DEGREES OF INCOMPLETENESS IN NEUTRALIZATION:

PARADIGM UNIFORMITY IN A PHONETICS WITH WEIGHTED CONSTRAINTS

by

AARON BRAVER

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Linguistics

written under the direction of

Shigeto Kawahara

and approved by

____

____

____

____

New Brunswick, New Jersey

October 2013

ABSTRACT OF THE DISSERTATION

Degrees of Incompleteness in Neutralization: Paradigm Uniformity in a Phonetics with Weighted Constraints

by AARON BRAVER

Dissertation Director:

Shigeto Kawahara

This dissertation presents two case studies of incomplete neutralization: flapping in American English and monomoraic vowel lengthening in Japanese. Experimental evidence is provided showing that the underlying contrast in each case is, indeed, only partly neutralized. I argue that these cases represent two distinct points on a continuum of completeness of neutralization: monomoraic vowel lengthening in Japanese results in a surface distinction that is plausibly perceptible, while flapping in American English results in /d/ flaps and /t/ flaps which cannot be distinguished—a claim supported by a series of perception experiments (Experiments 3, 4, and 5).

First, I argue that incomplete neutralization is not solely due to experimental artifacts or task effects, as some scholars have claimed. Experiments 1 and 2, both on flapping in American English, consist of two task types—one designed to increase the potential for these effects, and one designed to reduce them. I show that the degree of neutralization remains constant between these two task types, suggesting that results of incomplete neutralization cannot all be reduced to these extragrammatical factors.

Second, I address the types of contrasts which can be incompletely neutralized. The vast majority of studies on incomplete neutralization have thus far centered on feature-

and segment-level contrasts (especially final devoicing). I show in Experiments 6 and 7 that the Japanese short/long vowel length contrast is incompletely neutralized in monomoraic noun lengthening contexts. I suggest that the typology of processes that can lead to incomplete neutralization must be expanded to include those that operate on suprasegmental or prosodic contrasts.

Third, given the results of the experiments, I claim that incomplete neutralization is best accounted for in a model of phonetics based on a weighted-constraint grammar (Legendre et al. 1990, Zsiga 2000, Flemming 2001). I propose two types of constraints: the first pressures segments to match a target value for a given phonetic measure (Flemming 2001). The second type of constraint, based on the notion of transderivational identity (Benua 1997, Steriade 2000), requires candidates to match a base form for a given phonetic measure. I argue that basehood for this type of transderivational identity is best determined by type frequency within a candidate's inflectional paradigm.

These phonetic constraints differ from the familiar phonological constraints of, e.g., Optimality Theory (Prince and Smolensky 1993) in two major ways. First, phonetic constraints may refer to both phonological structures as well as raw, quantitative phonetic information. Second, constraint conflict is resolved by compromise, rather than strict domination as in a ranked-constraint grammar. This ability to compromise, found in weighted-constraint grammars, combined with access to quantitative phonetic detail, allows the model to generate languages at both the plausibly perceptible end of the incomplete neutralization spectrum (like Japanese) and the imperceptible end (like American English), as well as the continuum in between.

Finally, I show that this model best accounts for the Directionality Observation of incomplete neutralization. The Directionality Observation reflects the fact that the realization of two incompletely neutralized categories is predictable: on a continuum of the acoustic cue(s) that differentiate the categories, an incompletely neutralized segment must fall somewhere between the canonical realizations of the two categories.

Dedication

To my parents, Lane and Andrea,
who taught me the value
of learning and knowledge.

...I occasionally wondered, "Where is the normal phonology that I was trained to study?"
—Bruce Hayes (1992/1995:9)

Acknowledgments

This dissertation benefited enormously from a number of individuals, to whom I am immensely grateful.

First and foremost, I would like to thank my committee—Shigeto Kawahara, Bruce Tesar, Kristen Syrett, and Alan C.L. Yu—for their direction, insight, and mentoring. In particular, I would like to thank my chair, Shigeto Kawahara, for guiding me through my graduate career—even when we were separated by more than 6,000 miles. His intellectual inspiration, counseling, and friendship have made this dissertation possible. I owe a great intellectual debt to both Bruce Tesar and Kristen Syrett, for pressing me on the details of my experiments and analysis, and for bringing new perspectives to my ideas. Alan Yu deserves my gratitude for our many discussions throughout the years—on topics related to the dissertation and otherwise. I appreciate his willingness to serve as the external member of my committee.

In addition to my committee, I would also like to thank Donca Steriade, Eric Baković, Jason Riggle, Edward Flemming, John Kingston, Ricardo Burmúdez-Otero, Wendell Kimper, Jonathan North Washington, and Brian W. Smith for their useful comments and discussions on work related to the dissertation. Additional helpful feedback came from audiences at Rutgers, UMass, MIT, Manchester, WCCFL 31, and NELS 42. (NB: preliminary versions of portions of Chapters 2 and 3 (will) appear in the WCCFL and NELS proceedings. See the unnumbered footnote at the start of each chapter for details on previously published and pending work.)

My fellow grad students at Rutgers have had a major impact on this dissertation, and on my life. First, I am indebted to William G. Bennett for his comments on this work, for our discussions on all matters phonological, and for a friendship that survived sharing an apartment for four years. I also thank Seunghun Lee, Jeremy Perkins, Ryan Denzer-King, Luca Iacoponi, Hope McManus, and Nick Danis for sharing their insights on this project. My classmates, The Best Ling Grads Ever: Teresa Torres Bustamante, Veřa Dvořák, Todor Koev, and Jeremy Perkins, have been a source of support throughout the years, as have my colleagues Mateus Barros, Jimmy Bruno, Sara O'Neill, and Vandana Bajaj.

The experiments in this dissertation would not have been possible without the help of the undergraduate assistants in the Rutgers Phonetics Laboratory: Melanie Pangilinan, Natalie Dresher, Jess Trombetta, Chris Kish, Sarah Korostoff, Michelle Marron, Megan Moran, Kelly Garvey, and Melanie Kelliher. Thanks go to them, and to the participants in all of the experiments.

For their moral support and guidance in the ways of academia, I am grateful to Veneeta Dayal and Roger Schwarzschild. Their value as friends and allies helped me to overcome my fear of semanticists.

I would also like to thank my non-linguist friends, whose contributions to this dissertation are immeasurable. First and foremost, I thank Christopher Farrell — the Holmes to my Watson—for reminding me that sometimes a starry sky signifies nothing more than a stolen tent. Thanks also to Vanita Neelakanta for not letting me get away with anything, and for providing just the right mixture of sass, warmth, and (tough) love. I am grateful, too, to Albert Cahn, who should probably get a couch and start billing by the hour. Also, thanks to my oldest friend, Brad Taylor, for providing a much needed dose of comedy and a perspective from outside the ivory tower.

The gratitude I feel for my family exceeds my skill with words, so the following will have to suffice. My parents have supported me in every way imaginable, and continue to be a source of strength. Their emphasis on education, hard work, and perseverance has led me to where I stand today. I am grateful to my brothers, Alex and Zachary, for their support over the years (even when it came in the form of some 'friendly' roughhousing).

Finally, New York City has served as an escape from the dissertating bubble. In particular, I will miss the live music in the city. I thank The Drums, Bombay Bicycle Club, The Pains of Being Pure at Heart, Cold Fronts, Girls, of Montreal, Neon Indian, Cola Jet Set, The Mary Onettes, Johnny Marr, The Ballet, Making Marks, The School, French Films, and Camera Obscura, among others, for separating me temporarily from thoughts of incomplete neutralization.

