

Phonological Mismatch Errors:
An Acoustic Analysis

by

Andrea Gormley

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Institute of Cognitive Science

Carleton University

Ottawa, Ontario

© 2010, Andrea Gormley



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-69918-8
Our file *Notre référence*
ISBN: 978-0-494-69918-8

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Abstract

Phonological speech errors are generally said to occur at a level of production prior to the application of phonological rules. This phenomenon, called accommodation, is validated by several transcription studies. The transcription methodology is not the best choice for detecting errors at this level, however, as this type of error can be difficult to perceive. This dissertation presents an acoustic analysis of speech errors that uncovers non-accommodated or mismatch errors. A mismatch error is a sub-phonemic error that results in an incorrect surface phonology. The locus of this type of error is investigated by two further studies that explore whether mismatch errors arise during the processing of phonological rules or if they are made at the motor level of implementation. These studies reveal that mismatch errors occur after phonological rules but before motor level processing. The results of this work have important implications for both experimental and theoretical research. For experimentalists, it validates the tools used for error induction and the acoustic determination of errors free of the perceptual bias. For theorists, this methodology can be used to test the nature of the processes proposed in language production.

Table of Contents

| | |
|---|------|
| Abstract | ii |
| List of Tables | v |
| List of Figures | viii |
| List of Appendices | xi |
| Chapter 1: Background | 1 |
| 1.1 Competence and performance | 3 |
| 1.2 Phonological theories of competence | 6 |
| 1.3 Models of language use | 20 |
| 1.4 Previous speech error studies | 27 |
| Chapter 2: Research Framework | 40 |
| Chapter 3: Experiment 1 | 46 |
| 3.1 Method | 47 |
| 3.2 Results | 49 |
| 3.3 Discussion | 54 |
| Chapter 4: Experiment 2 | 57 |
| 4.1 Method | 57 |
| 4.2 Results | 60 |
| 4.3 Discussion | 71 |
| Chapter 5: Experiment 3 | 74 |
| 5.1 Method | 76 |
| 5.2 Results | 78 |

| | |
|------------------------------------|-----|
| 5.3 Discussion | 84 |
| Chapter 6: Python ACT-R models | 87 |
| 6.1 Model 1 | 87 |
| 6.2 Model 2 | 92 |
| 6.3 Discussion | 96 |
| Chapter 7: Summary and Conclusions | 98 |
| 7.1 Theoretical Implications | 99 |
| 7.2 Limitations and Future Work | 113 |
| 7.3 Concluding Remarks | 115 |
| References | 117 |
| Appendix A | 129 |
| Appendix B | 130 |
| Appendix C | 130 |
| Appendix D | 131 |

List of Tables

| | |
|---|----|
| Table 1. Example derivation for 'writing' and 'riding', correct rule order | 9 |
| Table 2. Example derivation for 'writing' and 'riding', incorrect rule order | 9 |
| Table 3. Example of speech error types from Dell (1986) | 28 |
| Table 4. Percent voicing results for [s] and [z] from Frisch and Wright (2002) | 33 |
| Table 5. Analysis of an error as a syntactic transformation error | 37 |
| Table 6. Steps assumed in speech production | 44 |
| Table 7. Error distribution by type | 53 |
| Table 8. Error classification | 53 |
| Table 9. Mismatch errors by process error | 55 |
| Table 10. Proportion of length mismatch errors for target voiced and voiceless tokens | 55 |
| Table 11. Sets of stimuli used in Experiment 2 | 59 |
| Table 12. Error count for <i>futter</i> and <i>fubber</i> that are more voiceless than expected | 64 |

| | |
|---|----|
| Table 13. Error count for <i>futter</i> and <i>fubber</i> that are longer than expected | 64 |
| Table 14. More voiceless than expected errors broken down by participant | 64 |
| Table 15. Longer than expected errors broken down by participant | 65 |
| Table 16. Error count for <i>futter</i> and <i>fubber</i> that are both longer and more voiceless than expected | 67 |
| Table 17. Total <i>kawdid</i> and <i>kawded</i> errors | 69 |
| Table 18. <i>Kawdid</i> and <i>kawded</i> errors by participant | 70 |
| Table 19. Vowel Duration - Experiment 1 | 79 |
| Table 20. Vowel Duration - Experiment 3 | 79 |
| Table 21. Coda Percent Voiceless by target voicing – Experiment 1 | 80 |
| Table 22. Coda Percent Voiceless by target voicing - Experiment 3 | 80 |
| Table 23. Error types for Experiments 1 and 3 | 82 |
| Table 24. Frequency of duration and voicing errors within the mismatch errors for Experiments 1 and 3 | 82 |
| Table 25. Errors as rule misapplication for Experiments 1 and 3 | 83 |
| Table 26. Example of correct processing | 89 |

| | |
|--|-----|
| Table 27. Example of processing an accommodation error | 90 |
| Table 28. Example of processing a mismatch error | 90 |
| Table 29. Error types from Experiment 1 and Model 1 | 91 |
| Table 30. Example of processing an accommodation error | 93 |
| Table 31. Example of processing a mismatch error | 93 |
| Table 32. Asymmetry of vowel length mismatch errors among Experiment 1 tokens | 95 |
| Table 33. Error rates and probabilities | 102 |
| Table 34. Examples of selection errors at various levels of production | 110 |

List of Figures

| | |
|--|----|
| Figure 1. A derivation from phonemic to allophonic representation | 7 |
| Figure 2. Feature bundle for the phoneme /k/ | 8 |
| Figure 3. An OT tableau showing the selection of the candidate with a nasal vowel (Kager, 1999) | 12 |
| Figure 4. Illustration used for the wug test from Berko (1958) | 16 |
| Figure 5. Model from Indefrey and Levelt (2004) showing the time course and location information for various components of speech production | 18 |
| Figure 6. Model of phonological encoding (from Levelt, 1993) | 23 |
| Figure 7. Dell's spreading activation model (from Dell, 1986) | 25 |
| Figure 8. Two possible processing orders of the target utterance 'sad cat'. Errors occur at the speech planning level | 41 |
| Figure 9. Spectrogram and waveform of a token of <i>tiss</i> showing segmentation | 50 |
| Figure 10. Mean durations for vowels preceding voiceless and voiced codas for a single participant | 51 |
| Figure 11. Mean durations for vowels before voiced and voiceless codas for a single participant showing considerable overlap | 52 |

| | |
|--|-----|
| Figure 12. Spectrogram and waveform of the sentence 'I am a futter now' showing a normal duration (24 ms) for the flap | 63 |
| Figure 13. Spectrogram and waveform of the sentence 'I am a futter now' showing a flap with a duration that is long (47 ms) | 63 |
| Figure 14. Spectrogram and waveform of the sentence 'Last week, I kawded outside' | 68 |
| Figure 15. Spectrogram and waveform of the sentence 'Last week I kawded outside' produced without -ed | 69 |
| Figure 16. Rates for model 2 showing accommodated errors, correct forms and mismatch errors (+SE) from the model and Experiment 1 data | 94 |
| Figure 17. Error rates (+SE) from model 2 for <i>tis</i> produced as <i>tiis</i> and <i>tiiz</i> produced as <i>tiz</i> | 96 |
| Figure 18. Bar graph showing the relative error rates compared with the probabilities for the four vowels | 103 |
| Figure 19. Bar graph showing the voicing error rates for four coda fricatives compared to their probability | 104 |
| Figure 20. Scatter plot showing the error rate and probability for the eight segments in Experiment 1 | 105 |

| | |
|--|-----|
| Figure 21. Error rates for the 'b' and 't' in Experiment 2 compared with their probabilities | 106 |
| Figure 22. Fromkin's (1971) model of speech production | 109 |

List of Appendices

| | |
|------------|-----|
| Appendix A | 129 |
| Appendix B | 130 |
| Appendix C | 130 |
| Appendix D | 131 |

Chapter 1: Background

This dissertation investigates phonological mismatch errors that were uncovered with an acoustic analysis of induced errors. A mismatch error is a type of error that has incorrect surface phonology. Previous studies revealed few mismatch errors, instead classifying the majority of errors as accommodated segmental errors in which the surface phonology adjusts to suit the error (Boomer & Laver, 1968; Fromkin, 1971; MacKay, 1972). This pattern of results may have been due to the transcription technique traditionally used to collect the data. With a transcription methodology the researcher writes down speech errors as they hear them, along with the intended form. This method relies on the researcher's perception of the error which becomes less reliable as the error unit becomes smaller. While an error switching entire words is easy to perceive, for example, '*a laboratory in our own computer*' for intended '*a computer in our own laboratory*' (Fromkin, 1971), errors become more difficult to perceive on smaller units such as segments, for example, '*reek long race*' for intended '*week long race*' (Fromkin, 1971). The type of error needed to test the idea that accommodation at the phonological level always holds would be very difficult to perceive. In order to perceive an error of this type, the researcher would need to hear, for example, that a vowel before a voiced consonant is too short. This difficulty in perceiving phonological mismatch errors is what

motivated the current study to use acoustic analysis to test the claim that errors are typically accommodated.

The mismatch errors discovered in the first experiment were explored further to determine their locus within the production process. The results from the first experiment suggest that some mismatch errors could be the result of the misapplication of phonological rules. Another possibility is that they are the product of the articulation process at the motor level. The third possible source is after phonological rules but before motor-level processing. Two additional studies were completed to determine if mismatch errors arise during the processing of phonological rules, during articulatory implementation at the motor level, or if they arise at a level of speech planning between these two processing levels. Speech planning is a term used by psycholinguists studying speech errors to describe the level at which content merges with structure and where errors are made. For example, in producing the tongue twister 'She sells seashells', the segments /ʃi sɛlz/ are added to the syllable structure; onset + nucleus, onset + nucleus + coda. It is during this merger that the error 'She shells ...' can arise as the /ʃ/ in the initial onset is also assigned to the second onset. Speech planning is assumed to be completed before phonological rules are processed. However, speech planning must also be present after syntactic and morphological processing in order to account for the speech errors of lexical and morphological processing. I will therefore extend the coverage of

speech planning to include a stage after phonological processing but before motor level implementation.

Phonological theorists are concerned with the systematic regularities in the processing of sounds in spoken language (Kenstowicz & Kisseberth, 1979). The study of regularities does not generally include situations that stress the system and produce exceptional results such as speech errors. Although speech errors are cited as evidence for the psychological reality of phonological units, they are not routinely used in testing phonological theories. The instrumental study of errors is relatively new and can be used to address theoretical issues. One goal of this research is to encourage the experimental study of phonological processes. The study of cognitive systems through the errors it produces is a potentially fruitful avenue of research that could be used to gain insight into phonological processes.

This chapter describes the background of phonological theory followed by models of speech production.

1.1 Competence and performance

The competence-performance distinction is a crucial concept in understanding the difference between how psychologists and phonologists approach the study of language. Saussure's concepts of *langue* and *parole* (de Saussure, 1916/

1966) predate the terms competence and performance in linguistic theory. The concept *langue* is the idealized set of signs attributed to a speech community and *parole* is the actual speech act itself. Chomsky's (1965) view of competence differs from the concept of *langue*, however, in that competence is an individual's knowledge of their language, not the speech community's common knowledge. Chomsky's notion of competence separates the speaker-hearer's knowledge of his language from his use of language.

A record of natural speech will show numerous false starts, deviations from rules, changes of plan in mid-course, and so on. The problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance.
(Chomsky, 1965, p.4)

For example, the competence system may allow a speaker-hearer to generate an infinitely long sentence; 'Mary said that John thought that Bill wrote ... ', on and on forever. But the performance system will be unable to remember all the players involved and the sentence will fail to be understood. This is a memory limitation, a limitation of performance but not a limitation of competence.

Both the concept of *langue* and competence abstract away from irregularities in order to determine the properties that underlie the language system. These irregularities are seen as products of the performance system through which language competence is necessarily expressed. Certain aspects of speech attributed to the performance system can safely be ignored when discussing underlying linguistic knowledge. Just as cognitive factors such as memory can limit the expression of competence, there can also be physical limitations. Speaking with a mouth full of peanut butter distorts a speaker's production but does not faithfully reflect their underlying knowledge of the sounds of their language.

While linguistic theorists tend to abstract away from the effects of the performance system, experimentalists tend to study language processing by manipulating the performance system. These studies use the methodologies developed in psychology to study speech production and perception, for example, through measures of reaction time in a picture naming task (Oldfield & Wingfield, 1965) or tracking eye movements to explore the constraints of binding theory (Sturt, 2003) or syntactic dependencies (Sussman & Sedivy, 2003).

The philosophical distinction between competence and performance can help to clarify the difference between a theorist's frame of reference and an experimentalist's. The next sections present models of linguistic competence developed by phonological theorists and models of language use developed by experimentalists.

1.2 Phonological theories of competence

This section presents three major theories of phonological competence: that contained in the *Sound Pattern of English* (Chomsky & Halle, 1968), *Optimality Theory* (Prince & Smolensky, 1993) and probabilistic models (Pierrehumbert, 2003). While the first two theories differ in detail, they both illustrate that the goal of a phonological theory is to capture the transformation from an underlying, phonemic representation to a form that specifies phonetic details. The third incorporates frequency data and makes predictions that are testable by error studies such as the present study. Because some probabilistic models are not generative, a lack of evidence for generative processing would support this view.

The theory in the *Sound Pattern of English* (SPE hereafter) lays out a formal system for characterizing the phonological component. Using sets of ordered rules, an underlying phonemic representation is transformed into its surface, phonetic form. For example, the derivation of the word 'cat' begins with the underlying form consisting of the phonemes /kæt/ and the application of a rule that aspirates initial voiceless stops, yielding the form [k^hæt]. In this derivation, shown in Figure 1, the phoneme /k/ is transformed into the allophone [k^h].

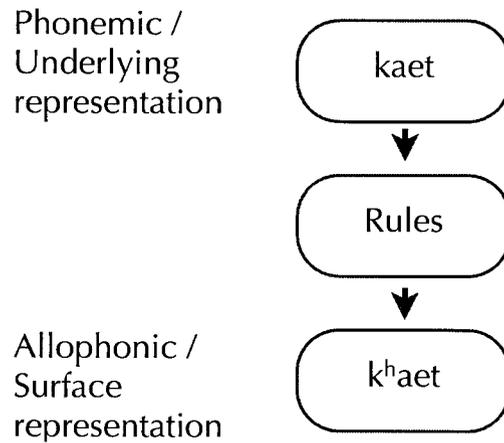


Figure 1. A derivation from phonemic to allophonic representation.

Phonemes like /k/ consist of a bundle of features as shown in Figure 2. In this SPE feature system, features reflect articulatory properties. The binary value they take indicates the presence or absence of this property. The difference between the phonemes /k/ and /g/, for example, is the value for the feature voice where /k/ has a negative value for this feature and /g/ has a positive value.

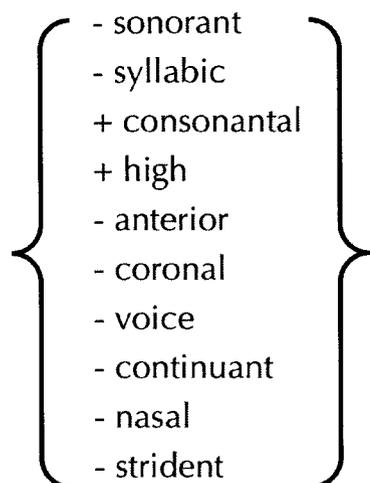


Figure 2. Feature bundle for the phoneme /k/.

The transformation from a representation consisting of phonemes to an allophonic representation may involve several rules. These rules must be ordered as evidenced by cases such as the rider/writer distinction in Canadian English. In this case, there are two rules that must be ordered with respect to each other; Canadian Raising changes the first vowel in the /aɪ/ diphthong to [ʌ] before voiceless segments and a second rule that changes the /d/ and /t/ to a flap [ɾ]. The raising rule must apply before the flapping rule since the raised vowel is maintained in the output as seen in Table 1. If they were ordered the other way, as in Table 2, the flapping rule would destroy the conditions under which the raising rule applies since it turns a voiceless 't' into a voiced flap.

Table 1. Example derivation for 'writing' and 'riding', correct rule order

| | | |
|---------------------------|----------|----------|
| Underlying representation | /raɪtɪŋ/ | /raɪdɪŋ/ |
| 1. Canadian Raising | /rʌɪtɪŋ/ | N/A |
| 2. Flapping | /rʌɪɾɪŋ/ | /raɪɾɪŋ/ |
| Surface representation | [rʌɪɾɪŋ] | [raɪɾɪŋ] |

With the ordering of rules in Table 1, the Canadian Raising rule transforms the diphthong /aɪ/ into [ʌɪ] before the flapping rule applies changing the /t/ into a flap [ɾ], yielding the output [rʌɪɾɪŋ] 'writing'.

Table 2. Example derivation for 'writing' and 'riding', incorrect rule order

| | | |
|---------------------------|----------|----------|
| Underlying representation | /raɪtɪŋ/ | /raɪdɪŋ/ |
| 1. Flapping | /raɪɾɪŋ/ | /raɪɾɪŋ/ |
| 2. Canadian Raising | N/A | N/A |
| Surface representation | [raɪɾɪŋ] | [raɪɾɪŋ] |

In the ordering presented in Table 2, the flapping rule applies first, changing the /t/ into a flap [ɾ] which then destroys the condition under which the Canadian Raising rule operates. The segment following the diphthong is no longer voice-

less and the Canadian Raising rule does not apply. The incorrect form [raɪrɪŋ] is derived.

The model presented in SPE provided a notational framework to capture the various phonological processes in the languages of the world. One shortcoming is the perception that, as a theory, generative phonology was not constrained enough and did not capture certain intuitions that phonologists had about language. One of these intuitions involved the notion of conspiracies in phonological processes (Kisseberth, 1970). For example, the devoicing and deletion of codas (syllable-final consonants) are both common processes in a number of languages. Phonologists felt that these similar types of processes needed to be explained by the theory itself. Thus, Optimality Theory became widely accepted.

Optimality Theory (Prince & Smolensky, 1993; hereafter OT) changed some of the SPE mechanics, for example, changing rules to constraints and taking advantage of parallel processing. The OT process begins with an input, the mechanism called GEN produces a candidate set that is evaluated against the set of constraints, yielding the optimal candidate as the output. This output is the candidate that incurred the fewest violations against the constraint set. While the set of constraints is universal, languages differ as to how the constraints are ranked. Two types of constraints are necessary: faithfulness constraints and markedness constraints. Faithfulness constraints require that the winning candidate does not deviate from the input. Markedness constraints require that the winning candidate is unmarked.

Markedness constraints include universally disfavoured structures as well as language-specific phonotactic restrictions.

Phonotactics refers to the co-occurrence restrictions of sounds in a language. In English, for example, no words begin with an [ŋ] though they occur word-finally in words such as 'sing', [sɪŋ]. This fact about English is captured with the markedness constraint *ŋ-ONSET. While phonotactic restrictions do not need to be derived they are necessary in an OT grammar since GEN produces an unrestricted set of candidates and those that violate the phonotactics of a language must not become the winning candidate.

OT also accounts for allophonic variation with constraints. In English, vowels that precede a nasal consonant are nasalized. This is captured by the constraint *VORALN which penalizes an oral vowel before a nasal consonant. The tableau in Figure 3 shows this process in OT. The markedness constraint that penalizes an oral vowel before a nasal consonant is balanced by the constraint *VNASAL, another markedness constraint that penalizes nasal vowels generally. The faithfulness constraint IDENT-IO(nasal) penalizes a candidate if it differs from the input in terms of nasal features. The ranking of these constraints is crucial for selecting the form [sæ̃nd] as the winning candidate.

| input /sænd/ | *VORALN | *VNASAL | IDENT-IO(nasal) |
|--------------|---------|---------|-----------------|
| a. [sænd] | *! | | |
| b. [sæ̃nd] | | * | * |

Figure 3. An OT tableau showing the selection of the candidate with a nasal vowel (Kager, 1999).

In addition to these mechanical differences, OT formalized the concept of markedness.¹ Universal markedness states that there are certain segments, sequences of segments or positions that are disfavoured by languages and are therefore avoided in the surface representation. Whether a constituent is considered marked is determined through its cross-linguistic and language-specific frequency of occurrence as well as order of acquisition and whether or not it tends to be lost through historical change. Markedness constraints are assumed to be innate and therefore exist in every grammar; languages vary only in how these constraints are ranked. Markedness thus acts as a force that results in languages using various phonological processes to avoid marked structures. For example, because the final element of a syllable, the coda, is marked, German devoicizes segments in the coda position, Japanese allows only certain segments in the coda, while some languages, such as Hawaiian, have no elements in the coda position at all.

¹ Previous discussions of markedness include Jakobson (1941), Trubetzkoy (1931, 1939), Greenberg (1966), Chomsky & Halle (1968:ch.9).

More recently, some phonological theories have incorporated concepts from probability theory (Coleman & Pierrehumbert, 1997; Pierrehumbert, 2003). Probabilistic approaches to phonology account for various types of frequency effects in language. Unlike markedness theory in OT which describes universal frequency effects in language typologies, a probabilistic approach describes speaker-specific frequency and its effects on language change and acquisition.

The crucial difference between SPE and OT theories and probabilistic models is that the latter are not generative. Probabilistic elements have been incorporated into generative theories, however, for example, by assigning probability scores to constraints in Stochastic OT (Boersma, 1997; Hayes & MacEachern, 1998; Boersma & Hayes, 2001). Exemplar theory is an example of a non-generative probabilistic model. In this theory, phonotactic co-occurrence restrictions are extended to account for allophonic alternations. In this model, a speaker-hearer stores a highly detailed representation of the sound combinations in their language. Categories such as syllables and vowels as well as phonological alternations such as vowel harmony (Cole, 2009) are emergent properties of statistical computation over the stored representations. Each token that a speaker-hearer perceives varies slightly and each of these representations is stored as an exemplar in a cloud of exemplars associated with that token. Speech production proceeds by randomly selecting an exemplar from the cloud and averaging over n number of that exemplar's neigh-

bours, yielding an output. This output is produced with a certain amount of noise to allow for deviations from the target (Pierrehumbert 2001, 2002).

These models of phonology are important as they make certain predictions relevant to the experiments presented in this dissertation. Despite their differences, both SPE and OT are considered generative. They have as their input a phonemic form that must be processed to yield an output that has the phonetic details necessary for the articulatory system to produce an utterance. In both models, the form /kæt/ is transformed into [k^hæt] even if the details of the process vary. This processing can be tested by overloading the system in order to see if the resulting output is the result of a missed step in processing.

