	-2.87	0.0044	-597.4	307.7	3.7	0.0550
word frequency	-0.0112	-3.03	0.0024	-601.3	310.7	6.0	0.0145
residualized rime density	0.008	4.84	0.0001	-621.2	321.6	21.89	0.0001
residualized onset density	0.0002	-0.14	0.9020	-619.2	321.6	0.02	0.8918

Table 4.12: Significant predictors of VOT in Experiment 2. The non-significant effect of onset density is included for reference. The residualized measures of rime density and onset density have had variation associated with the rime duration removed. This model includes random intercepts for participant and word.

random effect	SD	MCMC median	HPD95lower	HPD95upper
word (intercept)	0.054	0.051	0.041	0.062
speaker (intercept)	0.189	0.142	0.115	0.177
residual	0.213	0.214	0.209	0.219

Table 4.13: Random effects in the second fitted model predicting VOT in Experiment 2.

The fitted model predicting log-transformed VOT is presented in Table 4.12. Log-transformed rime duration, consonant, vowel height, trial number, word frequency, and residualized rime density (with variation predicted by rime duration partialled out) were all significant predictors of VOT. All effects were in the predicted direction: longer rime duration was associated with longer VOT; /p/ had the shortest VOT, and /k/ the longest; low vowels were associated with shorter VOT than high vowels; VOT decreased over the course of the experiment; high frequency words were produced with shorter VOT; and importantly, residualized rime density was associated with longer VOT. Words with many neighbors overlapping in the rime were produced with significantly longer voice onset times, and this effect was highly significant.

The second strategy to reduce multicollinearity was to enter rime density as a predictor, and to residualize rime duration and onset density, removing the variation associated with rime density from these predictors. This model is qualitatively identical to the one presented in Table 4.12; residualized rime duration and rime density were highly significant predictors ($\beta = 0.175$, $t = 6.93$, pMCMC < 0.0001 for residualized rime duration; $\beta = 0.007$, $t = 4.10$, pMCMC < 0.0001 for rime density), while residualized onset density did not even approach significance ($\beta =$ -0.0002, $t =$ -0.14, pMCMC $= 0.8926$).

In contrast, when rime duration was residualized from onset density, and onset density was entered as a predictor, its coefficient was not significantly different from zero ($\beta = 0.001$, t

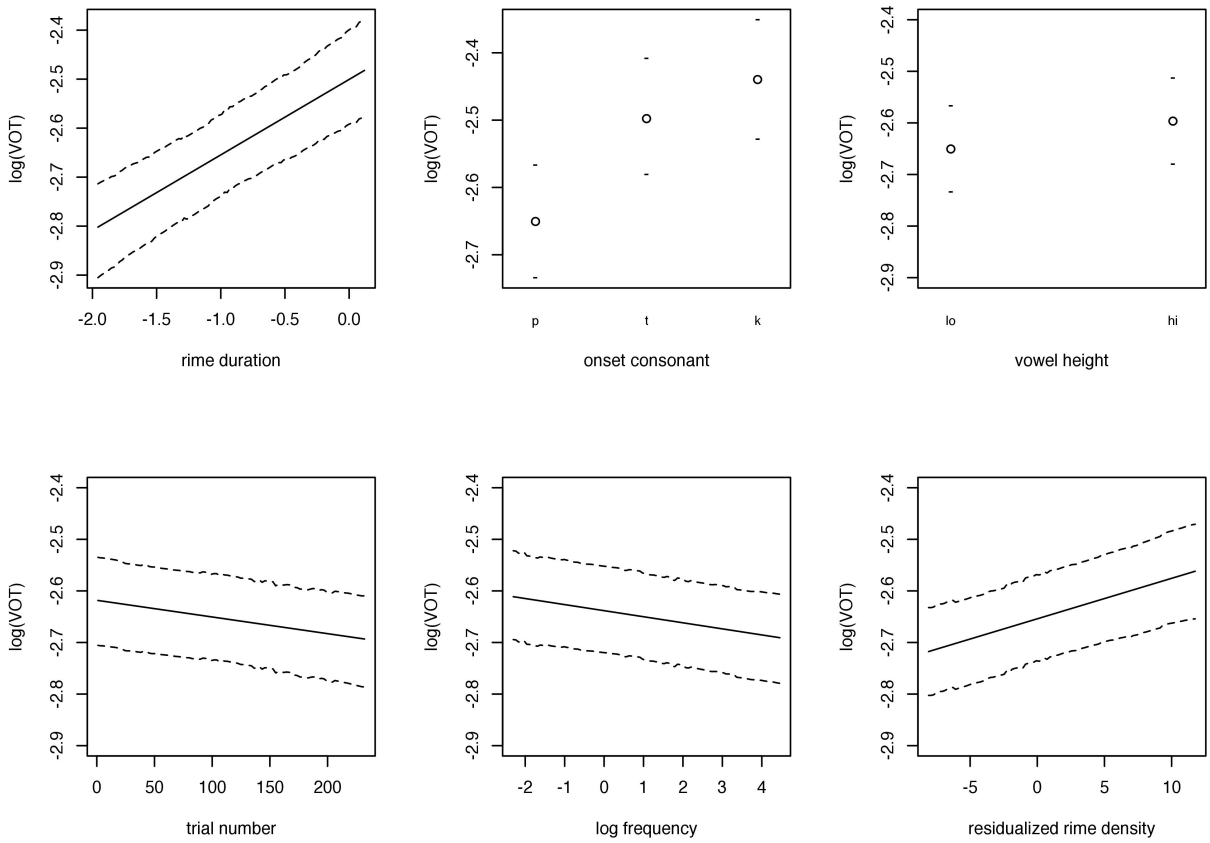


Figure 4.3: Partial effects plot for the second fitted model of VOT in Experiment 2.

= 0.80, pMCMC = 0.39), but when onset density and a residualized measure of rime density (removing variation due to the correlation between rime density and onset density) were added, the latter was significant ($\beta = 0.008$, $t = 4.29$, pMCMC < 0.0001). Both strategies of residualization therefore suggest that rime density is a better predictor of VOT than onset density.

Figure 5.3 provides a partial effects plot depicting log-transformed VOT as a function of the significant predictors in Table 4.12. The total amount of variance predicted by the fitted VOT model (after outliers have been removed) was 57.4%. To better understand the effect size of residualized rime density, we can compare the mean VOT for words with greater than the median number of rime neighbors to the mean VOT for words with less than the median number of rime neighbors; these means are 87.5 ms and 83.2 ms, respectively – roughly the same effect size as Baese-Berk and Goldrick found using their minimal pair metric.

	RimeDur	BaseRime	Trial	WordFreq	BiProb	SegProb	B1Prob	TNbors	ONbors	RNbors	CCNbors
BaseRime	0.268										
Trial	-0.011	0.010									
WordFreq	0.002	0.159	-0.120								
BiProb	0.129	0.238	-0.178	0.134							
SegProb	0.017	-0.135	-0.176	0.138	0.600						
B1Prob	0.121	0.152	-0.154	0.195	0.706	0.357					
T-Nbors	-0.212	-0.620	0.062	0.066	-0.028	0.241	-0.048				
O-Nbors	-0.233	-0.570	0.028	0.075	-0.149	0.192	-0.033	0.846			
R-Nbors	-0.138	-0.522	0.101	0.032	0.056	0.204	-0.083	0.864	0.524		
CC-Nbors	-0.238	-0.510	0.059	-0.037	-0.246	0.165	-0.258	0.765	0.899	0.508	
CV-Nbors	-0.123	-0.475	0.007	0.208	0.069	0.209	0.317	0.674	0.765	0.413	0.435

Table 4.14: Pairwise Spearman correlations among numerical predictors in the models of VOT and rime duration. RimeDur = log-transformed rime duration, BaseRime = log-transformed baseline rime duration, Trial = trial number, WordFreq = log-transformed word frequency from the SUBTLEX-US corpus, BiProb = log-transformed mean biphone probability, SegProb = mean positional segment probability, B1Prob = log-transformed probability of the first biphone only, T-Nbors = total number of neighbors, O-Nbors = number of onset neighbors, R-NBors = rime neighbors, CC-Nbors = number of neighbors different by a vowel change, CV-Nbors = number of neighbors differing in their last consonant.

Regression analysis 3: rime duration results

The fitted control model for the rime duration data only included fixed effects for the baseline rime duration and syllable structure; while all trends were in the predicted direction (e.g. higher word frequency and higher trial number were associated with shorter rimes, on average), no other predictors returned significant coefficients or contributed significant predictive power to the model, as determined by chi-squared tests of model log likelihood.

The same strategy for reducing multicollinearity was then followed as in the analysis of VOT: residualized and non-residualized measures of different neighborhood metrics were examined, on their own and in combination with one another. In addition to examining rime neighbors and onset neighbors, three additional neighborhood metrics were also examined: total number of neighbors, number of neighbors differing only in their vowel ("CC neighbors"), and number of neighbors differing only in their final consonant ("CV neighbors").

Each neighborhood density metric was residualized from the baseline rime duration, and the log likelihoods for models with versus without each metric were compared. Greater total neighborhood density, onset density, and CC density all added significant predictive power to the control model ($\chi^2 = 5.4$, $p = 0.02$ for total density; $\chi^2 = 5.4$, $p = 0.02$ for onset density; and $\chi^2 = 6.8$, $p = 0.009$ for CC density), and all were associated with shorter rime durations. Rime density and CV density did not significantly increase the model log likelihood relative to the control model.

To better understand the variation associated with each type of neighbor, the total number of neighbors plus a residualized measure of baseline rime duration were first entered as predictors. Residualized measures of each neighborhood metric were then examined in turn to see whether any type of neighbor could contribute significant predictive power over and above the effect of the total number of neighbors. The only metric that approached significance following this procedure was CC density; a greater total number of neighbors was significantly associated with shorter rime durations (β = -0.028, t = -5.87, pMCMC < 0.0001), and a greater number of CC neighbors was marginally associated with even shorter rime durations (β = -0.006, t = -1.43, $p = 0.09$). The CC density predictor did not significantly improve the log likelihood, however ($\chi^2 = 2.2$, $p = 0.14$). The model using total neighborhood density as a predictor is shown in Table 4.15, with the estimated coefficients for all other (non-significant) neighborhood metrics included at the bottom.

Overall, the analyses of rime duration suggest that words with many neighbors, and perhaps especially many neighbors contrasting in their vowel, tend to be produced with shorter rime durations.

Summary of results

The VOT and word duration data presented here were originally collected and analyzed in Baese-Berk and Goldrick (2009), but were re-measured and re-analyzed for the purposes of the present investigation. When only the voice onset time was examined, no significant effects of minimal pair status were found. It was hypothesized that the effect of minimal

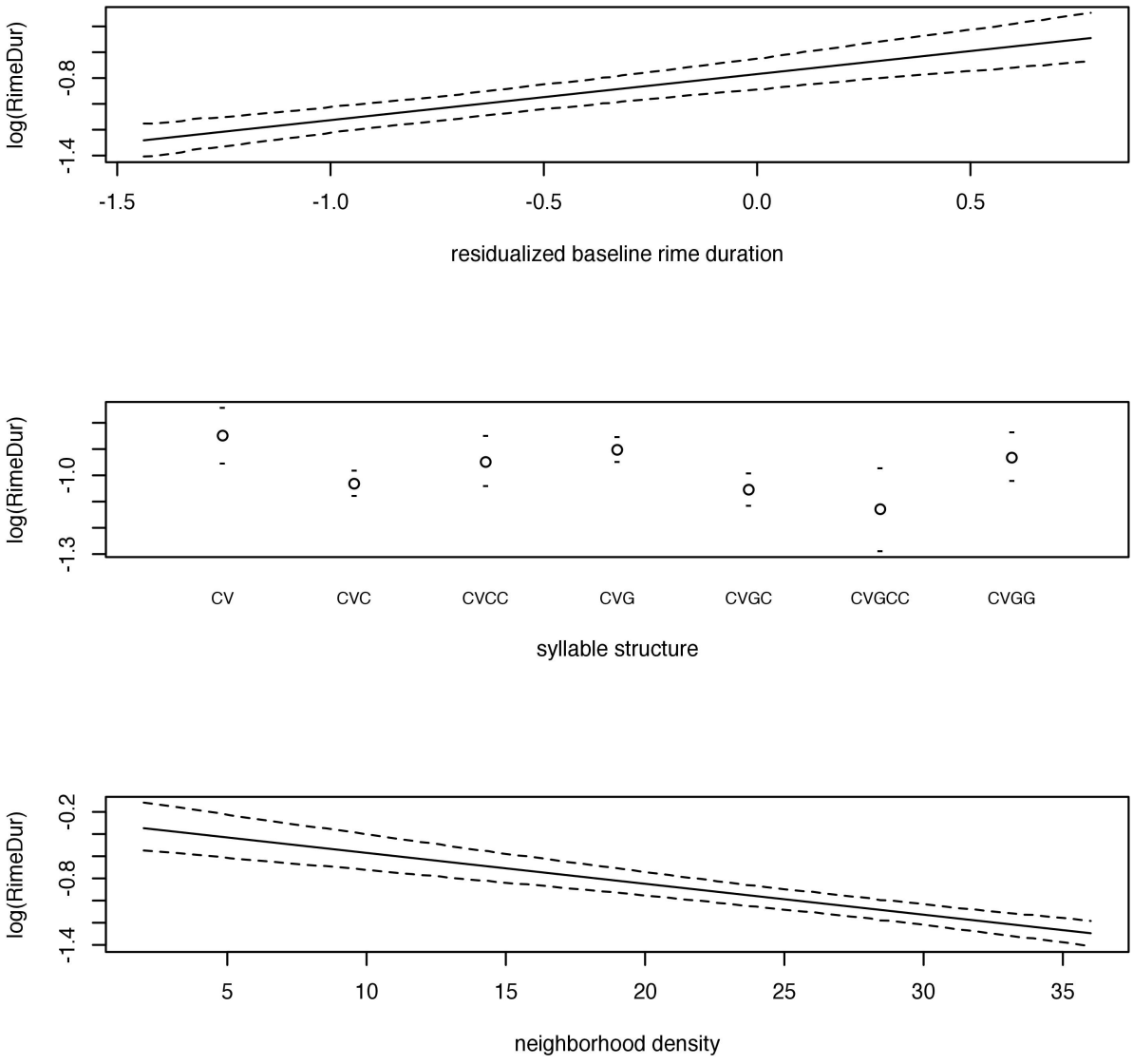


Figure 4.4: Partial effects plot for the fitted model of rime duration in Experiment 2.

predictor	β	t	p**MCMC**	**AIC**	**loglik**	χ^2	$p(\chi^2)$
intercept	-0.212	-1.39	0.0998	-3605.7	1806.8		
residualized baseline rime duration	0.357	5.26	0.0001	-3613.2	1811.6	9.6	0.0020
syllable structure				-3633.0	1827.5	31.7	0.0001
syll structure = CVC	-0.183	-3.12	0.0006				
syll structure = CVCC	-0.101	-1.11	0.2056				
syll structure = CVG	-0.054	-0.97	0.2594				
syll structure = CVGC	-0.206	-2.72	0.0022				
syll structure = CVGCC	-0.280	-2.31	0.0078				
syll structure = CVGG	-0.084	-1.03	0.2390				
neighborhood density	-0.028	-5.87	0.0001	-3662.7	1843.3	31.7	0.0001
residualized CC density	-0.006	-1.43	0.0910	-3662.9	1844.4	2.2	0.1359
residualized onset density	-0.004	-0.93	0.2834	-3661.6	1843.8	0.9	0.3314
residualized rime density	0.005	1.03	0.2292	-3661.8	1343.9	1.2	0.2809
residualized CV density	0.003	0.60	0.4794	-3661.0	1843.5	0.4	0.5318

Table 4.15: Significant predictors of rime duration in Experiment 2 (top portion), with coefficients for non-significant predictors included for reference (bottom portion). The residualized measure of baseline rime duration has had variation associated with neighborhood density removed. This model includes random intercepts for participant and word.

random effect	**SD**	**MCMC median**	**HPD95lower**	**HPD95upper**
word (intercept)	0.081	0.068	0.059	0.078
speaker (intercept)	0.135	0.095	0.077	0.117
residual	0.133	0.134	0.131	0.137

Table 4.16: Random effects in the fitted model predicting rime duration in Experiment 2.

pair status on VOT originally reported in Baese-Berk and Goldrick (2009) may have been a by-product of overall expansion in word duration caused by the association between minimal pair words and overall neighborhood density for the stimuli in that study. The prediction was that words with more phonological neighbors would be produced with longer durations overall.