Highland Park, New Jersey
July 2013

Contents

Abstract **ii**

Dedication **iv**

Acknowledgments **v**

Contents **viii**

List of Tables **xiv**

List of Figures **xvii**

Chapter 1 Introduction **1**

1.1 (In)complete Neutralization 3

1.1.1 A Taxonomy of Incomplete Neutralization and Subphonemic Distinctions 9

1.2 Previous Work 12

1.2.1 Production 13

1.2.2 Perception 14

1.2.3 Modeling Incomplete Neutralization 16

1.3 Issues in (In)complete Neutralization 21

1.3.1 Experimental Artifacts and Task Effects 21

1.3.2 Byproducts of Phonetic Implementation 22

1.3.3 Motivations and Causes for Incomplete Neutralization 22
1.3.4 Predicting the Direction of Incomplete Neutralization 23
1.3.5 Summary 25

Chapter 2 Flapping in American English **28**
2.1 Previous Work 29
2.2 Remaining Issues and Motivations for the Current Studies 32
2.3 Experiment 1: Production of Flaps in American English (I) 35
2.3.1 Methods 35
2.3.1.1 Speakers and Recording 35
2.3.1.2 The Minimal Pair Task 36
2.3.1.3 The Wug (Paradigm Completion) Task 37
2.3.1.4 Measurements 39
2.3.1.5 Statistical Analyses 41
2.3.2 Results 42
2.3.2.1 The Effect of Underlying Voicing Status (/t/ vs. /d/) . . . 42
2.3.2.2 The Effect of Task 44
2.3.2.3 Summary of Overall Effects 48
2.3.2.4 Variation Among Speakers 48
2.3.2.5 Summary of Results 53
2.3.3 Discussion 56
2.3.3.1 Incomplete Neutralization of Voicing in Flapping 56
2.3.3.2 Task 57
2.3.3.3 Variation 58
2.3.3.4 Effect Size and Just Noticeable Differences 58
2.3.3.5 Preceding Context 59
2.3.3.6 Actual English Roots 60
2.3.4 Conclusion 60

2.4 Experiment 2: Production of Flaps in American English (II) 61
2.4.1 Stimuli 61
2.4.2 Participants and Equipment 63
2.4.3 Methods 63
2.4.3.1 The Minimal Pair Task 63
2.4.3.2 The Wug Task 64
2.4.3.3 Acoustic Measurements and Statistical Analysis 65
2.4.4 Results 67
2.4.4.1 The Effect of Task 67
2.4.4.2 The Effect of Underlying Voicing Status 68
2.4.5 Discussion 68
2.5 Experiment 3: Identification Perception of Flaps in American English . . . 69
2.5.1 Participants and equipment 70
2.5.2 Tokens 70
2.5.3 Procedure 72
2.5.4 Analysis 72
2.5.5 Results 73
2.5.6 Discussion 74
2.6 Experiment 4: ABX Perception of Flaps in American English 74
2.6.1 Participants and equipment 75
2.6.2 Stimuli 75
2.6.3 Procedure 76
2.6.4 Analysis 77
2.6.5 Results 77
2.6.6 Discussion 78
2.7 Experiment 5: 2AFC Perception of Flaps in American English 80
2.7.1 Participants and Equipment 80

2.7.2 Stimuli 80
2.7.3 Procedure 81
2.7.4 Analysis 81
2.7.5 Results 81
2.7.6 Discussion 82
2.8 Chapter Discussion and Conclusions 83

Chapter 3 Bimoraicity-Related Vowel Lengthening in Japanese 85

3.1 Introduction 85
3.2 Background 86
3.3 Experiment 6: Vowel Lengthening in Monomoraic Nouns (I) 90
3.3.1 Method 90
3.3.1.1 Stimuli 90
3.3.1.2 Participants 91
3.3.1.3 Recording 91
3.3.1.4 Acoustic Measurements 92
3.3.1.5 Statistical Analysis 92
3.3.2 Results 93
3.3.3 Discussion 94
3.3.3.1 A New Case for Phonological Incomplete Neutralization 94
3.3.3.2 Discussion of the Current Stimuli 97
3.3.3.3 Conclusion 99
3.4 Experiment 7: Vowel Lengthening in Monomoraic Nouns (II) 100
3.4.1 Method 100
3.4.1.1 Stimuli 100
3.4.1.2 Participants 101
3.4.1.3 Procedure 101
3.4.1.4 Statistical Analysis 103

3.4.2 Results 103
3.4.3 Discussion 104
3.4.3.1 Looking Deeper into the Data 104
3.4.3.2 General Implications 109
3.5 General Discussion and Conclusion 111

Chapter 4 Modeling Incomplete Neutralization with Frequency-Based Transderivational Identity in a Weighted Constraint Grammar 113
4.1 Overview of the Model 115
4.1.1 Weighted Constraints and Phonetics: Zsiga (2000) and Flemming (2001) 115
4.1.2 Transderivational Identity and Paradigm Uniformity 117
4.1.3 Basehood and Canon 118
4.1.3.1 Defining and Evaluating Bases 120
4.1.3.2 Canon-hood 122
4.1.3.3 Correspondence within the Paradigm 122
4.1.3.4 Relation to Prototype and Exemplar Models 124
4.1.4 Exordium 127
4.2 Japanese 127
4.2.1 Weighted vs. Ranked Constraints 139
4.3 Flapping 143
4.4 Discussion and Conclusion 156
4.4.1 Speakers, Hearers, and Constraints Out of Their Motivating Context 159
4.4.2 Conclusion 160

Chapter 5 Conclusion 162

Appendix A Summary of All Measures from Experiment 2 164

Appendix B Stimuli from Experiment 6 165

Appendix C Stimuli from Experiment 7 168

Appendix D Application of the Model to Unrounded Experiment 7 Results 170

Appendix E R code 172

References 178

Curriculum Vitae 188

List of Tables

Table 2.1 Experiment 1: Stimuli from the Minimal Pair Task 37

Table 2.2 Experiment 1: Stimuli from the Wug Task 38

Table 2.3 Experiment 1: Results of linear mixed models (Section 2.3.1.5): effect of underlying voicing status (/d/ vs. /t/) 44

Table 2.4 Experiment 1: Mean and standard deviation of vowel durations (in ms) for each category 44

Table 2.5 Experiment 1: Results of linear mixed models (Section 2.3.1.5): effect of task (wug vs. minimal pair) 46

Table 2.6 Experiment 1: Results of linear mixed models (Section 2.3.1.5): effect of interaction between task and voicing. 48

Table 2.7 Experiment 1: The effect of underlying voicing status on each of the measurements for speaker 6 51

Table 2.8 Experiment 1: The effect of underlying voicing status on each of the measurements for speaker 12 53

Table 2.9 Experiment 2: Stimuli 62

Table 2.10 Experiment 2: Mean and standard deviation of vowel durations (in ms) for each category 68

Table 2.11 Experiment 2: Mean pre-flap vowel duration for all speakers in the acoustic experiment, across both tasks 69

Table 2.12 Tokens used in Experiments 3, 4, and 5. 71

Table 3.1 Experiment 6: Sample stimulus set 91

Table 3.2 Experiment 6: Mean and standard deviation of vowel durations (in ms) for each category 94

Table 3.3 Experiment 7: Sample stimulus set 100

Table 3.4 Experiment 7: Mean and standard deviation of vowel durations (in ms) for each category 103

Table 3.5 Experiment 7: Results from sets containing orthographic diphthongs in long vowel conditions, and those without orthographic diphthongs. 109

Table 3.6 Experiment 7: Results from sets with accent mismatches, and those without accent mismatches. 110

Table 4.1 Mean and standard deviation of vowel durations (in ms) for each category in Experiments 6 and 7 128

Table 4.2 Data to be modeled: Average vowel durations (in ms), based on an idealized version of the figures from Experiment 7. 128

Table 4.3 Costs for given short/Ø vowel durations, where w_{ID}=1, $w_{\mu\mu}$=3, *TargetDur(μ)*=50ms, and *TargetDur($\mu\mu$)*=150ms 138