Probability theory is relevant to the studies presented in this dissertation as it makes certain predictions about the robustness of speech segments based on their relative frequency. More frequent segments should be erroneous relatively less often than more infrequent segments. Second, since exemplar models are not generative, a lack of evidence to support the processing view of generative theories could support a non-generative model.

1.2.1 Phonological productivity

Phonology is the study of the underlying system that governs the use of sound in a language (Kenstowicz & Kisseberth, 1979). In English, the plural marker

is different depending on its context. After the voiced [g] in 'dog', it is voiced, [z]. After the voiceless [t] in 'cat', it is voiceless, [s]. These context-dependent sound alternations are abundant in spoken language. But do these context effects really exist as a process in the minds of speakers? And if they do, how are they represented?

One can argue that sound alternations are simply memorized by speakers. An English speaker has heard 'cats' enough to know that its plural form has a voiceless [s] while 'dogs' has a voiced [z] sound. This option is countered by the argument from productivity made famous by the wug test (Berko, 1958). In this test children are presented with an image of a creature called a wug as shown in Figure 4. Next, they are presented with two of them and are asked to complete the sentence "There are two ____". By the time they are in primary school children have no trouble responding 'wugs' with the plural marker a voiced [z] as is appropriate following a voiced [g].



Figure 4. Illustration used for the wug test from Berko (1958).

Given that *wug* is a nonsense word, none of the children would have heard it before and could not have memorized its plural form. This test and other phenomena, such as loan word adaptations show that speakers have some sort of knowledge of the combination of sounds in their language that they are able to use when faced with new forms they've not heard or spoken before.

There is also evidence for the existence of a phonological processing component from neuroscience. A meta-analysis of 82 experiments by Indefrey and Levelt (2004) reveal the location and time course for various components of speech production. Figure 5 shows the time course on the right as well as the corresponding brain regions on the left. The phonological code retrieval stage through to the generation of an articulatory score is the area of processing implicated by the current work. Phonological code retrieval begins the phonological stage of production by selecting the phonological information stored in the brain between 200-400 ms after stimulus onset in picture naming tasks. Syllabification proceeds between 330 and 455 ms after stimulus onset in the left posterior inferior frontal gyrus. From this point forms undergo the segmental level phonological processing, such as the assimilation in voicing of the plural marker. The form is then converted into an articulatory representation.

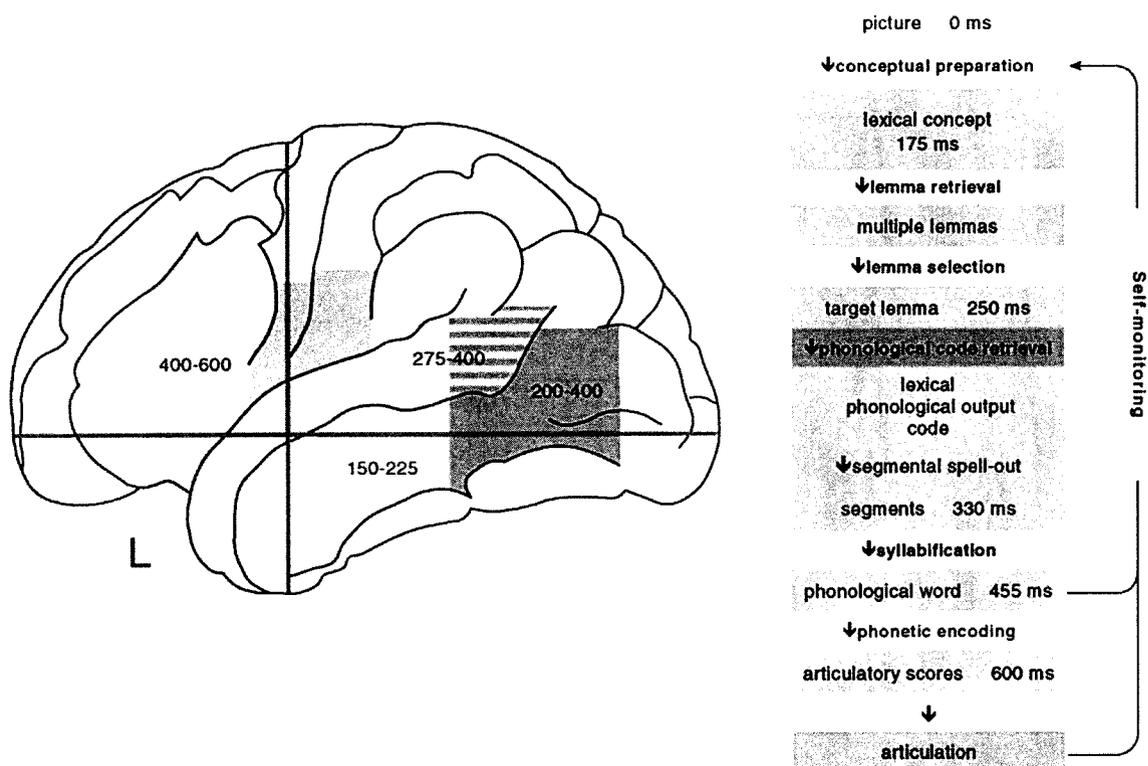


Figure 5. Model from Indefrey and Levelt (2004) showing the time course and location information for various components of speech production.

The study reported in Sahin, Pinker, Cash, Schomer, and Halgren (2009) provides additional neurological evidence for the productivity of phonological alternations. Electrodes were implanted in the brains of three patients with epilepsy who were undergoing clinical evaluation. Local field potentials were recorded while the participant was engaged in a speech production task. Three distinct regions of processing emerged from the study. Brain regions were differentially activated during tasks that required lexical, inflectional and phonological processing. The activity that was recorded at ~450 ms after stimulus onset is attributed to pho-

nological transformations such as the addition of schwa, [ə], between the past tense marker [-d] and the end of the verb stem in forms such as 'patted', [pærəd]. The time course of phonological activity found in the Sahin et al. (2009) study supports the time course information determined by the meta-analysis in Indefrey and Levelt (2004).

The cases cited above provide evidence that phonological alternations are not simply memorized but are instead processed during speech production. The harder question is how are these processes represented in the mind of the speaker? The convention adopted in this dissertation is to present phonological processes in terms of rules that take the form *A becomes B in the environment of C*. For example, /p/ > [p^h]/ #_V, the phoneme 'p' becomes aspirated word-initially before a vowel. This format is used in order to remain as agnostic as possible with respect to the details of phonological theory. The rule format requires little theoretical background to discuss. It may not, however, be the way in which phonological processes are encoded in the mind of the speaker. Other options that have been explored by researchers include constraint-based models and probabilistic models. While any or none of these models may be an accurate representation of phonological processing, they make certain predictions with respect to what a processing error would look like. Because OT and probabilistic models incorporate the concept of markedness and frequency, errors should result in an output that is less

marked or with a higher frequency of occurrence in that language. For example, all other influences being equal, the target utterance 'bad' would be more likely to be produced in error as 'bat' instead of 'bag' since the voiceless segment [t] in the coda is less marked than a voiced segment [g]. Without the concept of markedness, OT would predict that an error in processing should result in a random output. Rule based theories predict that there should be errors resulting from a rule that failed to apply. For example, the rule aspirating voiceless stops in English should fail, producing errors of the form [pɪt] for target [p^hɪt].

It is, however, entirely possible that phonological processes are represented in a way that is yet unknown and that an error in processing could be difficult or impossible to detect by the methods used in this dissertation.

Although the concept of an abstract, underlying form of a word that gains context-based phonetic information is fundamental to a model of phonology, it is less of a focus in psychological models of speech production, as seen in the next section.

1.3 Models of language use

Models of language production have been developed by psychologists on the basis of behavioural evidence such as speech errors, recall, and lexical decision

tasks. This differs from linguistic evidence, which is gathered from speakers as they would normally produce spoken language in addition to their judgements of what is an acceptable utterance. In addition, psychological models do not account for the /kæt/ > [k^hæt] type of derivation that linguistic theories address.

Despite these differences, the function of the component that processes speech sounds, termed phonological encoding by Levelt (1989), is the same for each field: "Its function is to retrieve or build a phonetic or articulatory plan for each lemma and for the utterance as a whole." (Levelt, 1989, p.12). This section outlines models of phonological processing developed within the field of psychology. The first is a procedural model developed by Levelt and the second is a spreading activation model developed by Dell (1986). Both models stress two independent streams of information: the structure of the speech sounds and the content within that structure. Structures include positional information within a word or a syllable while content refers to segments such as [b] or [n] and features such as [voiced] or [nasal]. This separation is necessary to account for the patterns found in speech error data, for example, that errors tend to involve the confusion of two similar items, either content or structure. For example, while the sounds [t] and [d] interfere with each other because they differ only by the feature *voice*, the syllable position onset may also interfere with another onset.

Levelt (1989) presents a thorough look at speech production through all levels of processing from concept to articulation. For my purposes, I will concen-

trate on the section relevant to the processing of speech sounds called phonological encoding illustrated by Figure 6. According to Levelt's model, information flows through several processing stages including a morphological level that assembles word chunks, a metrical level that assigns stress, and a segmental level that assembles the required phonemes. Each level then sends its output to a component called the prosody generator, which produces an address frame appropriate for the utterance. The address frame is a structure that defines syllable and metrical structure at the word and utterance level. This address frame and the segmental information are then fed into the phonetic spellout component, which combines the segmental contents to the address frame resulting in the phonetic plan. These levels of processing correspond to the linguists' view of a modular grammar.

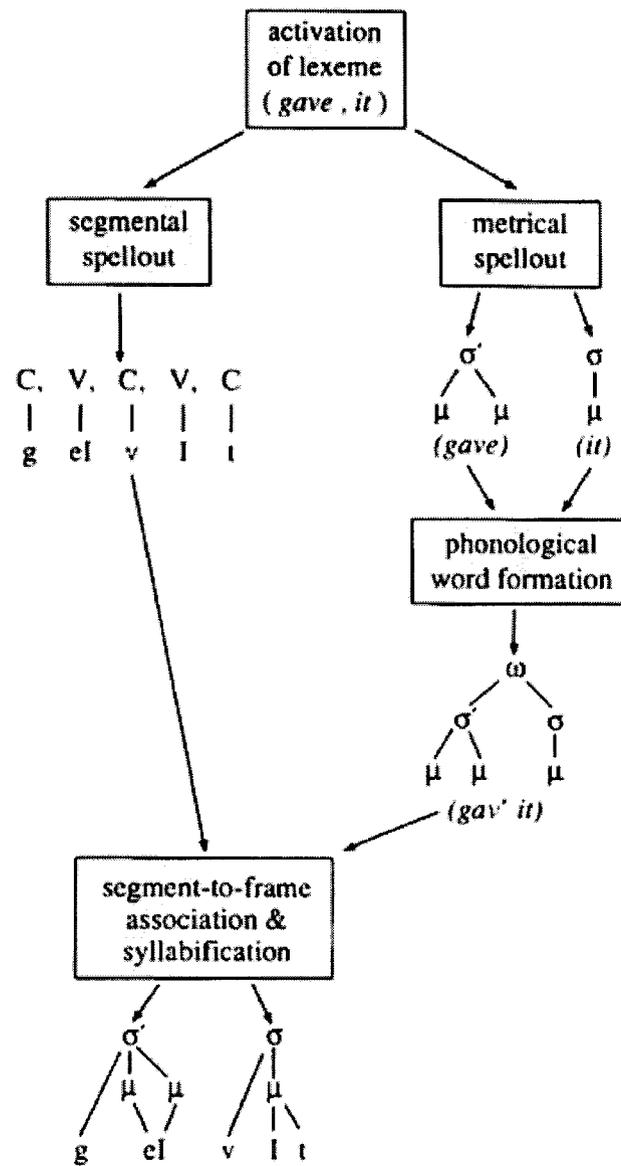


Figure 6. Model of phonological encoding (from Levelt, 1993).

The separation of address frame and segmental content is fundamental to psychological models of speech production and is the result of evidence from

speech errors and tip of the tongue phenomena. The spreading activation model presented by Dell (1986) differs from Levelt's procedural model in mechanics but the separation of content from frame is maintained.

Spreading activation processes information over a network of connected nodes. Dell's model, illustrated by Figure 7, has word level, semantic, syntactic, and morphological level nodes. Within the phonological stratum, there are phoneme, syllable structure, and feature nodes. During phonological processing, contents are requested by the syllable's onset node and the most highly activated onset then spreads activation to its feature nodes. The activation for the onset node falls to zero and the nucleus node becomes the active node. The evidence that shows errors conforming to syllable structure, that is, onsets are confused with other onsets, can be accounted for by the activation level of a recently selected syllable structure nodes. This model also captures the similarity constraint in speech errors, in which phones are confused with similar phones, i.e., phones that share features. Similar phones are closely connected in the network and the activation of one will spread to the other. Occasionally, this spread of activation will result in the selection of an incorrect phone.

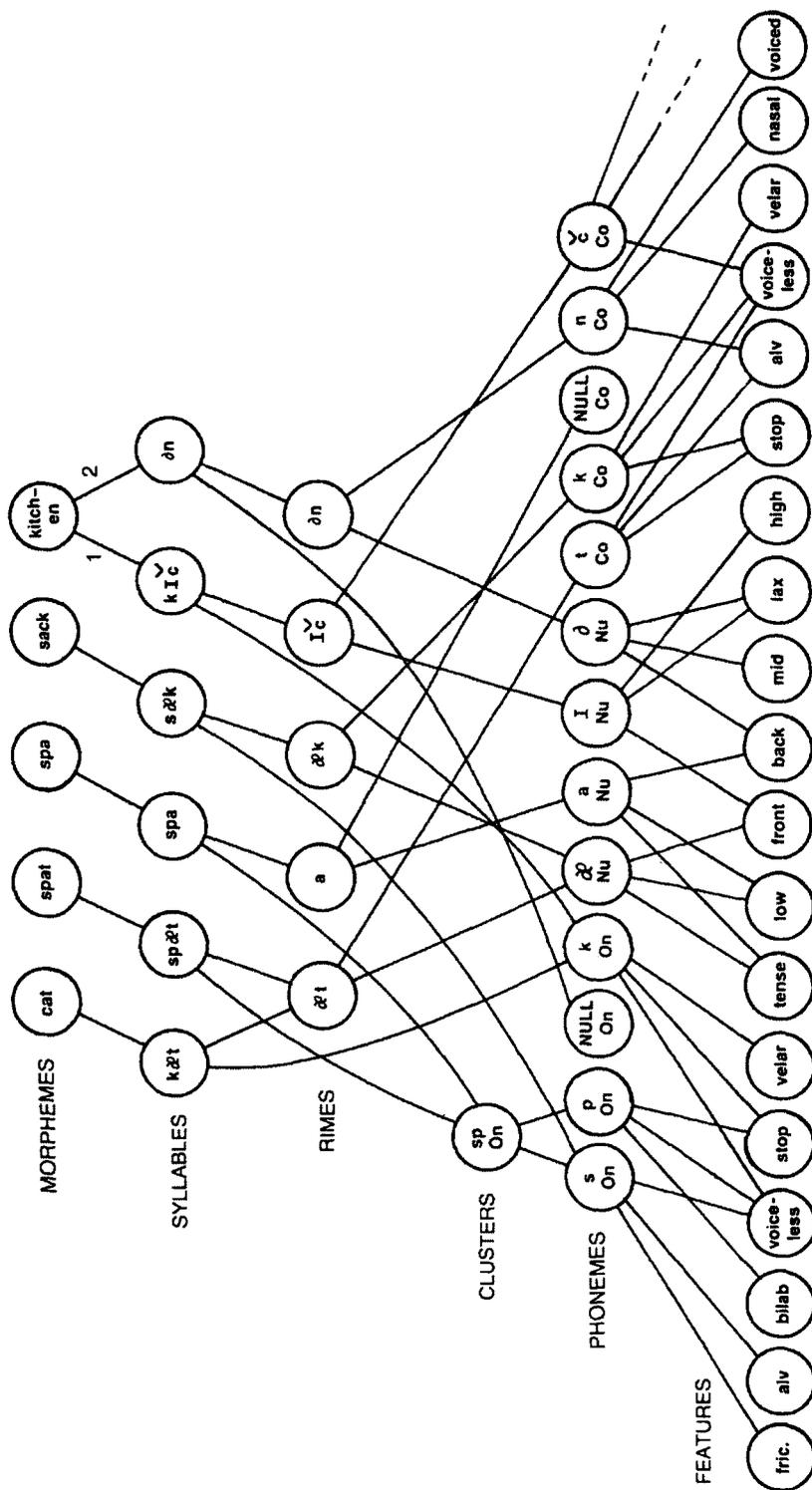


Figure 7. Dell's spreading activation model (from Dell, 1986).

More recent work expands Dell's spreading activation model to account for the influence of different levels of processing on speech errors through a mechanism of cascading activation (Goldrick, 2006; Goldrick & Blumstein, 2006). An example of a phenomenon that can be accounted for by cascading activation is the lexical bias effect in speech errors. Several studies report that the outcome of a speech error is more likely to result in a real word (Dell & Reich, 1981; Baars, Motley, & MacKay, 1975; Dell, 1986). Cascading activation accounts for the lexical bias effect through the interaction of activated forms. In attempting to produce the form 'cat' other, similar lexical items such as 'cab' are also activated. In failing to produce the target 'cat', the speaker is more likely to produce the partially activated word 'cab' than a form that was not partially activated such as the nonword 'cag'.

The model presented in Goldrick and Blumstein (2006) extends the concept of cascading activation between lexical selection and phonological processing to include cascading activation between phonological and articulatory processing. The results of their study reveal that in errors of target [k] produced as [g], there remains a phonetic trace of the intended [k]. Voice onset time (VOT) is the measure that distinguishes [k] from [g]. In these errors, the resulting [g] has a longer mean VOT, closer to that of a typical [k]. They conclude that the intended [k] was partially activated during production.

Both types of model presented here are able to account for the patterns found in speech errors by separating structure from content. Levelt's is a procedural model that processes information in sequential stages whereas Dell's spreading activation model uses associations to fuse the various streams of information. More recent work extends the spreading activation to include the blending of information from various processing levels (Goldrick & Blumstein, 2006). These psychological models are similar to linguistic models of sound production in their stated purpose: to take input from the lexicon and the syntactic processing stage and specify the phonetic information of an utterance. They differ in the type of evidence used to produce the models. Linguists seek language data that reflect how it would normally be produced. Psychologists, on the other hand, use evidence from speech errors, half-remembered words, the confusability of sounds, and reaction times. These distinct types of evidence have led to differences in linguistic and psychological models of speech production. Notice that psychological models do not account for the /kæt/ > [k^hæt] type of derivation and that the linguistic theories that account for such derivations do not account for speech errors.

1.4 Previous speech error studies

The study of speech errors has uncovered many regularities in speech production. Common error types include anticipations, perseverations, exchanges,

deletions and additions. Speech errors tend to fall into specific patterns; similar segments interfere with each other, as do items in the same syllable position. These patterns have been taken to reflect the units involved in the planning of speech. Examples of all of these types of errors as well as the units involved in errors are reproduced in Table 3 from Dell (1986, p. 285).

Table 3. Example of speech error types from Dell (1986)

| Error type | Example | Unit involved |
|----------------------------|-------------------------------------|-------------------|
| Exchange | York library → lork yibrary | Phoneme |
| Exchange | spill beer → speer bill | Rime constituent |
| Exchange | snow flurries → flow snurries | Consonant cluster |
| Exchange | clear blue → glear plue | Feature |
| Anticipation | reading list → leading list | Phoneme |
| Anticipation | couch is comfortable → com/is . . . | Syllable or rime |
| Perseveration | beef noodle → beef needle | Phoneme |
| Anticipatory addition | eerie stamp → steerie stamp | Consonant cluster |
| Perseveratory addition | blue bug → blue blug | Phoneme |
| Shift | black boxes → back bloxes | Phoneme |
| Deletion | same state → same sate | Phoneme |
| Noncontextual substitution | department → jepartment | Phoneme |

| | | |
|------------------------|-------------------------|----------|
| Noncontextual addition | winning → winnding | Phoneme |
| Noncontextual deletion | tremendously → tremenly | Syllable |

Early studies relied on the transcription of errors by the researcher as they were heard in everyday speech (Meringer & Mayer, 1896; Boomer & Laver, 1968; Fromkin, 1971). Some speech error collections are extensive, such as the Meringer and Mayor (1896) corpus with over 8000 examples, as well as Fromkin's corpus, with close to 4000 entries (Fromkins Speech Error Database, 2000). Although collecting natural speech errors provided a useful starting point for uncovering the basic patterns in speech errors listed above, as well as other patterns, such as the lexical bias (errors are more likely to result in a real word), and an anticipatory bias (anticipation errors outnumber perseveration errors) (Dell, Burger & Svec, 1997), this approach has several problems associated with it. First, a lot of time and patience is required to collect natural speech errors this way. Second, potential biases in the perception of errors exist due to the tendency of language perceivers to "repair" speech errors produced by others (Goldstein, 1980).

In response to these issues associated with the collection of natural speech errors, methodologies were developed to induce speech errors in the lab. Baars and Motley (1974) and Motley and Baars (1976) introduce the SLIP tech-

nique², which uses priming to induce errors. Word pairs are presented that prime segments in certain contexts and then the participant is asked to produce a new word pair that either matches the primed pattern or doesn't. For example, the initial segments [b] and [p] are primed with pairs such as 'Book Pole' and the reverse pattern is shown 'Pear Boot'. Errors result when the participant produces the new word pair with the primed pattern, resulting in the utterance 'Bear Poot'. Tongue twisters are also used to induce speech errors in the lab (Wilshire, 1999). The errors that result from these laboratory techniques were found to be comparable to natural speech errors (Stemberger, 1992; Wilshire, 1999). Not only do these techniques speed up the collection process, it also allows for control in targeting specific processes or units and for speech errors to be recorded for instrumental analysis.

Instrumental analysis of speech errors

Instrumental analysis is responsible for another important advance in speech error research; the discovery of sub-phonemic errors. Sub-phonemic errors are errors involving units that are smaller than a phoneme such as a feature. Feature errors were first reported by Fromkin (1971) with the error 'clear blue sky > 'glear plue sky' where the voicing feature of the [b] is anticipated on the [k] in 'clear' resulting in a voiced [g]. However, they were considered too rare to be relevant for

² Spoonerisms of Laboratory Induced Predisposition

theory building. In fact, sub-phonemic errors are difficult to perceive naturally and instrumental techniques must be used to reliably detect any error in their production.