However, when total word durations were examined as a function of overall neighborhood density, the exact opposite pattern was found: words with more phonological neighbors had significantly shorter total word durations than would otherwise be expected, once variation associated with their segmental content and syllable structure was brought under statistical control. If words with more neighbors have longer VOT but shorter total word durations, this suggests that a target word's phonological neighbors have different effects on its onset versus its rime.

The next set of analyses explored this possibility by examining the effect of different types of phonological neighbors on the duration of different portions of the words produced. In the analysis of VOT, the observed rime duration of each production was used as a statistical control, in order to determine whether VOT was disproportionately longer or shorter for

words with different types of neighbors. Longer rime durations were associated with longer VOT, but once this association was taken into account, VOT was disproportionately longer for words with more rime neighbors. The number of onset neighbors, however, was not reliably associated with longer (or shorter) VOT.

The analysis of rime duration was not as straightforward. The relationship between five different neighborhood metrics and the observed rime duration was examined: the total number of phonological neighbors, rime neighbors, onset neighbors, CC neighbors, and CV neighbors were each entered as predictors of rime duration. The strong correlations between the different neighborhood metrics made it difficult to determine which was the best predictor of rime duration. When variation associated with the baseline rime duration was removed from each of the metrics, and models with versus without each metric were compared, three neighborhood metrics contributed significant predictive power to the baseline model: the total number of phonological neighbors, number of onset neighbors, and number of CC neighbors were all significantly associated with shorter rime duration.

A second model was fit to the rime duration data, using the total number of neighbors as a predictor, plus residualized measures of each of the neighborhood metrics, in turn. Comparisons between models with only the total neighborhood density versus models with total neighborhood density plus each of the metrics showed that no metric was able to contribute significant predictive power beyond the effect of total number of neighbors. However, the number of "CC" neighbors (neighbors differing from the target word in their vowel only) did reach marginal significance, indicating that on average, words with more neighbors contrasting in their vowel had shorter rime durations.

4.4 Discussion

The results of the present investigation were not as expected. It was originally hypothesized that Baese-Berk and Goldrick's effect of minimal pair status on VOT could be attributed to an effect of total number of neighbors on total word duration. However, a series of regression models indicated that the best characterization of the data was that words with many neighbors contrasting in the initial segment were produced with longer VOT, while words with many total neighbors, and/or more neighbors contrasting in their vowel, were produced with shorter rimes.

In one sense, this re-analysis constitutes a rather more complicated replication of Baese-Berk and Goldrick's original analysis. Both analyses indicated that when a word has neighbors overlapping in the rime, it will be produced with a (very slightly) longer initial consonant. The original interpretation given to the data was that such lengthening could be attributed to competition between minimal pair neighbors, requiring a boost in lexical activation for the target words to be selected and encoded for production. However, the results of this study, in conjunction with previous studies looking at effects of "positional" neighborhood density, suggest a slightly different interpretation.

Vitevitch et al. (2004) examined the effect of onset density on repetition latency and speech errors. Words with so-called "dense onsets" – that is, many neighbors differing in their initial consonant – were produced with longer latencies and more speech errors. In other words, all else being equal, having many neighbors that "disagree" in their initial segment seems to interfere with the process of phonological encoding.

Vitevitch et al. (2004) rightly point out that it would be difficult to explain their findings without invoking feedback processes. The very concept of onset density (and, for that matter, neighborhood density in general) assumes that it is possible to define a lexical neighborhood based on phonological overlap with a given target word. The fact that studies controlling for phonotactic probability, either experimentally or statistically, have consistently found an effect of phonologically defined lexical neighborhood on speech production indicates that feedback between the target word and its phonologically related neighbors is important for understanding the dynamics of lexical selection and phonological encoding.

Dell's (1986, 1988) spreading activation model of speech production, along with Sevald and Dell (1994)'s data on CVC syllable production, provide a framework for understanding how positionally defined neighbors can affect the relative ease or difficulty of phonological encoding. In Sevald and Dell (1994), participants produced pairs of CVC syllables (some words, some nonwords) as many times as possible in eight seconds. When the syllable pairs were identical or overlapped in their final segments (e.g. *pick* and *tick*), participants produced relatively more syllables in the allotted time, but when pairs disagreed in their final sounds (e.g. *pick* and *pin*), participants produced fewer syllables. Sevald and Dell argued that the production of syllable pairs with discrepant final segments was slower because phonological encoding proceeds from left to right, a notion they referred to as the *sequential cuing effect*; by the time participants were ready to encode the final segment of each syllable, increased competition for the final slot made selecting the correct phoneme more difficult, thereby slowing articulation.

The sequential cuing effect underscores the notion that phonological encoding takes place in time. At a given moment during articulation, speakers are simultaneously executing an already prepared articulatory plan *and* encoding upcoming material for production in the very near future. While they do not explicitly state it, Sevald and Dell's findings imply that the articulatory duration of a given phonological segment can be directly impacted by the difficulty of encoding. The present findings suggest that this difficulty can manifest itself in two distinct ways.

First, two similar patterns in the data suggest that the ease of encoding a given segment can directly impact that segment's duration. Words with many neighbors overlapping in their rimes were produced with longer VOT, while words with many total neighbors were produced with shorter rimes. These findings suggest that competition between active segments for a given position within the word can slow down articulation, or alternately, that reinforcing activation from phonologically overlapping neighbors can facilitate articulation. If positional competition is associated with longer articulatory duration, then a reasonable locus for the effect would be the ease of phonological encoding.

The second way in which difficulty in encoding may affect duration is illustrated by

the finding that all else being equal, words with many CC neighbors were produced with marginally shorter vowels. When total neighborhood density was taken into account, the number of neighbors contrasting with the target in their vowel was *negatively* correlated with the duration of the vowel. Because the pattern is only marginally significant, it is perhaps prudent not to over interpret this association. However, a possible explanation could be that positional competition between segments is also related to the latency with which a given segment is initiated. In this case, the ease of encoding the final consonant for words with many CC neighbors could cause the final consonant to be initiated relatively sooner, effectively shortening the duration of the vowel.

One appealing aspect of this interpretation is that it may potentially be understood in terms of motor control processes. Rosenbaum (Rosenbaum, 1980, cited in Roelofs, 1999 and reviewed in Chapter 2) found that participants were slower to initiate arm movements when the number of possible movement trajectories was increased. On analogy with the present results, relative confusion over which segment to initiate next could increase the latency of that segment's production. In this case, then, having many active neighbors with contrasting final consonants could have led the final consonant of the target word to be initiated more slowly, causing the duration of the preceding vowel to lengthen.

A second appealing aspect is that this interpretation potentially assists in tying the present results together with previous findings. In Vitevitch et al. (2004), words with dense onsets were initiated more slowly, and in Baese-Berk and Goldrick's data, the same sort of words were produced with longer VOT. Taken together, this suggests that the longer VOT in the present data is indicative of relative planning difficulty. At first, this would seem difficult to reconcile with Sevald and Dell (1994)'s findings, that participants produced *more* word pairs of the *pick–tick* type, and *fewer* pairs of the *pick–pin* type. However, if it is the case that speakers' articulation of difficult to encode segments is both executed and initiated more slowly, then all three sets of findings can be reconciled. Since Sevald and Dell do not provide detailed durational data on participants' productions – only the number of word pairs they were able to produce in eight seconds – this account makes the prediction that one of the reasons participants produced fewer repetitions for *pick–pin* pairs is that the vowels in such pairs had longer durations, due to the final consonant segments being more difficult to initiate.

As a side note, the idea that easy to encode segments may be initiated more quickly may be consistent with Scarborough's findings regarding nasal coarticulation. If it is the case that greater neighborhood density generally facilitates phonological planning, and that relative ease of phonological planning is generally related to a faster onset of articulatory gestures, then words in dense neighborhoods might be expected to be produced with greater segmental coarticulation, since the onset of gestures would generally be earlier than for words in sparse neighborhoods.

Finally, it bears mentioning that the findings reported here are unlikely to be the result of online, listener-oriented adaptation for several reasons. First and foremost, the observed differences in rime duration and in total word duration would, if anything, reduce the perceptibility of the words in question. Words with more phonological neighbors are likely to

be more confusable than words with fewer neighbors, and shortening them would seem to do little to improve listeners' chances of successful recognition. Second, while the effect of rime density on VOT is in the proper direction for improving the perceptibility of more easily confusable words, the magnitude of the effect is so small that it is extremely unlikely to be perceptible by listeners, and as such is unlikely to be a listener-oriented adaption. And third, recall that the participants in the present experiment were unaware of the minimal pair manipulation, and were not engaged in any meaningful interactional context.

To be clear, this is certainly not to say that speakers never engage in listener-oriented adaptation, or that speech perception is not relevant to processes of speech production. On the contrary. With respect to the first point, there is an entire literature documenting the characteristics and causes of "clear speech" phenomena (c.f. Smiljanić & Bradlow, 2009). This experiment, and others in which participants sit alone in a sound booth reading single words from a computer screen, is simply not likely to encourage such adaptations. With respect to the second point, many studies have argued that the lexicon itself is the product of many generations of shaping by processes of speech perception (De Cara & Goswami, 2002; Gupta & Dell, 1999; Pierrehumbert, 2002). It seems likely that the reason more words tend to differ in their onsets is largely perceptual in nature – a lexicon in which the primary cues to word differentiation occur in the initial position is likely to result in more efficient word recognition (Gupta & Dell, 1999).

4.5 Conclusion

In this chapter, several different metrics of neighborhood density were examined for their ability to predict durational differences between words. Words with more neighbors overlapping in the rime (and therefore *not* overlapping in the onset) were found to be produced with significantly longer VOT. Words with more total neighbors, and perhaps especially more neighbors overlapping in their final consonant, were found to be produced with shorter rimes. The results were interpreted in the context of a model of speech production that incorporates both feedback and sequential encoding. It was argued that the combination of these two factors most parsimoniously accounts for the present findings, and also allows them to be tied to previous findings in the literature. Since previous studies have found that positional neighborhood density affects both production latency and accuracy, it is likely that the durational effects reported here can be attributed to the same underlying cause: the fast and accurate assembly of an articulatory plan is made more difficult when feedback from phonologically related words provides conflicting instructions for the encoding of segments.

Chapter 5

Experiment 3: Phonetic duration in conversational speech

5.1 Background to Experiment 3

Relation to Experiments 1 and 2

The results of Experiment 1, the novel word learning study reported in Chapter 3, indicated that the existence of phonologically related neighbors is not a significant predictor of articulatory duration in children's single word productions. There was no evidence that competition between lexical representations had an appreciable effect on children's production of either known or novel words. Rather, the significant durational differences that were found in Experiment 1 could apparently be attributed to articulatory practice and to contrastive focus; more frequent phonotactic patterns were produced with shorter durations, and words on the second and third day of testing were produced with longer overall durations as compared to Day 1. Both findings held across the board, irrespective of the target words' phonological relationships to one another, to their neighbors in the lexicon, or to the novel words.

The goal of Experiment 2 was to determine whether feedback between lexical and phonological representations affects articulatory duration in adult single word productions. In their study of adults' production of minimal pair words, Baese-Berk and Goldrick (2009) found that voiceless stop-initial words that had a minimal pair neighbor (e.g. *cod*, which has a neighbor *god*) were produced with significantly longer voice onset time than words without such a neighbor (e.g. *cog*, which has no neighbor *gog*). The authors hypothesized that competition between minimal pair neighbors could be responsible for such an effect, and proposed that the boost in lexical activation that allows the target to out-compete its neighbor could give rise to longer VOT for such minimal pair words.

It was argued in Chapter 4, however, that the longer VOT for minimal pair words was not a direct result of competition between lexical representations, but rather of competition between phonological representations during the process of phonological encoding. A reanalysis of Baese-Berk and Goldrick's original data indicated that words with voicing-initial

minimal pair neighbors also tended to have more rime neighbors overall; that is, minimal pair words (such as *cod*) tended to have many *additional* neighbors that differed by their onset consonant (around nine, on average).

A set of regression models showed that targets with more neighbors differing in their onset consonant were produced with significantly longer VOT in single word productions. Moreover, the total number of neighbors and the number of neighbors differing only in their vowel were significantly related to the target's rime duration; words with more total neighbors were produced with shorter rimes than words with fewer neighbors, and words with more "CC neighbors" tended to be produced with even shorter rimes than words with fewer CC neighbors (although the latter effect was statistically marginal).

The conclusion drawn was that the activation of phonologically related words in memory gives rise to competition between phonological segments for a given position in the articulatory plan. It was tentatively argued that such positional competition between segments could have an effect on articulatory duration by way of the speed of phonological encoding. If articulation is initiated and executed more slowly for difficult to encode segments than for easily encoded segments, this could explain the results of Experiment 2 as well as several related findings in the literature.

The present chapter seeks to determine whether the results obtained for adults' single word productions will also hold for conversational speech. A more complete review of studies examining duration in conversational speech is provided in Chapter 2, but a selection of the most relevant findings is also summarized in the next section.

Summary of previous findings

Studies that have investigated the relation between articulatory duration and production planning have typically examined word duration as a function of lexical predictability. Early corpus studies in this vein include Jurafsky et al. (2001), Bell et al. (2003), and Aylett and Turk (2004), all of which reported data linking contextual predictability to the duration of words produced in connected speech[1].

Jurafsky et al.'s Probabilistic Reduction Hypothesis, and Aylett and Turk's Smooth Signal Redundancy Hypothesis are both intended to capture the relationship between contextual predictability and phonetic reduction. The Probabilistic Reduction Hypothesis assumes that contextual predictability is related to lexical accessibility, and that more accessible words tend to be produced with reduced articulations because they are in some sense easier for the speaker to plan. The Smooth Signal Redundancy Hypothesis remains mute on the subject

[1]More precisely, Aylett and Turk (2004, 2006) examined syllable durations as a function of what they called "language redundancy", a term that encompasses measures of predictability at the syllabic, word, and discourse levels of representation. Because these studies focused on the relationship between language redundancy and prosody, modeling syllable duration rather than word duration allows factors such as stress and prosodic emphasis to be examined in more depth. The important point for the present discussion is simply that corpus studies examining duration have typically focused on units larger than the phonological segment.

of accessibility, but rather makes reference to the amount of information available to the listener at any given moment. Given the assumption that speakers strive to maintain a constant flow of information for the listener, in cases where the amount of abstract linguistic information (or "redundancy") available is relatively low, the acoustic channel must pick up the slack. The result is that high language redundancy is associated with phonetic reduction.

Both hypotheses are thus largely based on the idea that the predictability of words in discourse is in some way related to their acoustic realization. But both hypotheses stop somewhat short of offering an explicit account of how such acoustic differences are implemented. The Smooth Signal Redundancy Hypothesis is more easily tied to a listener-oriented account of phonetic reduction, in which speakers are on some level aware of their listeners' needs, and adjust their articulation accordingly, online. The Probabilistic Reduction Hypothesis, by contrast, offers a speaker-oriented account, in which reduction is the result of easier planning, but it is not entirely clear how the accessibility of words in discourse can be mapped to their phonetic realization.

Several studies reporting effects of higher level factors on articulation in laboratory speech have suggested mechanisms that may be relevant to understanding phonetic reduction in conversational speech. Balota and Chumbley (1985) and Balota et al. (1989) found significant differences in word duration related to lexical frequency and semantic priming, respectively. These authors suggested that cascading activation flow from lexical to phonological representations was consistent with their results; if a given segment is articulated as it reaches a threshold level of activation, and if phonological encoding proceeds from left to right, this provides a potential link between higher lexical activation and faster articulation.

With respect to cascading activation flow, Bell et al. (2003) in fact incorporate this idea into their discussion of the gradient effects of contextual predictability on phonetic reduction in connected speech. They suggest that a cascading model of speech production can potentially account for the gradient effect of predictability on word duration. The argument is that unless varying amounts of lexical activation can somehow be passed down to the level of articulatory processes, it is difficult to explain the continuous (that is, non-binary) differences in duration that have been observed.