Table 4.4 Predicted short/Ø vowel durations for sample weightings of OO-ID-DUR (w_{ID}) and DUR($\mu\mu$)=TARGETDUR($\mu\mu$) ($w_{\mu\mu}$), where *TargetDur(μ)*=50 ms and *TargetDur($\mu\mu$)*=150 ms. 139

Table 4.5 Mean and standard deviation of vowel durations (in ms) for each category in Experiments 1 and 2 143

Table 4.6 Data to be modeled: Average vowel durations (in ms), based on Experiment 2, and extrapolated target durations for vowels in non-flapping contexts. 145

Table 4.7 Combined frequency of all words in the top 100 and top 1,000 most frequent words for each verb inflection type, from COCA's spoken data 152

Table A.1 Experiment 2: Summary of results from all measured correlates of voicing. 164

Table D.1 Mean and standard deviation of vowel durations (in ms) for each category in Experiments 7 170

Table D.2 Costs for given short/Ø vowel durations, where w_{ID}=1, $w_{\mu\mu}$=2.1$\overline{5}$, *TargetDur(μ)*=54.99 ms, and *TargetDur(μμ)*=157.45 ms 170

List of Figures

Fig. 1.1 Schema of the Directionality Observation. 6

Fig. 1.2 Graphs of a hypothetical case of incomplete neutralization, illustrating the Directionality Observation 7

Fig. 2.1 Experiment 1: Representative spectrograms (from Speaker 13, minimal pair task) 40

Fig. 2.2 Experiment 1: Overall means by underlying voicing status across minimal pair and wug tasks combined 43

Fig. 2.3 Experiment 1: Mean duration of pre-flap vowels by underlying voicing status 45

Fig. 2.4 Experiment 1: Mean flap closure duration by underlying voicing status . 45

Fig. 2.5 Experiment 1: Mean intensity dip by underlying voicing status 45

Fig. 2.6 Experiment 1: Mean vowel duration by task 47

Fig. 2.7 Experiment 1: Mean F0 at flap offset by task 47

Fig. 2.8 Experiment 1: Mean F1 at flap onset by task 47

Fig. 2.9 Experiment 1: Mean F1 at flap offset by task 47

Fig. 2.10 Experiment 1: Pre-flap vowel duration for each speaker, by underlying voicing status. 50

Fig. 2.11 Experiment 1: Mean pre-flap vowel duration for speaker 6, by underlying voicing status 51

Fig. 2.12 Experiment 1: Mean intensity for speaker 6, by underlying voicing status 51

Fig. 2.13 Experiment 1: Mean pre-flap vowel duration for speaker 12, by underlying voicing status 52
Fig. 2.14 Experiment 1: Mean intensity dip for speaker 12, by underlying voicing status, n.s. 52
Fig. 2.15 Experiment 1: Mean pre-flap vowel duration for speaker 12, by underlying voicing status 52
Fig. 2.16 Experiment 1: Pre-flap vowel duration for each speaker, by underlying voicing status and task. 54
Fig. 2.17 Experiment 1: Mean /d/-flap F0 at flap offset – mean /t/-flap F0 at flap offset for speaker 1, by task 55
Fig. 2.18 Experiment 1: Mean /d/-flap intensity dip – mean /t/-flap intensity dip for speaker 10, by task 55
Fig. 2.19 Experiment 2: Representative spectrograms, speaker 6, minimal pair task 66
Fig. 2.20 Experiment 2: Mean pre-flap vowel duration by underlying voicing status and task 67
Fig. 2.21 Experiment 2: Mean pre-flap vowel duration by underlying voicing status (averaged across both tasks) 68
Fig. 2.22 Experimeng 3: Results of the identification experiment 73
Fig. 2.23 Experiment 4: Results of the ABX experiment 78
Fig. 2.24 Experiment 5: Results of the 2AFC experiment 82

Fig. 3.1 Experiment 6: A representative segmented spectrogram. Speaker 14, *su ga nai* (short/prt), repetition 9. 92
Fig. 3.2 Experiment 6: Mean vowel durations, averaged over all speakers and all items. 93
Fig. 3.3 Experiment 6: Mean vowel durations, by item. 95
Fig. 3.4 Experiment 6: Mean vowel durations, by speaker. 95
Fig. 3.5 Experiment 6: Distribution of vowel duration by condition, for each speaker. 96

Fig. 3.6 Experiment 7: A representative segmented spectrogram. Speaker 43, *kii nakushita yo* (long), repetition 7. 102

Fig. 3.7 Experiment 7: Vowel duration over all speakers and all items. 104

Fig. 3.8 Experiment 7: Vowel duration by speaker, averaged across items. 105

Fig. 3.9 Experiment 7: Vowel duration by item, averaged across speakers. 106

Fig. 3.10 Experiment 7: Distribution of vowel duration by condition, for each speaker. 107

Fig. 4.1 Schematic showing $\mathfrak{R}_{IO(cand)}$, $\mathfrak{R}_{IO(base)}$, and $\mathfrak{R}_{OO}$ for hypothetical input /pata/, candidate [pada], and base [past]. 123

Fig. 4.2 Experiment 7: Predicted short/Ø vowel duration for the idealized data, given weights for Dur(μμ)=TargetDur(μμ) and OO-ID-Dur. *Target-Dur(μ)*=50 ms, and *TargetDur*(μμ)=150 ms. 141

Fig. 4.3 Experiment 2: Predicted pre-flap vowel duration for /d/-flaps (top sheet) and /t/-flaps (bottom sheet), given weights for w_{ID} and $w_{(\mathrm{V/\_ɾ})}$. *Target-Dur(V/_C̥)*=138.50 ms, and *TargetDur(V/_$C_{[+voi]}$)*=145.90. 156

Chapter 1

Introduction

For decades, undergraduate linguistics students have learned that various classes of segments neutralize in certain prescribed positions. Whether expressed in a theory based on rules, representations, or violable constraints, the average undergraduate is taught that contrasts are completely obliterated in neutralizing contexts. Indeed, this view was held by many linguists before the 1970s and 1980s (Trubetzkoy 1939/1969, p. 235; Bloomfield 1933/1984, pp. 218–219; Jakobson et al. 1952/1975, p. 9; Hyman 1975, pp. 29, 71–72). However, with developments in the technology of acoustic analysis, the often small distinctions associated with incomplete neutralization—the partial preservation of an underlying contrast—became possible to measure on a wide scale.

This dissertation has three major aims, supported by two case studies. First, it aims to add to the growing body of experimental work on (in)complete neutralization. To this end, the first case study, consisting of production and perception experiments on flapping in American English, and the second case study, consisting of production experiments on monomoraic lengthening in Japanese, are described in Chapter 2 and Chapter 3, respectively. I show that flapping in American English results in incomplete neutralization of the /t~d/ voicing contrast, while at the same time remaining imperceptible to listeners (see, e.g., Fisher and Hirsh 1976, Fox and Terbeek 1977, Zue and Laferriere 1979, Joos 1942, Port 1976, Huff 1980, Herd et al. 2010, for related studies). Further, I show that in Japanese monomoraic nouns, vowel lengthening in response to a prosodic bimoraic minimality requirement (Poser 1990, Itô 1990, Mori 2002) results in incomplete neutralization.

The second major aim of this dissertation is to address a number of issues about (in)complete neutralization that have remained unresolved in the literature. Among these issues, I address the phonological 'reality' of incomplete neutralization by showing that incomplete neutralization holds irrespective of hyperarticulation and orthography, and that not every case of incomplete neutralization can be relegated to a byproduct of phonetic implementation. I also show which types of processes can lead to incomplete neutralization and which cannot, thereby developing a typology of incomplete neutralization. Finally, I address the Directionality Observation—arguing that theories of incomplete neutralization should capture the fact that the direction of an incompletely neutralized surface distinction is predictable from the canonical realization of the underlying contrast.