Instrumental techniques have been used in several studies to show the existence of sub-phonemic level errors. Mowrey and MacKay (1990) investigated induced speech errors using electromyographic (EMG) recordings of tongue movements. The results showed a high number of sub-phonemic intrusions and gradient productions as well as gradient deletions. The EMG recordings of the tongue twister, 'She sells seashells by the seashore', showed a labialized [s], which is not a feature normally associated with this segment. The labial component was interpreted as an intrusion from the labial component of an [ʃ] elsewhere in the tongue twister. This labialization was not perceived in the audio recordings, however, supporting the view that errors at this level are difficult to perceive unaided.

Of the 150 tokens analyzed, 48 contained errors. Gradient deletion or transposition errors accounted for 43 errors while only 5 were characterized as fully deleted segments. This result indicates that not only are sub-phonemic errors possible but they are far more common than phoneme or feature-level errors.

This study presents several important findings. First, sub-phonemic and even sub-featural errors occur. Second, these errors are not generally perceived without instrumental analysis. Third, errors can be made after phoneme selection

and phonological rule implementation, contrary to the models presented in MacKay (1970), Fromkin (1971), Shattuck-Hufnagel (1983) and Stemberger (1983).

Articulatory techniques were also used in the study reported by Pouplier (2003) that support the results of Mowrey and MacKay (1990) by demonstrating that a gesture from one segment can intrude on the production of another segment. Tongue, lip and jaw movements were recorded using an electromagnetic midsagittal articulometer (EMMA). Data from nine participants producing two-word combinations, e.g., *top cop*, was included in the analysis. The normal production of the [k] in 'cop' includes raising the tongue dorsum (TD) while the production of the [t] in 'top' includes the raising of the tongue tip (TT). Errors were determined by calculating the mean height of a gesture during the control trials. Gestures that fell two standard deviations outside of this mean were counted as errors. The results showed considerable intrusion errors. For example, the raised tongue back (TD) was frequently produced during the [t] and tongue tip (TT) gesture was produced during the [k]. The error rate for [t] was 39% and 32% for [k]. Of these errors 76% were gesture intrusion errors during the production of [t] and 83% during [k]. Gesture reduction errors accounted for 10% of the [t] errors and 8% of the [k] errors. As in Mowrey and MacKay (1990), most of these errors were found to be partial rather than full intrusion or deletion errors.

A similar study using acoustic analysis is reported in Frisch and Wright (2002). This study measured the percent voicing from productions of word-initial [s]

and [z] tokens that were induced using tongue twisters. All productions were included in the analysis, not only those deemed to be errors. Values for voicing fell along the continuum between these two sounds. These values are reproduced in Table 4.

Table 4. Percent voicing results for [s] and [z] from Frisch and Wright (2002)

| Percent voicing | Intended segment | |
|-----------------|------------------|-----|
| | /s/ | /z/ |
| 0% | 324 | 56 |
| 0-5% | 25 | 8 |
| 5-10% | 13 | 24 |
| 10-30% | 12 | 39 |
| 30-60% | 4 | 23 |
| 60-100% | 6 | 33 |
| 100% | 13 | 252 |
| Total | 397 | 435 |

The percent voicing measure typical of [s] is 0% while a [z] can have partial devoicing, it is normally produced with a percent voicing of 60% or higher (Haggard, 1978; Smith, 1997). The results for percent voicing show that both [s] and [z] were produced with various values of voicing along this continuum, con-

firming through acoustic measures the results found using articulatory techniques in Mowrey and MacKay (1990) and Pouplier (2003). Further, because of the high number of tokens that were produced at the end points, [z] produced with 0% voicing and [s] produced with 100% voicing, this study indicates that errors are categorical as well as gradient.

Goldrick and Blumstein (2006) presented further evidence to support the existence of sub-phonemic errors. Tongue twisters were designed to induce errors on word-initial stops, e.g., *keff geff geff keff*. Forms that were determined perceptually to be errors were paired with perceived correct productions. Data from five participants yielded 2400 tokens. Voiceless consonants were produced as voiced in 58 tokens and voiced consonants were produced as voiceless in 40 tokens. An analysis of voice onset time (VOT), the acoustic measure that distinguishes a voiced from a voiceless stop, was performed. This analysis revealed that the VOT of the error tokens was skewed towards the VOT of the target. The correctly produced voiced consonants had a mean VOT of 21.03 ms while the voiceless consonants incorrectly produced as voiced had a mean VOT of 25.63 ms, which was significantly different, $p < .005$. Voiced consonants incorrectly produced as voiceless had a mean VOT of 68.28 ms which was significantly different, $p < .05$, from the voiceless tokens with a mean VOT of 76.31 ms.

Accommodation

Before these studies showed that sub-phonemic errors do, in fact, occur, it was believed that all errors nonetheless resulted in the correct surface phonology. This accommodation is obtained when the phonology is appropriate for the new, erroneous environment instead of the intended environment (Boomer & Laver, 1968; Fromkin, 1971). For example, in the error 'track cows' [træk kæwz] for target 'cow tracks' [kæw træks], the slipped [s] assimilates to its new environment as [z] (Fromkin, 1973). Morphology level errors have been found that fail to accommodate, for example, 'Der (masculine) Christa und die (feminine) Helmut' for intended 'der Helmut und die Christa' (Berg, 1987). However, accommodation is considered the norm by speech error researchers (Boomer & Laver, 1968; Fromkin, 1971; Garrett, 1975; Shattuck-Hufnagel, 1979; Stemberger, 1982). Accommodation is cited as evidence for the order of processing of the two levels involved: 1) speech planning, the level of phonological encoding where errors are generated, and 2) the phonology, where the rules are applied. This ordering has been formalized in the speech production models of Fromkin (1971) and Shattuck-Hufnagel (1979, 1983).

Because sub-phonemic errors can only be reliably detected through instrumental analysis, it could be that accommodation does not always hold since an incorrect surface phonology would be difficult to perceive. The question of phonol-

ological accommodation and the processing order of the two components must be readdressed.

1.4.1 Errors in Syntax

Errors that are attributed to the syntactic level are usually lexical selection errors. Lexical selection errors provide evidence for the psychological reality of syntactic categories in the same way that they do for phonological units. In lexical selection errors, the target and the errors are generally of the same syntactic category, so that a noun is replaced by another noun, a verb by another verb and so on (Noo-teboom, 1973; Fromkin, 1973). An example from Fromkin (1973) illustrates an error in which two nouns in the same sentence are switched, '*a computer in our own laboratory* > *a laboratory in our own computer*'.

More rare are discussions about syntactic transformation errors. This type of error would be the result of a syntactic transformation failing to apply or applying in the wrong context. Attempts have been made to determine if errors of syntactic rule processing occur. Transformational errors have been analyzed by Fay (1980) and Chen and Baars (1992). While some errors do seem to reflect transformational rule failure and over-application, there seem to be alternative explanations by mechanisms already known to exist.

Fay (1980) describes cases from a collection of over 4000 naturally occurring speech errors collected over a five year period. Various transformational errors are presented, including an error of Subject Auxiliary Inversion (SAI), listed in the error analysis in Table 5. The analysis suggests that the error is due to the failure of SAI to create the conditions necessary for the Affix-Hopping rule to apply.

Table 5. Analysis of an error as a syntactic transformation error

Target - And what did he say?

Error - And what he said?

| | |
|------------------------|----------------------|
| and Q he PAST say WHAT | Underlying Structure |
| and WHAT he PAST say | Wh-Fronting |
| and WHAT he PAST say | *SAI (omitted) |
| and WHAT he say+PAST | Affix-Hopping |
| and what he said | Morphophonemics |

Fay concludes by presenting a counter argument to one of the cases of apparent transformational rule error. In some cases, tense placement rules can be explained by phonological errors that are already known to exist. For example, the error '*He always know whats to say*' for the target '*He always knows what to say*' could be analyzed as a displaced segment [s] instead of a mis-applied tense formation rule.

A second problem with this approach pointed out both by Chen and Baars (1992) and Fromkin (1980) is that the analysis of transformational rule errors is dependent on the theory that describes the structures of the rules themselves. As the theory changes, so does the description of the rule and the predicted errors that would arise from their failing.

Chen and Baars (1992) designed an experiment to induce transformation errors. Sixty participants were asked to memorize two sentences that paraphrase the targeted error. For example, the two paraphrased sentences for the target error *'What could have I done with the check?'* are *'What have I done with the check?'* and *'What could I have done with the check?'*. They were then shown one of the two paraphrased sentences and cued to either repeat this sentence or to produce the other paraphrased sentence they had originally memorized.

The results were presented as successful. Nine of the ten error target sentences were elicited using this technique. The reasons for the errors, however, are not clear. Although the stimuli were constructed to induce errors, the results seem to be in line primarily with inducing sentence blends. The error sentence was constructed, it seems, based on what the authors thought the two component sentences could have been to create a sentence that was a blend. While it is impossible to say that none of the errors induced were transformational, priming a participant with two similar sentences logically could result in a blend of those two sentences.

These two studies do not provide conclusive evidence of transformational errors. In fact, they highlight that what appear to be transformational errors can be classified as other, more well known types of errors. The existence of transformational errors, therefore, has yet to be demonstrated.

Chapter 2: Research Framework

Recent work in instrumental analysis of speech errors has uncovered the existence of sub-phonemic errors through articulatory studies (Mowrey & MacKay, 1990; Pouplier, 2003) that have revealed gestural intrusions and gradient productions that are too difficult to detect without instrumental analysis. Acoustic studies have confirmed gradient productions and identified feature errors (Guest, 2002; Gormley & Thomson, 2007) as well as determined that gradient errors alone are not sufficient to account for error patterns (Frisch & Wright, 2002). Further, it has been found that a phonetic trace of the target persists in the error (Goldrick & Blumstein, 2006).

Given the success of instrumental analysis for detecting sub-phonemic speech errors, researchers are in a position to test assumptions that had been decided based solely on transcription studies. One of these assumptions is that errors are accommodated so that the surface phonology adjusts to suit the error. If errors are generated in a separate speech planning component, then accommodation has implications for the ordering of these components. If the surface phonology is always correct, then phonological processing must occur after the error. If accommodation does not always hold, then errors can be made after phonological processing.

A further implication of accommodation is that phonological processing is always correctly executed and that errors at this level never occur. Both of these assumptions are tested in this dissertation.

Experiment 1 tested the long-held assumption of phonological accommodation: that phonological rules suit the outcome of an error. For example, consider the English rule of vowel lengthening before word-final voiced consonants.³ In the hypothetical error 'sat cad' [sæt k^hæ:d] for intended 'sad cat' [sæ:d k^hæt], illustrated by Figure 8, it is assumed that the vowel in 'cad' should be long, having adjusted to the voiced coda 'd' even though it would have been short due to the voiceless 't' of the intended utterance.

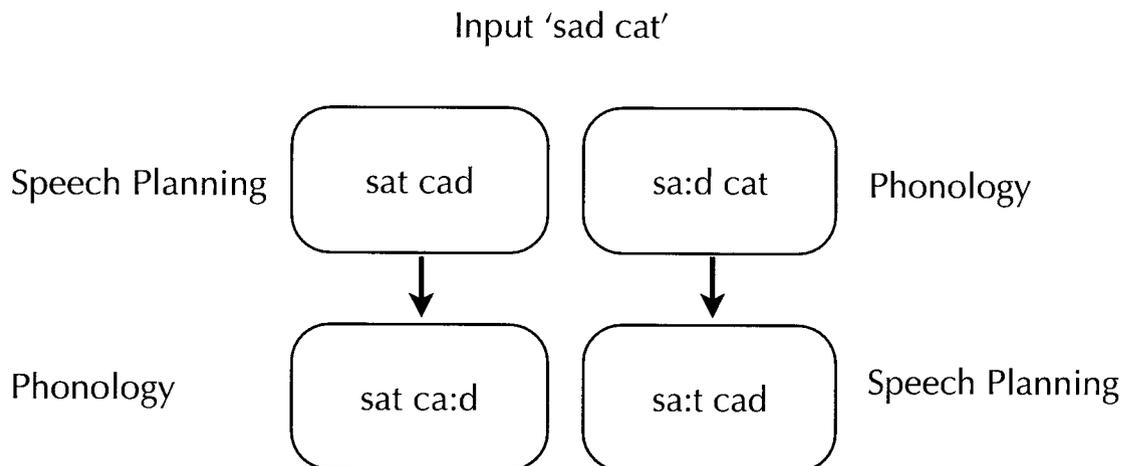


Figure 8. Two possible processing orders of the target utterance 'sad cat'. Errors occur at the speech planning level.

³ This rule could also be formulated as vowel shortening before a voiceless coda.

If it is true that phonological rules always accommodate to the new, erroneous output, then we have evidence for two separate processing components since the phonological rule component is always processed after the error is made and no errors should arise during phonological processing since the surface phonology is always correct. An acoustic analysis of speakers producing tongue twisters reveals that this assumption is wrong and that the phonology does not always match the error. These mismatch errors have three possible sources; they could occur during phonological rule processing, during motor-level articulatory implementation, or at a level of speech planning between these two stages.

What does an error during phonological processing look like? An error made during phonological processing could result in a failure of a phonological rule to apply. For example, in the processing of 'cat' from underlying /kæt/ to fully specified [k^hæt], the rule that aspirates the /k/ could fail to apply, resulting in the form [kæt]. If this were to occur, it could be perceived by a listener as [gæt] or, the error may not even be noticed. Due to the difficulty in perceiving this hypothesized type of error, it would not be surprising to find that they exist but have never been detected. Experiment 2 tested the possibility that mismatch errors are errors of phonological processing by determining whether these errors are more frequent in forms that involve the processing of phonological rules compared with forms that

lack the rule. Experiment 3 tested another possible source of mismatch errors, motor-level articulatory implementation.

The third possible source is a speech planning level after phonological rules are processed but before the motor level implementation. The term *speech planning* is used to describe a level at which errors occur in models of language use. Errors are associated with various levels of production, including lexical access and morphological processing so speech planning levels can be assumed to be associated with each of these error types.

The hypotheses to be tested in this dissertation are:

1) All errors are accommodated.

This is the current default assumption based on the results of transcription studies.

2) Phonological transformations are processed during speech production and are prone to errors that are detectable by acoustic analysis.

There are converging lines of evidence to support the view that phonological alternations such as the voicing assimilation of the English plural marker /-s/ are not simply memorized but are processed during production. This process should be vulnerable to error especially during increased cognitive load.

3) Mismatch errors are motor-level errors.

If mismatch errors are similar to known motor-level errors, they may also be motor level errors and be produced during articulatory implementation.

The present investigation assumes a model of speech production that consists of the different processing levels shown in Table 6. It assumes that one of these levels is responsible for phonological transformations. One or both of these assumptions may prove to be false.

Table 6. Steps assumed in speech production

| Step | Process |
|------|--------------------------------|
| 1 | Phoneme selection and ordering |
| 2 | Phonological rules |
| 3 | Output Level |
| 4 | Motor level implementation |

The remainder of this dissertation is organized as follows; Chapter 3 presents a tongue twister experiment that demonstrates errors do not always accommodate to the surface phonological context, contrary to the view that accommodation dominates. Potential sources for these errors are after phonological processing or at the level of phonological processing itself. An experiment is presented in Chapter 4 that tests if these errors are made during the processing of phonological

rules. If the phonology is a source of errors, then more errors should be present for forms that require phonological processing. This experiment can be viewed as a test of phonological processing itself; if phonological processing occurs using the mechanisms described by transformational theories, then we should detect errors of the implementation of those mechanisms. A third experiment is presented in Chapter 5 that tests whether mismatch errors are made after phonological processing (i.e., during articulatory implementation). This experiment uses delayed auditory feedback to induce motor-level errors that are then compared to the mismatch errors induced in Experiment 1. These experiments are designed to pinpoint the source of mismatch errors. Chapter 6 presents a computational model that illustrates the proposed steps in processing necessitated by the results of Experiment 1. The results of these experiments show that mismatch errors do not occur during phonological processing, or at the motor level and must therefore be the result of the processing step after phonological rules but before motor implementation.

Chapter 3: Experiment 1

Speech errors are generally assumed to accommodate to the unintended environment. This accommodation provides evidence for the ordering relationship between the processing component responsible for phonological assimilation and the component that is the locus of errors. When a segment accommodates, it means that the error must have been made before the level at which phonological processes are implemented. Previous work on accommodation relied on transcription to detect whether an error has accommodated to its new environment (Boomer & Laver 1968; Meara & Ellis, 1981). Given that transcription-based studies are prone to perceptual bias errors (Goldstein, 1980; Cutler, 1981; Pouplier & Goldstein, 2005), the conclusion that accommodation is the norm is questionable.

An instrumental analysis of speech errors was conducted to re-address the question of phonological accommodation. Thirty-two nonword tongue twisters, e.g., *tiff tivv tivv tiff*, were designed to induce voicing errors on the coda. Because vowel lengthening before voiced codas is a phonological process in English, vowel length can be measured to see if phonological accommodation has occurred. Errors are determined for each participant based on comparison with measures for percent voicing and vowel duration in a control condition.

Transcription studies on phonological accommodation are at a disadvantage given the difficulties in perceiving errors at a phonetic level. Acoustic analysis

removes this perceptual issue, yielding a more objective result. Given that errors occur at all levels of speech production – semantic, morphological and phonological, it would not be surprising to find that errors can occur after phonological processing but before the formation of an articulatory plan.

3.1 Method

Several tongue twisters were designed to induce voicing errors in the coda of nonwords. Nonwords were used to minimize any semantic or frequency effects that might be associated with the use of words. Given the phonological rule that lengthens vowels before voiced codas, vowels were measured to determine if their length was appropriate for the voicing of the coda. For example, in the twister *tiff tivv tivv tiff*, all vowels and coda fricatives were measured to determine if vowels were long in *tivv* [tɪ:v] but short in *tiff* [tɪf].

3.1.1 Participants

Thirty Carleton University students received course credit for participating in the study. Due to the many fine-grained analyses required for the acoustic measurements, only data from the first six male students are included in this dissertation. Each participant was a native speaker of English with no reported history of speech or hearing problems.

3.1.2 Materials

Recordings were made in a sound attenuated booth through a head-mounted microphone to a Sony HiMD mini disc recorder. Trials were presented on a laptop screen using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993). Acoustic analysis was done using Praat (Boersma & Weenink, 2008) and statistical analysis was performed in SPSS.

Stimulus sets were created by combining [t] or [k] as the initial segment, [ɛ], [ɪ], [ʌ] and [æ] as the vowels and [v], [f], [s] and [z] as the final segment. Three blocks were created; the control block combined stimuli in a non-alternating AAAA pattern (*tivv tivv tivv tivv*) whereas the two identical experimental conditions created tongue twisters by combining stimuli in an alternating ABBA pattern (*tivv tiff tiff tivv*).

3.1.3 Procedure

Each trial began with 12 metronome clicks played at a rate of three clicks per second.⁴ This set the pace by which the tongue twisters were to be produced. Next, a four-‘word’ tongue twister was presented on the computer screen. Each of the 96 four-‘word’ sequences was repeated three times each at the rate established by the metronome. The participant then pressed a key to advance to the next trial. There was a 4 second lag between utterances during which time the metronome

⁴ This procedure was adapted from Goldrick and Blumstein (2006).

played. A practice session of three trials preceded the experimental trials and two optional breaks were provided. The entire experimental procedure lasted approximately twenty minutes.

3.2 Results

Unlike other speech error studies, all data were analyzed, not only those tokens that were perceived by the researcher as errors. This method was also used by Frisch and Wright (2002) and eliminates the perceptual bias inherent in the perception of speech. The combined spectrogram/waveform in Figure 9 shows how the segments were delineated. The uppermost panel shows the waveform, the middle panel shows the spectrogram and the bottom panel shows the text grid with the segmentation for the token *tiss*.

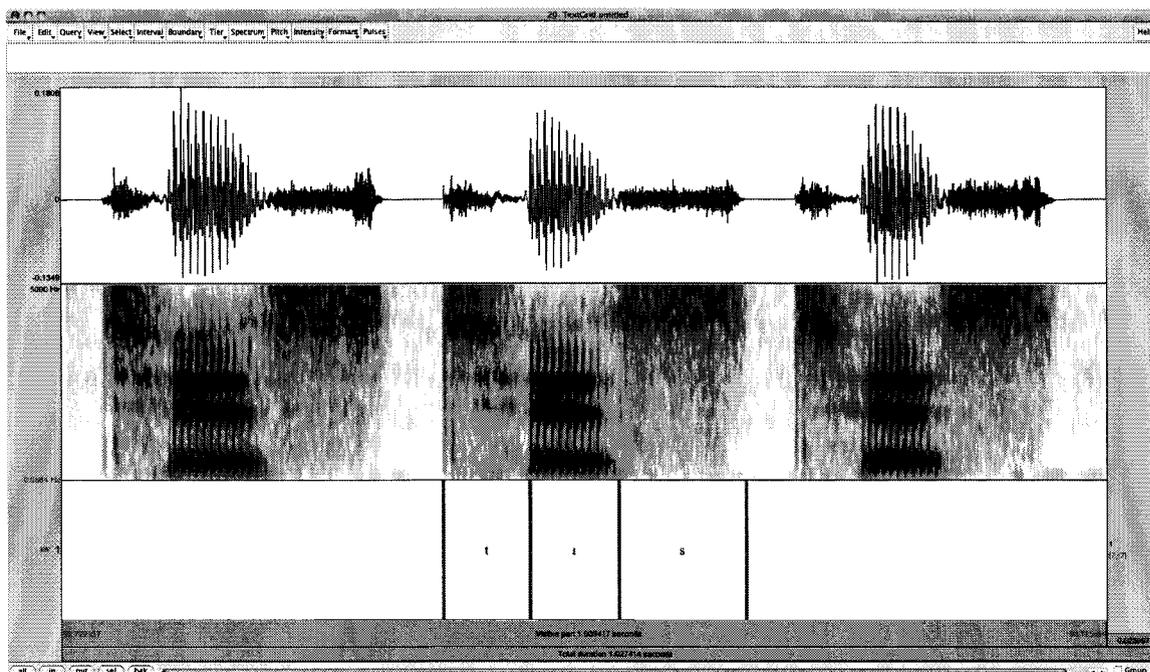


Figure 9. Spectrogram and waveform of a token of *tiss* showing segmentation.

Scripts were run in Praat to extract the percent voiceless measures from the fricatives and the duration for the vowels for both conditions for each participant. As expected, the initial consonant did not make a significant difference in vowel duration or voicing of codas and these data were not analyzed separately. Due to the process used to design the stimuli, one of the tokens used in the experiment was the word 'kiss' and others were phonological words such as [tɪf], 'tough', [kʌf], 'cuff' and [kʌs], 'cuss'. No differential effect was found on error rates between the nonwords and the orthographic and phonological words.