With respect to the stipulation that encoding proceeds from left to right, Sevald and Dell (1994) argued for what they termed a *sequential cuing effect* in phonological planning. In a set of experiments involving factorial manipulations of various types of phonological overlap, Sevald and Dell provided evidence that the position and amount of segmental overlap between two co-activated words was directed related to how easily they could be encoded for production. Participants were tasked with repeating as many alternating CVC syllable pairs as possible in eight seconds. Analyses of the number of syllables produced accurately and the type of speech errors committed indicated that segments planned earlier in the word can "cue" or "miscue" segments being planned for later in the word. Speakers produced relatively more CVC syllable pairs with overlapping coda consonants (e.g. *pick–tick*), whereas syllables with non-overlapping codas (*pick–pin*)were produced more slowly, and with more word-final errors.

The idea that interaction between lexical and phonological representations can have an

effect on the relative ease of phonological encoding is consistent with studies that have demonstrated an effect of phonological neighborhood density on speech production processes. In laboratory speech, studies such as Vitevitch and Sommers (2003), Vitevitch et al. (2004), and Goldrick et al. (2010) have provided evidence that the number and type of phonological connections between a target word and its neighbors affect the ease with which the target word can be encoded for production. In conversational speech, Gahl et al. (2012) recently demonstrated that words with more phonological neighbors are also produced with shorter durations and more reduced vowels, all else being equal. Both sets of findings would seem to be consistent with the idea that increased activation due to feedback from phonological neighbors is generally facilitatory for production.

The primary goal of Experiment 3 is therefore to better understand how feedback from phonological neighbors affects phonological encoding in conversational speech. The assumption is that a better understanding of how neighborhood structure affects the low-level phonetic details of words can usefully contribute to the question of whether and how lexical accessibility is related to articulation. If phonological overlap between words can predict the durations of sublexical units in conversational speech, then it stands to reason that neighborhood structure – and by extension, feedback between lexical and phonological representations – can affect the encoding of individual segments via cascading activation.

If positional phonological overlap is reliably related to segmental duration, then the same mechanism could potentially explain the relationship between greater contextual predictability and phonetic reduction. Higher lexical activation cascading down to phonological representations, plus left-to-right sequential cuing, would provide a viable account of the relationship between lexical accessibility and phonetic duration; the hypothesis is that the relative availability of phonological segments (and not of lexical representations *per se*) is the operative factor.

Research questions

1. Does positional overlap between phonological neighbors affect the duration of words and segments in conversational speech?

2. Does the contextual probability of a word have a consistent effect on the duration of all of its constituent segments, or are certain positions in the word more or less affected than others?

5.2 Methods

The Corpus

The Buckeye Corpus of Conversational Speech

The Buckeye Corpus of Conversational Speech (Pitt et al., 2007) contains audio recordings and time aligned transcriptions of sociolinguistic-style interviews with 40 speakers from Columbus, Ohio. Each interviewee was recorded speaking conversationally for approximately one hour, and transcriptions were later time aligned at the utterance, word, and phone levels. The full corpus contains approximately 300,000 words.

Speakers were recruited so as to balance the number of men and women in each of two stratified age ranges; "younger" speakers were those under age 30, and "older" speakers were those over age 40 (no speakers between the ages of 30 and 40 were recorded). A total of 20 men and 20 women were recorded, for 10 participants of each gender in each age group. All speakers were natives of central Ohio and spoke with a Midwestern accent.

Inclusion criteria and data reduction

To address the present research questions concerning voice onset time, rime duration, and any differences between the two, a list of candidate English words was first generated using the MRC Psycholinguistic Database (Coltheart, 1981). Candidate word types were those that started with the segments /p t k/ immediately followed by a vowel (i.e. no word-initial consonant clusters), were monosyllabic and monomorphemic, not proper names, and not function words (e.g. *can* was not included, because even though it has a common noun meaning, its most common usage is as a modal verb). No restrictions were placed on the number or type of segments appearing in the coda. A total of 173 words met the inclusion criteria; however, not all words were present in the corpus, and the number of unique word types ultimately included in the analysis was further diminished due to the data reduction procedures.

Only utterance-medial tokens were included in the analysis. Any tokens occurring at the beginning or end of a conversational turn were excluded, as were tokens immediately preceding or following pauses longer than 500 ms. Tokens abutting "non-speech" sounds such as laughter, coughing, and filled pauses (*uh, umm*) were also excluded. An additional three tokens were excluded because they were calculated to have negative voice onset time (see below for details on measuring VOT). The final data set included 4,151 tokens of utterance-medial, monosyllabic, monomorphemic /p t k / words, produced by all 40 of the speakers in the corpus.

Final data set

Following data reduction, 4,151 tokens remained for the durational analyses, comprising 159 unique word types. Table 5.1 provides summary statistics for these 4,151 tokens. A very

		mean	median	sd	range
corpus	tokens/word	26.1	4.0	59.3	1 – 384
	tokens/speaker	103.8	94.5	46.2	37 – 234
lexical	word dur (ms)	259.5	248.0	82.6	70.6 – 743.4
	VOT (ms)	56.2	53.7	24.0	0.1 - 231.3
	rime dur (ms)	150.1	140.0	67.6	17.4 – 558.3
	# phon segments	3.0	3.0	0.5	2 – 4
	freq/million	886.6	699.6	847.2	1.5 – 3141.0
	mean biph prob	0.0038	0.0021	0.0034	0.0002 – 0.0164
	mean seg prob	0.055	0.054	0.015	0.024 – 0.089
	orth length	3.8	4.0	0.6	3 – 6
contextual	cond prob, prev	1.5×10^{-7}	1.8×10^{-8}	5.5×10^{-7}	3.3×10^{-10} – 4.2×10^{-6}
	cond prob, follow	1.9×10^{-7}	2.9×10^{-8}	5.6×10^{-7}	3.3×10^{-10} – 4.2×10^{-6}
	rate before (syll/s)	6.3	6.0	2.2	1.2 – 26.3
	rate after (syll/s)	5.4	5.3	1.8	1.0 – 41.0
neighborhood	total # nbors	21.5	22.0	6.5	3 – 39
	# rime nbors	8.5	8.0	4.2	0 - 17
	# onset nbors	10.6	12.0	3.5	1 – 18
	# CC nbors	5.6	5.0	3.0	0 – 10
	# CV nbors	5.0	6.0	2.5	0 – 10
	sum freq of nbors	2946.3	1688.0	3940.3	20 – 53033

Table 5.1: Summary statistics for data used in the analysis of word and segment durations in the Buckeye corpus. See text for details.

small number of word types made up the majority of the data; note that the mean number of tokens per word is 26.1, while the median is only 4. The top twelve most frequent words each contributed over 100 tokens to the analysis, while the 90 least frequent words each contributed five tokens or less.

5.3 Acoustic analysis

All acoustic analyses were performed automatically, by computer script. Since the word and segment boundaries were hand corrected by the developers of the corpus (Pitt et al., 2007), word and rime durations were calculated directly from the existing label files. However, since the segment boundaries for stops included the stop closure, and the dependent measure of interest for the current study is voice onset time, determining VOT required further acoustic analysis. A burst detector developed by Yao (2007, 2009) was therefore used to calculate VOT. Yao (2007) reports a margin of error of approximately 3 ms for the burst detector's performance on a subset of the Buckeye corpus.

5.4 Statistical analysis and results

The modeling strategy and predictor variables for the analysis of the Buckeye data are largely the same as for the analysis presented in Chapter 4. The main difference, obviously, is that the present analysis is concerned with connected speech rather than isolated words, requiring the inclusion of several control variables not relevant for single-word production. Each predictor will first be described, in the order in which it was initially entered during model fitting. The modeling procedure – the same as that implemented in the analysis of isolated words – will then be described, followed by the results for the analyses of total word duration, rime duration, and voice onset time.

Predictor variables

Random effects. No matter how well the fixed effects are able to describe the observed differences in duration, some amount of variation will be due to the fact that individual speakers and words have their own idiosyncratic properties. To capture some of this random variation, random intercepts for speaker and word, and random by-speaker slopes for the effects of speech rate, conditional probability, and minimal pair status were examined.

Baseline duration. Each phonological segment has its own intrinsic duration, and word and rime durations will vary depending on the particular segments they contain. As in the analysis of isolated words, a grand mean segment duration was calculated for each phonological segment: the average duration for each segment, for each speaker, was calculated over the entire Buckeye corpus, and these by-segment, by-speaker means were then averaged together. Mean baseline word and rime durations were then calculated by adding together the mean durations of segments comprising the word or rime. As noted in Chapter 4, this is by no means a perfect baseline measure, but in combination with the random effect for word, the predictors related to syllable structure, and the fact that all words under examination are relatively similar, it should provide a good starting point. Baseline word and rime durations were log transformed to improve the normality of their distribution.

For the VOT analysis, the observed rime duration (rather than a theoretical baseline duration derived from the grand segment means) was again used as a control parameter. Note that this strategy is motivated by the finding that changes in VOT have repeatedly been found to track with changes in rime duration, when measured proportionally; Boucher (2002) found that best-fit regression lines for ratios of VOT and total syllable duration had a slope of zero, indicating that they were constant across speech rates, and several studies by Smiljanic and Bradlow have demonstrated that proportional VOT is consistent across speaking rates, styles, and even languages (Smiljanić & Bradlow, 2008a, 2008b). All else being equal, VOT is expected to increase or decrease in proportion to rime duration. Observed rime durations were also log transformed.

Initial consonant segment. In the analysis of VOT, the initial consonant segment ('p' vs. 't' vs. 'k') was entered as a factor, since VOT is known to vary systematically according

to place of articulation for physiological reasons (Cho & Ladefoged, 1999). It was expected that /p/ would have the shortest VOT, and /k/ the longest.

Vowel height. VOT also varies according to the height of the following vowel, with high vowels typically being associated with longer VOT. For the sake of simplicity, and to avoid overfitting of the data, vowel height was treated as a binary variable; the vowels /i ɪ e o ʊ u/ were considered high, and all other vowels were considered low.

Syllable structure. In the analysis of isolated words in Chapter 4, the stimuli had been carefully chosen to balance syllable structure as closely as possible across stimulus types, and all words belonged to only one of seven syllable shapes (Baese-Berk & Goldrick, 2009; see Table 4.6 for reference). In the present analysis, however, relatively fewer restrictions were placed on syllable structure in order to include as many word types as possible; recall that all monosyllabic, monomorphemic words beginning with /p t k/ and followed by a vowel were included in the analysis, regardless of the number or type of consonants in the coda. This resulted in a total of 13 syllable types, when voicing and manner of articulation are taken into account.

For the present analysis, then, it makes little sense to include a fixed effect of syllable structure with 13 levels. Rather, three fixed effects rooted in well-known phonological principles were examined as predictors of phonetic duration: the number of consonants in the coda, the voicing of the final consonant, and the manner of the consonant immediately following the vowel. All two-way interactions between these predictors were also examined.

Speech rate. Word and segment durations will be longer or shorter depending on how quickly the speaker is talking, and speech rate will vary throughout a given interaction. For each token in the dataset, the speech rate immediately preceding and following it was calculated by determining the duration of the utterance before and after the token (excluding the token itself), counting the number of syllables in that portion of the utterance, and dividing the syllable count by the duration. This resulted in two measures, "speech rate before" and "speech rate after", which were log transformed to improve their normality.

Contextual probability. Many studies have demonstrated that words with greater contextual probability tend to be produced with reduced articulations, including shorter overall durations (Bell et al., 2003, 2009; Jurafsky et al., 2001; Gahl, 2008, among others). Following Bell et al. (2009), the conditional probability for each word given the previous word, and given the following word, were log transformed and entered as control variables. Also following Bell et al. (2009), the two-way interactions between previous and following conditional probability, and word frequency were also examined as potential predictors.

Lexical frequency. More frequent words also tend to be produced with reduced articulations (e.g. Jurafsky et al., 2001). As in Chapter 4, the measure of word frequency used was the number of occurrences per million words in the SUBTLEX-US database of English movie subtitles (Brysbaert et al., 2012). Word frequencies were log transformed.

First mention. The first mention of a word in a given speaking situation tends to be less reduced than subsequent mentions (?, ?). First mention was entered as a binary variable corresponding to whether the word had been previously produced by the speaker earlier in the recording session.

Phonotactic probability. Just as more frequent words tend to be produced with shorter durations, more frequent segments and sequences of segments may also be produced with shorter durations. To control for this possibility, three measures of phonotactic probability were examined: the mean probability of all biphone sequences in the word, the mean positional probability of all segments in the word, and in the VOT analysis, the positional probability of the first biphone alone was also examined. All measures of phonotactic probability were drawn from the Online Phonotactic Probability Calculator (Vitevitch & Luce, 2004). Mean biphone probabilities and initial biphone probabilities were log transformed, but mean segmental probabilities were already relatively normally distributed and therefore not transformed. Phonotactic predictors were examined one at a time, such that none were competing with one another for variance. Phonotactic measures residualized with neighborhood metrics were also examined, in combination with the neighborhood metrics. This procedure is described in more detail below.

Part of speech. Words of different syntactic categories may be produced with different durations on average, for example due to their average position in the sentence or the probability that they will receive additional prosodic stress. For each word type, the part of speech with the highest frequency of occurrence in the SUBTLEX database was assigned to all tokens of that word. For example, the word *coach* appears nine times in the dataset. According to the SUBTLEX counts, *coach* is used as a noun 62% of the time, so all nine tokens of *coach* that appear in the corpus were coded as nouns.

Speaker characteristics. Bell et al. (2003) found that speaker sex and age were both significant predictors of word duration in conversational speech. These predictors were both coded as binary variables ('male' vs. 'female' and 'younger' vs. 'older', respectively). It may also be the case that age interacts with other predictors. In light of this possibility, the interactions between previous and following conditional probability, and speaker age were also examined.

Orthographic length. Warner, Jongman, Sereno, and Kemps (2004) found that the orthographic length of a word, in letters, can affect word duration above and beyond any effects of phonemic length. Orthographic length was therefore entered as a predictor.

Neighborhood metrics. As in the analysis of single word productions presented in Chapter 4, several different neighborhood density metrics were examined as possible predictors in the present analyses. Total number of phonological neighbors (words differing from the target by a single phoneme addition, substitution, or deletion), number of rime neighbors (words differing only in their onset consonant), number of onset neighbors (words differing by a single rime segment), number of "CC" neighbors (words differing only in their vowel), number of "CV" neighbors (words differing by a single coda consonant), log transformed summed lexical frequency of all neighbors, and minimal pair status were examined in turn.

Also as in Chapter 4, the number of phonological neighbors was based on the Hoosier Mental Lexicon (Nusbaum et al., 1984), and was obtained using the Child Mental Lexicon online calculator (Storkel & Hoover, 2010), because the latter provides a simple online interface for obtaining positional neighbors in the HML. Minimal pair status was determined using the SUBTLEX-US database (Brysbaert et al., 2012), and was coded as a binary vari-

able contingent on the existence of a non-proper name minimal pair (based on the voicing of the first segment) in SUBTLEX. The requirement that minimal pairs be non-proper names was only relevant for *torn* (cf. *Dorn*, which occurs in SUBTLEX as a proper name nine times), *tease* (cf. *Dees*; five times), and *cup* (cf. *Gup*; one time).

To help distinguish any possible effects of minimal pair status from more general neighborhood effects, the voicing minimal pair neighbor was not included in the count of total neighbors or rime neighbors. (That is, the counts of total neighbors and rime neighbors based on the HML were each reduced by one, although whether the minimal pair neighbor was included or not ultimately had no effect on the results.)

Modeling procedure

Predictor variables were entered into the model in the order in which they were expected to yield significant model improvement: random effects, baseline durations, factors related to syllable structure/phonology, speech rate, contextual probability, lexical frequency, first mention, phonotactic probability, part of speech, speaker characteristics, and orthographic length. Neighborhood metrics were examined last, after the control model was fit.

Significant model improvement was determined by a chi-squared test comparing the log likelihoods for a model with vs. without a given predictor. Predictors that did not yield significant improvement at an alpha level of 0.15, and/or whose coefficients were not significantly different from zero according to pMCMC simulations (at an alpha level of 0.10) were dropped from the model before additional predictors were added. Once the control model was fit, leave-one-out comparisons were used to ensure that each predictor still resulted in a significant gain in log likelihood, with all other predictors in the model.

Each of the neighborhood metrics was then examined, one at a time, to determine whether it contributed any predictive power to the fitted control model. As in Chapter 4, a strategy of predictor residualization was used to minimize multicollinearity. Because the neighborhood metrics were all significantly correlated with the baseline durations (and with one another), simple linear models predicting neighborhood density from the baseline duration were constructed, effectively removing variation associated with baseline duration from each of the neighborhood metrics. The residuals from each simple linear model were then entered as "residualized" predictors, and the log likelihood of models with versus without each neighborhood metric were compared.