The third aim of the dissertation is to model these observations. Chapter 4 describes a two-level system: a more or less standard Optimality Theoretic (Prince and Smolensky 1993) phonology, and a weighted-constraint phonetic grammar that refers to both phonetic and phonological properties (see, e.g., Flemming 2001 for a weighted-constraint grammar, and Zsiga 2000 for a model combining ranked constraints for categorical phenomena, which feeds a system of weighted constraints for gradient phenomena). A gradient version of transderivational faithfulness (Benua 1997) provides pressure at the phonetic level for neutralized forms to maintain some similarity to other forms in their morphological paradigm—thus resulting in incomplete neutralization. Crucially, and unlike traditional transderivational faithfulness, the forms to which candidates become similar are the most frequent (typewise) in their inflectional paradigm (Steriade 2013).

Chapter Contents

1.1 (In)complete Neutralization 3

1.2 Previous Work 12

1.3 Issues in (In)complete Neutralization 21

1.1 (In)complete Neutralization

Traditionally, phonologists did not consider incomplete neutralization as a possibility. Under such a view of neutralization, two underlyingly distinct segments would become completely identical on the surface. Take as an example the 'classic' case of German final devoicing—often cited as completely neutralizing (Trubetzkoy 1939/1969, p. 235; Bloomfield 1933/1984, pp. 218-219; Jakobson et al. 1952/1975, p. 9; Hyman 1975, pp. 29, 71-72). Assuming complete neutralization, German's underlying voicing contrast is not preserved on the surface in codas, where only voiceless segments are allowed. As diagrammed in (1), underlyingly voiced-final /ʁɑd/ 'wheel' and underlyingly voiceless-final /ʁɑt/ 'advice' surface identically.

(1) The traditional picture of German coda devoicing

/ʁɑt/ 'advice'

[ʁɑt]

/ʁɑd/ 'wheel'

The picture, however, is not quite so simple. Keeping with the German devoicing case for the moment, a number of more recent studies have shown that while the contrast between underlyingly voiced and voiceless segments is indeed reduced on the surface, it is not *completely* neutralized. In other words, in devoicing positions, a trace of the underlying voicing distinction still remains on the surface (Port and O'Dell 1985, Mitleb 1981a, Mitleb 1981b, Taylor 1975 (for some places of articulation), Dinnsen and Garcia-

Zamor 1971 (in disyllables only)). Port and O'Dell (1985), for example, show that a vowel preceding a devoiced segment is in fact longer than a vowel preceding an underlyingly voiceless segment (by about 15 ms), among other surface distinctions.[1] This is all to say that neutralization can sometimes be incomplete.

To put a more concrete character to these notions, let us define *complete* and *incomplete* neutralization. In *complete* neutralization, the surface acoustic cues to two underlyingly distinct segments in a given context are completely identical. In *incomplete* neutralization, however, the surface acoustic cues to two underlyingly distinct segments in a given context are less distinct than the segments' canonical realizations in non-neutralizing contexts, but are not completely identical. Formal definitions are provided in (2).

(2) For:

$$\underset{[\alpha\text{F}]}{/\text{X}/} \rightarrow [\text{Z}^{(\alpha F)}] \ / \ (\text{Context A})$$

$$\underset{[\beta\text{F}]}{/\text{Y}/} \rightarrow [\text{Z}^{(\beta F)}] \ / \ (\text{Context A})$$

Where:

- Distinctness(x, y) is the distinctness between cues of two phonological units, with 0 being identity and with increasing values corresponding to increasing distinctness
- Canon(/*x*/) is the canonical realization of a phonological unit in a non-neutralizing context

a. **Complete neutralization**:
$\text{Distinctness}([\text{Z}^{(\alpha F)}]), [\text{Z}^{(\beta F)}]) = 0$

b. **Incomplete neutralization**
$\text{Distinctness}([\text{Z}^{(\alpha F)}]), [\text{Z}^{(\beta F)}]) < \text{Distinctness}(\text{Canon}(\underset{[\alpha\text{F}]}{/\text{X}/})), \text{Canon}(\underset{[\beta\text{F}]}{/\text{X}/}))$

[1] Fourakis and Iverson's (1984) study of German final devoicing argues that when an experiment's focus is on pronunciation, incomplete neutralization is more likely than when this focus is disguised. See Section 1.3.1 for further discussion.

Under complete neutralization, as in (2a), /X/ and /Y/, which differ in their value for feature F, are both realized with identical acoustic cues on the surface: there is no trace of their underlying distinction in F. In incomplete neutralization, as in (2b), the output segments [Z$^{(\alpha F)}$] and [Z$^{(\beta F)}$] differ in their acoustic cues for the underlying distinction in feature F.

Note that this 'surface trace' of the underlying contrast trends in the same direction as the canonical realization of the contrast being incompletely neutralized—a pattern I will call the 'Directionality Observation', defined in (3).

(3) **Directionality Observation**

Summary: Incompletely neutralized segments fall somewhere between the canonical realizations of the two neutralized segments.

Definition: For an underlying segment /α/ in an incompletely neutralized contrast with /β/:

- Given a continuum $\mathbb{C}$ along the acoustic cue(s) separating the canonical realization of [α] from the canonical realization of [β]:
  - Incompletely neutralized [α] will fall closer to [β] than the canonical realization of [α] is, and
  - Incompletely neutralized [α] will not surpass the canonical realization of [β].

This definition is schematized in Figure 1.1, where the arrows represent potential points along a continuum that an incompletely neutralized [α] might potentially fall. Those with an ✗ are illicit, and those with a ✓ are licit. Given the position of the canonical realizations of segments α and β along a continuum of the degree of the acoustic cue(s) that separate them, the Directionality Observation makes two main claims.

First, the incompletely neutralized version of [α] will fall closer to [β] than the canonical realization of [α]. Looking at Figure 1.1, we see that the licit incompletely

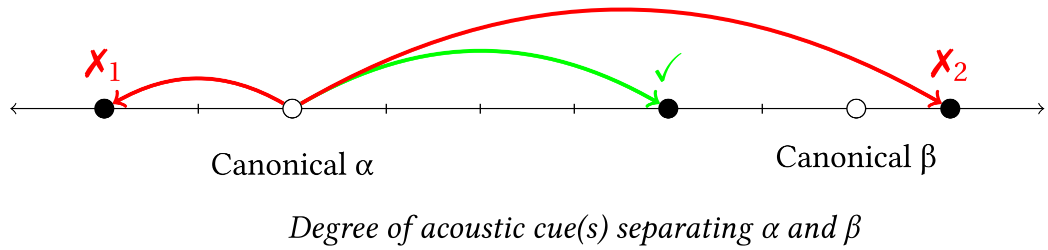
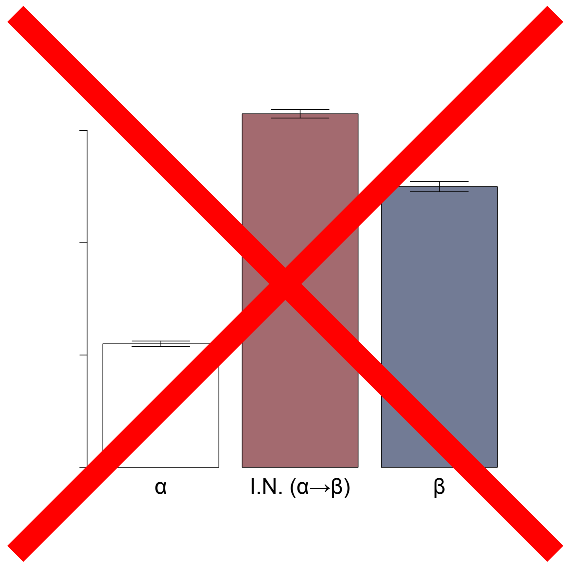