A range of normal values was defined to be within two standard deviations of the mean computed for the control condition for each participant. All tokens in the experimental condition were categorized as normal or erroneous based on this range, as illustrated in Figure 10.

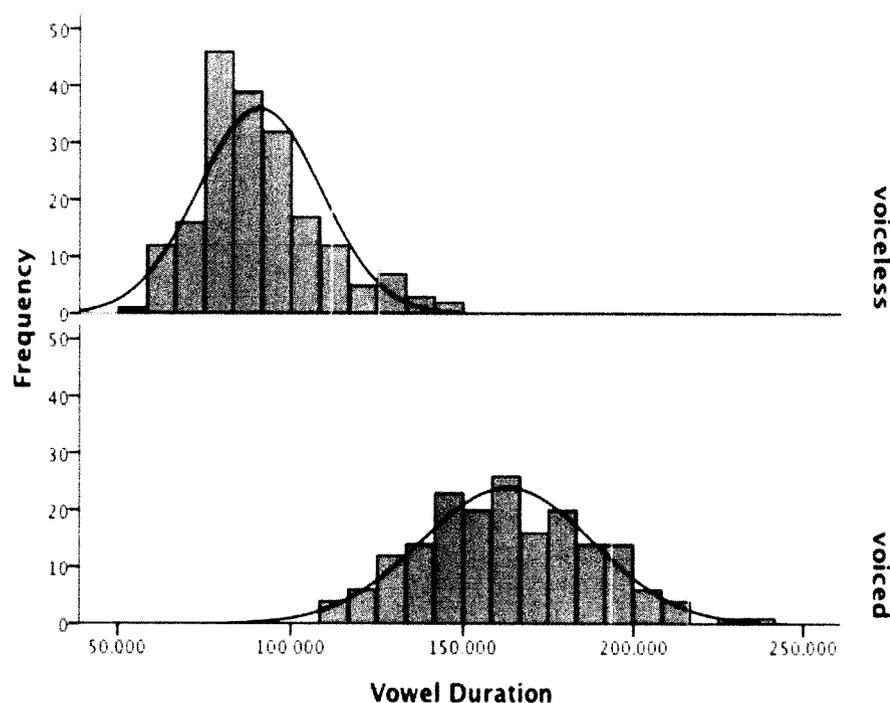


Figure 10. Mean durations for vowels preceding voiceless and voiced codas for a single participant. The mean duration of vowels before voiceless codas is 91 ms with a SD (standard deviation) of 18 ms yielding, a range of $91 \pm (18 \times 2) = 55 - 127$ ms. The mean vowel length before voiced codas is 163 ms with a SD of 25, yielding a range of $163 \pm (25 \times 2) = 113 - 213$ ms.

Tokens with values that were more than two standard deviations from the mean were considered errors. Mismatch errors were defined as when the vowel

length did not correspond to the voicing value of the coda fricative, such as when the token *tiff* [tɪf] was determined to have a long vowel and the coda fricative was determined to be voiceless [tɪ:f]. There was a considerable grey area in the determination of the range of normal values for each participant. Because values for the categories voiced and voiceless as well as vowel duration were on a continuum, a vowel's duration could be a value that falls within a normal range for either long or short, as shown in Figure 11. These indeterminate cases were not classifiable.

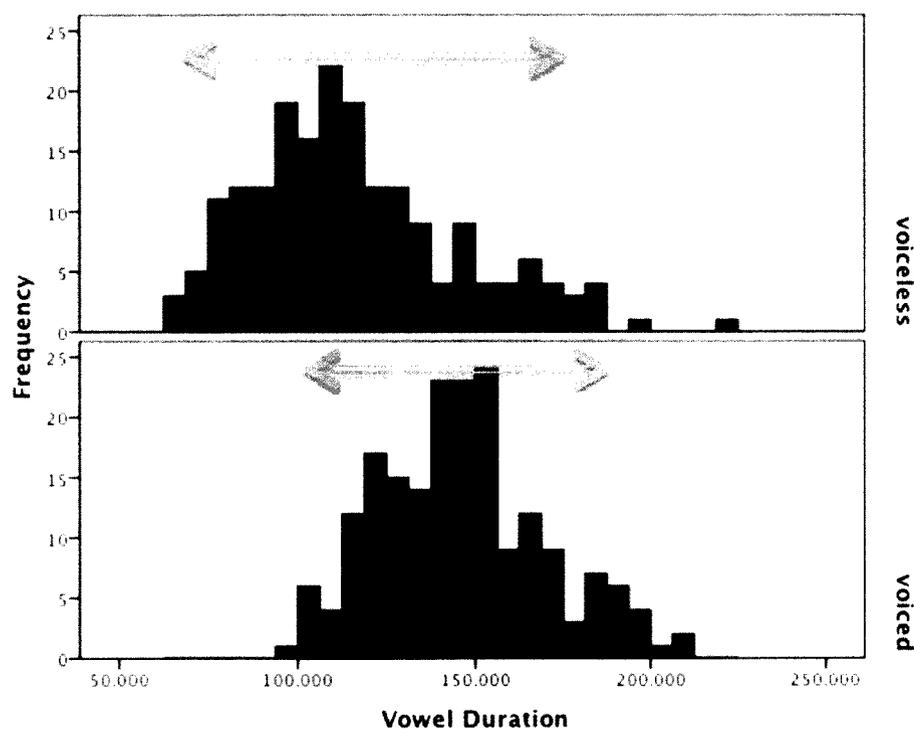


Figure 11. Mean durations for vowels before voiced and voiceless codas for a single participant showing considerable overlap. The mean duration for short vowels is 116 ms, SD 30. Long vowel duration is 146 ms, SD 24.

As shown in Table 7, of the 4790 tokens produced by participants in the experimental conditions, 872 were errors. Of those, 513 were vowel length errors, 293 were fricative voicing errors, and 66 were both length and voicing errors.

Table 7. Error distribution by type

| Error type | N | % |
|-------------------|-----|------|
| Vowel Length | 513 | 10.9 |
| Fricative Voicing | 293 | 6.2 |
| Both | 66 | 1.4 |
| Total Errors | 872 | 18.5 |

As shown in Table 8, of those 872 errors, 355 are mismatch errors, 66 are accommodated, and 451 are unclassifiable.

Table 8. Error classification

| Error class | N | % |
|----------------|-----|------|
| Mismatch | 355 | 40.7 |
| Accommodated | 66 | 7.6 |
| Unclassifiable | 451 | 51.7 |
| Total Errors | 872 | |

3.3 Discussion

The results of Experiment 1 show that not all errors accommodate to the erroneous environment. By analyzing all tokens and determining the normal range of production for each participant, an unbiased view of speech errors is obtained. That errors can be phonologically inappropriate for their environment shows that phonological rules are not consistently processed after the error is made (see Figure 6). This suggests that the assumption made by transcription-based research that speech errors must occur before the phonological processing component is wrong. There is, however, another possible interpretation of this result. Phonologists do not study errors and as a result they have no explanation for their occurrence within phonological competence models. What if the notion that errors can arise during the processing of phonological rules was considered? What would an error of this type look like? The answer is that an error of phonological rule processing, (i.e., the failure of a phonological rule to apply) would look just like the data collected in Experiment 1. Table 9 indicates the division of mismatch errors into process errors.

Table 9. Mismatch errors by process error

| | Count | % |
|--|-------|------|
| a. t _Λ f > t _Λ v | 158 | 44.6 |
| b. t _Λ :v > t _Λ :f | 56 | 15.8 |
| c. t _Λ :v > t _Λ v | 102 | 28.8 |
| d. t _Λ f > t _Λ :f | 38 | 10.7 |

Under this interpretation, the phonology is inappropriate to the environment in the mismatch cases not because the error happened after the phonology was processed but because the error was the failure of the phonological rule to apply. Assuming that the vowel length rule applies to underlying short vowels to produce long vowels and not the other way around, we would predict that there should be more length errors for target forms that should undergo this rule. As is shown in Table 10, this is, in fact, what was found.

Table 10. Proportion of length mismatch errors for target voiced and voiceless tokens

| | Target voiced | Target voiceless |
|------------------------|---------------|------------------|
| Mismatch/Length Errors | 102/158 | 38/197 |
| Percent | 65% | 19% |

The results of Experiment 1 show that the surface phonology does not always conform to its erroneous environment. This result can be interpreted two ways. First, this could show that errors can, contrary to previous assumptions, be made after phonological rules, perhaps due to an implementation error at the motor level. Second, that the errors observed could be attributed to the phonological rule component itself (i.e., the result of a rule failing to apply). Data from Experiment 1 were insufficient to allow a choice between these possibilities and so two further experiments were designed to tease these possibilities apart.

Chapter 4: Experiment 2

Experiment 1 revealed mismatch errors, demonstrating that the phonology does not always conform to the surface environment. This result could indicate that errors are made after phonological processing, during motor-level implementation, or it could be evidence that errors occur during phonological processing. While the results of Experiment 1 are consistent with the second view, they cannot be used to conclusively choose between the alternatives.

Experiment 2 was designed to test the fallibility of phonological rules. This assumes that there is some kind of process that operates on an input, producing a distinct output. Although different theories assume different mechanisms - rules in generative phonology and constraints in optimality theory - this test can generalize to any procedural theory. Experiment 2 was designed to determine if mismatch errors could be the result of a breakdown in phonological processing by comparing two sets of tokens; one set that contained a rule and a second set that did not.

4.1 Method

Two sets of stimuli were administered. One set consisted of forms that undergo a phonological process while the other set lacked this process. The frequency of errors was then compared to see if the condition that involves the application of a phonological rule had more errors than the tokens with no phonological

processing. If there are more errors in the phonological rule condition, it suggests that errors from within the phonological processing component are possible.

4.1.1 Participants

Speech samples were obtained from sixteen native speakers of North American English with no history of speech or hearing problems. Participants were Carleton University students and received course credit for participating.

4.1.2 Materials

The recording equipment and configuration used in this experiment was the same as that used in Experiment 1.

Experiment 2 contained two sets of stimuli: one set targeted the phonological process of flapping and the other targeted the phonological process of schwa-insertion in past tense formation.⁵ In each set, the stimulus that involves a process was paired with a stimulus that does not involve a process but is as close in form as possible. The division of these sets is provided in Table 11. This allowed for a direct comparison of the frequency of errors in each case.

Table 11. Sets of stimuli used in Experiment 2

⁵ The form *futter* could arguably have a flap as its input instead of deriving the flap from an underlying /t/.

| Set | Process | Token | Sentence |
|-----|-----------------|--------|-----------------------------|
| A | Flapping | futter | This is a futter now |
| | None | fubber | This is a fubber now |
| B | Schwa insertion | kawded | Last week, I kawded outside |
| | None | kawdid | Hey Look, a kawdid is here |

In the flapping group, the nonword containing a flap, *futter*, was compared to the form *fubber* which does not undergo such a process. Each nonword was produced within the carrier sentence '*This is a _____ now*'.

The second group targeted the phonological phenomenon of schwa insertion in the past tense when the base morpheme ends with the alveolar stops /t/ or /d/. This process is illustrated by forms such as *divided* [dɪvəjɹəd] and *knitted* [nɪtəd] as opposed to forms like *pushed* [pʊʃt] or *signed* [sajnd]. The semantic context for the past tense for the nonword *kawded* was cued by the carrier sentence '*Last week, I _____ outside*'. In the non-past tense sentence, context was given for a monomorphemic, nonword noun, *kawdid* by the carrier sentence '*Hey look, a _____ is here*'.

4.1.3 Procedure

Each participant was tested individually in a sound attenuated booth in front of a laptop on which the stimuli were presented using SuperLab 4 (Abboud, Shultz, & Zeitlin, 2008). After a short practice session, the stimuli were presented and recording began. In the control condition, participants produced each of the 200 target sentences at their own pace. The participants were responsible for pressing a key to begin each trial and were able to take a break and a sip of water whenever they wished. The experimental condition, on the other hand, automatically presented each of the 480 sentences every 1500 ms in order to speed up the participant's production and to encourage errors. This condition was divided into three four-minute sessions to allow for a break in between each session. The entire experimental procedure lasted approximately twenty-five minutes.

4.2 Results

As in Experiment 1, acoustic measurements were made using Praat and statistical analysis was performed using SPSS.

Experiment 2a compared the error rates of forms that undergo the rule of flapping to forms that do not, for example, *futter* vs. *fubber*. Experiment 2b compared forms that undergo the phonological rule of schwa insertion in past tense formation to those that do not, for example, *kawded* vs. *kawdid*. They both tested the hypothesis that forms which undergo an additional process are more prone to

errors. This hypothesis was tested by comparing the error rates of tokens that are closely matched in form, but that differ in the presence or absence of a process.

4.2.1 Experiment 2a, flapping

An acoustic analysis was performed on data from thirteen participants to determine the normal range of values from the control condition for measures of percent voicing and duration.⁶ Measures of these variables in the control condition determined normal production values for each participant. As in Experiment 1, tokens that fell two standard deviations from the mean were considered errors. The number of errors in each condition was compared to determine if the tokens that undergo the phonological process of flapping have more errors than tokens that did not.

Experiment 2 focused on determining whether any errors could be the result of a failure of the flapping rule. If an underlying /t/ fails to flap the result would be voiceless and longer in duration than a flap, more characteristic of a [t]. A [t] has a typical duration of ~102 ms while a flap has a duration of ~29 ms (Byrd, 1993). In order to focus on errors of this type, only errors that were more voiceless or longer than a typical flap and [b] as produced in the control condition were counted. All errors that were the result of a target flap or [b] that were shorter or

⁶ Three of the sixteen participants failed to flap consistently.

voiced were, therefore, not counted. Spectrograms illustrating a token with an average duration as well as a token with a duration that was longer than expected are shown in Figure 12 and Figure 13.

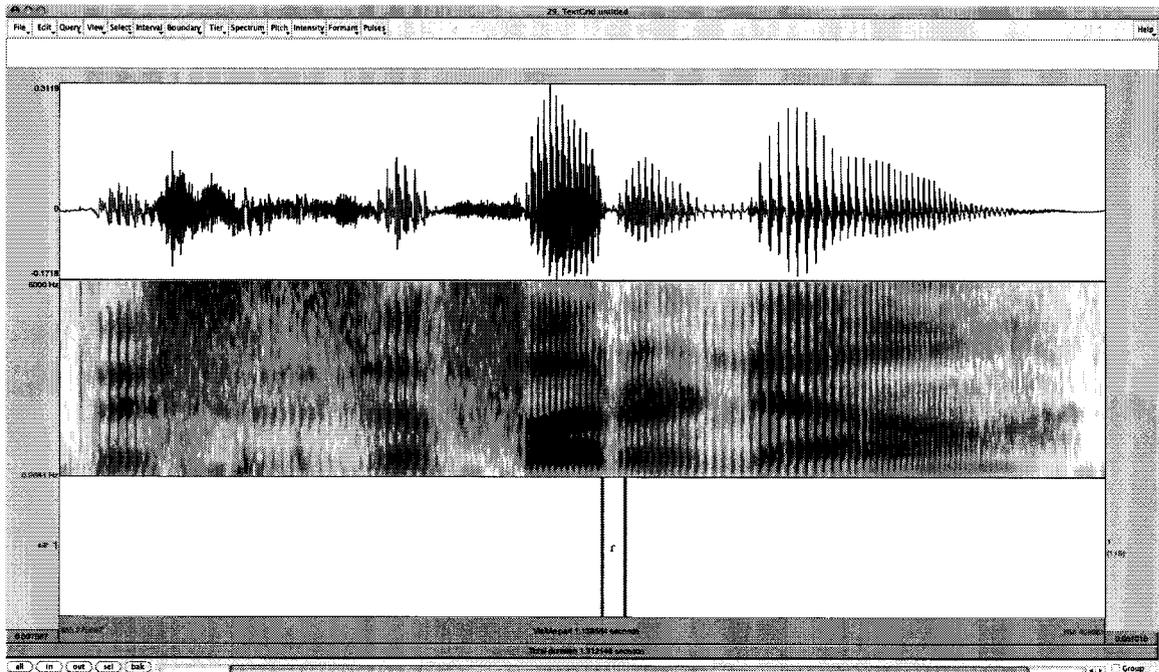


Figure 12. Spectrogram and waveform of the sentence 'I am a futter now' showing a normal duration (24 ms) for the flap.

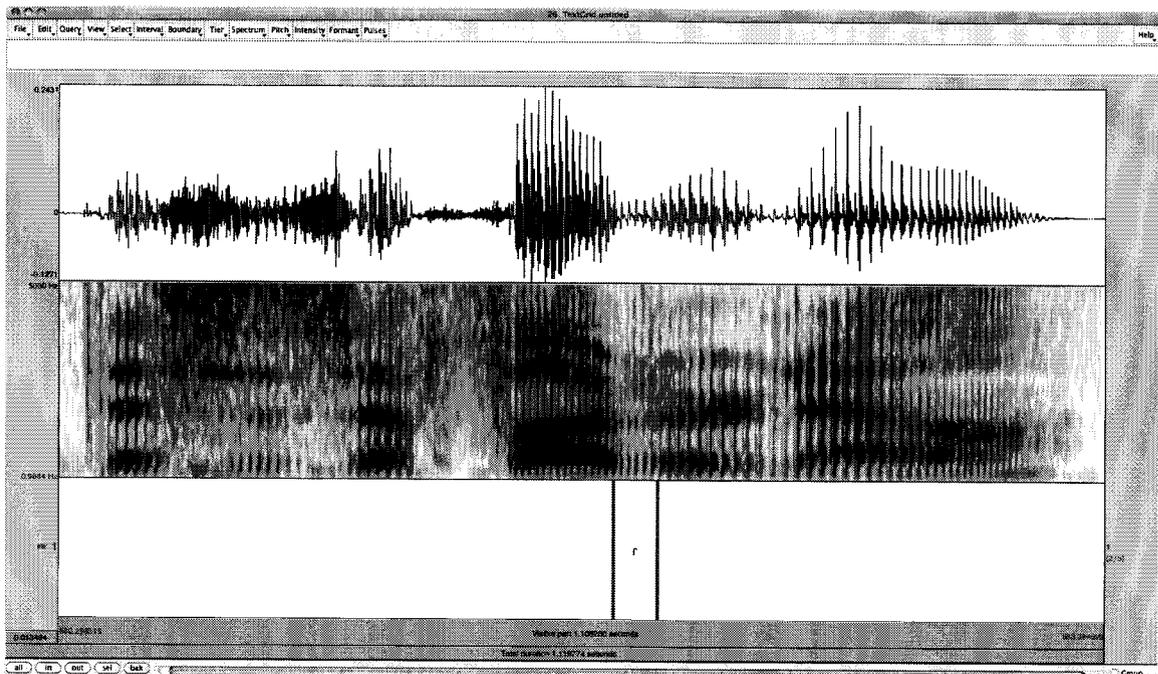


Figure 13. Spectrogram and waveform of the sentence 'I am a futter now' showing a flap with a duration that is long (47 ms).

Given that the experimental manipulation in this case was speed, it was assumed that duration would not be a usable measure as it would necessarily be shorter in the experimental condition as an artefact of the error induction technique used. However, a number of duration errors were detected that were too long despite the quickened pace of the experimental condition. The combined results from all participants for both percent voiceless and duration are therefore presented in Table 12 and Table 13.

Table 12. Error count for *futter* and *fubber* that are more voiceless than expected

| Token | Voicing Error | Total |
|-------|---------------|-------|
| b | 66 | 1552 |
| flap | 55 | 1547 |

Table 13. Error count for *futter* and *fubber* that are longer than expected

| Token | Length Error | Total |
|-------|--------------|-------|
| b | 29 | 1552 |
| flap | 48 | 1547 |

The number of errors that were more voiceless was not significantly different for *futter* and *fubber*, with 66 out of the total 121 errors classified as [b] errors (binomial test, two-tailed, $p = .36$). The results for each participant are shown in Table 14. There were no significant differences in errors of this type for any of the participants.

Table 14. More voiceless than expected errors broken down by participant

| Participant | b | flap |
|-------------|---|------|
| 53 | 3 | 4 |
| 55 | 6 | 16 |
| 56 | 1 | 0 |
| 57 | 1 | 0 |

| | | |
|-------|----|----|
| 58 | 0 | 1 |
| 59 | 5 | 3 |
| 60 | 7 | 1 |
| 62 | 10 | 2 |
| 64 | 8 | 8 |
| 65 | 7 | 10 |
| 66 | 4 | 1 |
| 67 | 3 | 2 |
| 69 | 11 | 7 |
| Total | 66 | 55 |

The number of errors that were longer than expected was significantly greater for the flap than for [b], with 29 out of the total 77 errors classified as [b] errors (binomial test, two-tailed, $p = .04$). The participant breakdown in Table 15, however, shows that one participant contributed 15 of the total 48 flap errors.⁷

Table 15. Longer than expected errors broken down by participant

| Participant | b | flap |
|-------------|---|------|
| 53 | 0 | 4 |
| 55** | 0 | 15 |
| 56 | 0 | 0 |
| 57 | 0 | 0 |

⁷ A single asterisk indicates significance at $p < .05$. A double asterisk indicates significance at $p < .01$.

| | | |
|-------|----|----|
| 58 | 8 | 7 |
| 59 | 3 | 4 |
| 60 | 0 | 1 |
| 62 | 5 | 3 |
| 64 | 2 | 3 |
| 65 | 1 | 0 |
| 66 | 7 | 2 |
| 67* | 1 | 9 |
| 69 | 2 | 0 |
| Total | 29 | 48 |

If participant 55 is removed from the analysis, the difference is no longer significant, with 29 of the 62 total errors classified as [b], (binomial test, two-tailed, $p = .7$).

These errors are based on voicing and duration separately. It could also be the case that the process for flapping does not operate over features separately but that an underlying /t/ becomes a flap at the level of the segment, without reference to the features. In this case, both voicing and duration errors would be expected. Specifically, the errors would consist of a duration that is longer than expected and more voiceless than expected. The count for this type of error is shown in Table 16.

Table 16. Error count for *futter* and *fubber* that are both longer and more voiceless than expected

| Token | Error | Total |
|-------|-------|-------|
| b | 3 | 1552 |
| flap | 6 | 1547 |

The number of errors that were both more voiceless and longer than expected were not significantly different for *futter* than *fubber*, with 3 out of the total 9 errors classified as [b] errors, (binomial test, two-tailed, $p = .5$).

4.2.2 Experiment 2b, schwa insertion in the past tense

Experiment 2b differs from Experiments 1 and 2a in the method used for detecting errors. In this case, the presence or absence of the final vowel and coda in the forms *kawdid* and *kawded* was determined by visual inspection of the spectrogram and waveform and with careful listening. Segmented spectrograms are presented in Figures 14 and 15 that show a target sentence with *kawded* fully produced and a sentence with the final *-ed* in *kawded* missing.