Because phonotactic probability and the neighborhood metrics were in some cases significantly correlated, residualized measures of phonotactic probability were also examined in combination with the neighborhood metrics. However, in no case was there any indication that the measures of phonotactic probability contributed even marginal predictive power to the model, whether residualized and entered along with neighborhood density, or given a chance to explain variability on their own.

Finally, because the majority of the data comes from a very small number of word types, the same procedure was followed for a reduced version of the dataset; ten words contributing

predictor	β	t	p**MCMC**	**AIC**	**loglik**	χ^2	$p(\chi^2)$
intercept	-0.480	-2.79	0.0036	326.5	-159.3		
baseline word dur	0.584	5.43	0.0001	307.1	-148.6	21.4	0.0001
# coda cons = 2	-0.180	-3.50	0.0002	295.9	-140.9	15.3	0.0001
final cons = voiceless	-0.070	-1.88	0.0328	141.0	-63.5	154.8	0.0001
manner post-V cons				132.2	-57.1	12.9	0.0016
post-V cons = sonorant	-0.043	-1.19	0.1544				
post-V cons = fricative	0.089	2.60	0.0050				
# coda cons = 2:final cons = vcless	0.094	1.68	0.0546	129.3	-54.6	4.9	0.0273
speech rate before	-0.061	-4.96	0.0001	101.0	-39.5	30.3	0.0001
speech rate after	-0.143	-10.82	0.0001	7.0	8.5	96.0	0.0001
following cond prob	-0.022	-7.85	0.0001	-121.0	73.5	130.1	0.0001
previous cond prob	-0.009	-3.55	0.0004	-131.6	79.8	12.6	0.0004
word frequency	-0.056	-3.19	0.0004	-143.7	86.9	14.2	0.0002
part of speech				-144.0	91.0	8.3	0.0805
PoS = adj	0.040	0.73	0.4962				
PoS = adv	0.086	1.03	0.2784				
PoS = number	0.062	0.66	0.4226				
PoS = verb	-0.043	-1.70	0.0402				
age = young	-0.183	-2.36	0.0166	-144.2	92.1	2.2	0.1407
age = young:following cond prob	-0.008	-1.93	0.0570	-145.9	94.0	3.7	0.0547
min pair = y	0.055	2.18	0.0064	-149.5	96.8	5.6	0.0178

Table 5.2: Significant predictors of total word duration in Experiment 3. This model includes random intercepts for participant and word.

random effect	**SD**	**MCMC median**	**HPD95lower**	**HPD95upper**
word (intercept)	0.086	0.067	0.054	0.082
speaker (intercept)	0.097	0.089	0.070	0.113
residual	0.229	0.230	0.224	0.235

Table 5.3: Random effects in the model of total word duration in Experiment 3.

over half the tokens in the dataset were removed, and the analysis was re-run. This procedure is described in more depth below.

Analysis of total word duration

Significant fixed effects in the fitted model of word duration are shown in Table 5.2. The top portion of the table gives the control model, and the bottom portion gives the neighborhood metrics that resulted in a significant gain in the model's log likelihood; for total word durations, minimal pair status was the only neighborhood-related metric whose addition significantly improved model fit. Table 5.3 summarizes the random effects included in the fitted model.

In general, effects were in the predicted direction. Longer baseline word durations were

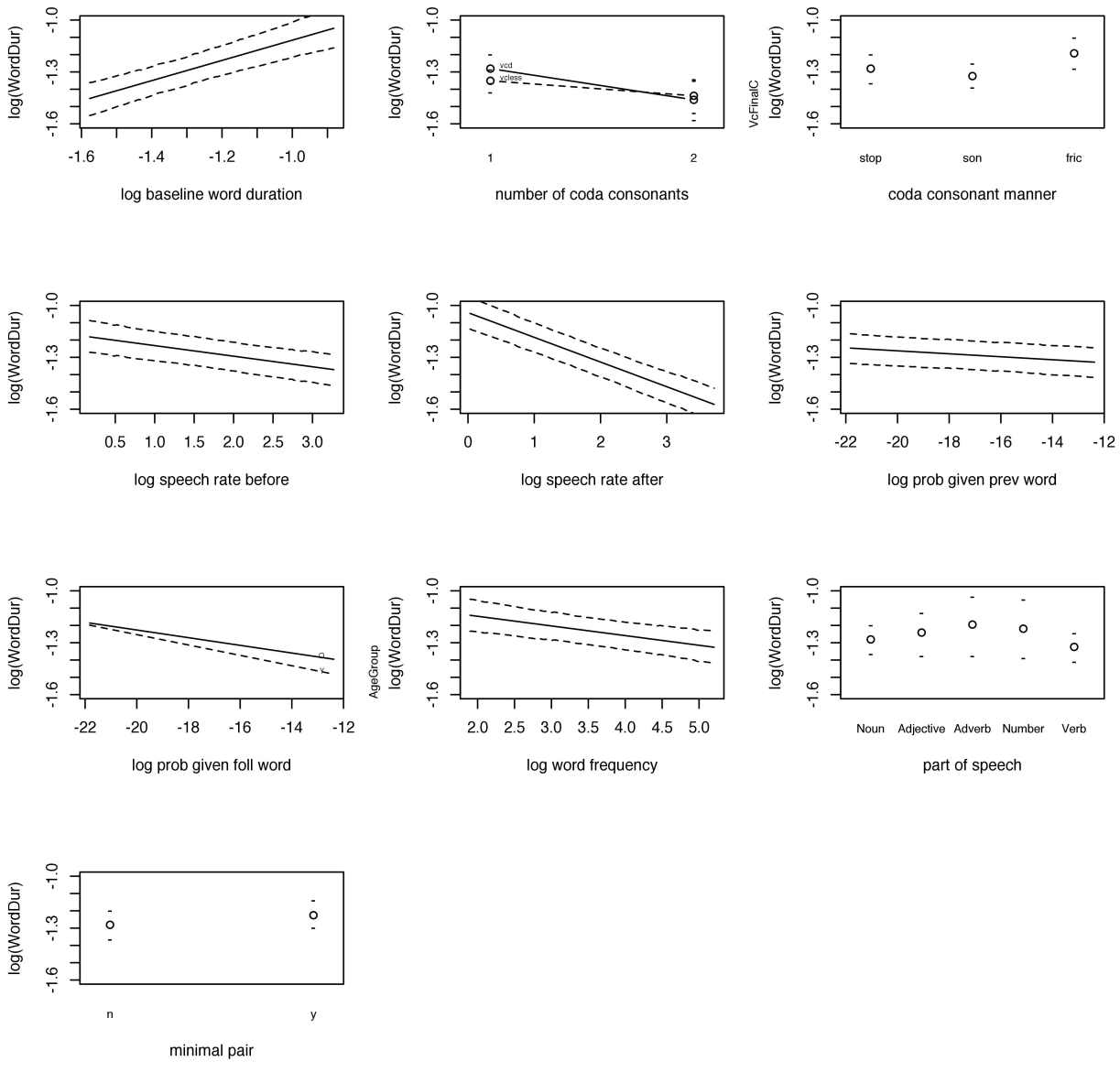


Figure 5.1: Partial effects plot for the fitted model of total word durations in Experiment 3.

associated with longer observed durations, and the negative coefficient for the number of coda consonants indicates that for longer words, the baseline durations overestimated the observed durations. Several phonological factors were also highly significant. Words with voiceless final consonants were produced with significantly shorter durations than words with voiced final consonants. When a fricative consonant immediately followed the vowel, the observed word duration was significantly longer than when a stop followed the vowel (the reference level). Vowel-sonorant sequences were associated with marginally shorter word duration than vowel-stop sequences. The number of coda consonants also interacted significantly with the voicing of the final consonant; words with complex codas ending in a voiceless consonant (e.g. *coast*, *tank*) were longer than would otherwise be predicted by the main effects. There were no significant interactions involving post-vocalic consonant manner.

With respect to context-related predictors, preceding and following (log transformed) speech rate, as well as preceding and following (log transformed) conditional probability were all highly significant predictors, and all were associated with shorter word durations; faster speech rate and a more predictable context resulted in shorter articulation time. Higher word frequency was also associated with significantly shorter durations.

Syntactic category, or part of speech, significantly increased model log likelihood, with the gain in predictive power coming mostly from the fact that all else being equal, verbs were produced with shorter durations than nouns. Word durations were also significantly shorter for younger speakers; the main effect of speaker age alone is only marginally significant ($p(\chi^2) = 0.14$), but once the interaction between speaker age and following conditional probability is added (also contributing significant predictive power, $p(\chi^2) = 0.05$), the coefficient for the main effect of speaker age is found to be significantly different from zero, with a pMCMC value around 0.01.

The only neighborhood-related predictor that reached significance was minimal pair status. Words with onset voicing minimal pairs were produced with shorter durations, and this effect was quite statistically reliable ($\chi^2 = 5.6$, $p(\chi^2) = 0.0178$, when added to the control model). It also persisted when random by-speaker slopes for the effect of minimal pair status were added (these random slopes resulted in a marginally significant improvement in model log likelihood, but no qualitative changes to the model). That minimal pair status would be reliably associated with word duration in conversational speech is rather surprising for several reasons, however.

First, as in Chapter 4, minimal pair status was highly correlated with all of the neighborhood metrics under investigation, and yet none of the neighborhood metrics even approached statistical significance. This could suggest a non-neighborhood-related explanation for the association between minimal pair status and word duration. For example, it could be the case that words with minimal pairs were on average more likely to receive contrastive focus precisely due to the existence of their onset voicing minimal pairs. However, this explanation seems rather unlikely given the nature of the speaking situation and the minimal pairs in question. The SUBTLEX database was used to determine whether words that met the inclusion criteria for the study had existing minimal pairs in the English lexicon. While some words had relatively high frequency minimal pairs, many of the minimal pair words are low

min pair = no	# occurrences	min pair = yes	# occurrences
take	283	two (do)	384
put	204	time (dime)	351
keep	106	come (gum)	240
took	95	talk (dock)	162
kind	76	tell (dell)	159
point	57	part (bart)	142
type	56	came (game)	141
teach	45	care (gare)	127
case	28	kid (gid)	113
kept	22	ten (den)	106

Table 5.4: Top ten highest frequency words in each minimal pair category in the analysis of word duration in the Buckeye corpus.

frequency (e.g. *paid/bade*), of a different syntactic class (*come/gum*), and/or have absolutely no semantic relation to their counterparts (*pass/bass*), making it unlikely that the words under investigation would be contrastively focused in a conversational context simply due to the existence of an onset voicing minimal pair in the lexicon.

Second, since the majority of the data comes from a small number of word types, the "effect" of minimal pair status largely comes down to a comparison of the top ten highest frequency words with versus without a minimal pair. Table 5.4 gives the top ten highest frequency words in each minimal pair category, along with the number of times they appear in the corpus and their minimal pair word in SUBTLEX, where appropriate.

Arguments could be made for excluding several of the highest frequency words in the corpus. Two of the most frequent words with minimal pairs are numbers (*two* and *ten*), and it is conceivable that the behavior of number words is qualitatively different from that of other syntactic categories. Several of the words in Table 5.4 are also likely to be part of fixed phrasal chunks and/or may be partially grammaticalized or behave along the same lines as function words (e.g. *take/took, keep, come/came*). It is also striking that the highest frequency words without minimal pairs almost all end in voiceless consonants (the only exception is *kind*), while the highest frequency words with minimal pairs tend to end in voiced consonants. All of these issues raise the possibility that the "minimal pair effect" that is significant over the entire dataset may actually be due to some otherwise uncontrolled-for properties of the highest frequency words; it is reasonable to ask whether the observed effect is actually generalizable over the rest of the data, or whether it is due to some idiosyncratic properties of the highest frequency words.

To investigate this possibility, the same analysis was run on a reduced version of the full dataset. All tokens representing the top five highest frequency words in each minimal pair category were removed, thereby reducing the number of unique word types from 159 to 149, but reducing the size of the dataset from 4,151 total observations to 2,091 (a 50% reduction).

Using the reduced dataset, all predictors that were significant in the analysis of the full

dataset remained significant, except for minimal pair status; the fitted control model is qualitatively exactly the same as that shown in Table 5.2, but a model that also contains minimal pair status fares no better than the control model in predicting total word duration ($\chi^2 = 1.3$, $p(\chi^2) = 0.25$). The apparent effect of minimal pair status on total word duration therefore appears to be due to some unknown property or properties of the highest frequency words, and is not generalizable to all word types.

To evaluate model fit, the squared correlation between fitted and observed values was calculated. For the fitted model of the full dataset (including the effect of minimal pair status), the squared correlation was 0.430. For the reduced dataset (and therefore not including minimal pair status), it was 0.403.

Analysis of rime duration

Because the majority of the total word duration *is* the rime duration, the same fixed effects that are significant predictors of total word duration should be expected to be significant in the analysis of rime duration. This is indeed exactly what we find, with one very interesting exception: conditional probability given the previous word is *not* a significant predictor of rime duration ($\chi^2 = 0.5$, $p(\chi^2) = 0.47$, when added to the fitted control model), suggesting that the significant effect of previous conditional probability in the analysis of word duration was due to a difference in the duration of the initial consonant. This finding will be discussed at length below. The fitted rime duration model is given in Table 5.5, and Table 5.6 provides a summary of the random effects.

Other than the lack of effect for previous conditional probability, the results for rime duration are the same as the results for total word duration: baseline rime duration, post-vocalic fricatives, final voiceless consonants in complex clusters, and minimal pair status are all associated with significantly longer rime durations, and all other significant predictors are associated with rime shortening.

Also as in the analysis of total word duration, the only neighborhood-related predictor that reaches statistical significance is minimal pair status, and this predictor is no longer significant when the top five highest frequency word types in each minimal pair category are excluded ($\chi^2 = 1.7$, $p(\chi^2) = 0.19$, comparing log likelihoods for models with vs. without minimal pair status, using the reduced dataset). This again suggests that the significant association between minimal pair status and rime duration is due to some property of the highest frequency words, and is not generalizable to all word types.

For the rime duration model, the squared correlation between fitted and observed values for the fitted model of the full dataset (including the effect of minimal pair status) was 0.482. For the reduced dataset (and therefore not including minimal pair status), it was 0.432.