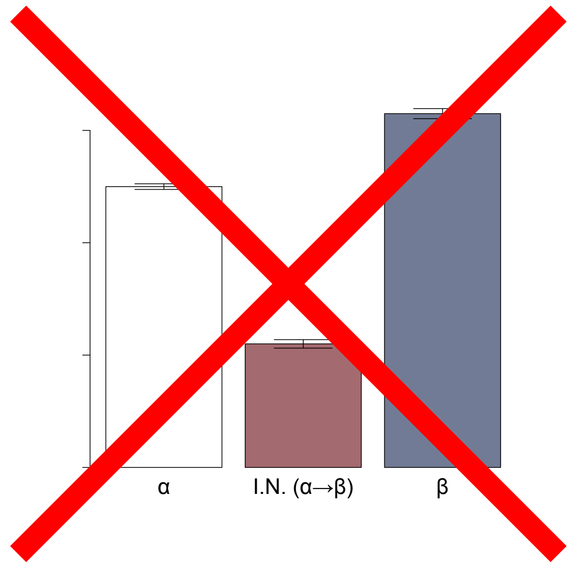
Figure 1.1: Schema of the Directionality Observation.

neutralized segment (the checkmark) is closer to the canonical realization of [β] than the canonical realization of [α] is. Further, leftmost illicit incompletely neutralized segment (✗1) is illicit specifically because it is from the canonical realization of [β] than the canonical realization of [α] is. Second, incompletely neutralized [α] cannot surpass the canonical realization of [β]. As can be seen in Figure 1.1, the rightmost illicit incompletely neutralized segment (✗2) violates this condition: it is further to the right than the canonical realization of [β]. Finally, the licit segment (✓) is licit specifically because it violates neither of these conditions.

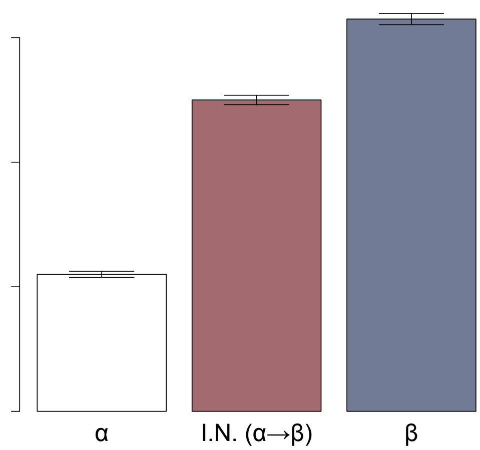
To visualize this idea in slightly more concrete terms, consider the graphs in Figure 1.2. The graphs show a hypothetical case of incomplete neutralization of segments [α] and [β] (the leftmost and rightmost bars, respectively). The middle bar shows the incompletely neutralized segment. The y-axis represents the degree of the acoustic cue(s) that separate [α] and [β]. In Figure 1.2(a), the Directionality Observation is violated because the incompletely neutralized segment (middle bar) surpasses the canonical realization of [β] (rightmost bar). Similarly, in Figure 1.2(b), the Directionality Observation is violated because the incompletely neutralized segment is further from [β] than the canonical realization of [α] is. Figure 1.2(c) shows a licit case of incomplete neutralization. The Directionality Observation, practically speaking, requires incompletely neutralized segments to fall somewhere between the canonical realizations of [α] and [β].



(a) The incompletely neutralized segment surpasses the canonical realization of [β]



(b) The incompletely neutralized segment is further from [β] than the canonical realization of [α] is



(c) Licit incomplete neutralization

Figure 1.2: Graphs of a hypothetical case of incomplete neutralization of segments [α] and [β] (leftmost and rightmost bars, respectively). The middle bar is the incompletely neutralized segment. The y-axis represents the degree of the acoustic cue(s) that separate [α] and [β]. (a) and (b) violate the Directionality Observation; (c) is a licit case of incomplete neutralization.

For example, in a canonical voicing distinction, vowels preceding voiced stops are longer than those preceding voiceless stops (Chen 1970). In the case of German final devoicing described above, this is precisely the direction of the vowel duration distinction found by Port and O'Dell (1985): on a continuum from [d]-like to [t]-like, [d̥] is more like the canonical realization of /d/ than [t] is. Similarly, Port and O'Dell (1985) also find that there is less aspiration, voicing into closure is longer, and closure duration is marginally shorter in underlyingly voiced stops—again matching the pattern of fully-realized voicing contrasts (Kingston and Diehl 1994, Lisker 1986, Kluender et al. 1988). This is a property of incomplete neutralization that any theory must explain; the model in Chapter 4 achieves this goal.

Incomplete neutralization has proved somewhat of a puzzle, since it seems to straddle the boundary between the binary contrasts normally associated with phonology, and the gradient realizations associated with phonetics. As Keating (1988, p. 287) summarized early theories of phonetics and phonology, "[t]he phonetic component consisted mostly of automatic, universal rules for implementing [the] feature matrices [provided by the phonology] as continuous physical events." Under these early models, phonetic implementation was privy only to a subset of phonological and morphological information. Incomplete neutralization was difficult to explain under these theories, since the phonetic module had no way to differentiate between phonologically neutralized segments and faithful ones. Taking the example of German devoicing, the phonetic module would have no way to know whether a [t] was underlyingly voiceless or if it had been devoiced by the phonology. As such, all [t]s that came from the phonology should be treated identically, resulting in complete neutralization.

Given the apparent link between phonetics and phonology in incomplete neutralization, its existence has been used to argue for varying degrees of interaction between these two modules. For example, van Oostendorp (2008) argues that final devoicing can be derived by allowing the phonetic module to distinguish between underlyingly voiceless and devoiced segments (by way of differing relations between a segment and its associated [voice] feature—devoiced segments maintain a 'projection' relation to their [voice] feature, while underlyingly voicless segments have no such relation.) On the other hand, Yu (2011) argues that incomplete neutralization is better analyzed in a model that allows the phonological module more control over the kinds of variation found in the phonetic implementation of contrasts. Under such a model (such as the one proposed by Kingston and Diehl (1994)), in implementing a given phonological feature, speakers choose from among multiple possible active articulations. The subphonemic differences that are the hallmark of incomplete neutralization, then, are "qualitatively not different from those observed between allophones appearing in different phonetic contexts" (Yu 2011, p. 311). In other words, incomplete neutralization is just like any other contrast, except for the fact that it is at the lower end of our range of perception.

1.1.1 A Taxonomy of Incomplete Neutralization and Subphonemic Distinctions

Many phenomena have been put forward as cases of incomplete neutralization. In this section, I provide a taxonomy of these cases, in order to deliniate the empirical scope of this dissertation. The cases fall broadly into five groups:

Featural contrasts Featural contrasts result in incomplete neutralization when an underlying featural or structural contrast is reduced in a given context. Examples include final devoicing (as above) and flapping in American English (see the following section). I will argue in Chapter 3 that contrasts based on suprasegmental categories also belong in this category.

Another example from this category is word-internal coda aspiration in Eastern Andalusian Spanish, which results in a surface distinction in aspiration duration (Gerfen and Hall 2001, Gerfen 2002). In Eastern Andalusian Spanish, word-internal codas undergo 's-aspiration'—the deletion of a coda segment, combined with aspiration on the preceding vowel and frequently lengthening of the following onset consonant (e.g., /kasta/ → [ka^{h}t.ta] 'caste'). Gerfen and Hall (2001) note that when coda /s/ undergoes s-aspiration, the result is distinct from forms in which codas with other segments undergo s-aspiration (e.g., /kasta/ 'caste' ≠ /kapta/ 's/he captures', since the coda segments {p, k} *incompletely* neutralize). Specifically, forms with coda /s/ have a longer aspiration duration than those with other codas. They argue that this result is in the predicted direction, as voiceless fricatives are realized with a greater glottal width than voiceless stops (p. 28–29).[2]

Epenthetic vs. underlying segments In some cases, epenthetic segments surface with different phonetic attributes than their lexical counterparts. For example, epenthetic [i] is shorter, backer, or both shorter and backer than lexical [i] in Levantine Arabic (Gouskova and Hall 2009). This situation follows the generalization that incomplete neutralization will surface in the canonical direction of the contrast: Gouskova and Hall (2009) argue that epenthetic [i] is "something *less* than [i]: the vowel is backer and shorter, all properties that would make this vowel closer to [ɨ] or [ə]—and, arguably, to zero" (p. 18, emphasis in original).