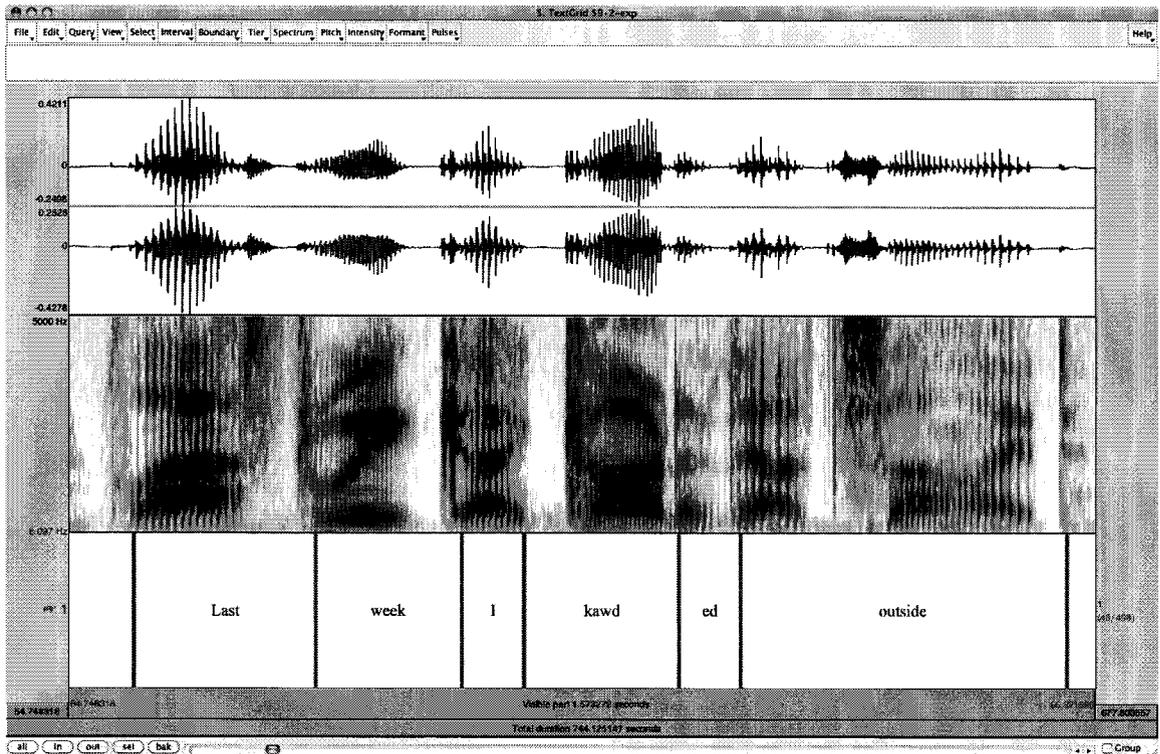


Figure 14. Spectrogram and waveform of the sentence 'Last week, I kawded outside'.

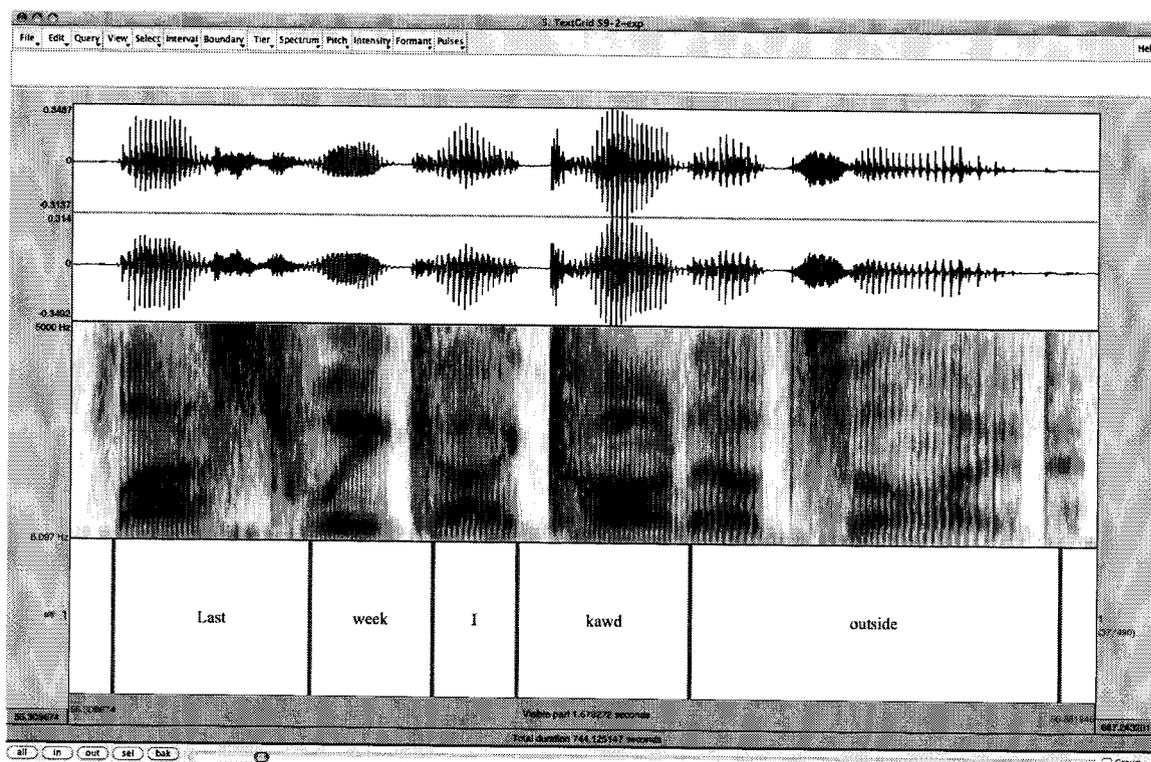


Figure 15. Spectrogram and waveform of the sentence 'Last week I kawded outside' produced without -ed.

Data from sixteen participants were included in the analysis. The total errors produced by all participants in both conditions are presented in Table 17.

Table 17. Total *kawdid* - *kawded* errors

| Token | Error | Total |
|-------|-------|-------|
| ed | 12 | 2718 |
| id | 23 | 2716 |

The number of errors produced was not significantly different for *kawded* and *kawdid*, with 12 out of the total 35 errors classified as -ed errors, (binomial test, two-tailed, $p = .09$). The participant break down shown in Table 18 reveals that one participant contributed ten -id errors and only one -ed error. Differences that are significant are marked with an asterisk.

Table 18. *Kawdid* - *kawded* errors by participant

| Participant | id | ed |
|-------------|----|----|
| 53 | 0 | 0 |
| 54* | 10 | 1 |
| 55 | 1 | 1 |
| 56 | 1 | 0 |
| 57 | 1 | 2 |
| 58 | 2 | 0 |
| 59 | 0 | 3 |
| 60 | 0 | 0 |
| 61 | 0 | 0 |
| 62 | 0 | 0 |
| 64 | 4 | 3 |
| 65 | 0 | 0 |
| 66 | 1 | 0 |
| 67 | 0 | 0 |
| 68 | 3 | 0 |

| | | |
|-------|----|----|
| 69 | 0 | 2 |
| Total | 23 | 12 |

When this participant was removed from the analysis, 11 of the total 24 errors were -ed errors, (binomial test, two-tailed, $p = .8$).

4.3 Discussion

Experiment 2a revealed a significant difference between in the number of duration errors of the ‘too long’ type for *futter* and *fubber*. This effect was very small, however, and must be interpreted cautiously. If rule application errors were more robust, one would also expect to find a distinction in the number of voicing errors which was not the case here. Although the duration errors for flap could be the result of the experimental manipulation, the speeded trials should lead to the opposite effect, producing errors that are too short, not errors that are too long. Further, we would expect that the [b] in *fubber* would be equally subject to any effect the experimental manipulation might have on duration.

Although the sample size of 1675 tokens is quite large, it could be that it would take an even larger sample size to find a differential effect for voicing errors and that a bigger distinction would be found for duration errors. The significant difference found here for duration errors in the form with a phonological rule is, due

to its small effect size, deemed to be suggestive of phonological rule errors, though not conclusive.

Experiment 2b tested sentences that either contained a phonological rule or did not and no significant difference was found in their error rates. The noun *kawdid* was produced without the final *-id* just as much as the past tense form *kawded* was produced without its final *-ed*. As in the previous case, although the sample size of 2717 tokens is quite large, it is still possible that a bigger sample size could reveal a distinction not found here. Based on these data, errors do not appear to occur during phonological processing.

Taken together, the results of Experiments 2a and 2b do not provide sufficient evidence for the existence of rule processing errors. It therefore cannot be concluded that the mismatch errors found in Experiment 1 are caused by phonological rule errors. Two potential sources for mismatch errors remain: after phonological rule processing but before motor execution or within the motor level itself.

The lack of evidence for errors of rule processing could indicate that there are no rules involved in speech production at all. Not all theorists assume a model that includes phonological transformations. For example, Levelt (1989) claims that rules governing alternations exist only for segment interactions between words, not within a word. The current investigation tested only the within-word type of alternation. Current probabilistic theories (Pierrehumbert, 2003) push phonological processes into the domain of phonotactic type co-occurrence restrictions where each

output variation is memorized and, therefore, rules governing alternations are not required.

Having found no evidence that mismatch errors are phonological rule errors, it remained to be seen if mismatch errors could be caused at the motor level of speech production. Experiment 3 was designed to test this possibility.

Chapter 5: Experiment 3

The connection between the abstract mental processing of speech and its physical output can be studied by inducing motor-level errors produced via delayed auditory feedback (DAF, Lee, 1950; Yates, 1963). This technique involves playing the participant's speech back to them amplified and with a delay. Speech produced under this condition demonstrated changes in rate and intensity in addition to repetitions and omissions. For example, Chapin, Blumstein, Meissner, and Boller (1981) report productions such as 'ha... ha... hand' and, 'envelo' for 'envelope'.

Yates (1963) presents a review of early studies that used the DAF technique. He reports on experiments that describe the effects of delayed feedback on speech rate, intensity, fundamental frequency and articulatory changes. Several researchers found that increasing the delay from between 0.06 to 0.18 seconds created a corresponding increase in speech rate (Black, 1951; Atkinson, 1953; Fairbanks, 1955). The intensity of speech was also found to increase with the increase in delay time, though the exact cutoff point was not consistent among all studies (Black, 1951; Atkinson, 1953; Fairbanks, 1955; Spilka, 1954). DAF was also found to increase fundamental frequency (Fairbanks, 1955). Various articulatory changes were described by several studies and include the repetition of syllables, mispro-

nunciations, omissions, substitutions, and additions (Lee, 1951; Atkinson, 1953; Tiffany & Hanley, 1956; Fairbanks & Guttman, 1958).

More recent work demonstrates that speakers compensate and adapt to both pitch shifted feedback (Jones & Munhall, 2000, 2002, 2005; Purcell & Munhall, 2006) and to the manipulation of vowel formants (Houde & Jordan, 1998, 2002).

The nature of articulatory changes that occurred during DAF was investigated in Chapin et al. (1981). This study examined the speech of twenty aphasic and ten control participants. The authors were interested in determining if the articulatory changes found in DAF speech implicated the phonological level or the motor level of production. Errors were transcribed and categorized based on whether they involved phonological units such as the syllable and phoneme or phonetic properties such as the duration of vowels or fricatives. The results showed that only vowel length was significantly affected by DAF. Errors involving phonological units were not significant. This result indicates that DAF does not cause errors at the phonological level but at the level of articulatory implementation.

Further evidence in support of the motor-level nature of DAF errors is derived from the generalizability of errors from other modalities such as finger tapping while whistling (Finney, 1999), and playing a musical instrument while clapping hands (Kalmus, Denes, & Fry, 1955). Due to the motor nature of the errors induced by the DAF procedure, a comparison can be made between DAF-induced errors

and the mismatch errors found in Experiment 1 to determine if they are attributable to the motor level of articulatory implementation.

5.1 Method

Delayed auditory feedback was used during the production of a subset of the stimuli used in Experiment 1. The requirement in Experiment 1 that participants produce the stimuli at the speeded rate set by a metronome, however, was not implemented so that the effects of DAF on speech production could be clearly identified as distinct from the effects of DAF plus paced speech.

5.1.1 Participants

Production data were collected from five native speakers of English with no history of speech or hearing problems. Participants were Carleton University students who received course credit for participating.

5.1.2 Materials

The recording equipment used in this experiment was the same as in Experiment 1. The stimuli used were *kuff* [kʌf], *kuvv* [kʌ:v], *tuff* [tʌf] and *tuvv* [tʌ:v] which were chosen from the stimuli used in Experiment 1 in order to allow a direct comparison of error types. Experiment 1 revealed no significant difference in vowel

length and coda voicing measures between k-initial and t-initial tokens, therefore, these forms were collapsed for the analysis.

5.1.3 Procedure

The procedure in Experiment 3 was the same used in Experiment 1 except that the trials were not paced (i.e., no trials were set to a metronome's pace). Instead, the experimental manipulation consisted solely of delayed auditory feedback. Participants had their own speech played back to them via headphones, amplified and delayed by 200 ms using SpeechMonitor (Arenas, 2009). Stimuli were presented using SuperLab 4. In the control condition, the four nonword set appeared on the screen and the participant was required to repeat this three times. When finished, they pressed a key to call up the next set of nonwords. There was no set lag time in between trials participants initiated each trial and were instructed that they could take a break between trials. Nonwords were always presented in an AAAA pattern, for example, *tuff tuff tuff tuff*. The experimental condition was the same as the control condition with the addition of the delay. There were 16 trials presented in both the control and the experimental conditions. The entire experimental procedure lasted approximately ten minutes.

5.2 Results

Measurements of the data collected in Experiment 3 were made using Praat and statistical analyses were performed in SPSS. Section 5.2.1 describes the general characteristics of delayed auditory feedback on vowel duration and coda voicing by comparing the measures obtained in Experiment 1 and 3. Section 5.2.2 directly compares the types of errors observed in Experiment 3 with those from Experiment 1.

5.2.1 General characteristics of DAF speech

This section compares the duration of vowels and the voicing of codas in Experiment 1 and 3. Experiment 1 consisted of a control condition wherein participants produced the target word 12 times in time to a beat played before each trial. The experimental condition increased the task demand by having participants produce two similar targets in an ABBA pattern, e.g. *tuvv tuff tuff tuv*. Experiment 3 consisted of a self-paced task wherein participants produced the target twelve times and the experimental condition was the same as the control condition except that participants heard their own speech played back to them with a 200 ms delay. This section will compare the results of Experiment 1 and 3 for two measures, vowel duration and coda voicing.

Vowel duration

The mean values for vowel duration for both conditions for voiced and voiceless targets in the two experiments are presented in Tables 19 and 20.

Table 19. Vowel Duration - Experiment 1

| Target | Condition | Mean | Std. Dev. | N |
|-----------|--------------|-------|-----------|------|
| Voiceless | Control | 113.9 | 33.1 | 1150 |
| | Experimental | 111.7 | 41.0 | 2395 |
| Voiced | Control | 153.4 | 29.1 | 1134 |
| | Experimental | 156.8 | 51.1 | 2395 |

Table 20. Vowel Duration - Experiment 3

| Target | Condition | Mean | Std. Dev. | N |
|-----------|--------------|-------|-----------|-----|
| Voiceless | Control | 103.3 | 27.7 | 472 |
| | Experimental | 121.7 | 42.5 | 471 |
| Voiced | Control | 135.3 | 44.5 | 484 |
| | Experimental | 150.0 | 52.9 | 479 |

There was no significant difference in vowel duration for voiceless targets in Experiment 1, $t(2748.794) = 1.664$, $p = .096$ (two-tailed). However, there was a small but significant difference in vowel duration for voiced targets, $t(3414.85) = -2.494$, $p = .013$ (two-tailed), $d = -.09$. There was a significant increase in duration

in the experimental condition of Experiment 3 for both voiceless, $t(807.021) = -7.901$, $p < .001$ (two-tailed), $d = -.56$, and voiced targets, $t(930.619) = -4.668$, $p < .001$ (two-tailed), $d = -.31$.

Coda voicing

The mean values for coda voicing for both conditions in the two experiments are presented in Tables 21 and 22.

Table 21. Coda Percent Voiceless by target voicing – Experiment 1

| Target | Condition | Mean | Std. Dev. | N |
|-----------|--------------|------|-----------|------|
| Voiceless | Control | 89.8 | 13.8 | 1147 |
| | Experimental | 88.5 | 13.6 | 2394 |
| Voiced | Control | 71.4 | 18.7 | 1137 |
| | Experimental | 72.1 | 19.1 | 2395 |

Table 22. Coda Percent Voiceless by target voicing - Experiment 3

| Target | Condition | Mean | Std. Dev. | N |
|-----------|--------------|------|-----------|-----|
| Voiceless | Control | 88.4 | 9.4 | 472 |
| | Experimental | 88.5 | 9.8 | 471 |
| Voiced | Control | 57.0 | 18.4 | 484 |
| | Experimental | 56.9 | 21.9 | 479 |

There was no significant difference in the voicing of the coda for voiced tokens in Experiment 1, $t(3530) = -.921, p = .357$ (two-tailed). Voiceless targets had codas that were significantly more voiced in the experimental condition, $t(3539) = 2.658, p = .008$ (two-tailed), $d = .09$.

There was no significant difference in coda voicing in Experiment 3 for either voiceless targets, $t(941) = -.138, p = .89$ (two-tailed) or for voiced targets, $t(929.311) = .084, p = 0.933$ (two-tailed).

5.2.2 DAF errors

This section compares the types of errors induced by the DAF technique to those induced in Experiment 1 using the tongue twister technique.

The total error rate for Experiment 1 was 8.8% while in Experiment 3 it was 20%. These errors can be broken down into the following types; mismatch errors, accommodation errors, vowel duration errors and coda voicing errors. Accommodated errors are errors for which both the vowel duration and the coda voicing are errors, for example, when *tuvv* [tʌ:v] with a long vowel and voiced coda is produced as *tuff* [tʌf] with a short vowel and voiceless coda.⁸ Mismatch errors are those for which only one of the two components in the rhyme is an error, for exam-

⁸ These errors could also be interpreted as full word errors.

ple, when *tuff* [tʌf] is produced as *tuvv* [tʌv] with a short vowel and voiced coda.

Table 23 presents the error types for both Experiment 1 and 3.

Table 23. Error types for Experiments 1 and 3

| Error Type | Experiment 1 | | Experiment 3 | |
|--------------|--------------|-----|--------------|------|
| | Count | % | Count | % |
| Mismatch | 354 | 7.4 | 184 | 19.4 |
| Accommodated | 66 | 1.4 | 6 | 0.6 |
| Total Errors | 420 | 8.8 | 190 | 20.0 |

Mismatch errors can be further broken down into two types; vowel duration errors and coda voicing errors. The frequency of these types for both experiment are presented in Table 24.

Table 24. Frequency of duration and voicing errors within the mismatch errors for Experiments 1 and 3

| Error Type | Experiment 1 | | Experiment 3 | |
|-----------------|--------------|------|--------------|------|
| | Count | % | Count | % |
| Vowel duration | 140 | 39.5 | 154 | 83.7 |
| Coda voicing | 214 | 60.5 | 30 | 16.3 |
| Total MM Errors | 354 | 100 | 184 | 100 |

Experiment 1 mismatch errors were made up largely of coda voicing errors while Experiment 3 mismatch errors were made up largely of vowel duration errors.

If we consider that these mismatch errors could be the result of rule failure, then the errors can be divided into types that reflect the misapplication of a rule. When the form *tuvv* [tʌ:v] is produced as [tʌ:f], with a long vowel but voiceless coda, the vowel rule is appropriate for the target but the voicing of the coda is not. In contrast, when *tuvv* [tʌ:v] is produced as [tʌv], with a short vowel but voiced coda, it could be the result of the vowel lengthening rule failing to apply. Table 25 shows the distribution of errors based on rule misapplication.

Table 25. Errors as rule misapplication for Experiments 1 and 3

| | Experiment 1 | | Experiment 3 | |
|----------------|--------------|------|--------------|------|
| | Count | % | Count | % |
| a. tʌf > tʌv | 158 | 44.6 | 26 | 14.1 |
| b. tʌ:v > tʌ:f | 56 | 15.8 | 4 | 2.2 |
| c. tʌ:v > tʌv | 102 | 28.8 | 26 | 14.1 |
| d. tʌf > tʌ:f | 38 | 10.7 | 128 | 69.6 |

Error types a and b are cases that cannot be explained by the vowel lengthening rule misapplying as it is the voicing of the coda that differs from the

target, not the vowel length. Type c reflects under application of the rule as the rule fails to apply when it should. Type d reflects over application as the rule applies where it should not. The mismatch errors in Experiment 1 consist mostly of type a errors where the rule misapplication is not a factor and type c errors that indicate rule under application. The mismatch errors in Experiment 3 consist mostly of what could be categorized as rule over application errors.

5.3 Discussion

The productions in Experiment 1 differed from Experiment 3 both generally and in specific error types. Generally speaking, the largest difference in results between control and experimental conditions was in the duration of vowels in the DAF experiment.

While the error rates and proportion of accommodated to mismatch errors were similar for the two experiments, other types of errors were different. Within the mismatch errors, Experiment 1 had more coda voicing errors while Experiment 3 had far more vowel duration errors.

The error pattern of Experiment 3 was quite different than that seen in Experiment 1. Type d errors were by far the most common in Experiment 3, however, given the general lengthening property of DAF speech discussed in section 1, it is likely that the type d errors produced in Experiment 3 were not caused by rule

over application. Since the errors induced in Experiment 1 did not share the general property of lengthening found when using the DAF technique, it suggests that the type a, b and c errors found in Experiment 1 were not the result of motor level issues. The motor level errors in Experiment 3 were not consistent with those found in Experiment 1. The conclusion drawn here is that the mismatch errors in Experiment 1 were not induced at the motor level of speech production.

This experiment, however, can not conclusively exclude a motor level account. While the errors induced using the DAF technique are reported to be motor-level errors, they are not necessarily the only kind of motor level errors. DAF tends to slow speech, lengthening speech sounds. Errors of the motor level can presumably occur as the result of the failure of a mechanism other than the one triggered by the DAF technique. Further, Mowrey and MacKay (1990) provide evidence suggesting that what are described as feature and phoneme errors are also consistent with a motor-level account. Results from an electromyographic study show that even in productions that are judged as error-free, there are anomalous motor level intrusions. These motor level intrusions could be responsible for those errors that are categorized as feature or phoneme errors.