Analysis of voice onset time

Contrary to the predictions for rime duration, there is good reason to suspect that a different set of predictors may be relevant in the analysis of voice onset time. For one, using the

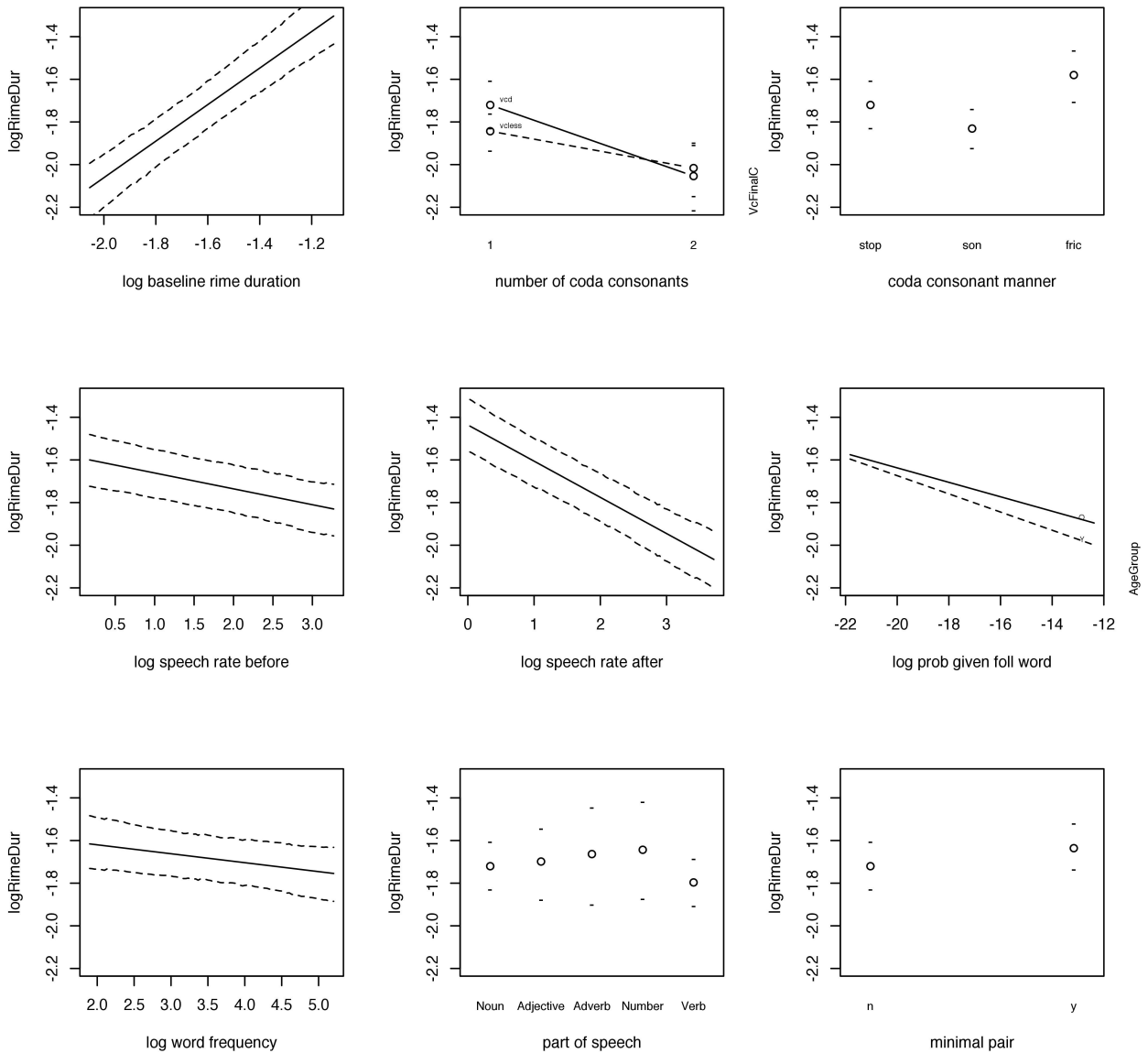


Figure 5.2: Partial effects plot for the fitted model of rime durations in Experiment 3.

predictor	β	t	p**MCMC**	**AIC**	**loglik**	χ^2	$p(\chi^2)$
intercept	-0.361	-1.66	0.1224	2955.9	-1474.0		
baseline rime dur	0.806	7.71	0.0001	2916.7	-1453.4	41.2	0.0001
# coda cons = 2	-0.332	-4.44	0.0001	2894.8	-1440.4	25.9	0.0001
final cons = voiceless	-0.138	-2.61	0.0020	2143.6	-1064.8	751.3	0.0001
manner post-V cons				2125.7	-1053.9	21.9	0.0001
post-V cons = sonorant	-0.098	-1.90	0.0268				
post-V cons = fricative	0.154	3.12	0.0010				
# coda cons = 2:final cons = vcless	0.159	1.98	0.0172	2120.5	-1050.3	7.2	0.0072
speech rate before	-0.074	-4.56	0.0001	2095.6	-1036.8	26.9	0.0001
speech rate after	-0.169	-9.77	0.0001	2022.9	-999.4	74.7	0.0001
following cond prob	-0.034	-9.10	0.0001	1862.4	-918.2	162.5	0.0001
word frequency	-0.044	-1.78	0.0274	1857.7	-914.8	6.7	0.0096
part of speech				1857.7	-910.9	7.94	0.0937
PoS = adj	0.047	0.61	0.5890				
PoS = adv	0.108	0.92	0.2730				
PoS = number	0.080	0.58	0.4478				
PoS = verb	-0.074	-2.02	0.0188				
age = young	-0.213	-2.12	0.0388	1857.2	-909.6	2.6	0.1099
age = young:following cond prob	-0.009	-1.65	0.1052	1856.5	-908.2	2.7	0.1004
min pair = y	0.085	2.44	0.0036	1851.6	-904.8	6.9	0.0087

Table 5.5: Significant predictors of rime duration in Experiment 3. This model includes random intercepts for participant and word.

random effect	**SD**	**MCMC median**	**HPD95lower**	**HPD95upper**
word (intercept)	0.120	0.092	0.072	0.114
speaker (intercept)	0.112	0.105	0.081	0.134
residual	0.301	0.303	0.296	0.310

Table 5.6: Random effects in the model of rime duration in Experiment 3.

observed rime duration as a baseline measure in the VOT analysis means that any significant predictors can be interpreted as affecting VOT above and beyond their effects on the rest of the word. Put another way, if a given predictor does not reach significance in the VOT analysis, this does not necessarily indicate that the variable in question has no effect on VOT; rather, it indicates that its effect on VOT is proportionally the same as its effect on the rime.

Second, as described in the section on predictor variables, several additional factors are potentially relevant for predicting VOT; namely, consonant place of articulation, vowel height, and the biphone probability of the first biphone alone.

And finally, the somewhat surprising finding that conditional probability given the previous word did not have a significant effect on rime duration (while it did have a significant effect on total word duration) suggests that previous conditional probability will be a significant predictor of voice onset time.

predictor	β	t	p**MCMC**	**AIC**	**loglik**	χ^2	$p(\chi^2)$
intercept	-2.791	-23.85	0.0001	6271.7	-3131.9		
rime dur	0.160	7.01	0.0001	6213.8	-3101.9	60.0	0.0001
onset consonant				6195.4	-3090.7	22.3	0.0001
onset = /p/	-0.183	-4.00	0.0001				
onset = /k/	0.067	1.60	0.0472				
speech rate before	-0.039	-1.55	0.1060	6193.7	-3088.8	3.7	0.0530
speech rate after	-0.094	-3.52	0.0002	6182.1	-3082.1	13.6	0.0002
previous cond prob	-0.021	-4.34	0.0001	6165.9	-3072.9	18.3	0.0001
first mention = y	0.027	1.65	0.0918	6165.4	-3071.7	2.5	0.1150
part of speech				6163.9	-3067.0	9.4	0.0511
PoS = adj	0.148	1.66	0.0710				
PoS = adv	-0.073	-0.51	0.6076				
PoS = number	0.194	1.96	0.0242				
PoS = verb	-0.036	-0.98	0.2522				
min pair = y	0.062	1.69	0.0520	6162.5	-3065.3	3.4	0.0655
residualized rime nbors	0.010	2.12	0.0274	6161.1	-3064.6	4.8	0.0280
residualized sum lex freq	0.047	3.49	0.008	6153.3	-3060.7	12.6	0.0004

Table 5.7: Significant predictors of voice onset time in Experiment 3, using the full dataset. This model includes random intercepts for participant and word.

random effect	**SD**	**MCMC median**	**HPD95lower**	**HPD95upper**
word (intercept)	0.113	0.094	0.064	0.127
speaker (intercept)	0.170	0.162	0.125	0.202
residual	0.496	0.497	0.487	0.508

Table 5.8: Random effects in the model of voice onset time in Experiment 3.

	RimeDur	RateBef	RateAft	PrevProb	VCDens	LFSum	CCDens
RateBef	-0.130						
RateAft	-0.124	0.163					
PrevProb	0.047	-0.051	-0.014				
VCDens	0.020	-0.035	-0.109	0.011			
LFSum	0.016	-0.013	-0.075	0.029	0.792		
CCDens	-0.018	0.007	0.024	-0.097	-0.009	-0.013	
CVDens	0.035	0.002	-0.043	0.095	0.225	0.377	-0.224

Table 5.9: Pairwise Spearman correlations among numerical predictors in the fitted model of VOT. RimeDur = log transformed rime duration; RateBef = log transformed speech rate before the target word; RateAft = log transformed speech rate after the target word; PrevProb = conditional probability given the previous word; VCDens = number of VC neighbors, residualized with rime duration; LFSum = summed lexical frequency of the target word's neighbors, residualized with rime duration; CCDens = number of CC neighbors, residualized with rime duration; CVDens = number of CV neighbors, residualized with rime duration.

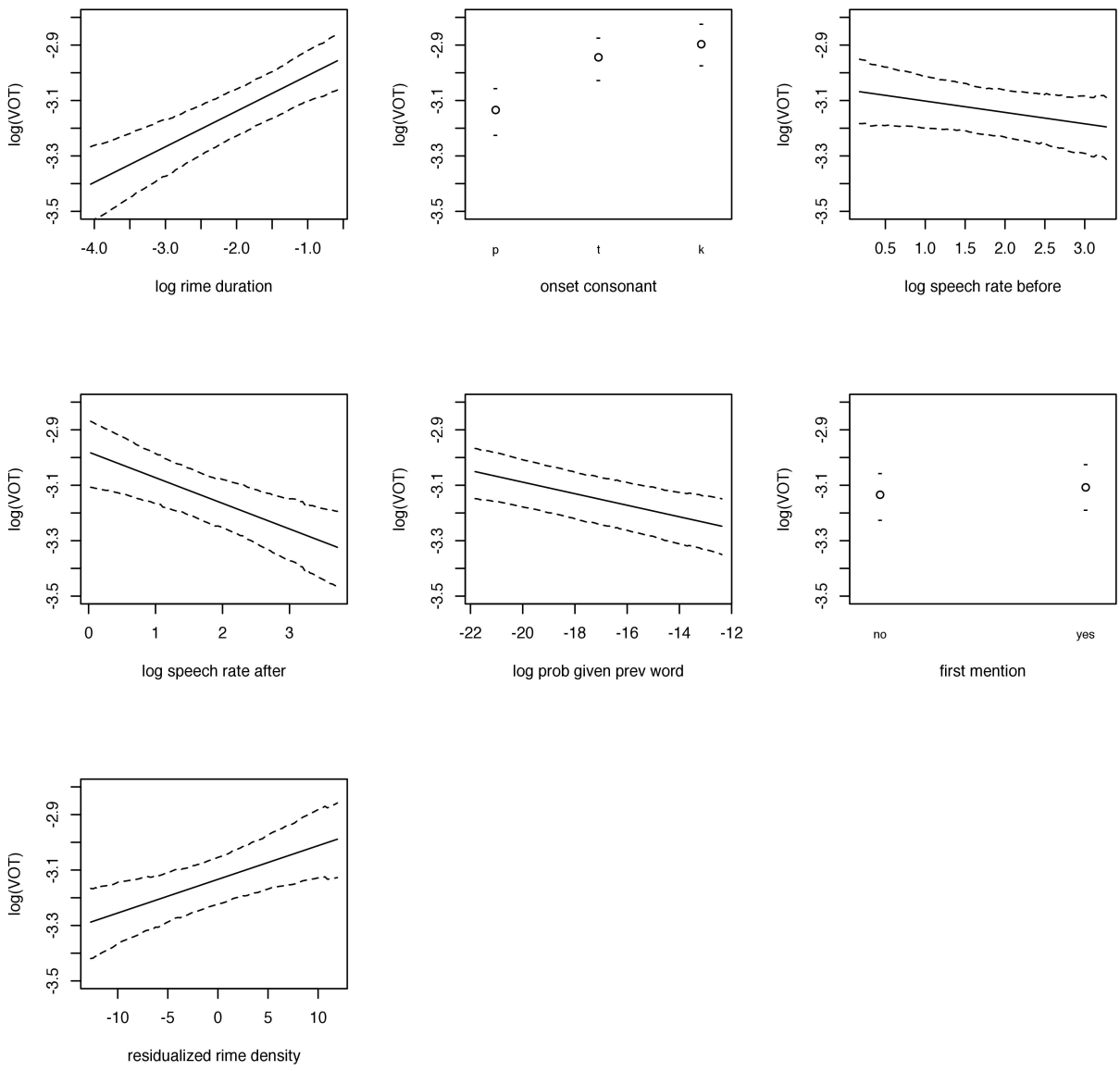


Figure 5.3: Partial effects plot for the fitted model of voice onset time in Experiment 3.

Table 5.7 gives the fitted model for voice onset time, again with the control model on top and significant neighborhood metrics listed separately, below. A summary of the random effects is given in Table 5.8. As expected, onset consonant was a significant predictor; words beginning with /p/ had significantly shorter VOT than words beginning with /t/ (the reference level), while words beginning with /k/ had significantly longer VOT. The effect of vowel height was in the expected direction (high vowels were associated with longer VOT, on average), but it did not reach statistical significance and was dropped from the model.

Preceding and following (log transformed) speech rate were each associated with additional shortening of VOT that was not captured by the observed rime duration predictor. As predicted, greater conditional probability given the previous word was associated with significantly shorter VOT. Conditional probability given the following word did not contribute significant predictive power to the model. Part of speech also resulted in significant model improvement, though its predictive power seems to come largely from the fact that numbers had significantly longer VOT than nouns. It should be noted that the only numbers in the dataset were *two* and *ten*, making it difficult to say whether this effect is generalizable to all numbers, and further motivating the use of the reduced dataset as in the analyses of total word duration and rime duration. Finally, in contrast to the findings for total word duration and rime duration, the first mention of a word was associated with significantly longer VOT than subsequent mentions.

With respect to the neighborhood metrics, three neighborhood-related predictors were significant in the analysis of the full dataset: minimal pair status, number of rime neighbors, and the log transformed sum of the lexical frequencies of all neighbors (the latter two predictors having been residualized from the observed rime duration, which was highly correlated with all of the neighborhood metrics). Words with minimal pairs had longer VOT, as did words with more rime neighbors and words with higher summed lexical neighborhood frequency.

A correlation matrix providing pairwise Spearman correlations between numerical variables in the final model is given in Table 5.9. The condition number of these predictors, κ, was calculated to be 11.3, which is considered an unproblematic level of multicollinearity (Baayen, 2008).

The analyses of word duration and rime duration suggested that some idiosyncratic properties of the highest frequency words in the dataset may have had undue influence on the model. To determine whether this was also the case when voice onset time was the dependent variable, the same analysis was run using the reduced dataset (that is, excluding the five highest frequency words that had a minimal pair, and the five highest frequency words that did not have a minimal pair). Table 5.10 gives the fitted model for this reduced version of the dataset.

The fitted model predicting (log transformed) VOT in the reduced dataset is considerably different from the model for the full dataset. With respect to the control model (top portion of Table 5.10), part of speech was not a significant predictor when the ten high frequency words were excluded ($\chi^2 = 3.6$, $p(\chi^2) = 0.46$). With respect to the neighborhood metrics, minimal pair status was also not a significant predictor ($\chi^2 = 1.4$, $p(\chi^2) = 0.24$), but the

predictor	β	t	p**MCMC**	**AIC**	**loglik**	χ^2	$p(\chi^2)$
intercept	-2.644	-17.43	0.0001	3187.9	-1590.0		
rime dur	0.167	5.06	0.0001	3158.3	-1574.2	31.6	0.0001
onset consonant				3141.0	-1563.5	21.3	0.0001
onset = /p/	-0.211	-3.96	0.0001				
onset = /k/	0.036	0.71	0.2990				
speech rate before	-0.065	-1.87	0.0508	3138.5	-1561.3	4.5	0.0347
speech rate after	-0.075	-2.01	0.0350	3136.2	-1559.1	4.3	0.0373
previous cond prob	-0.016	-2.60	0.0070	3131.2	-1555.6	7.0	0.0081
first mention = y	0.041	1.79	0.0762	3130.0	-1554.0	3.2	0.0730
residualized rime nbors	0.011	2.12	0.0248	3127.3	-1551.7	4.6	0.0313
residualized sum lex freq	0.048	3.26	0.0004	3121.3	-1548.6	10.7	0.0011
residualized CC nbors	0.014	1.96	0.0302	3128.0	-1551.5	3.98	0.0461
residualized CV nbors	-0.017	-2.20	0.0114	3127.0	-1551.5	4.96	0.0260

Table 5.10: Significant predictors of voice onset time in Experiment 3, using the reduced dataset. This model includes random intercepts for participant and word.

residualized number of CC neighbors and CV neighbors were significant predictors, with effects in opposite directions. Words with more CC neighbors (neighbors contrasting in the vowel) were produced with significantly longer VOT than words with fewer CC neighbors. Words with more CV neighbors (neighbors contrasting only in one of their coda consonants) were produced with shorter VOT than words with fewer CV neighbors. The effects of rime density and (log transformed) summed lexical frequency of neighbors are consistent with the analysis of the full dataset; both predictors were associated with significantly longer VOT in the reduced dataset as well.