Similarly, vowel epenthesis in non-native clusters by speakers of English results in shorter vowels than lexical ones (Davidson 2006) and English nasal-fricative intrusive stops are shorter in duration than lexical stops (Fourakis and Port 1986)—both following the Directionality Observation.[3]

[2]Gerfen and Hall (2001, p. 29–30) also note that stop closure duration is shorter in forms with /s/ than other forms. At first glance this seems counter to expectations, but they argue that this is due to the fact that the relatively long duration in /s/ forms due to increased aspiration duration leads to shorter closure duration to compensate.

[3]Davidson (2006) also found that inserted vowels had lower F1 and F2 midpoints. She argues that this indicates the lack of a gestural target. Arguably, lacking a gestural target makes a vowel more similar to zero, and as such, this, too, follows the Directionality Observation.

Deleting a segment affects its neighbors The mirror of the previous category, deleting a segment can result in different realizations of that segment's neighbors. For example, schwa deletion in French results in a distinct pronunciation from forms in which there was no schwa to begin with (Fougeron and Steriade 1997). As Fougeron and Steriade (1997) show, segments preceding a deleted schwa may show characteristics of its underlying form (which contains schwa) including greater linguopalatal contact and longer lingual occlusion.

Similarly, /VV/ sequences in Turkish arising from the deletion of an intervocalic /g/ (e.g., /VgV/ →[VV]) differ in duration from underlying /V:/ (Rudin 1980, Dinnsen 1985). Specifically, the long vowels derived by /g/-deletion are about 13% longer than underlying long vowels. This pattern again follows the Directionality Observation: when present, /g/ would normally increase the overall duration of the /VgV/ sequence.

Lexical-level near merger In this category, minimal pairs differ based on their lexical entry, rather than phonological context. (Note that this is context-free, unlike the *featural contrasts* class above). As an example, Labov et al. (1972) describe the case of 'sauce' vs. 'source' in New York City English. In this dialect, final and pre-consonantal /r/ vocalize, creating an apparent merger between words like 'sauce' and 'source'. Labov et al. (1972) show, however, that upon examination of spectrographic evidence, the vowel in 'source' is reliably higher and/or further back than the vowel in 'sauce'. This pattern does, indeed, conform to the Directionality Observation, as (Labov et al. 1972, p. 234) point out: "the relations between /ohr/ [as in 'sauce'] and /oh/ [as in 'source'] are basically the same in all dialects, whether or not r is pronounced."

Morphological conditioning In this category, a morphologically-derived structure varies from a lexical counterpart. In Cantonese, morphologically-derived mid-rising tone has a higher F0 maximum than lexical mid-rising tone (Yu 2007b). As Yu (2007b) points out, it is worth noting that this type of subphonemic distinction is different from traditional cases of incomplete neutralization in that the underlying distinction is not made apparent on the surface by phonemic alternations. Given the fact, then, that the morphologically-derived mid-rising tone never occurs in such an alternating context, i.e., where its realization outside the influence of neutralization can be seen, it is not possible to assess such cases for conformity to the Directionality Observation.

Here and throughout I set aside those subphonemic distinctions that fall under the categories of *epenthetic vs. underlying segments* and *deleting a segment affects its neighbors*. There are two reasons for this. First, it is impossible to rule out that the distinctions found in phenomena belonging to these categories are due to byproducts of phonetic implementation, rather than incomplete neutralization *per se*. For example, intrusive stop in English has been analyzed as articulatory mistiming (Ohala 1974, Fourakis and Port 1986). The categories *lexical-level near merger* and *morphological conditioning* receive relatively less attention in the dissertation. As Yu (2007b) argues, near merger operates over the lexicon or morphology, whereas incomplete neutralization is context-dependent and operates in phonologically-defined environments. Since, *lexical-level near merger* and *morphological conditioning* fit better under the umbrella of near merger than incomplete neutralization, they are not the focus of study.

1.2 Previous Work

In this section, I describe much of the previous literature on incomplete neutralization, with a focus on experimental studies of the phenomenon. The aim of this section is to lay out the current situation in the field. In Sections 1.2.1 and 1.2.2 below, I outline previous production and perception studies of incomplete neutralization. In Section 1.2.3, I describe previous attempts to model these results. The following Section (1.3) discusses the questions that remain unanswered.

1.2.1 Production

Perhaps the most commonly-cited cases of incomplete neutralization are those that result from final devoicing. Production studies have found distinctions between underlying voiceless and devoiced segments in a number of languages, including German (Port and O'Dell 1985), Catalan (Dinnsen and Charles-Luce 1984), Polish (Slowiaczek and Dinnsen 1985, Slowiaczek and Szymanska 1989), Russian (Dmitrieva 2005), and Dutch (Warner et al. 2004, though see that article itself and Warner et al. 2006 for caveats).

On the other hand, some production studies have either found complete neutralization in such contexts, or suggest that subphonemic distinctions may be the result of extra-grammatical factors such as hyperarticulation or orthography. For example, Fourakis and Iverson (1984) argue that the subphonemic distinctions between underlyingly voiceless and devoiced segments in German final devoicing is due to 'hypercorrect spelling pronunciation' in the laboratory setting, and that neutralization is complete in more natural tasks. More specifically, they found subphonemic distinctions in a reading task in which speakers read a randomized list of conventionally spelled words from index cards, but not in a morphological paradigm elicitation task in which speakers were given the infinitival form of a verb and were asked to conjugate it. In the second task, the experiment's focus on pronunciation is masked, and speakers do not see the written form of the target word. Similarly, Warner et al. (2006) show that in Dutch, an underlying /t/ vs. /t-t/ distinction is completely neutralized in words where this distinction is not represented orthographically.

While not having received as much attention as final devoicing, flapping in American English has also been put forward as a case of incomplete neutralization.[4] In flapping, underlying /d/ and /t/ become output [ɾ] in certain prosodic configurations (Kahn 1980). While this situation looks, at first glance, like complete neutralization, some experimental

[4]For additional background on incomplete neutralization in flapping, see Section 2.1.

studies have shown that—at least for some speakers—flaps stemming from underlying /d/ (‘/d/ flaps’) and those stemming from underlyingly /t/ (‘/t/ flaps’) are in fact different from one another on the surface. These reported differences include longer preceding vowels, smaller intensity dips during closure, and shorter closure duration for /d/ flaps (Fisher and Hirsh 1976, Fox and Terbeek 1977, Zue and Laferriere 1979, Herd et al. 2010) (these distinctions all go in the direction of a canonical voicing contrast).

Some studies, though, have failed to find incomplete neutralization in flapping—at least for some speakers (Joos 1942, Port 1976, the latter of which examined only flapping contexts including [ɪ]). Huff (1980) examined New York City speakers, and found mixed results: the F1 and F2 of /æ/ and /ay/ preceded by /d/ flaps and /t/ flaps patterned with those preceded by non-flapped /d/ and /t/, though this was not the case for /aw/.[5] In his study, all pre-flap vowels in monosyllabic words were significantly longer before /d/ flaps than before /t/ flaps, but in disyllabic words, only /ay/ was significantly longer before /d/ flaps than before /t/ flaps.