The next chapter presents a Python⁹ ACT-R model that will operationalize the processing steps discussed in this work. The first model consists of a system that can generate errors after phonological processing. The second model can generate errors both during phonological rule implementation and after this, at the speech planning level.

⁹ Python is a programming language. ACT-R is a cognitive modelling architecture used to model many different kinds of cognitive processes. "Python ACT-R" is a version of ACT-R written in the Python programming language. More detail is provided in Section 6.1.

Chapter 6: Python ACT-R models

This section presents a model of production (Gormley & Stewart, 2009) that follows the information flow necessitated by the findings of Experiment 1. Two versions of the model will show the two options that explain mismatch errors; that errors can occur after phonological processing or that errors occur during the processing of phonological rules. The model was created using Python ACT-R (Stewart & West, 2007) and uses noisy recall to produce speech errors.

6.1 Model 1

This model is constructed in Python ACT-R (Stewart & West, 2007) which was developed to make ACT-R more accessible and therefore more broadly used by researchers. Python ACT-R is an adaptation of ACT-R, a cognitive architecture developed by Anderson and colleagues (e.g., Anderson & Lebiere, 1998). ACT-R is intended to be a general purpose system for modelling any cognitive process.

The model uses the two fundamental components of an ACT-R system, the memory module and the production module. The memory module stores long-term information such as phonemes and their features. The production component defines the operations and the sequence in which they take place. In contrast with standard ACT-R, Python ACT-R allows for four separate special-purpose memory

components, each of which can act in parallel. The feature store is a long-term static memory that holds the features associated with each phoneme. The first and second speech planning modules are short-term memories that hold the phonemes currently being processed. Finally, the rule memory stores the rules to be applied.

Errors can be made during either or both of the two speech planning components. Two mechanisms shape these errors. The first is the noise level that interferes with proper recall resulting in the retrieval of an incorrect item. The second mechanism is a partial similarity matching mechanism that drives the confusability of similar items such as [s] and [z].

Information is processed through the model procedurally beginning with the first speech planning module and then on to the phonological component and finally to the second speech planning module. A subset of the nonwords used in Experiment 1 are input into the system and converted to a string of phonemes. The features associated with each phoneme are then retrieved from the feature store. These features are associated with the proper syllable position and order information and are then placed in the first speech planning module. This speech planning module is an extremely short-term memory that stores the output of the initial processing and then makes it available as input to the next processing component: the phonological rule module. If an incorrect item is passed into the phonological rule module, the result will be an accommodated error since the incorrect informa-

tion will still be passed through the phonological system acquiring the appropriate surface phonology.

The phonological rule module applies the rule of vowel lengthening to forms that fit the condition of having a voiced coda. Each segment is retrieved from the first speech planning module and the voicing of the codas are determined. If the coda has the feature [voiced], the rule is retrieved from the rule buffer and the preceding vowel is lengthened. The output of the rule module is then passed to the second speech planning module in sequence based on the associated order information. It is during this retrieval process that mismatch errors may occur. Because the phonological rule has already applied, the retrieval of an [s] instead of a [z] will yield the form *tiis* [tɪ:s] with a long vowel that is inappropriate before the voiceless [s]. An example of error-free processing is illustrated in Table 26. Examples of the procedure that results in an accommodated error are shown in Table 27 and the procedure that results in a mismatch error is shown in Table 28.

Table 26. Example of correct processing

| Step | | Stored Value |
|------|-------------------|--------------|
| 1 | Input phonemes | tiztis |
| 2 | Retrieve features | tiztis |
| 3 | Phonological rule | tiiztis |
| 4 | Output | tiiztis |

Table 27. Example of processing an accommodation error

| Step | | Stored Value |
|------|--|--------------|
| 1 | Input phonemes | tiztis |
| 2 | Retrieve features (error: got features for z instead of s) | tiztiz |
| 3 | Phonological rule | tiiztiiz |
| 4 | Output | tiiztiiz |

Table 28. Example of processing a mismatch error

| Step | | Stored Value |
|------|-------------------------------------|--------------|
| 1 | Input phonemes | tiztis |
| 2 | Retrieve features | tiztis |
| 3 | Phonological rule | tiiztis |
| 4 | Output (error: got i instead of ii) | tiztis |

An incorrect retrieval from the feature store or from the speech planning module is not random. The model incorporates a partial similarity matching mechanism to reflect the tendency for speech errors to confuse elements that are similar in form and structure such as two onsets or two alveolar stops.

6.1.1 Model 1 Results

To test the accuracy of the model, error rates were compared to the human results from Experiment 1 which are reproduced in Table 29. The model's overall error rate should be comparable to that of the Experiment 1 participants. The percent correct rate from Table 10 is 81.8%. However, there were also a large number of unclassified responses. The conservative hypothesis is that the unclassified tokens could be from any category and as such the percentage of correct responses is between 81.8% and 91.2%.

Table 29. Error types from Experiment 1 and Model 1

| | Exp 1 % | Possible range % | Model 1 % |
|--------------|---------|------------------|-----------|
| Correct | 81.8 | 81.8-91.2 | 84.6 |
| Mismatch | 7.4 | 7.4-16.8 | 9.8 |
| Accommodated | 1.4 | 1.4-10.8 | 5.5 |
| Unclassified | 9.4 | - | - |
| Total Errors | 18.2 | - | - |

The model produces results that do not differ significantly from the human data when the parameter setting for noise was set to between -0.82 and -0.8 and the similarity parameter between 's' and 'z' was set to between -0.9 and -1. The results shown in Table 29 are from the noise parameter set to 0.18 and the similarity

parameter set to -0.9. Although this model fits most of the human data, it misses a distinction made within the mismatch error category. This distinction was captured by the second model presented below.

6.2 Model 2

A second possible explanation for mismatch errors is that they are derived from an error in the processing of phonological rules. The idea that errors can arise from within the phonological component itself has not been widely explored, however, the failure of a phonological rule to apply would result in an output that looks like a mismatch error. The second model presents this possibility.¹⁰

This model is the same as the first except that the potential for error during the application of phonological rules is introduced. This additional type of error is triggered by adding random noise during rule retrieval. If rule retrieval fails, the phonological rule will not be applied resulting in a mismatch error. The probability of a rule failing was kept at the same value used for random noise in the rest of the model in order to maintain simplicity. Examples of the procedures that lead to an accommodation or a mismatch error in this model are illustrated in Tables 30 and 31.

¹⁰ The full model is provided in Appendix D.

Table 30. Example of processing an accommodation error

| Step | | Stored Value |
|------|--|--------------|
| 1 | Input phonemes | tiztis |
| 2 | Retrieve features (error: got features for z instead of s) | tiztiz |
| 3 | Phonological rule | tiiztiiz |
| 4 | Output | tiiztiiz |

Table 31. Example of processing a mismatch error

| Step | | Stored Value |
|------|--|--------------|
| 1 | Input phonemes | tiztis |
| 2 | Retrieve features | tiztis |
| 3 | Phonological rule (error: the rule fails to apply) | tiztis |
| 4 | Output | tiztis |

6.2.1 Model 2 Results

The model produced a pattern of results similar to the data from Experiment 1 when the noise parameter was set between -1 and 0.86 and the similarity rating for 's' and 'z' were set to between -1 and -0.4. Figure 16 shows results from the model with noise set to 0.12 and similarity to -0.8. The percent of resulting ac-

accommodated errors is 3%, the percent mismatch errors is 11.3% and percent correct is 85.6%

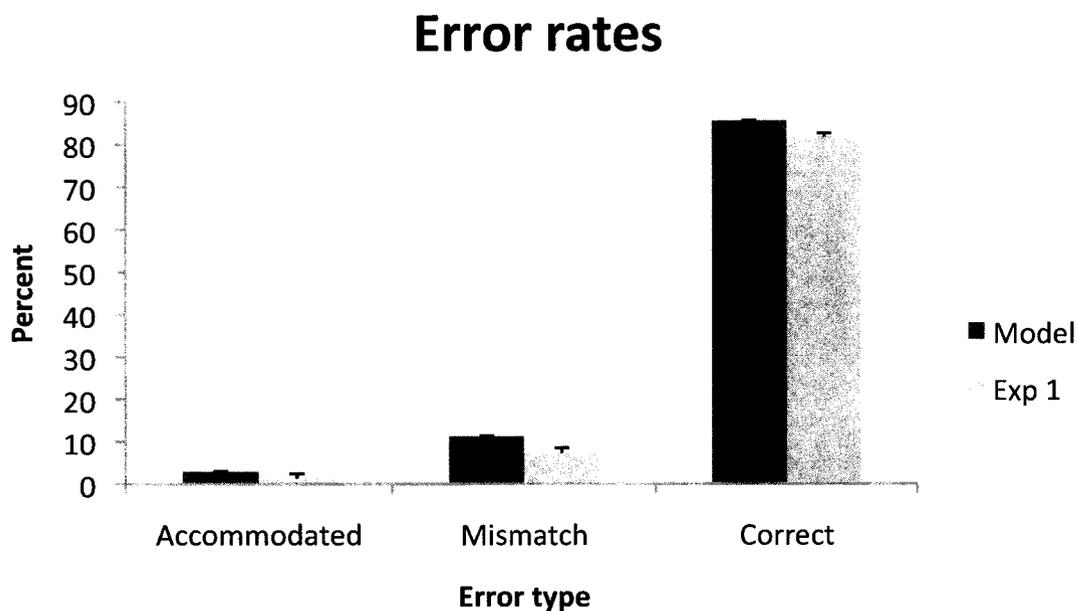


Figure 16. Rates for model 2 showing accommodated errors, correct forms and mismatch errors (+SE) from the model and Experiment 1 data.

This model produces an asymmetry in the data. The errors it generates can be classified as rule under application rather than rule over application. There could not be a case where the rule over applies, for example, the lengthening of the vowel in *tis* [tɪs] to *tiis* [tɪ:s]. This asymmetry corresponds to the error data from Experiment 1 in which a strong tendency was found for *tiiz* [tɪ:z] to be produced as *tiz*

[tɪz] more often than *tis* [tɪs] was produced as *tiis* [tɪs]. The error counts and the percent out of a total 4790 tokens produced are shown in Table 32.

Table 32. Asymmetry of vowel length mismatch errors among Experiment 1 tokens

| | N | % |
|-------------------------------------|-----|------|
| Rule under application - tɪ:z > tɪz | 102 | 2.13 |
| Rule over application - tɪs > tɪ:s | 38 | 0.79 |

Model 2 produces error rates that match the data from Experiment 1 and, unlike the first model presented, this version was able to capture the asymmetry in the error rates shown in Table 32 for *tiiz* [tɪ:z] versus *tis* [tɪs]. The Experiment 1 data shows that errors of the form *tiiz* [tɪ:z] produced as *tiz* [tɪz] outnumber the errors of the form *tis* [tɪs] produced as *tiis* [tɪs]. This asymmetry can be explained if errors were made during the processing of phonological rules since a failure of this rule to apply would result in a short vowel before a voiced coda but not in a long vowel before a voiceless coda. Figure 17 shows the error rates for *tiis* [tɪs] (0.52%) versus *tiz* [tɪz] (10.8%) in the second model.

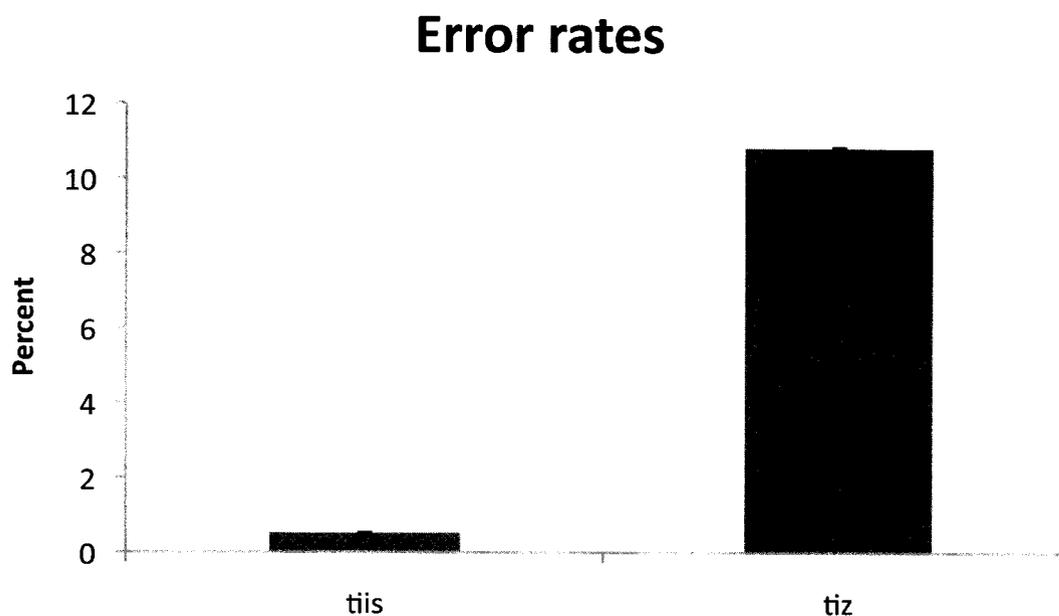


Figure 17. Error rates (+SE) from model 2 for *tis* produced as *tiis* [tɪ:s] and *tiiz* [tɪ:z] produced as *tiz* [tɪz].

The second model better captures the data from Experiment 1 as it allows for errors at all three levels of processing as opposed to restricting errors to the speech planning components.

6.3 Discussion

The goal of these models was to proceduralize the claims made about the location of speech errors made in this dissertation. Results from Experiment 1

showed that the surface phonology does not always conform to its erroneous environment. First, this could show that speech errors can occur after phonological rules are processed, either in a speech planning component or during motor level implementation. Second, it could also show that the errors can be due to the failure of a phonological rule to apply. The two models presented here were created to evaluate these possibilities. While both models could match the general pattern of data from Experiment 1, only the model in which errors could result from the failure of the rule to apply in addition to retrieval errors that occur after phonological processing can capture the asymmetry in the empirical results. This finding was not borne out by Experiment 2 which found no evidence for errors during rule processing. Therefore, until evidence of rule level errors are found, the first model presented here is assumed.

Chapter 7: Summary and Conclusions

The goal of Experiment 1 was to test the assumption that speech errors accommodate. Once mismatch errors were found, the second and third experiments were designed to determine the level of processing at which these errors are generated. Mismatch errors could be caused by a failure of phonological rule implementation, they could arise at the motor level, or between these two stages.

Experiment 2 was designed to determine if mismatch errors could be generated during the processing of phonological rules. By comparing the frequency of errors in conditions that differ only by the presence or absence of a phonological rule, it can be determined if errors of phonological processing contribute to the overall error count. No evidence was found that would indicate that errors of phonological rule application are possible.

Experiment 3 was designed to test articulatory implementation as a possible source of mismatch errors. A delayed auditory feedback task was administered that compared the nature of errors with the mismatch errors discovered in Experiment 1. The results showed different characteristics for DAF and mismatch errors leading to the conclusion that mismatch errors are not motor-level errors.

The results from Experiment 1 suggest that errors can be made after phonological processing. Further, the skewed nature of the results indicates that errors could be made during phonological rule implementation as well. A computational

model based on this result was presented in Chapter 6. It demonstrated how errors made during phonological rule processing and after could account for the error data. Another version of this model was presented that makes errors only after phonological processing. This was the version of the processing component indicated by the results of Experiment 2.

7.1 Theoretical Implications

Phonological Theory

One goal of this research was to test the predictions made by various phonological theories. Generative theories process an underlying representation producing a distinct output. While the theories vary in detail, they should all show errors attributable to their processing. Rule based, procedural theories such as SPE might even show errors at the different stages of rule implementation. Experiment 2 failed to show errors of phonological processing. This result could indicate that phonological processing levels are not vulnerable to errors of this kind, or that errors of this level were not properly induced by the methods used in Experiment 2, or that the generative theories do not accurately describe the processing involved in speech production. The failure to detect phonological transformation errors weakens the generative claim of transformational processing. This is consistent with the

view presented by probabilistic models that maintain that phonological processing is not transformational.

Optimality Theory shares predictions with both SPE and Probability Theory. Like SPE, OT processes apply to an input, producing the appropriate output. In this sense, it is generative. If the OT mechanism that selects a winning candidate via a constraint ranking were to fail, we would expect random outputs. For the input /kæt/, we could expect outputs ranging from [kæt], without the expected aspirated [k^h], to gibberish [babababa]. While no cases of complete gibberish were produced, under the OT framework any error could be attributed to a missed constraint.

The OT concept of markedness makes another prediction similar to the one made by probabilistic models. Markedness states that certain sounds and combinations of sounds are less common in the languages of the world because they are 'marked' within the universal language faculty. For example, [t] is considered unmarked because it is common in languages and its position is less restricted. Markedness theory explains this distribution by stating that [t] is universally favoured by the language faculty. One method for determining markedness is the frequency of use within a particular language. Applying this concept to speech errors, errors of marked segments should be more prevalent than unmarked segments (see Goldrick & Daland, 2009, for a similar analysis). All else being equal, OT predicts

that errors of marked structures will emerge as unmarked structures (McCarthy & Prince, 1994).

The concept of probability also accounts for the observation that infrequent forms are more likely to err, resulting in a more frequent form (Hay, Pierrehumbert & Beckman, 2004). This idea differs from markedness as it does not claim that the frequency distribution is caused by the nature of the language faculty, rather that an individual's speech environment influences their speech production. Sounds that are heard and said more frequently become more consistently represented in a speaker's grammar and are therefore more resistant to error. The concepts of markedness and probability are very different conceptually, yet they both predict that error rates will be higher for less frequent forms.

Data from Experiment 1 can address this prediction by comparing the number of duration errors for the four vowels and voicing errors for the four coda fricatives. Probabilities were calculated using a phonotactic probability calculator (Vitevich & Luce, 2004). The probabilities generated by this calculator are based on the positional frequency of each segment derived from the approximately 20,000 words in the 1964 Merriam-Webster Pocket Dictionary.

Table 33 presents the frequency data for these segments in their respective positions as well as the error rates for these segments as determined by Experiment 1.

Table 33. Error rates and probabilities

| Segment | Probability | Error rate % |
|---------|-------------|--------------|
| ε | 0.0729 | 8.5 |
| ɪ | 0.0962 | 13.8 |
| ʌ | 0.0392 | 15.2 |
| æ | 0.0794 | 10.9 |
| s | 0.0788 | 12.7 |
| z | 0.0201 | 5.0 |
| f | 0.0197 | 7.9 |
| v | 0.0236 | 4.4 |

Previous studies have found that error rates are lower for more frequent forms (Kupin, 1982; Dell, 1990), or that errors are more likely to result in a higher frequency form (Levitt & Healy, 1985; Goldrick, 2002). Figure 18 presents the same data in a bar chart for easy comparison. Error rates should correspond to the inverse of the probabilities as more frequent segments should err less often.

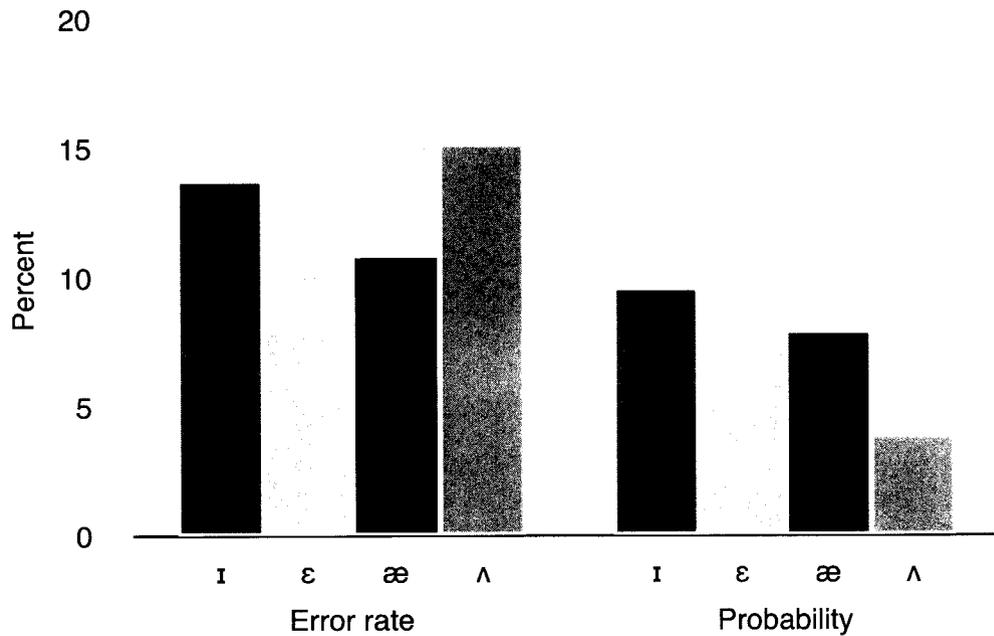


Figure 18. Bar graph showing the relative error rates compared with the probabilities for the four vowels.

The error rates for the voicing of the four coda fricatives as determined by Experiment 1 along with the probabilities are presented in the bar graph in Figure 19. The error rates should correspond to the inverse of the probabilities as more frequent segments should err less often.

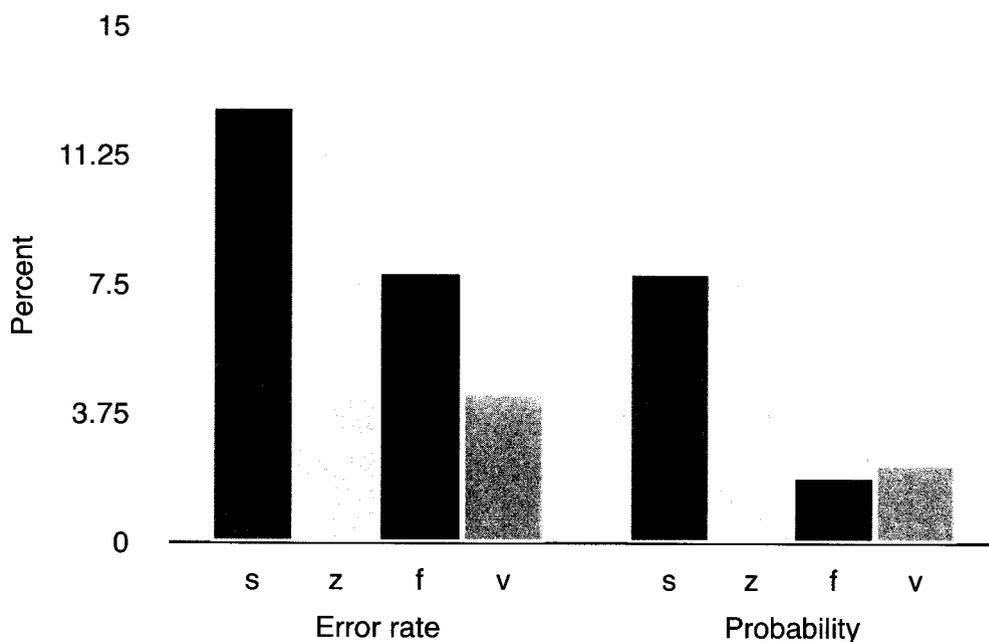


Figure 19. Bar graph showing the voicing error rates for four coda fricatives compared to their probability.