It should be noted that while these neighborhood metrics resulted in significant improvements in model log likelihood, none of them had a noticeable effect on the proportion of variance accounted for by the model. For example, the squared correlation between fitted and observed values for the fitted model of VOT, using the reduced dataset, was 0.235; using the reduced dataset and including residualized rime density as a predictor resulted in a squared correlation of 0.234.

Summary of results

Analyses of all three dependent variables (total word duration, rime duration, and voice onset time) initially indicated that minimal pair status – whether a word has a minimal pair based on the voicing of the initial consonant – was a significant predictor of articulatory duration. However, when the five highest frequency words in each minimal pair category were excluded from the analyses, minimal pair status was no longer significantly associated with articulatory duration, for any of the three dependent measures. This suggests that the durations of the highest frequency words were affected by some other factor or factors correlated with minimal pair status, but that minimal pair status itself was not significantly

	high ND	low ND	high LF Sum	low LF Sum	high rime dens	low rime dens	minpair = y	minpair = n
VOT	59.2	56.9	59.1	57.1	59.1	56.9	58.1	58.3
rime	154.6	178.2	153.3	177.8	153.5	179.8	159.3	189.5
word	268.3	291.7	265.8	292.5	266.6	294.0	272.9	303.5

Table 5.11: Summary of average duration measures for words in the Buckeye Corpus.

associated with durational differences in the rest of the dataset.

No neighborhood-related metrics were found to be significant predictors of total word duration, or of rime duration. Several neighborhood metrics were significant predictors of voice onset time, however; words with more rime neighbors (neighbors differing only in their onset consonant) and words whose neighbors had higher summed lexical frequency were produced with significantly longer VOT than would otherwise be expected, when all other factors are held constant. Additionally, the analysis of the reduced dataset (excluding the ten high frequency words) returned significant effects of CC neighborhood density and CV neighborhood density. Words with many neighbors contrasting in the vowel were produced with longer VOT, while words with many neighbors contrasting in the coda were produced with shorter VOT.

Table 5.11 gives the average VOT, rime duration, and total word duration as a function of a median split on several of the neighborhood-related variables found to be significant in the analysis of VOT in the reduced dataset. The effect size of neighborhood density on VOT is similar across neighborhood metrics. Words with more than the median number of neighbors were produced with a VOT approximately 2-3 ms longer than words with less than the median number of neighbors, and the effects of summed lexical frequency and rime density are comparable. This small effect size is consistent with the results of Baese-Berk and Goldrick (2009), and with the analysis presented in Chapter 4, which indicated that the VOT for words with more than the median number of rime neighbors was also approximately 3 ms greater.

While no neighborhood metrics were found to be significant predictors of rime duration or total word duration in the present analysis, the difference in the averages is in the same direction as in Chapter 4; on average, words with more neighbors were produced with shorter rimes.

In addition to the neighborhood-related findings, which addressed the primary research questions that were posed, an unanticipated finding was that the effects of conditional probability were quite local in their scope. Conditional probability given the previous word was associated with VOT shortening, but not with rime shortening, while conditional probability given the following word was associated with rime shortening, but not with VOT shortening. This finding is in fact perfectly consistent with the idea that effects of contextual predictability on articulatory duration occur at a representational level below the word, at the level of phonological encoding, a hypothesis which will be fully elaborated below.

5.5 Discussion

As summarized in Chapter 2 and in the introduction to the present chapter, the motivation behind the present investigation was to better understand how feedback processes affect articulatory duration. The analysis explored the extent to which different types of phonological neighbors affect the durational properties of words produced in natural, spontaneous, connected speech. Perhaps surprisingly, however, none of the neighborhood metrics were found to be significant predictors of total word duration or of rime duration.

This would seem to be inconsistent with Gahl et al. (2012)'s recent finding – also examining the Buckeye corpus – that words with more phonological neighbors were produced with shorter durations in conversational speech, all else being equal. However, it should be noted that the inclusion criteria for words examined in that study were at once stricter in some respects and more lax in others, as compared to the present inclusion criteria. Gahl et al. (2012) only examined CVC words, and no restrictions were placed on the segmental makeup of word types. Recall that the present study examined only monosyllabic words beginning with /p t k/, but that no restrictions were placed on syllable shape (that is, words with a wide variety of complex codas were examined). Gahl et al. (2012)'s dataset was therefore much larger than the current dataset (534 word types and 12,414 tokens, versus only 159 word types and 4,151 tokens in the present analysis).

Differences in both size of the dataset and in the efficacy of the control parameters in capturing variation due to phonological factors can likely explain the fact that phonological neighborhood density was not significantly associated with total word duration or rime duration in the present study. As compared to Gahl et al. (2012), the present dataset may have suffered from a relative lack of statistical power and statistical control, resulting in the lack of a significant association between neighborhood density and total word duration.

While including a wide variety of syllable shapes may have made the analyses of word and rime duration more difficult, using the observed rime duration as a statistical control in the analysis of VOT appears to have been quite effective. Consistent with previous work indicating that changes in syllable-initial VOT are proportionally commensurate with changes in the rest of the syllable (Boucher, 2002; Smiljanić & Bradlow, 2008a, 2008b), observed rime duration was highly significant as a predictor of VOT, with longer rime duration being strongly associated with longer VOT. Contrary to previous studies, however, several additional factors were found to improve the model's predictive power above and beyond the association between rime duration and VOT: speech rate, conditional probability given the previous word, first mention, and several neighborhood metrics (as well as the identity of the onset consonant) all increased model log likelihood significantly.

The least surprising of these factors is the onset consonant, since the size of the cavity behind the occlusion location is known to affect the time needed for the vocal folds to resume vibration (Cho & Ladefoged, 1999). The finding that increased speech rate was associated with VOT shortening even when rime duration is taken into account suggests that contrary to previous work, VOT does appear to shorten (slightly) disproportionately more than the rest of the word as a function of local speech rate. This difference between

the present study and previous work may be related to the fact that the present study is the first to compare VOT and rime duration in truly conversational speech. Previous studies have typically focused on isolated words and/or read speech, and have obtained differences in speaking rate and style by explicitly instructing participants to adjust these parameters. It may be that the differences in rate and style obtained through conscious manipulation of read speech are not as great as, or are qualitatively different from, conversational speech, such that VOT increases in perfect proportion to rime duration in read speech but not in spontaneous connected speech.

Based on previous work, it is also to be expected that greater contextual predictability would be associated with shorter articulatory durations. Many studies have found that more probable words are produced with shorter durations, all else being equal (Jurafsky et al., 2001; Bell et al., 2003, 2009). However, previous studies have examined total word duration, while the unique aspect of the present study is that it examines the durational properties of units smaller than the word. In doing so, the present results provide evidence that differences in articulatory duration due to contextual predictability may be primarily due to processing at the segmental (rather than lexical) level; examining VOT and rime duration separately revealed that greater previous contextual probability was associated with shorter VOT (but not rime duration), while greater following contextual probability was associated with shorter rime duration (but not VOT).

These findings, in combination with the significant association between neighborhood structure and VOT, motivate a view of phonological encoding that is grounded in interactive models of language production, and that assumes sequential (left to right) encoding of phonological segments (Sevald & Dell, 1994). The following section elaborates on the mechanisms and assumptions needed to relate such a model to the findings in the present study. A discussion of the implications of the proposed model for previous findings is postponed until the General Discussion in Chapter 6.

The "Articulate As Soon As Possible Principle"

Two major assumptions are necessary in order to relate Sevald and Dell (1994)'s sequential cuing model to the present results. In the discussion that follows, the combination of these assumptions will be referred to as the "Articulate As Soon As Possible Principle", or the AASAPP. They are as follows:

1. The articulatory plan for a given segment is initiated and executed as quickly as possible.

2. The time course for the initiation and execution of an articulatory plan is directly related to the activation level of the target segment at the time of selection.

Each of these assumptions will be elaborated in turn before the link between the proposed principle and the present findings is made explicit.

The first assumption – that the articulatory plan for a given segment is initiated and executed as quickly as possible – obviously carries with it the implication that the psychological representation for each segment is associated with an articulatory plan. It should be noted that at present, this is not intended as a strong claim that the segment is the only level of phonological representation relevant during speech production. The current model does not necessarily exclude the existence of, or reliance on, syllable-sized units, either in English or in any other language. What is of primary importance is the relation between (sequential) activation and speed of articulation of units smaller than the word.

The second assumption – that speed and latency of articulation are directly related to the activation level of the target segment – requires further specification in two respects. First, it is assumed that the articulation of a given segment is initiated as soon as the segment reaches some threshold level of activation. This implies that in general, the articulation of words with many phonological neighbors will be initiated more quickly, because feedback from phonologically related words will quickly increase the activation of the target segments beyond the threshold level. Second, it is assumed that the *relative* activation level of a target segment determines the speed with which its articulatory plan can be executed. In general, then, the segments of words with many phonological neighbors will be articulated more quickly, because they will receive more "support" from overlapping phonological representations.

However, as will be made clear below, if a word has a disproportionate number of neighbors that do *not* overlap in a given syllable position, then the activation of phonological neighbors can lead to positional competition between segments. Thus for a word with many neighbors differing in their onset consonants, feedback from the neighbors leads to higher activation of the target lexical representation, but also relatively greater competition for the onset slot. In this case, the AASAPP predicts the durations of the rime consonants to be relatively short, but the duration of the onset consonant and the initiation of its articulation to be relatively long, due to positional competition.

The second assumption can usefully be operationalized by linking the relative activation level of a segment to the time required for phonological encoding. If a segment can only be selected for encoding once its activation level reaches some relative threshold above all competitor segments, then target segments with many competitors will take longer to encode. Longer encoding time may then be reflected in the time it takes to initiate and execute the articulatory gestures associated with the target, perhaps in part because the system is slower to proceed to the encoding of subsequent segments, effectively lengthening the duration of difficult to encode segments. When encoding time is very short, as when there is little competition between segments for a given position in the word, the target segment can be planned and executed, and the encoding for the following segment can be initiated, all in rapid succession.

Together, these assumptions entail that the duration of a given segment will be primarily determined by the ease with which the target itself is encoded, but also by the ease with which the immediately following segment can be encoded. A highly active segment followed by another highly active segment will have a particularly short duration, since the articulation of the target segment will be fast and the immediately following segment will be initiated

quickly. A highly active segment followed by a difficult to encode segment will have relatively longer duration, since the initiation of the following segment will be relatively later. A difficult to encode segment followed by an easy to encode segment will be even longer, if segmental activation primarily affects articulatory speed. And the longest segment will be one that is itself difficult to encode and is followed by another segment that is difficult to encode.

The AASAPP and the present findings

Recall that in the present study, rime duration was found to be shorter when conditional probability with the following word was high. If the following word is highly probable in a given context, then during the articulation of the current word, lexical activation for the following word will be relatively higher than that of other candidate words. Higher activation of the following word will extend to higher activation of the following word's segments; then following the AASAPP, the upcoming phonological segments will be planned and initiated more quickly, effectively shortening the articulatory duration of the current word's rime.

Voice onset time was not affected by conditional probability given the following word, but it was affected by probability given the previous word, with more probable previous contexts being associated with shorter VOT. This finding can be accounted for by the assumption that all else being equal, the duration of a given segment is directly related to the ease with which that segment was encoded. The explanation then follows the logic outlined above: if the current word is highly probable given the previous word, then the lexical activation of the current word is relatively higher than it would be in other contexts, resulting in higher activation of the segments associated with the current word, allowing them to be planned and executed more quickly, thereby causing the duration of the current segments to be relatively shorter in the current context.

This set of findings was somewhat unexpected, but is in fact perfectly consistent with the account proposed here. Assuming 1) a spreading activation model of speech production, 2) cascading activation flow from the lexical to the phonological level, 3) left to right encoding of segments, and 4) that the relative activation level of a segment is directly related to the speed with which it is initiated and executed (the AASAPP), it makes perfect sense that segments abutting more probable words would be shorter.

What is perhaps surprising is that the shortening in articulatory duration did not extend beyond the one or two segments immediately abutting more probable transitions. That it did not, however, should be taken as evidence that such durational shortening is due to increased availability of units at the phonological level, and not at the lexical level, *per se*. (Of course, this account assumes that a relative increase in a phonological unit's activation level can originate at the lexical level, so the two are obviously intimately related.) The proposal is that articulatory durations are shortened primarily for segments immediately abutting more probable words because phonological encoding is incremental; if conditional probability with the previous word is high, but probability given the following word is low, then segments at the end of the target word *will* receive some benefit from the previous

word, but they will *also* undergo some lengthening due to the relatively slow initiation of the following word. The result will be that segments occurring at word boundaries are most likely to show effects of conditional probability given the adjacent word.

Another important set of findings emerging from the present analysis concerned the significant associations between VOT and several different neighborhood metrics. The number of rime neighbors (neighbors differing only in their onset consonant), number of CC neighbors (neighbors differing only in their vowel), number of CV neighbors (neighbors differing only in one of their coda consonants), and summed lexical frequency of all phonological neighbors were all significant predictors of VOT. Higher values for summed lexical frequency, rime neighbors, and CC neighbors were associated with longer VOT, while higher values for the number of CV neighbors were associated with shorter VOT. These findings will be discussed in turn.

First, let us maintain the assumptions of the AASAPP: that articulatory gestures are initiated and executed as quickly as possible, and that the time course of an articulatory plan depends on the relative activation level of the target segments.

Focusing first on rime density, a greater number of rime neighbors implies a greater degree of competition for the onset slot of the target word. As phonological neighbors differing only in their onset consonant become active, the activation of the target onset consonant is relatively less active as compared to the set of possible competitors. This competition for the onset position causes the encoding of the word initial segment to be relatively more difficult. Following the current logic, if relative ease of encoding the current segment is related to faster execution of articulatory gestures, then relative difficulty of encoding is obviously associated with slower execution of articulatory gestures. In this case, feedback from a set of phonological neighbors with discrepant initial consonants leads to longer duration in the articulation of the word initial consonant.

Turning now to the effect of CC neighbors on voice onset time, words with many neighbors differing only in their vowel were also produced with significantly longer VOT. This finding follows from the assumption that a given segment's duration is also affected by the ease with which the following segment can be encoded, the idea being that if the following segment is difficult to encode, its articulation will be initiated later, effectively lengthening the current segment. For words with many neighbors differing only in their vowel, then, the encoding of the vowel will be relatively more difficult, since the activation of the vowel segment as compared to its many competitors will be relatively lower. And if the articulation of the vowel segment is initiated later, then the articulation of the preceding segment – in this case, the abduction gesture associated with VOT – will take longer, resulting in longer VOT.

The explanation of the effect of CV neighbors on voice onset time follows exactly the same logic. Again assuming that a given segment's duration is affected by the ease with which it is encoded, as well as the ease with which the following segment is encoded, then words with many neighbors overlapping in their initial two segments should be produced with shorter VOT; feedback from the many CV neighbors makes both segments comprising the initial CV sequence easy to encode, resulting in especially short duration for the articulatory gestures associated with the initial consonant. This is exactly what we find: having a greater number

of CV neighbors was associated with shorter VOT.

Finally, the summed lexical frequency of a word's phonological neighbors was associated with longer VOT. A possible explanation for this finding is that of all the neighborhood metrics, summed lexical frequency is most highly correlated with rime density (Pearson's product moment correlation of 0.69, vs. 0.57, -0.06, and 0.25 for total number of neighbors, CC neighbors, and CV neighbors, respectively). The explanation for why greater summed lexical frequency would lead to longer VOT is therefore exactly the same as the explanation of the effect of rime density: because most phonological neighbors in English differ in their onset consonants, the activation of a target's neighbors causes competition for the onset position, making the encoding of the initial consonant more difficult, which is hypothesized to lead to longer articulation time.

The fact that summed lexical frequency resulted in the greatest amount of model improvement of all the neighborhood metrics (see Table 5.10 for reference) supports the idea that the amount of activation in the system during production planning carries explanatory power for understanding the speed of phonological encoding and – by extension – articulatory duration. Summed lexical frequency is a measure that attempts to quantify the magnitude of potential activation in a given lexical neighborhood, by taking into account both the number of a word's neighbors and the frequency of those neighbors. In this case, if higher lexical frequency is associated with higher resting activation, and if the activation levels of neighbors determine the strength with which their segments can compete for a given phonological position in the target word, then because most neighbors in English differ in their onset consonant, it follows that higher summed lexical frequency would be associated with longer VOT. Higher summed lexical frequency means a greater magnitude of competition for the onset position, and therefore results in increased difficulty in encoding the initial segment.