To summarize, then, production studies of (in)complete neutralization have shown mixed results. Some have found incomplete neutralization, while others have not. Still others, such as Fourakis and Iverson (1984) and Warner et al. (2006) raise questions about extragrammatical factors as the source of incomplete neutralization. Further, it should be noted that the majority of these studies made use of actual words, rather than nonce words, which are susceptible to effects of lexical frequency (e.g., Gahl 2008) —a real concern, especially when dealing with such small distinctions.

1.2.2 Perception

Since the subphonemic distinctions at play in incomplete neutralization are often so small, a number of linguists have investigated the perceptibility of such reduced contrasts.

[5] I use Huff’s (1980) vowel transcriptions here, which appear to conform to Americanist system.

A perceptible contrast lends itself to an analysis in terms of contrast-preservation. On the other hand, if a hearer cannot tell the difference between two minimally different tokens, the speaker's motivation for maintaining this contrast is less obvious (see Section 4.4.1; Lindblom 1990, Diehl and Kluender 1989, Flemming 2001, Flemming 2010, Syrett and Kawahara under revision).

In an early test of incomplete neutralization's perceptibility, Port and O'Dell (1985), paired with their production experiment on German final devoicing referenced above, conducted a forced-choice identification task. Listeners were given an answer sheet with minimal pairs written out for each token, and were asked to circle which word they had heard on each trial. Listeners performed better than chance (59% correct), leading the authors to conclude that German listeners can, in fact, make use of the incompletely neutralized voicing contrast.

Also conducted as a package with a production experiment, Warner et al. (2004) present several experiments on the perception of the incompletely neutralized voicing contrast in Dutch final devoicing. In their production experiment, vowels preceding devoiced stops were on average 3.5 ms longer than those preceding underlyingly voiceless stops (though see Warner et al. 2006, which argues that effects of orthography may be responsible for at least some cases of incomplete neutralization). In one identification task, Warner et al. (2004) find that listeners perform better than chance in distinguishing devoiced from underlying voiceless tokens from two speakers, but that listeners' accuracy is still quite poor (in the condition in which listeners performed best, $d' = 0.33$).

Perception studies have also examined the ability of listeners to distinguish the reduced voicing contrast in /d/ flaps and /t/ flaps in American English. Sharf (1960), for example, presents an identification task with flapped minimal pairs from a male and a female speaker. For tokens from the female speaker, listeners were accurate more than 86% of the time, while for the male speaker, they were accurate 61% of the time. Sharf notes, however, that the female speaker may have produced a [t] rather than a flap in /t/

words, whereas the male speaker used a flap in these words,[6] perhaps accounting for the difference across speakers.

On the other hand, Malécot and Lloyd (1968), who used a similar procedure as Sharf (1960), found that 50 listeners performed at only 56.6% accuracy. Fisher and Hirsh (1976), in addition to a production study, found that five out of six phonetically-trained judges were able to correctly categorize /d/ flaps and /t/ flaps to some degree, though some judges were better than others (χ^2 values ranged from 4.24 to 24.52).[7] In a more recent study, Herd et al. (2010) present a forced-choice identification task which suggests that naive listeners cannot distinguish between pairs of words like 'leader' /lidɚ/ and liter /litɚ/, both of which surface with [ɾ] (this study gets further discussion in Section 2.1).

As with the production studies, perception studies on incomplete neutralization have found mixed results. Some studies have shown relatively good detectability of incompletely neutralized contrasts, while others have found that listeners cannot detect such contrasts. It should also be noted that the majority of the studies cited above (Warner et al. 2004 being a notable exception) make use of the percent-correct measure of accuracy, which does not take bias into account.

1.2.3 Modeling Incomplete Neutralization

A number of theories have been proposed to model incomplete neutralization. This section gives a brief overview of this literature; Chapter 4 develops a new model of incomplete neutralization, which I argue more accurately accounts for the details of incomplete neutralization phenomena.

[6]Sharf (1960) uses the term 'voiced t', rather than 'flap', describing it as 'the post-stress, intervocalic sound which occurs in American speech to represent orthographic *t* and *d* in such words as *ladder* and *latter*' (p. 105).

[7]Fisher and Hirsh (1976) note that one judge could not confidently determine whether a flap was voiced or voiceless, and so he labeled all flaps as 'flap, but I can't say whether voiceless or voiced' (p. 187).

Anderson (1975) presents an early theory to deal with an incompletely neutralized contrast. Without calling it incomplete neutralization *per se*, Anderson notes that flapping in American English manipulates the binary [voice] feature, but that the resulting differences in preceding vowel duration are gradient (p. 53-54). Working in a rule-based framework, he argues that a phonological lengthening rule (4a) followed by a phonological flapping rule (4b) cannot capture the data.

(4) a. V → [+long] / ___ [+voice]

b. {t,d} → [ɾ]/ V ___ [V, -stress]

Rather, he argues that the vowel lengthening rule must be phonetic, in order to capture its gradient nature. Given the required rule ordering, this then implies that (some) phonetic rules must be allowed to precede (some) phonological ones.

More recently, van Oostendorp (2008) argues that incomplete neutralization can be accounted for by providing limited access to a segment's derivational history to the phonetic module. Under this analysis, based on Turbidity Theory (Goldrick 2001), features can stand in two possible relations with a segment: (i) the *projection* relation, which is an "abstract, structural relationship", and (ii) the *pronunciation* relation, which "describes the output realisation of structure" (van Oostendorp 2008, p. 1368). The Turbidity Model analyzes incomplete neutralization in coda devoicing under this theory by distinguishing underlyingly voiceless coda consonants from devoiced coda consonants in terms of which relations hold between the coda consonant in question and the [voice] feature. More specifically, underlyingly voiceless coda consonants have no relation to the [voice] feature, while devoiced coda consonants are related to the [voice] feature by the projection relation — but not the pronunciation relation (underlyingly voiced segments that do not undergo devoicing have both relations with the [voice] feature). Since voiceless segments are distinguished from devoiced segments by this theory, the phonetic module can differentiate between them, allowing for, e.g., longer vowels preceding devoiced segments. This idea is schematized in (5).

(5) Schema of voiceless, voiced, and devoiced segments in Turbidity Theory[8]

/t/ → [t] (voiceless)	**/d/ → [d] (voiced)**	**/d/ → [d̥] (devoiced)**
R a t [voice]	R a d *pronunciation* *projection* [voice]	R a d *projection* [voice]

One drawback of this proposal is that it does not guarantee adherence to the Directionality Observation. That is, there is no mechanism requiring a segment with only a projection relationship to be realized in any particular way. Sticking with the example of devoicing, there is no mandate for the phonetics to realize /d/ → [d̥] with longer preceding vowels (as it does with /d/ → [d]). Such a theory, then, allows languages in which the incompletely neutralized segment does not fall within the boundaries of the canonical realizations of the two categories being neutralized, which would violate the Directionality Observation. One approach for the Turbidity Model is to argue that a projection relationship with a [voice] feature might yield a reduced version of the acoustic cues to voicing (van Oostendorp 2008, p. 1371), though this would seem to violate the spirit of the projection relationship as an "abstract, structural" one. The model developed in Chapter 4 overcomes this issue.