The relative probability of the segments does not match the overall error rates as predicted by the concepts of markedness or probability. The most infrequent vowel of the four, [ʌ] does have the most errors, and [æ] has the third highest error rate, however, the other two vowels do not match the prediction. The vowel [ɪ] is the most common and should have the fewest errors, however it has the second highest error rate. None of the coda segments are consistent with the prediction. The most frequent coda consonant of the four is [s] which has the most errors.

The error rates in Experiment 1 correspond to the frequency of the segments, however, not in the expected direction. This relationship is illustrated by the

scatter plot in Figure 20. A positive correlation, $r(6) = .63$, $p = .09$, was found between the error rates and probabilities; as the frequency increases so does the error rate.¹¹

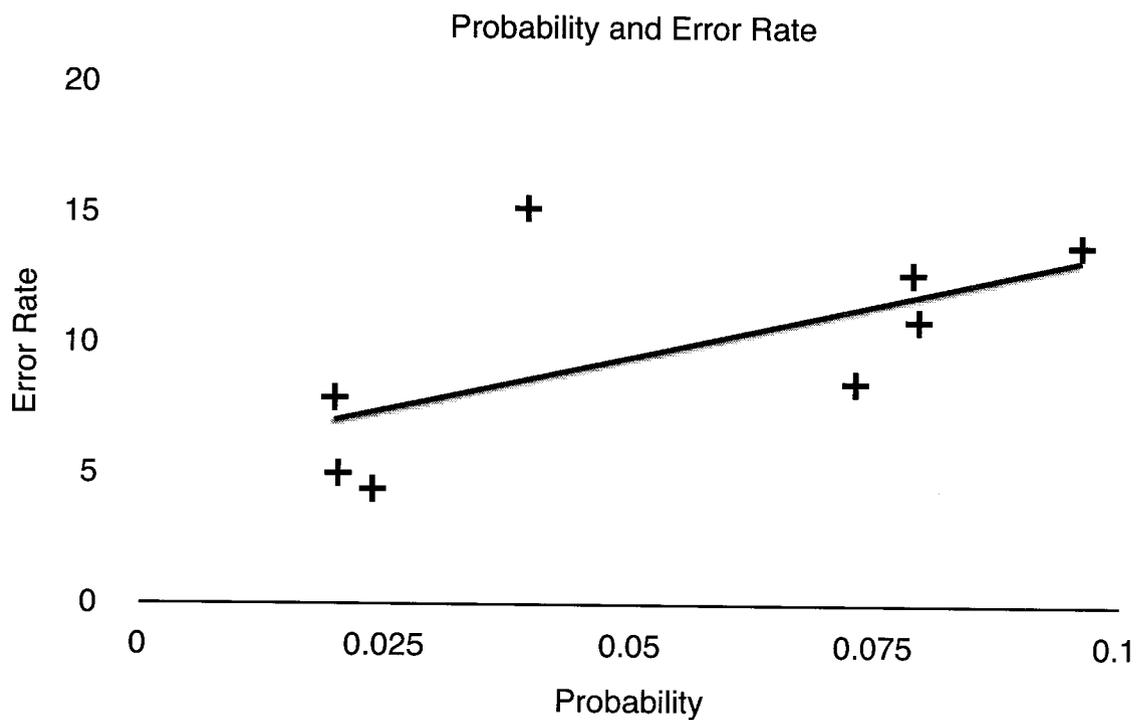


Figure 20. Scatter plot showing the error rate and probability for the eight segments in Experiment 1.

The set of stimuli from Experiment 2 allows a comparison of 'b' and 't'. Both markedness and probability theories predict that errors should be more common in the form *fubber* than in *futter* as 'b' in this context is less frequent. According to the phonotactic probability calculator (Vitevich & Luce, 2004), the 'b' in

¹¹ That the correlation is not significant at the $p < .05$ level may be due to the small sample size.

fubber has a probability of .026 while the 't' in *futter* has a probability of .066. The overall error rates for both voicing and duration determined for these segments in Experiment 2 are 9.1% for 'b' and 11.9% for 't'. A bar graph showing the error rates and probabilities are shown in Figure 21. The error rates should correspond to the inverse of the probabilities.

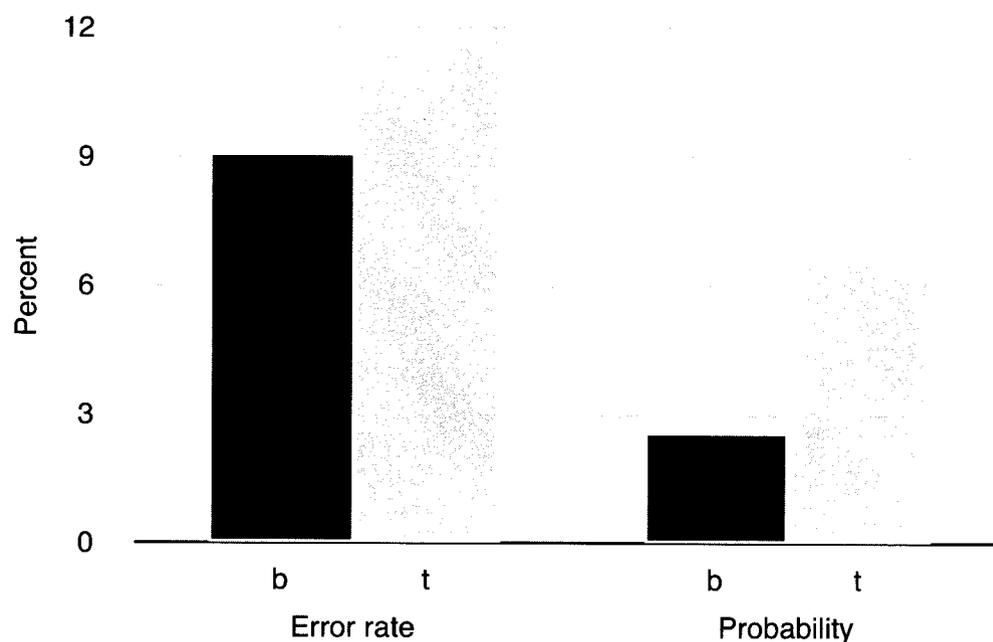


Figure 21. Error rates for the 'b' and 't' in Experiment 2 compared with their probabilities.

The prediction from markedness and probability is not borne out by the error rates for any of the three groups of segments. The relative probabilities of the vowels indicate that [ɪ] should have the fewest errors as it is the most frequently occurring vowel in this context. Likewise for the coda fricatives where [s] should have

the fewest errors as it is the most frequent. The 't' in the *futter* set should have fewer errors than 'b' as it is more frequent in this context. Collectively, the forms in Experiment 1 show a positive correlation whereas Markedness and Probability both predict a negative correlation.

One of the goals of this work was to demonstrate how speech error data can be used to test aspects of phonological theories. In the two cases presented here, the data contradicts the predictions made by the concepts of markedness and probability for the relative frequency of errors. Markedness draws on several sources to calculate markedness including cross-linguistic and language specific frequencies, rate of acquisition and language change. The relative frequencies calculated here only use language specific frequency and do not, therefore, represent a complete picture of markedness. The comparison presented in this chapter of relative frequencies with speech error data is best viewed as a demonstration of how speech error data can be used to assess various elements of phonological theory.

Models of Language use

The models of language use described in the first chapter differ from phonological models in terms of the specific phenomena they describe. Models of lan-

guage use do not consider phonological processes, such as the [t] - flap alternation in English, but account for the kinds of processing that can be inferred by speech errors. Both the procedural model presented in Levelt (1989) and the spreading activation model presented in Dell (1986) account for the association of content and structure. During speech production, segments are united with syllable structure to produce an utterance. Failures in the association of these two types of representation are appealed to in the description of the processes involved in speech production.

The results of the three experiments described in this dissertation have implications for the processing levels described by these models. Fromkin (1971) describes a production model, part of which is shown in Figure 22, wherein errors are said to occur before the morphological rule stage. This model is typical of other procedural models such as the one described by Levelt (1993) shown in Figure 6. Errors were said to occur at this level because morphological and phonological level rules appeared to be respected in the speech error data collected through transcription. The results of the current research suggest that there is a stage after phonological rule processing where mismatch errors can occur.

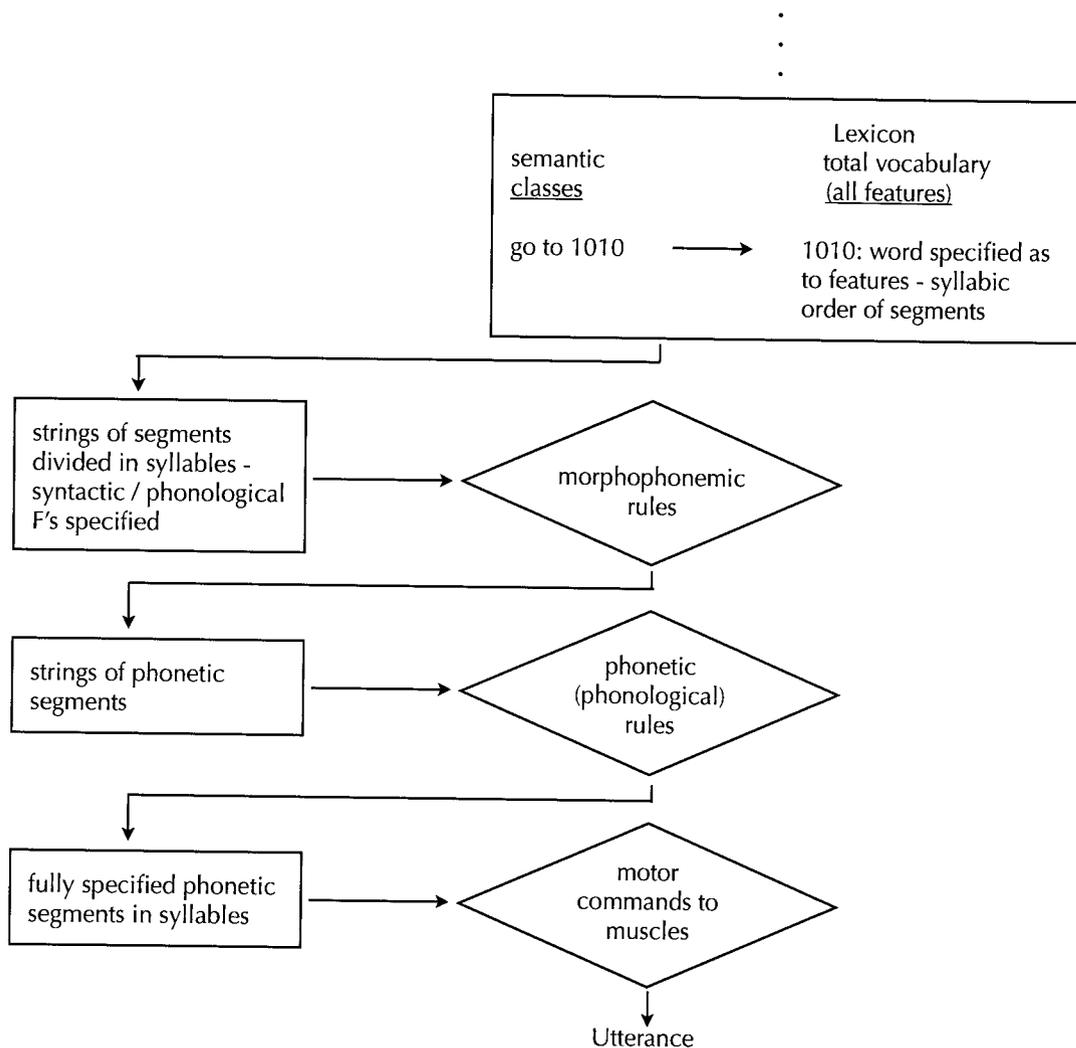


Figure 22. Part of Fromkin's (1971) model of speech production. Based on transcription studies, errors were assumed to occur only before the morphological rules stage (from Fromkin, 1971).

Various levels of speech planning are currently assumed to account for speech errors at lexical, morphological and phonological stages of production. In addition to these levels, we need to consider a stage after phonological processing where errors can occur.

Selection versus Transformation errors

Evidence was found in Experiment 1 suggestive of errors for phonological rule implementation. Of the phonologically mismatched errors, 40% could be categorized as rule under or over-application errors. The remaining 60%, however, can not be explained by rule failure. No further evidence of phonological rule errors was found in Experiment 2 even though this experiment was designed specifically to uncover such errors. An alternative explanation for the mismatch errors detected in Experiment 1 is that they are a type of selection error. An error is a selection error when one unit is replaced by another. Typically, the switched units share content or structural properties or both, such as the error 'leading list' for target 'reading list' (Dell 1986). In this error, the [l] and [r] have shared features and the structural position of the elements, syllable onset, is the same. Examples of possible selection errors at each level of production are given in Table 34.

Table 34. Examples of selection errors at various levels of production

| Level | Target | Error | Unit |
|---------------|--------------------------------------|--------------------------------------|--------------|
| Lexical | your grandmother's <i>funeral</i> | your grandmother's <i>wedding</i> | lexical item |
| Morphological | an important person | a important person | morpheme |
| Phonological | that's a sad cat | that's a sat cad | phoneme |

| | | | |
|-------------------|---------------------|---------------------|-------------------------|
| Post-phonological | [tu:v] (long vowel) | [tuv] (short vowel) | articulatory feature |
|-------------------|---------------------|---------------------|-------------------------|

The mismatch errors found in Experiment 1 can be described as selection errors. The stage of speech production that follows phonological processing is articulatory implementation. Errors at a post-phonological level could be the misselection of the units required by the articulatory processing stage. A short vowel before a voiced coda could be caused by the selection of the articulatory feature that results in a short vowel instead of a long vowel. This level corresponds to what Levelt (1989) refers to as the articulatory buffer. The articulatory buffer is a storage level between the phonological encoding stage and the articulator. The buffer stores the phonetic plan outputted by phonological encoding which is retrieved incrementally by the articulator, creating motor commands.

This account of mismatch errors as selection errors instead of phonological rule errors is supported by the results from syntax. As discussed in Chapter 1, studies have failed to show errors of syntactic transformations that could not be explained by selection errors.

Three hypotheses

The three hypotheses investigated in this dissertation were:

1) All errors are accommodated

The results from Experiment 1 showed that this hypothesis is not supported. Non-accommodated errors were found.

2) Phonological transformations are processed during speech production and are prone to errors that are detectible by acoustic analysis

The results from Experiment 2 did not support this hypothesis. While more data may be required to determine whether or not phonological rules can err, there is currently no evidence to support such a view.

3) Mismatch errors are motor-level errors

The results from Experiment 3 revealed a distinction between the motor level errors induced using the DAF technique and mismatch errors.

These results reveal the conceptually murky boundary between the phonological/speech planning level and the motor level of speech production. Mismatch errors could reflect either the over application or under application of phonological rules. However, since Experiment 2 indicated that they do not arise during phonological rule processing, it must be concluded that they occur after this stage. The speech planning level implicated by these results must, therefore, be processed after phonological rules but before the level that is disrupted by the DAF

technique since the characteristics of mismatch errors did not correspond to the errors induced in Experiment 3. Taken together, the results point to a level between the phonological and motor levels. It could be the case that these errors are articulatory selection errors that are processed at the beginning of motor implementation but before the stage that is impacted by the DAF technique used in Experiment 3. This interpretation is the simplest as there is currently no evidence to support the need for a transformational account of errors in addition to a selectional account.

7.2 Limitations and Future Work

The results of the three experiments presented in this dissertation are based on assumptions about the nature of speech production. It is possible that no evidence of phonological rule errors were detected in Experiment 2 because phonological alternations are not processed as is currently described by phonological theory. It is also possible that the alternations selected for testing are exceptions and their distinct outputs are simply memorized by speakers. Finally, it is possible that not enough data were collected in order to reveal significantly more errors for tokens that involve phonological rules.

The results of Experiment 3 showed that mismatch errors do not have the same characteristics as the errors induced using the DAF technique. While this

technique is said to induce motor level errors, it is not necessarily the case that DAF-induced errors are the only errors attributable to the motor level.

Future studies could be carefully designed to test the interpretation of mismatch errors as articulatory selection errors. The idea that sub-phonemic errors do not need an explanation in higher level processing such as phonology or speech planning was proposed by Mowrey and MacKay (1990). Mismatch errors may belong to the level of articulatory implementation whose features resemble more closely those described by Articulatory Phonology (Browman & Goldstein, 1986). These features are articulatory in nature and are arranged over time in a gestural score. This avenue of research may provide a more elegant explanation for mismatch errors.

Finally, the methodology for error detection in Experiment 1, 2a and 3 should be validated in future experiments. A validation experiment could be designed to test errors of various acoustic measures by calculating the normal range for a segment and determining if known errors are detected. For example, once the normal range for various measures of [t] is determined, a speaker could be asked to produce a set of forms containing [d]. The [d] forms should be calculated as errors with respect to the normal range for [t]. Various segments can be tested this way to determine if certain segments or acoustic measures are more reliable than others in these testing conditions.

7.3 Concluding Remarks

This dissertation presented an acoustic analysis intended to readdress the claim that speech errors always accommodate. Experiment 1 provided data from induced speech errors that indicated the presence of mismatch errors. These sub-phonemic errors could be caused either during phonological processing or, at the motor level of articulatory implementation or, at an intermediary level.

Experiment 2 was designed to assess if mismatch errors are the product of the level that processes phonological rules. No difference was found, however, between the number of errors in the set that contained a rule versus the set that did not contain a rule. The presence of a phonological rule did not, therefore, contribute to the error count. Experiment 3 was designed to compare the mismatch errors of Experiment 1 to known motor-level errors. The errors induced using delayed auditory feedback had different properties than the mismatch errors, suggesting that mismatch errors are not errors of the motor level. The source of mismatch errors appears to be a level between phonological rule processing and the motor level that is affected by the DAF technique. It was suggested that there may be a speech planning level where articulatory features are selected, and in some cases mis-selected in the creation of an articulatory plan.

The results of this study are important for both psychology and linguistics. Psychologists are not generally concerned with phonological processing. However, if further research determines that errors can arise from this level, then phonological processes would become relevant. This dissertation presented speech error data that refuted the claims made by three phonological theories. The study of speech errors should become a more commonly used tool to test the validity of linguistic theories.

References

- Abboud, H., Shultz, H., & Zeitlin, V. (2008). SuperLab 4: Stimulus Presentation Software (Version 4.0.6) [Computer Software]. San Pedro, CA: Cedrus Corporation.
- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.
- Arenas, R. (2009). Speech Monitor [Computer Software]. Retrieved August 1, 2009, from <http://myweb.uiowa.edu/rarenas/index.htm>
- Atkinson, C. J. (1953). Adaptation to delayed sidetone. *Journal of Speech and Hearing Disorders, 18*, 386-391.
- Baars, B. J., & Motley, M. T. (1974). Spoonerisms: Experimental elicitation of human speech errors. *Abstract in Catalogue of Selected Documents in Psychology, 4*, 118.
- Baars, B. J., Motley, J. T., & MacKay, D. (1975). Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14*, 382-391.
- Berg, T. (1987). The case against accommodation: Evidence from German speech error data. *Journal of Memory and Language, 26*, 277-299.
- Berko, J. (1958). The child's learning of English morphology. *Word, 14*, 150-177.

- Black, J. W. (1951). The effect of delayed sidetone upon vocal rate and intensity. *Journal of Speech and Hearing Disorders, 16*, 56-60.
- Boersma, P. (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam, 21*, 43–58.
- Boersma, P., & Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry, 32*, 45-86.
- Boersma, P., & Weenink, D. (2006). Praat: doing phonetics by computer (Version 4.4.33) [Computer software]. Retrieved October 5, 2006, from <http://www.praat.org>
- Boomer, D., & Laver, J. (1968). Slips of the tongue. *British Journal of Disorders of Communication, 3*, 1-12.
- Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219–252.
- Byrd, D. (1993). *54,000 American Stops*. UCLA Working Papers in Phonetics (83)
- Chapin, C., Blumstein, S. E., Meissner, B., & Boller, F. (1981). Speech production mechanisms in aphasia: A delayed auditory feedback study. *Brain and Language, 14*, 106-13.
- Chen, J. Y., & Baars, B. J. (1992). General and specific factors in "transformational errors": An experimental study. In B. J. Baars (Ed.), *Experimental slips and hu-*

- man error: Exploring the architecture of volition* (pp. 217-233). New York: Plenum Press.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25, 257-271.
- Cole, J. (2009). Emergent feature structures: Harmony systems in exemplar models of phonology. *Language Sciences*, 31, 144-160.
- Coleman, J., & Pierrehumbert, J. (1997). Stochastic phonological grammars and acceptability. In *Computational phonology: Third Meeting of the ACL Special Interest Group in Computational Phonology*. Somerset, NJ: Association for Computational Linguistics, 49-56.
- Cutler, A. (1981). The reliability of speech error data. *Linguistics*, 19, 561-582.
- de Saussure, F. (1966). *Course in general linguistics*. (W. Baskin, Trans.). New York: McGraw-Hill. (Original work published 1916).
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321.
- Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech Errors. *Language and Cognitive Processes*, 5, 313-349.

- Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, 20, 611–629.
- Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, 104, 123-47.
- Fairbanks, G. (1955). Selective vocal effects of delayed auditory feedback. *Journal of Speech and Hearing Disorders*, 20, 333-345.
- Fairbanks, G., & Guttman, N. (1958). Effects of delayed auditory feedback upon articulation. *Journal of Speech and Hearing Research*, 1, 12-22.
- Fay, D. (1980). Transformational errors. In V. A. Fromkin (Ed.), *Errors in Linguistic Performance: Slips of the tongue, ear, pen and hand* (pp. 111-122). Academic Press.
- Finney, S. A. (1999). *Disruptive effects of delayed auditory feedback on motor sequencing*. Unpublished doctoral dissertation, Brown University, Providence, RI.
- Frisch, S. A., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, 30, 139-162.
- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 215–69). The Hague: Mouton.
- Fromkin, V.A. (Ed.) (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.