5.6 Conclusion

In this chapter, voice onset time and rime duration in spontaneous, conversational speech were each examined separately as a function of several different neighborhood metrics. None of the neighborhood metrics were significant predictors of rime duration, but several metrics were significant in the analysis of voice onset time. It was suggested that the lack of significant neighborhood effects in the rime duration analysis was due to a lack of statistical power, since there were relatively few tokens in the dataset as compared to previous analyses, and because durational variation due to the heterogeneity of the coda consonants may not have been sufficiently controlled for.

However, several of the neighborhood metrics were significant predictors of VOT: summed lexical frequency of phonological neighbors, number of rime neighbors, number of CC neighbors, and number of CV neighbors were all reliably associated with VOT. Also of note was the finding that higher conditional probability with the previous word was significantly associated with shorter VOT, while higher conditional probability with the following word was significantly associated with shorter rime duration.

These findings formed the basis of the "Articulate As Soon As Possible" Principle, which hypothesizes that the speed and latency of articulatory gestures is directly related to the ease with which phonological segments can be encoded. The present findings were then interpreted in the context of this hypothesis. In the General Discussion, the AASAPP will be further discussed with respect to a wider range of findings in the literature.

Chapter 6

General Discussion

The main question set forth in the Background to the dissertation was whether the specific phonological relationships between a target word and its neighbors are predictive of articulatory duration. Specifically, it was hypothesized that positional phonological overlap between neighbors – for example, whether a target has more neighbors overlapping in its initial consonant versus in its vowel – may be associated with differences in the duration of the target segments. If positional overlap is relevant for the duration of particular segments, then such durational differences should be attributed primarily to processes occurring at the phonological level. Such a finding would suggest that the speed articulation can be tied to the ease or speed of encoding a given segment.

Three experiments were undertaken to better understand the relationship between phonologically overlapping neighbors and phonetic duration. Experiment 1 asked whether learning novel words that entered into different types of minimal pair relationships with already known words would affect children's pronunciations of either the novel or known words. Experiment 2 further explored Baese-Berk and Goldrick's (2009) finding that words with a minimal pair based on the voicing of the first consonant were produced with longer VOT. A reanalysis of their data was undertaken to determine whether the more general neighborhood characteristics of the target words affected their pronunciation, and whether there was any evidence that aspects of the word other than word-initial VOT – in this case, rime duration and total word duration – were affected. Finally, Experiment 3 extended the analysis of VOT, rime duration, and word duration to voiceless stop words in the Buckeye Corpus (Pitt et al., 2007), asking whether neighborhood structure was predictive of sublexical durational patterns in spontaneous, conversational speech.

The structure of the chapter is as follows. First, the results of each of the three experiments will be summarized in turn. Following the summary of the most important findings, the "Articulate As Soon As Possible Principle", which was introduced to account for the findings in Experiments 2 and 3 of the dissertation, will be described. The chapter concludes by relating the proposed AASAP Principle to a selection of relevant findings in the literature.

6.1 Summary of main findings

Results of Experiment 1: phonetic duration in word learning

In Experiment 1, preschool-aged children were taught novel words that differed from known words by a one segment substitution in the onset consonant (*tog*, vs. the known word *dog*), or in the syllable nucleus (*keet*, vs. the known words *cat*, *coat*, and *feet*). The prediction was that if the particular phonological relationship between a target word and its neighbors was relevant for understanding feedback processes during phonological encoding, then the two novel words (and their already known minimal pair words) would pattern differently as they were incorporated into the lexicon.

There was no evidence suggesting that the particular phonological relationships between the novel words and the already known words were relevant for the durational changes that occurred over the course of the experiment. However, several durational differences were of note. For known words, total word duration increased significantly over the course of the experiment (from Day 1 to Day 2, and marginally again on Day 3), and rime duration and VOT increased approximately in proportion to total word duration, irrespective of minimal pair relationship. For example, the VOT for *cat* and *coat* (which were minimal pairs with the novel word *keet*) increased to the same extent as that of *toast*, which was not a minimal pair word in this experiment.

A post-hoc analysis indicated that three aspects of children's pitch also increased over the course of the experiment: the mean pitch, peak pitch, and change in pitch occurring on a single word (the latter being an index of pitch movement) were all significantly greater on Day 3 than on Day 1. Moreover, pitch was found to be weakly but significantly correlated with total word duration, suggesting that at least some of the across-the-board increase in total word duration could be attributed to contrastive focus on Days 2 and 3 of the experiment.

Interestingly, while the VOT for all known words increased proportionally to total word duration over the three days of the experiment, the VOT for the novel words actually showed a slight *decrease* from Day 1 to Day 3. This was argued to be consistent with the finding that on Day 1, the analysis of 20 baseline words (all produced before the novel words were introduced) indicated that VOT was related to each word's phonotactic probability; baseline words with more common phonotactic patterns were produced with significantly shorter VOT.

In general, then, low frequency consonant articulations in this experiment were produced with significantly longer articulatory duration than more familiar patterns. It was therefore argued that the coordination of articulatory gestures for children this age (approximately 4;3 in this experiment) was subject to a practice effect, which was evident in the production of both the known words and the novel words.

It remains unclear whether the observed practice effect should be located at the level of phonological encoding, or at the level of motor representations, since in this case the data do not allow us to distinguish between the two possibilities. That longer duration would

be associated with less practiced patterns is consistent with ideas in motor planning more generally; there is widespread agreement that the duration of highly practiced movements tends to reduce over time, as the motor system learns the optimal control parameters and becomes more efficient (Haith & Krakauer, 2013). However, it is equally possible that the efficiency or speed of phonological encoding also increases with practice, making it difficult to say whether or to what extent children's articulations in this study were affected by factors related to low-level motor control versus more abstract phonological planning.

The results of Experiments 2 and 3, however, were argued to be located at the level of phonological (rather than motor) planning; there was no evidence that the adult productions examined in those studies were sensitive to phonotactic pattern frequency, but they *were* subject to differences in the neighborhood structure of the target words. This suggests that feedback processes and the relative activation levels of the target segments – but not effects of motor practice – are relevant to the interpretation of these results.

Results of Experiment 2: phonetic duration in single word productions

In Experiment 2, the relationship between voicing-initial minimal pair neighbors and neighborhood density more generally was explored. In a reanalysis of Baese-Berk and Goldrick's (2009) data on voice onset time in minimal pair words, a set of regression models sought to determine whether minimal pair status, total neighborhood density, or positionally defined neighborhood density provided the best account of the differences in VOT observed in that study. The reanalysis additionally addressed the question of whether minimal pair status or neighborhood structure more generally was related to durational aspects of the target words other than word-initial VOT; as in Experiments 1 and 3, rime duration and total word duration were also examined.

Two primary hypotheses were pursued. The first hypothesis was that the longer VOT for words with voicing-initial minimal pair neighbors (e.g. *cod*, which has a neighbor *god*) was reflective of a general expansion in duration that affected all segments in the word. This hypothesis followed the logic of Baese-Berk and Goldrick's original interpretation; if words with minimal pair neighbors receive more feedback from related phonological representations, and if higher lexical activation is generally associated with hyperarticulation, then all aspects of a highly active word's pronunciation should be hyperarticulated. This predicts that rime duration and total duration should also be lengthened as compared to words without voicing initial minimal pairs.

This account also leads directly into the second hypothesis, which was that if durational differences are the result of increased lexical activation due to feedback from phonological neighbors, then the total number of a word's phonological neighbors could be a better predictor of any differences in VOT, rime duration, or word duration. Under this account, minimal pair status is important only inasmuch as it is reflective of the total amount of potential support from phonologically related neighbors (and it was observed that the minimal

pair words in the study had significantly more total neighbors than the non minimal pair words).

In sum, if a relationship between increased activation at the lexical level and global hyperarticulation was responsible for the VOT effect, then general neighborhood density should be a better predictor than minimal pair status *per se*, and aspects of the word other than the VOT should also be hyperarticulated. The alternative is that if the VOT effect was primarily due to processes occurring at the phonological (rather than lexical) level, then the particular types of phonological relationships between the target and its neighbors could be important for understanding durational differences. In this scenario, it would be possible for a voicing-initial minimal pair neighbor, or perhaps the total number of neighbors contrasting in their initial segment, to affect the production of only the initial segment of the target word.

The most challenging aspect of analyzing rime duration and total word duration is, of course, that each phonological segment has its own intrinsic duration, such that the particular phonological makeup of a word has an important effect on its total duration. To control for effects of phonological makeup, grand mean segment durations were calculated over the entire Buckeye Corpus, and a baseline duration for each word was included as a control parameter in all statistical models. Since phonological processes (such as vowel lengthening before voiced consonants, for example) also affect baseline durations, a measure of syllable type was also included as a control.

Once baseline durations were taken into account, neighborhood structure was found to have significant effects on the duration of sublexical units. In particular, VOT was significantly associated with the number of neighbors contrasting in their initial segment; words with more "VC neighbors" were produced with significantly longer VOT, all else being equal. Rime duration was significantly associated was total neighborhood density, and marginally associated with the number of "CC neighbors" (neighbors contrasting only in the vowel). Words with more total neighbors were produced with shorter rimes, and the effect of CC neighbors was additive, such that words with more CC neighbors had even shorter rimes, on average.

The interpretation of these results was that positional segmental overlap – and by extension, the activation level of particular phonological segments – is important for understanding the relationship between phonological encoding and articulatory duration. It was hypothesized that when neighbors "agree" on a given segment, that segment can be encoded and produced more quickly. Conversely, when neighbors disagree in the initial segment of a word, that segment is more difficult to encode, and this difficulty is associated with longer duration. To explain the tendency for words with many CC neighbors to be shorter, it was also stipulated that easily encoded segments may be initiated more quickly; this would result in shorter vowel durations (and therefore shorter rime durations) for words with many neighbors overlapping in their final consonant.

The finding that positional phonological overlap was significantly associated with sublexical articulatory duration suggests that the ease of phonological encoding is directly related to segmental duration. If this is true – that small fluctuations in segmental duration fall out naturally from the encoding process – then effects of phonological encoding should also be

evident in conversational speech, a prediction that was directly tested in Experiment 3.

Results of Experiment 3: phonetic duration in conversational speech

Experiment 3 undertook an analysis similar to the previous experiments, asking whether VOT, rime duration, or total word duration were subject to feedback from positionally defined phonological neighbors, this time in spontaneous, conversational speech. The strategy and findings were similar to that of Experiment 2; once baseline durations and factors such as local speech rate and contextual probability were taken into account, the neighborhood structure of CVC words had a significant influence on word-initial VOT.

Several neighborhood metrics were in fact relevant; words with more VC neighbors were again produced with relatively longer VOT, as were words with more CC neighbors and higher summed lexical frequency. CV neighbors, on the other hand, were associated with significantly shorter VOT. The findings with respect to CC and CV neighbors were argued to result from the same principles set forth in the interpretation of Experiment 2; namely, that positional phonological overlap between co-activated words affects the ease of phonological encoding, and that ease of encoding is related to both the speed and latency of articulatory gestures.

For words with many CV neighbors, the initial consonant and vowel representations of the target word both receive support from other words in the lexicon. Phonological encoding of the initial CV is therefore relatively easy and proceeds quickly, with the result that the articulatory gestures associated with the initial consonant can be produced particularly quickly. For words with many CC neighbors, however, the initiation of the vowel is relatively slowed due to competition from non-overlapping phonological segments. This positional segmental competition is hypothesized to interfere with the encoding of the vowel, resulting in longer latency for the vowel onset, and consequently longer VOT for the preceding segment.

Contrary to the findings for Experiment 2, no significant associations between neighborhood density and rime duration or total word duration were found for words in conversational speech. It was suggested that the relatively smaller size of the dataset, and perhaps a lack of adequate control variables, was responsible for the lack of association, since recent evidence suggests that greater neighborhood density is associated with shorter total word duration in conversational speech (Gahl et al., 2012).

A somewhat unexpected finding was that the effect of conditional probability on phonetic duration was quite local in scope; greater conditional probability with the preceding word was associated with significantly shorter VOT but not with shorter rime duration for the target word, while greater probability given the following word was associated with significantly shorter rime duration but not VOT. This suggests that the durational shortening associated with greater lexical accessibility can be primarily attributed to processes occurring at the phonological level, since different segments within the target word were affected differently.

To account for these findings, Chapter 5 introduced the Articulate As Soon As Possible

Principle, which states that the articulatory gestures associated with a given phonological segment are initiated and executed as quickly as possible, as soon as the target segment can be encoded for production. In the next section, the AASAPP is explained in more detail, and its relation to previous findings is also considered.

6.2 The Articulate As Soon As Possible Principle

The Articulate As Soon As Possible Principle was proposed as a way to tie together the findings for articulatory duration in both single word productions (Experiment 2) and conversational speech (Experiment 3). It was argued that positional segmental competition caused by the co-activation of phonologically related words was relevant for understanding the duration of individual segments, and that two main assumptions were needed to capture the data from these experiments. The two assumptions are as follows:

1. The articulatory plan for a given segment is initiated and executed as quickly as possible.

2. The time course for the initiation and execution of an articulatory plan is directly related to the activation level of the target segment at the time of selection.

The idea that articulatory movements are executed as quickly as possible is consistent with ideas in motor planning more generally. With respect to the control signals needed for the efficient planning of arm movements and eye saccades, Harris and Wolpert (1998, described in more detail in the Background) offer the following account of the tradeoff between speed and positional accuracy: "the temporal profile of the neural command is selected... to minimize the movement duration for a specified final positional variance determined by the task." In other words, movements are executed as quickly as possible while still maximizing positional accuracy.

With respect to the latency of movement initiation, Rosenbaum (1980, cited in Roelofs, 1999) showed that participants were slower to initiate arm movements when the number of possible movement trajectories was increased. This, too, can potentially be seen as consistent with the AASAPP; if the activation of competing segments is seen as increasing the number of possible movement trajectories, then it follows that greater positional segmental competition would be associated with longer articulatory latency for the target segment.

In sum, the AASAPP captures the relationship between ease of production planning and articulatory duration that was observed in the dissertation, and it does so by situating such effects at the level of phonological encoding. It should be noted that effects of phonological encoding were only unambiguously observed in Experiments 2 and 3, since the longer duration of articulatory movements in children's productions in Experiment 1 may have been due in part to relative difficulty with both phonological encoding and motor planning. In the adults' productions examined here, however, difficulty with motor planning was unlikely to be a significant factor, since phonotactic probability was never significantly related to the

duration of adults' articulations. Rather, the relative accessibility of individual segments due to feedback and competition processes was argued to be most relevant.

In conjunction with the assumptions laid out in the Background concerning the architecture and functioning of the speech production system – specifically, that the most appropriate model is one that incorporates interactive spreading activation, sequential encoding of segments, and cascading activation flow – the AASAPP can also usefully be extended to findings documented in the literature. In the final section, a selection of these findings is reviewed, with an eye toward situating the present results and the AASAPP with respect to the broader literature.

The AASAPP and results in the literature

Articulatory duration and contextual predictability

Consistent with previous studies examining articulatory duration in conversational speech, the results of Experiment 3 indicated that higher conditional probability was associated with shorter articulatory duration. Additionally, because the present study examined word onsets and rimes separately, it was also determined that the effect of conditional probability was rather local. The VOT for a given word was not significantly affected by conditional probability given the following word, and rime duration was not significantly affected by conditional probability given the preceding word. It was therefore argued that phonological encoding is both sequential and incremental, such that the duration of a given segment is affected by its own accessibility and that of the immediately adjacent segments.

A possible implication of this argument has to do with the finding that the duration of function words and content words is affected by following conditional probability, but only function words are affected by previous probability (Jurafsky et al., 2001; Bell et al., 2009). Since function words are typically shorter than content words (in segmental content and in duration), it is possible that previous results and the present results point toward a smaller "sphere of influence" for previous conditional probability than for following probability. Both sets of findings could be interpreted as indicating that the effects of previous probability extend only one to two segments ahead of the previous word (which would explain why previous probability was a significant predictor for VOT and for the duration of function words), while the effects of following probability seem to extend slightly farther back (potentially explaining why following probability was a significant predictor of rime duration in this study, and of the duration of content words in previous studies).

If this interpretation is correct, then it is consistent with the AASAPP, which entails that the main determinants of a given segment's duration are the ease with which it is encoded as well as the ease with which the immediately following segment is encoded. If a word is highly predictable given the following word, then by definition, both the current word and the following word are predictable by their context. This means that both the current segments and the immediately following segments (those at the beginning of the next word) will be easy to encode, resulting in an additive shortening effect on the current segments.

With respect to previous probability, however, if the current word is highly predictable given the word that precedes it, there is no guarantee that the following word will also be highly predictable. Previous conditional probability will therefore be associated with durational shortening to the extent that the encoding of the current segments is facilitated by higher activation of the current word, but this effect will not necessarily be compounded by the early initiation of articulation for the following word.

Of course, one caveat is the fact that in both the present study and in previous studies reporting effects of conditional probability (Jurafsky et al., 2001; Bell et al., 2003, 2009), local speech rate was used as a control variable. It should be noted that some amount of variation attributable to contextual predictability is likely to be captured by the local speech rate variable, since speech will tend to be faster for more predictable stretches of words. Thus the short span of contextual predictability effects on articulatory duration in these studies should not be interpreted as the absolute extent of phonological planning. Rather, because contextual predictability effects are partially confounded with local speech rate, the temporal and/or segmental extent of predictability effects on phonological planning may be considerably longer than that suggested by the duration measures in any study that uses local speech rate as a control variable.

Articulatory latency and phonological overlap

Many experiments have manipulated the phonological overlap between target words to gain more information about phonological encoding processes. In general, these studies have indicated that both the amount and type of overlap is relevant during encoding. Meyer (1990, 1991), using the form preparation paradigm, showed that pairs of words with initial phonological overlap were produced with significantly shorter latencies than pairs with final or no overlap, and moreover, that the magnitude of the effect was greater when more segments overlapped. Similarly, Damian and Dumay (2007, 2009) also provided evidence that initial segmental overlap is in some sense facilitatory. These studies indicated that production latencies were faster for word pairs with the same initial segment in both a picture-word interference paradigm and a colored picture naming task.

Complicating the picture somewhat, Damian and Dumay (2009) additionally found that trial-to-trial word-initial overlap resulted in slower latencies as compared to when words were non overlapping; the word *goat*, for example, was initiated more slowly following trials where participants named the color *green* than when it followed trials where they named the color *blue*. This suggests that in addition to the position and amount of phonological overlap, the demands of the particular speaking situation must be taken into account. The authors suggested two possible explanations for the facilitation versus inhibition effect. The first was that the prosodic position of the overlapping segments is relevant; from one trial to the next, residual activation of the previously produced word, in combination with Sevald and Dell's sequential cuing effect, could cause concurrently active segments to compete for the same "slots" in the word. When speakers plan a two-word phrase, however, each word is specified for its own position, potentially eliminating competition between segments for the same slot.

The second possible explanation offered by Damian and Dumay concerned the temporal profile of activation processes. This account assumes that phonological activation dissipates rapidly, while lexical activation is longer lasting. If this is the case, then word-initial phonological overlap in two-word phrases could be facilitatory due to the co-activation of the shared segment, whereas phonological overlap from trial to trial could be inhibitory due to the segmental cuing effect.

The results of the present Experiments 2 and 3 are perhaps relevant for adjudicating between these two proposals. The finding that positional overlap between phonological neighbors is facilitatory for production would seem to support the temporal profile account, since neighbors that become activated during the normal course of speech production *are* potentially competing for the same prosodic position within the phrase. If this interpretation is correct, then Damian and Dumay's (2009) findings and the present findings may both indicate that positional overlap between co-activated lexical representations leads to faster phonological encoding for the shared segments.

Finally, studies examining production latencies as a function of neighborhood density have demonstrated that in general, words with more phonological neighbors are articulated with shorter onset latencies (Vitevitch & Sommers, 2003), but words with so-called "dense onsets" (many neighbors differing in their onset consonant) are articulated with longer onset latencies than words with sparse onsets (Vitevitch et al., 2004). These findings are consistent with the results of the present experiments and the proposed AASAP Principle. With respect to total neighborhood density, support from phonological neighbors generally increases activation of the target lexical representation and its associated segments. Following the AASAPP, words with many phonological neighbors should in general be produced with shorter latencies, since their initial segments will typically reach the threshold level of activation sooner than words with relatively fewer phonological neighbors. However, when words are equated for their total number of neighbors but the proportion of neighbors overlapping in the initial consonant differs (as in Vitevitch et al., 2004), competition for the initial segmental position will cause words with "dense onsets" to be produced with longer latencies than words with relatively sparse onsets.

Articulatory duration and phonological overlap

Experiments 2 and 3 are not the first to report differences in articulatory duration due to phonological overlap between co-activated lexical representations. As discussed previously, Gahl et al. (2012) found that words with more phonological neighbors were produced with shorter overall durations (as well as reduced vowels). This is, of course, consistent with the AASAPP; if feedback from phonologically related neighbors is facilitatory for phonological encoding, and if more easily encoded segments are produced with shorter latencies and durations, then words with many neighbors should be produced with shorter overall durations, all else being equal.

A recent study by Goldrick, Vaughn, and Murphy (2013) can also be interpreted as showing a facilitative effect of positional segmental overlap. In a follow-up study to Baese-Berk

and Goldrick (2009), speakers produced words with and without minimal pair neighbors based on the voicing of their final consonant segment. Words with final minimal pair neighbors (e.g. *bud*, which has a minimal pair *but*) were produced with relatively shorter vowel durations than words without such neighbors (e.g. *thud*, which has no neighbor *thut*). Assuming that positional overlap of phonological segments is facilitatory for production, then a possible interpretation of this result is that the overlap between vowels in *bud*/*but* resulted in faster phonological encoding of the target vowel, reducing its duration relative to *thud*, which undergoes no such facilitation. Of course, similarly to Baese-Berk and Goldrick (2009), this account would only seem to hold to the extent that the specific minimal pair relationships investigated in the study are reflective of the targets words' neighborhood structure more generally.

Finally, ongoing modeling work by Buxo-Lugo, Simmons, and Watson (2013, in preparation) is also consistent with the AASAPP and with the results reported in the dissertation. Buxo-Lugo et al. trained simple recurrent networks to produce one of two possible word pairs: initially overlapping pairs that shared the same initial morpheme (e.g. *layover–layout*), and finally overlapping pairs that shared the same final morpheme (*outlay–overlay*). The networks assumed both interactive activation and serial encoding, and the models were trained and tested on just one of the overlapping word pairs. The summed squared error of the models indicated that in general, positional overlap should be facilitatory; shorter durations were thus predicted for the first morpheme of initially overlapping words, with relatively longer durations for the second morpheme, and the reverse pattern was predicted for finally overlapping words.

In a test of the model predictions, human participants were tasked with repeating the initially and finally overlapping pairs as many times as possible in eight seconds, following the same procedure as Sevald and Dell (1994). Human performance in that study (and, it should be noted, in the current studies) was consistent with the modeling results: positional phonological overlap in co-activated lexical representations was found to be facilitative for production.

One difference between Buxo-Lugo et al. (2013) and the present work, of course, is the nature of the speaking situation. It was noted above that repeating alternating pairs of phonologically overlapping words is quite a different task from spontaneously producing full utterances. However, it would seem that both tasks provide data on the effects of feedback from concurrently activated lexical neighbors, and both studies indicate that when such feedback results in increased activation levels for phonologically overlapping segments, phonological encoding is facilitated, and articulatory duration is reduced.

Chapter 7

Conclusion

The goal of the dissertation was to explore the relationship between lexical activation and phonetic duration by examining the effects of phonological neighborhood structure on voice onset time, rime duration, and total word duration in three different speaking contexts. A novel word learning study with children indicated that relative difficulty with phonological and/or motor planning may have overshadowed any effect of lexical activation levels on articulatory duration; in this experiment, less familiar phonotactic patterns were reliably produced with longer VOT, but no effects of minimal pair status on articulatory duration were found.

When adult productions were examined, however, feedback between phonologically related words and the target segmental representations played a significant role. Positional competition between phonological segments – as indexed by the number of lexical neighbors differing by a single segment in a given position of the word – was associated with longer phonetic duration. Words with many neighbors differing in only their onset consonant were produced with significantly longer VOT in both single word productions and in spontaneous, conversational speech.

Other, related effects were also found. In the analysis of single word productions, words with many total phonological neighbors were produced with significantly shorter rime durations. In the analysis of conversational speech, words with many neighbors overlapping in their initial consonant and vowel were produced with significantly shorter VOT. Taken together, these effects motivate an account of phonological encoding in which increased activation for phonological segments is associated with faster articulation in two respects: in the studies presented here, more accessible segments were both executed and initiated more quickly than less accessible segments.

To account for these results, the Articulate As Soon As Possible Principle was proposed. The AASAPP posits that the articulatory plan for a given segment is initiated and executed as quickly as possible, and that the time course for the production plan is related to the activation level of the target segment at the time of selection. When combined with an interactive spreading activation model of speech production that assumes left-to-right encoding of phonological segments and cascading activation flow, the AASAPP accounts for

the results presented in the dissertation.

More broadly, the idea that articulatory plans are initiated and executed as quickly as possible makes the link between lexical accessibility and phonetic reduction more explicit. Relative ease versus difficulty in phonological encoding is certainly not the only source of variation in phonetic duration, but it is *a* source, and this dissertation has argued that it may provide the link between the ease with which words are retrieved from memory, and the speed with which they are produced.

References

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56.

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, *119*, 3048–3058.

Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Baayen, R., Piepenbrock, R., & Van Rijn, H. (1995). The CELEX database. *Nijmegen: Center for Lexical Information, Max Planck Institute for Psycholinguistics, CD-ROM*.

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

Balota, D. A., Boland, J. E., & Shields, L. W. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language*, *28*(1), 14–36.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, *24*(1), 89–106.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, *113*, 1001.

Boersma, P., & Weenink, D. (2012). *Praat, version 5.5.*

Boucher, V. J. (2002). Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates. *Perception & Psychophysics*, *64*(1), 121–130.

Brownell, R. (2000). *Expressive one-word picture vocabulary test*. Academic Therapy Publications.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, *44*(4), 991–997.

Buxo-Lugo, A., Simmons, D., & Watson, D. (2013). *Modeling word duration in language production.* Presented at the CUNY Conference on Sentence Processing.

Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young childrens lexicons. *Journal of Child Language*, *17*(1), 205–215.

Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language*, *22*, 727–735.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207–229.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, *33*(4), 497–505.

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

Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, *57*(2), 195–209.

Damian, M. F., & Dumay, N. (2009). Exploring phonological encoding through repeated segments. *Language and Cognitive Processes*, *24*(5), 685–712.

De Cara, B., & Goswami, U. (2002). Similarity relations among spoken words: The special status of rimes in English. *Behavior Research Methods, Instruments, & Computers*, *34*(3), 416–423.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*(2), 124–142.

Dell, G. S., Martin, N., & Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language*, *56*(4), 490–520.

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language and Hearing Research*, *47*(2), 421.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *26*(5), 489–504.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, *62*(2), 151–167.

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

Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, 748–775.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*(4), 789–806.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132.

Goldinger, S. D., & Summers, W. V. (1989). Lexical neighborhoods in speech production: A first report. *Research on Speech Perception Progress Report*(15), 331–342.

Goldrick, M., Folk, J. R., & Rapp, B. (2010). Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language*, *62*(2), 113–134.

Goldrick, M., Vaughn, C., & Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *Journal of the Acoustical Society of America*, *134*, EL172–EL177.

Grigos, M. I. (2009). Changes in articulator movement variability during phonemic development: a longitudinal study. *Journal of Speech, Language and Hearing Research*, *52*(1), 164–177.

Grigos, M. I., Saxman, J. H., & Gordon, A. M. (2005). Speech motor development during acquisition of the voicing contrast. *Journal of Speech, Language and Hearing Research*, *48*(4), 739–752.

Gupta, P., & Dell, G. S. (1999). The emergence of language from serial order and procedural memory. In B. MacWhinney (Ed.), *The emergence of language* (pp. 447–481). Taylor & Francis.

Haith, A. M., & Krakauer, J. W. (2013). Theoretical models of motor control and motor learning. *Routledge Handbook of Motor Control and Motor Learning*, 7 – 28.

Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*(6695), 780–784.

Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824–843.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (Vol. 21). MIT Press.

Jurafsky, D., Bell, A., & Girand, C. (2002). The role of the lemma in form variation. *Laboratory Phonology*, *7*, 3–34.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Frequency and the Emergence of Linguistic Structure*, *45*, 229.

Katz, J., & Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language*, *87*(4), 771–816.

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech perception. *Journal of Experimental Psychology: General*, *129*(3), 340–360.

Kilanski, K. (2009). *The effects of token frequency and phonological neighborhood density on native and non-native speech production.* Unpublished doctoral dissertation, University of Washington, Seattle.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*, 1208–1221.

Klatt, D. H., & Cooper, W. E. (1975). Perception of segment duration in sentence contexts. In A. Cohen & S. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 69–89). Springer Verlag, Heidelberg.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990.

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive psychology*, *55*(4), 306–353.

Levelt, W., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling* (pp. 403–439). Springer.

Löfqvist, A. (1980). Interarticulator programming in stop production. *Journal of Phonetics*, *8*, 475–490.

Löfqvist, A., & Yoshioka, H. (1980). Laryngeal activity in Swedish obstruent clusters. *The Journal of the Acoustical Society of America*, *68*, 792–801.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, *19*(1), 1–36.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants. *Journal of Child Language*, *7*(01), 41–74.

Marslen-Wilson, W. (1989). Access and integration: Projecting sound onto meaning.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*(1), 1–86.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, *91*(1), 3–19.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, *29*(5), 524–545.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, *30*(1), 69–89.

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language and Hearing Research*, *44*(4), 778.

Munson, B. (2007). Lexical access, lexical representation, and vowel production. *Laboratory Phonology*, *9*, 201–228.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language and Hearing Research*, *47*(5), 1048.

Munson, B., Swenson, C. L., & Manthei, S. C. (2005). Lexical and phonological organization in children: Evidence from repetition tasks. *Journal of Speech, Language and Hearing Research*, *48*(1), 108.

Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech, Language and Hearing Research*, *36*(5), 959.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report*, *10*, 357–376.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory Phonology*, *7*, 101–139.

Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., et al. (2007). *The Buckeye Corpus of Conversational Speech (2nd release).* www.buckeyecorpus.osu.edu, Department of Psychology, Ohio State University (Distributor).

Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes*, *14*(2), 173–200.

Rosenbaum, D. A. (1980). Human movement initiation: Specification of arm, direction, and extent. *Journal of Experimental Psychology: General*, *109*(4), 444–474.

Scarborough, R. (2004). *Coarticulation and the structure of the lexicon.* Unpublished doctoral dissertation, University of California, Los Angeles.

Scarborough, R. (2012). Lexical similarity and speech production: Neighborhoods for nonwords. *Lingua*, *122*(2), 164–176.

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language*, *54*(2), 228–264.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition*, *53*(2), 91–127.

Smiljanić, R., & Bradlow, A. R. (2008a). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics*, *36*(1), 91–113.

Smiljanić, R., & Bradlow, A. R. (2008b). Temporal organization of English clear and conversational speech. *Journal of the Acoustical Society of America*, *124*, 3171–3182.

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass*, *3*(1), 236–264.

Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language*, *29*(02), 251–274.

Storkel, H. L., & Hoover, J. R. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods*, *42*(2), 497–506.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental psychology: Learning, Memory, and Cognition*, *28*(4), 735.

Vitevitch, M. S., Armbruster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 514–529.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408.

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 481–487.

Vitevitch, M. S., & Luce, P. A. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language*, *52*(2), 193–204.

Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition*, *31*(4), 491–504.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from dutch. *Journal of Phonetics*, *32*(2), 251–276.

Wright, R. (2004). Factors of lexical competition in vowel articulation. *Laboratory Phonology*, *6*, 75–87.

Yaniv, I., Meyer, D. E., Gordon, P. C., Huff, C. A., & Sevald, C. A. (1990). Vowel similarity, connectionist models, and syllable structure in motor programming of speech. *Journal of Memory and Language*, *29*(1), 1–26.

Yao, Y. (2007). Closure duration and VOT of word-initial voiceless plosives in English in spontaneous connected speech. *UC Berkeley Phonology Lab Annual Report*, 183–225.

Yao, Y. (2009). Understanding VOT variation in spontaneous speech. In *Proceedings of the 18th International Congress of Linguists (CIL XVIII).*