- Fromkin, V. A. (1980). Introduction. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp.1-12) San Francisco: Academic Press.
- Fromkins Speech Error Database, Max Planck Institute for Psycholinguistics. (2000). Nijmegen, The Netherlands.
- Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation*. (Vol. 1, pp. 133-177). New York: Academic Press.
- Goldrick, M. (2002). *Patterns in sound, patterns in mind: Phonological regularities in speech production*. Unpublished doctoral dissertation, Johns Hopkins University, Baltimore, MD.
- Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*, 21, 817-855.
- Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, 21, 649-683.
- Goldrick, M., & Daland, R. (2009). Linking speech errors and phonological grammars: Insights from Harmonic Grammar networks. *Phonology*, 26, 147-185.

- Goldstein, L. (1980). Bias and asymmetry in speech perception. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 241-261). San Francisco: Academic Press.
- Gormley, A., & Stewart, T. (2009). Errors in speech production: Explaining mismatch and accommodation. In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2692-7). Austin, TX: Cognitive Science Society.
- Gormley, A., & Thomson, R. (2007). Phonetic features are anticipatory. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (p. 1763). Austin, TX: Cognitive Science Society.
- Greenberg, J. H. (1966). *Universals of language*. Cambridge, MA: MIT Press.
- Guest, D. J. (2002). *Phonetic features in language production: An experimental examination of phonetic feature errors*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Haggard, M. (1978). The devoicing of voiced fricatives. *Journal of Phonetics*, 6, 95–102.
- Hay, J., Pierrehumbert, J., & Beckman, M. (2004). Speech perception, well-formedness and lexical frequency. In J. Local, R. Ogden and R. Temple, (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology, VI* (pp. 57-74). Cambridge, UK: Cambridge University Press.

- Hayes, B. P., & MacEachern, M. (1998). Quatrain Form in English Folk Verse. *Language*, 74, 473-507.
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279, 1213–1216.
- Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*, 45, 295–310.
- Indefrey, P., & Levelt, W.J.M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92, 101-104.
- Jakobson, R. (1941). *Kindersprache, Aphasie, und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell. [1962 reprint in *Selected Writings of Roman Jakobson*, vol. 1 (Berlin: Mouton de Gruyter). 1969 reprint (Frankfurt am Main: Suhrkamp). 1968 English translation by A. Keiler (The Hague: Mouton).]
- Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *Journal of the Acoustical Society of America*, 108, 1246–1251.
- Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: Studies of Mandarin tone production. *Journal of Phonetics*, 30, 303–320.
- Jones, J. A., & Munhall, K. G. (2005). Remapping auditory-motor representations in voice production. *Current Biology*, 15, 1768–1772.
- Kager, R. (1999). *Optimality theory*. Cambridge: Cambridge University Press.

- Kalmus, H., Denes, P., & Fry, D. B. (1955). Effect of delayed acoustic feed-back on some non-vocal activities. *Nature*, *175*, 1078.
- Kenstowicz, M., & Kisseberth, C. (1979). *Generative phonology: Description and theory*. San Diego, CA: Academic Press.
- Kisseberth, C. (1970). The functional unity of phonological rules. *Linguistic Inquiry*, *1*, 291-306.
- Kupin, J. J. (1982). *Tongue twisters as a source of information about speech production*. Bloomington: Indiana University Linguistics Club.
- Lee, B. S. (1950). Effects of delayed speech feedback. *Journal of the Acoustical Society of America*, *22*, 824-826.
- Lee, B. S. (1951). Artificial stutter. *Journal of Speech and Hearing Disorders*, *16*, 53-55.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1993). Timing in speech production with special reference to word from encoding. *Annals of the New York Academy of Sciences*, *682*, 283-295.
- Levitt, A.G., & Healy, A.F. (1985). The roles of phoneme frequency, similarity, and availability in the experimental elicitation of speech errors. *Journal of Memory and Language*, *24*, 717-733.
- MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, *8*, 323-350.

- MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, 3, 210-227.
- McCarthy, J., & Prince, A. (1994). The emergence of the unmarked: Optimality in prosodic morphology. *Proceedings of the North East Linguistics Society*, 24, 333-379.
- Meara, P., & Ellis, A. (1981). The psychological reality of deep and surface phonological representations: Evidence from speech errors in Welsh. *Linguistics. An Interdisciplinary Journal of the Language Sciences*, 19, 797-804.
- Meringer, R., & Mayer, K. (1896). *Versprechen und Verlesen*. Stuttgart: Goschen.
- Motley, M. T., & Baars, B. J. (1976). Laboratory induction of verbal slips: A new method for psycholinguistic research. *Communication Quarterly*, 24, 28-34.
- Mowrey, R. A., & MacKay, I. R. A. (1990). Phonological primitives: Electromyographic speech error evidence. *The Journal of the Acoustical Society of America*, 88, 1299.
- Nooteboom, S. G. (1973). The tongue slips into patterns. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 144-156). The Hague: Mouton.
- Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17, 273-281.
- Pierrehumbert, J. (2001) Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee and P. Hopper (Eds.), *Frequency effects and emergent grammar*. (pp.137-157). Amsterdam: John Benjamins.

- Pierrehumbert, J. (2002) Word-specific phonetics. In C. Gussenhoven and N. Warner (Eds.), *Laboratory Phonology VII* (pp. 101-140). Berlin: Mouton De Gruyter.
- Pierrehumbert, J. (2003), Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 97-138), Cambridge, MA: MIT Press.
- Poupplier, M. (2003). *Units of phonological encoding: Empirical evidence*. Unpublished doctoral dissertation, Yale University, New Haven, CT.
- Poupplier, M., & Goldstein, L. (2005). Asymmetries in the perception of speech production errors. *Journal of Phonetics*, 33, 47-75.
- Prince, A., & Smolensky, P. (1993). Optimality theory: Constraints in generative grammar. *Rutgers University Center for Cognitive Science Technical Report 2*. New Brunswick, NJ.
- Purcell, D.W., & Munhall, K.G. (2006). Compensating following real-time manipulation of formants in isolated vowels. *Journal of the Acoustical Society of America*, 119, 2288–2297.
- Python (1990-2010). [Computer software]. Wolfeboro Falls, NH: Python Software Foundation.
- Sahin, N.T., Pinker, S., Cash, S.S., Schomer, D., & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within broca's area. *Science*, 326, pp. 445-449.

- Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295-342). Hillsdale, NJ: Erlbaum.
- Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage, (Ed.), *The production of speech* (pp. 109–136). New York: Springer-Verlag.
- Smith, C. L. (1997). The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics*, 25, 471–500.
- Spilka, B. (1954). Some vocal effects of different reading passages and time delays in speech feedback. *Journal of Speech and Hearing Disorders*, 19, 37-47.
- Stemberger, J. P. (1982). The nature of segments in the lexicon: Evidence from speech errors. *Lingua*, 56, 235-259.
- Stemberger, J. P. (1983). *Speech errors and theoretical phonology: A review*. Bloomington, Indiana: Indiana University Linguistics Club.
- Stemberger, J. P. (1992). The reliability and replicability of naturalistic speech error data: A comparison with experimentally induced errors. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition* (pp. 195-215). New York: Plenum Press.
- Stewart, T.C., & West, R.L. (2007). Deconstructing and reconstructing ACT-R: Exploring the architectural space. *Cognitive Systems Research*, 8, 227-236.

- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542-562.
- Sussman, R., & Sedivy, J. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, 18, 143-163.
- Tiffany, W. R., & Hanley, C. N. (1956). Adaptation to delayed sidetone. *Journal of Speech and Hearing Disorders*, 21, 164-172.
- Trubetzkoy, N. (1931). Die phonologischen Systeme. *Travaux du cercle linguistique de Prague IV*, pp. 96-116.
- Trubetzkoy, N. (1939). *Grundzüge der Phonologie*. Göttingen: Vandenhoeck & Ruprecht.
- 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, and Computers*, 36, 481-487.
- Wilshire, C. E. (1999). The "tongue twister" paradigm as a technique for studying phonological encoding. *Language and Speech*, 42, 57-82.
- Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin*, 60, 213-32.

Appendix A

Tongue twisters used in Experiment 1

a. Control condition (non-altering, AAAA)

| | | | |
|---------------------|---------------------|---------------------|---------------------|
| tiff tiff tiff tiff | teff teff teff teff | taff taff taff taff | tuff tuff tuff tuff |
| tiss tiss tiss tiss | tess tess tess tess | tass tass tass tass | tuss tuss tuss tuss |
| tivv tivv tivv tivv | tevv tevv tevv tevv | tavv tavv tavv tavv | tuvv tuvv tuvv tuvv |
| tizz tizz tizz tizz | tezz tezz tezz tezz | tazz tazz tazz tazz | tuzz tuzz tuzz tuzz |
| kiff kiff kiff kiff | keff keff keff keff | kaff kaff kaff kaff | kuff kuff kuff kuff |
| kiss kiss kiss kiss | kess kess kess kess | kass kass kass kass | kuss kuss kuss kuss |
| kivv kivv kivv kivv | kevv kevv kevv kevv | kavv kavv kavv kavv | kuvv kuvv kuvv kuvv |
| kizz kizz kizz kizz | kezz kezz kezz kezz | kazz kazz kazz kazz | kuzz kuzz kuzz kuzz |

b. Experimental condition (alterating, ABBA)

| | | | |
|---------------------|---------------------|---------------------|---------------------|
| tiff tivv tivv tiff | teff tevv tevv teff | taff tavv tavv taff | tuff tuvv tuvv tuff |
| tiss tizz tizz tiss | tess tezz tezz tess | tass tazz tazz tass | tuss tuzz tuzz tuss |
| kiff kivv kivv kiff | keff kevv kevv keff | kaff kavv kavv kaff | kuff kuvv kuvv kuff |
| kiss kizz kizz kiss | kess kezz kezz kess | kass kazz kazz kass | kuss kuzz kuzz kuss |
| tivv tiff tiff tivv | tevv teff teff tevv | tavv taff taff tavv | tuvv tuff tuff tuvv |
| tizz tiss tiss tizz | tezz tess tess tezz | tazz tass tass tazz | tuzz tuss tuss tuzz |
| kivv kiff kiff kivv | kevv keff keff kevv | kavv kaff kaff kavv | kuvv kuff kuff kuvv |
| kizz kiss kiss kizz | kezz kess kess kezz | kazz kass kass kazz | kuzz kuss kuss kuzz |

Appendix B

Stimuli used in Experiment 2

This is a fubber now
This is a futter now
Hey look, a kawdid is here
Last week, I kawded outside

Appendix C

Stimuli used in Experiment 3

tuff tuff tuff tuff
tuvv tuvv tuvv tuvv
kuff kuff kuff kuff
kuvv kuvv kuvv kuvv

Appendix D

Model 2

```
dm_noise=0.07
slot_similarity=-0.4
```

```
r_noise=0
f_noise=0
dm_basenoise=0
dm_decay=0
```

```
iterations=10
```

```
# speech planning + phonology + speech planning2
# errors can occur in speech planning 1 = accommodated
# errors can occur during the phonology = mismatch
# errors can occur in speech planning 2 = mismatch
```

```
import ccm
done within the ccm suite
log=ccm.log()
```

```
# all of the modeling is
```

```
from ccm.lib.actr import *
agent
```

```
# import act-r to create the
```

```
test_words=[['tisstizz', 'tistiiz'], ['tizztiss', 'tiiztis']]
```

```
class MyAgent(ACTR):
```

```
    focus=Buffer()
```

```
    DMbuffer=Buffer()
```

```
# create a buffer for the de-
```

```
clarative memory (henceforth DM)
```

```
    DM=Memory(DMbuffer,threshold=None)
```

```
    Fbuffer=Buffer()
```

```
# feature buffer
```

```
    FB=Memory(Fbuffer,threshold=None)
```

```
    f_n=DMNoise(FB,noise=f_noise)
```

```
    RMbuffer=Buffer()
```

```
    RM=Memory(RMbuffer,threshold=None)
```

```

r_n=DMNoise(RM,noise=r_noise)

SPbuffer=Buffer() #speech planning buffer
SP=Memory(SPbuffer,threshold=None,latency=0.01)

SP2buffer=Buffer() #speech planning 2 buffer
SP2=Memory(SP2buffer,threshold=None,latency=0.01)

dm_n=DMNoise(SP,noise=dm_noise,baseNoise=dm_basenoise) # turn on for
DM subsymbolic processing
dm_bl=DMBaseLevel(SP,decay=dm_decay,limit=None)

dm_n=DMNoise(SP2,noise=dm_noise,baseNoise=dm_basenoise) # turn on for
DM subsymbolic processing
dm_bl=DMBaseLevel(SP2,decay=dm_decay,limit=None) # turn on for
DM subsymbolic processing

partial=Partial(SP,strength=1.0,limit=-1.0) # usually this is strength
divided by number of slots
partial.similarity('1','4',slot_similarity)
partial.similarity('2','5',slot_similarity)
partial.similarity('3','6',slot_similarity)

partial=Partial(SP2,strength=1.0,limit=-1.0)
partial.similarity('1','4',slot_similarity)
partial.similarity('2','5',slot_similarity)
partial.similarity('3','6',slot_similarity)

def init():
    FB.add ('s voiceless alveolar fricative') # segmental inventory
    FB.add ('t voiceless alveolar stop')
    FB.add ('i high front vowel')
    FB.add ('z voiced alveolar fricative')

    # lexicon
    DM.add ('tisstizz t i s t i z')
    DM.add ('tizztiss t i z t i s')

    RM.add ('rule v')

# retrieve lexical item and phonemes
def get_input(focus = 'start ?x'):

```

```

    DM.request (x)
    focus.set('sp1')
    self.outputp=""
    self.orderp=""
    self.order=""
    self.output=""

# Speech planning 1
# gets phonemes from FB and outputs to SP

# syllable 1
def getO(focus ='sp1', DMbuffer='?dw ?do1 ?dn1 ?dc1 ?do2 ?dn2 ?dc2'):
    Fbuffer.clear()
    FB.request('?do1 ?fa ?fb ?fc')
    focus.set('sp2')

def outO(focus ='sp2', Fbuffer='?fo ?fx ?fy ?fz'):
    SP.add ('onset 1 ?fo ?fx ?fy ?fz')
    focus.set('sp3')

def getN(focus='sp3',DMbuffer='?dw ?do1 ?dn1 ?dc1 ?do2 ?dn2 ?dc2'):
    Fbuffer.clear()
    FB.request('?dn1 ?fx ?fy ?fz')
    focus.set('sp4')

def outN(focus='sp4', Fbuffer='?fn ?fx ?fy ?fz'):
    SP.add ('nucleus 2 ?fn ?fx ?fy ?fz')
    focus.set('sp5')

def getC(focus='sp5',DMbuffer='?dw ?do1 ?dn1 ?dc1 ?do2 ?dn2 ?dc2'):
    Fbuffer.clear()
    FB.request('?dc1 ?fx ?fy ?fz')
    focus.set('sp6')

def outC(focus='sp6', Fbuffer='?fc ?fx ?fy ?fz'):
    SP.add ('coda 3 ?fc ?fx ?fy ?fz')
    focus.set('sp7')

# syllable 2
def getO2(focus ='sp7', DMbuffer='?dw ?do1 ?dn1 ?dc1 ?do2 ?dn2 ?dc2'):
    Fbuffer.clear()
    FB.request('?do2 ?fx ?fy ?fz')

```

```

    focus.set('sp8')

def outO2(focus='sp8', Fbuffer='?fo ?fx ?fy ?fz'):
    SP.add ('onset 4 ?fo ?fx ?fy ?fz')
    focus.set('sp9')

def getN2(focus='sp9', DMbuffer='?dw ?do1 ?dn1 ?dc1 ?do2 ?dn2 ?dc2'):
    Fbuffer.clear()
    FB.request('?dn2 ?fx ?fy ?fz')
    focus.set('sp10')

def outN2(focus='sp10', Fbuffer='?fn ?fx ?fy ?fz'):
    SP.add ('nucleus 5 ?fn ?fx ?fy ?fz')
    focus.set('sp11')

def getC2(focus='sp11', DMbuffer='?dw ?do1 ?dn1 ?dc1 ?do2 ?dn2 ?dc2'):
    Fbuffer.clear()
    FB.request('?dc2 ?fx ?fy ?fz')
    focus.set('sp12')

def outC2(focus='sp12', Fbuffer='?fc ?fx ?fy ?fz'):
    SP.add ('coda 6 ?fc ?fx ?fy ?fz')
    focus.set('p2')

# phonology
# syllable 1
def getOnset(focus='p2'):
    SP.request('onset 1 ?o ?x ?y ?z')
    focus.set('p3')

def outOnset(focus='p3', SPbuffer='?s ?w ?o ?x ?y ?z'):
    self.orderp+=w
    SP2.add('onset 1 ?o')
    SPbuffer.clear()
    SP.request('coda 3 ?fc ?fx ?fy ?fz')
    focus.set('p4')

# check coda for voicing
def Vcheck1(focus='p4', SPbuffer='?s ?w ?c voiced ?y ?z'):
    self.orderp+=w
    SP2.add('coda 3 ?c')
    SP.request('nucleus 2 ?fn ?fx ?fy ?fz')

```

```

    RM.request('?rule v')
    focus.set('p5')

def Vcheck2(focus='p4', SPbuffer = '?s ?w ?c voiceless ?y ?z'):
    self.orderp+=w
    SP2.add('coda 3 ?c')
    #SPbuffer.clear()
    SP.request('nucleus 2 ?fn ?fx ?fy ?fz')
    focus.set('p6')

def rule(focus='p5', SPbuffer = '?s ?w ?n ?x ?y ?z', RMbuffer = 'rule v'):
    n = n + n
    self.orderp+=w
    SP2.add('nucleus 2 ?n')
    focus.set('p7')

def NoRule(focus='p6', SPbuffer = '?s ?w ?n ?x ?y ?z'):
    self.orderp+=w
    SP2.add('nucleus 2 ?n')
    focus.set('p7')

# syllable 2
def getOnset2(focus = 'p7'):
    SP.request('onset 4 ?fo ?fx ?fy ?fz')
    focus.set('p8')

def outOnset2(focus='p8', SPbuffer = '?s ?w ?o ?x ?y ?z'):
    self.orderp+=w
    SP2.add('onset 4 ?o')
    SP.request('coda 6 ?fc ?fx ?fy ?fz')
    focus.set('p9')

# check coda for voicing
def Vcheck1_2(focus='p9', SPbuffer = '?s ?w ?c voiced ?y ?z'):
    self.orderp+=w
    SP2.add('coda 6 ?c')
    SP.request('nucleus 5 ?fn ?fx ?fy ?fz')
    RM.request('?rule v')
    focus.set('p10')

def Vcheck2_2(focus='p9', SPbuffer = '?s ?w ?c voiceless ?y ?z'):

```

```

self.orderp+=w
SP2.add('coda 6 ?c')
SP.request('nucleus 5 ?fn ?fx ?fy ?fz')
focus.set('p11')

def rule2(focus='p10', SPbuffer = '?s ?w ?n ?x ?y ?z', RMbuffer = 'rule v'):
    n = n + n
    self.orderp+=w
    SP2.add('nucleus 5 ?n')
    focus.set('2sp1')

def NoRule2(focus='p11', SPbuffer = '?s ?w ?n ?x ?y ?z'):
    self.orderp+=w
    SP2.add('nucleus 5 ?n')
    focus.set('2sp1')

# Speech planning 2
# gets phonemes from buffer and outputs

# syllable 1
def getsp2O(focus = '2sp1'):
    SP2.request('onset 1 ?o')
    focus.set('2sp2')

def out2O1(focus = '2sp2', SP2buffer='?onset ?w ?o'):
    self.output+=o
    self.order+=w
    focus.set('2sp3')

def getsp2N(focus='2sp3'):
    SP2.request('nucleus 2 ?n')
    focus.set('2sp4')

def out2N1(focus='2sp4', SP2buffer='?nucleus ?w ?n'):
    self.output+=n
    self.order+=w
    focus.set('2sp5')

def getsp2C(focus='2sp5'):
    SP2.request('coda 3 ?c')
    focus.set('2sp6')

```

```

def out2C1(focus='2sp6', SP2buffer='?coda ?w ?c'):
    self.output+=c
    self.order+=w
    focus.set('2sp7')

# syllable 2
def getsp2O2(focus='2sp7'):
    SP2.request('onset 4 ?p')
    focus.set('2sp8')

def out2O2(focus='2sp8', SP2buffer='?onset ?w ?o'):
    self.output+=o
    self.order+=w
    focus.set('2sp9')

def getsp2N2(focus='2sp9'):
    SP2.request('nucleus 5 ?n')
    focus.set('2sp10')

def out2N2(focus='2sp10', SP2buffer='?nucleus ?w ?n'):
    self.output+=n
    self.order+=w
    focus.set('2sp11')

def getsp2C2(focus='2sp11'):
    SP2.request('coda 6 ?c')
    focus.set('2sp12')

def out2C2(focus='2sp12', SP2buffer='?coda ?w ?c'):
    self.output+=c
    self.order+=w
    focus.set('stop')

def stopa(focus='stop'):
    self.stop()

miss=0      # error
miss1=0     # error in sp1
miss2=0     # error in sp2
mm=0        # mismatch

```

```

c=0          # correct
acc=0        # accommodated

for i in range(iterations):
    for word,target in test_words:
        agent=MyAgent()
        agent.focus.set('start %s'%word)
        agent.run()
        print word,target,agent.outputp,agent.orderp,agent.output,agent.order
        if agent.output!=target:
            miss+=1
        if agent.orderp!='132465':
            miss1+=1
        if agent.order!='123456':
            miss2+=1
        if agent.output==target:
            c+=1
            print 'c'
        if 'tiis' in agent.output:
            mm+=1
            print 'm'
        if 'tiz' in agent.output:
            mm+=1
            print 'm2'
        if target=='tistiiz' and agent.output in ['tiiztiiz','tiiztis','tistis']:
            acc+=1
            print 'a'
        if target=='tiiztis' and agent.output in ['tistis','tiiztiiz','tistiiz']:
            acc+=1
            print 'a2'

log.rate=float(miss)/(iterations*len(test_words))
log.rates1=float(miss1)/(iterations*len(test_words))
log.rates2=float(miss2)/(iterations*len(test_words))
log.correct_rate=float(c)/(iterations*len(test_words))
log.acc_rate=float(acc)/(iterations*len(test_words))
log.mm_rate=float(mm)/(iterations*len(test_words))

```