Another recent proposal, from Gouskova and Hall (2009), comes at the question of subphonemic distinctions from another angle, yet retains the idea of access to a segment's derivational history. They present a study showing that epenthetic vowels in Levantine Arabic are either shorter, backer, or both shorter and backer than their lexical counterparts. In order to model this phenomenon in a phonological grammar, Gouskova

[8] I have slightly modified van Oostendorp's (2008) notation in these examples.

and Hall argue that the phonetics must access an intermediate stage of phonological derivations. Following the assumptions of Optimality Theory with Candidate Chains (OT-CC McCarthy 2007), they take a candidate to consist of a derivational chain from the phonological input to the phonological surface form, with gradual, incremental steps along the way. Under their theory, the epenthetic vowel [i] goes through a number of steps—starting null, then gradually becoming more and more like a full vowel, forming a chain like that in (6) (Gouskova and Hall's (8)):

(6) Candidate chain for epenthesis of [i]:

/CC/ <CC, CɨC, CəC, CiC>

If the phonetics can access the entire chain, they argue, rather than just the last step in the chain, epenthetic vowels in Levantine Arabic might be realized as one of the other steps in the chain—explaining why some speakers produce the epenthetic vowel closer to [ɨ] or [ə] rather than [i]. Underlying /i/, however, has no such chain of changes, and therefore must surface as [i].

There is, though, a practical issue which complicates the adoption of this sort of theory in other cases of incomplete neutralization. While it is fairly apparent what should constitute a step along the chain in cases of epenthesis—decreasingly less sonorous segments—it is less clear in other cases. To again call upon our example of final devoicing, it is unclear what intermediate steps between voiced and unvoiced would correspond to the incompletely neutralized, devoiced segments.

A different class of analysis, on which I draw in Chapter 4, concerns paradigm uniformity among morphologically related forms (e.g., Steriade 2000, Yu 2007a for subphonemic differences, and Benua 1997 generally). Steriade (2000) argues, for example, that paradigm uniformity prefers words within a morphological paradigm to share certain phonological and phonetic properties. As will be familiar from Benua (1997), one observes the effects of paradigm uniformity when a given allomorph avoids a language-general pattern in favor of similarity to some given morphological neighbor. As an example, Steriade describes the (optional) schwa deletion process in French, which renders forms such as *bas retrouvé* [baʁətʁuve] 'stocking found again' → *bas r'trouvé* [ba**ʁ**tʁuve]. However, this latter form, with schwa deletion, is not identical to *bar trouvé* [baʁtʁuve] 'bar found'—the bold [**ʁ**] in the form with schwa deletion surfaces with "qualities that would only be appropriate if schwa was still present" (Steriade 2000, p. 327, Fougeron and Steriade 1997, Rialland 1986). Under Steriade's phonetic analogy analysis, forms that have undergone schwa deletion (e.g., *bas r'trouvé* [ba**ʁ**tʁuve]) mimic forms with surface schwa (e.g., *bas retrouvé* [baʁətʁuve]), thus accounting for their similarity.

A crucial assumption of this sort of analysis is that speakers have relatively fine-grained control over phonetic implementation. In the model proposed by Yu (2011), the phonology has just such control over phonetic implementation of contrast. Following Kingston and Diehl (1994, p.420 fn.2), allophones are simply "any phonetic variant of a distinctive feature specification or arrangement of such specification that occurs in a particular context." That is to say, contrastive features can vary considerably in their realization depending on their context. [+voice] in English may be realized with closure voicing intervocalically, but as voiceless unaspirated word-initially. Kingston and Diehl (1994) argue that speakers can therefore choose between the various methods to articulate a feature such as [+voice]. Yu (2011) claims that subphonemic differences such as incomplete neutralization and near merger should be seen in a similar light. The voiced 'allophone' and the voiceless 'allophone' are distinct, even if their phonetic cues are so impoverished as to "escape detection by traditional methods of linguistic data collection… Nonetheless, the contrast is maintained from the perspective of the native speaker, albeit covertly" (Yu 2011, p. 311). The model proposed in Chapter 4 builds on these general ideas.

1.3 Issues in (In)complete Neutralization

In spite of the considerable work already done on incomplete neutralization, a number of issues remain open. In this section, I catalog some of these questions, with a focus on those that will be discussed in the remainder of the dissertation. In addition to these issues, it should be noted that the facts of some proposed cases of incomplete neutralization are under dispute—as evidenced by the contradictory experimental evidence laid out in Section 1.2. Chapters 2 and 3 present several experiments which add to the empirical base from which conclusions on incomplete neutralization can be drawn.

1.3.1 Experimental Artifacts and Task Effects

Some linguists have suggested that incomplete neutralization is not a part of grammar *per se*, but rather that it is a task effect or an experimental artifact—an extragrammatical effect such as hyperarticulation in the laboratory setting or influence from orthography. Fourakis and Iverson (1984) conducted an experiment on German final devoicing with two tasks, designed to tease apart just these effects. Their first task, a word-list task, provided speakers with access to the orthographic form of each word and focused the speaker on pronunciation. Their second task, designed to reduce the influence of extragrammatical effects, was an elicitation task, in which speakers conjugated verbs based on oral cue. Since they found support for incomplete neutralization only in the word-list task, they conclude that incomplete neutralization results from 'hypercorrect spelling pronunciation', and that it does not occur in more natural contexts. In a similar vein, Warner et al. (2006) show that in Dutch, an underlying /t/ vs. /t-t/ distinction is completely neutralized in words where this distinction is not represented orthographically.

One question that remains, then, is whether these effects can be said to explain *all* cases of incomplete neutralization. Further, in those cases where extragrammatical effects do contribute to a subphonemic distinction, are they the *only* factor? (That is, could it not be that incomplete neutralization has origins both within and without the grammar?) To address these issues, Chapter 2 presents a series of experiments on the incomplete neutralization of the voicing contrast in flapping in American English. In particular, following the model of Fourakis and Iverson (1984), it reports paired tasks—one designed to encourage the effects of hyperarticulation and orthography, and one designed to repress them. Through two iterations of this experiment, Chapter 2 shows that there is no significant difference between these two task types, suggesting that hyperarticulation and orthography play at most a minimal role in this case of incomplete neutralization.

1.3.2 Byproducts of Phonetic Implementation

As discussed in Section 1.1.1, some putative cases of incomplete neutralization are susceptible to analyses as byproducts of phonetic implementation (e.g., coarticulation). For example, Ohala (1974) and Fourakis and Port (1986) treat the case of intrusive stops in English as a matter of (mis-)timing of articulatory gestures. If the phenomenon in question is simply a matter of phonetic implementation, it is not, strictly speaking, incomplete neutralization as two segments are not neutralized phonologically. In Chapter 3, I present experimental evidence from the case of Japanese monomoraic lengthening. I show that this process (i) is motivated by a clearly phonological, rather than phonetic, bimoraic minimality constraint, which governs many Japanese morphophonological patterns (Itô 1990, Poser 1990, Mester 1990, Itô and Mester 1992, Mori 2002), and (ii) can still result in incomplete neutralization.

1.3.3 Motivations and Causes for Incomplete Neutralization

One puzzle about incomplete neutralization has been its cause. On the one hand, preventing the complete obliteration of a contrast can be seen as an effort on the part of the speaker to aid the hearer in perceiving what has been said (Lindblom 1990, Syrett and Kawahara under revision). On the other hand, however, if such a goal were prioritized, no neutralization would occur at all. These competing demands—speaker's effort and maximization of contrast for the hearer's benefit—have received some attention at the inventory level (e.g., Lindblom 1986, Flemming 2001), and the phonetic level (e.g., Diehl and Kluender 1989, Kingston and Diehl 1994, Scarborough 2003, 2010, Flemming 2010). Under a view such as this, incomplete neutralization would be seen as a compromise between these two sets of priorities.

Building on Lindblom's (1990) H&H Theory, and the idea that there is a continuum between *hyper-* and *hypo*-speech, I argue that incomplete neutralization also runs along a continuum: from highly perceptible to imperceptible. While the incomplete neutralization of German final devoicing, for example, results in vowel duration distinctions on the order of 15 ms, and listeners can distinguish underlyingly voiced from devoiced tokens above chance (Port and O'Dell 1985), I show in Chapter 2 that listeners cannot distinguish /d/ flaps from /t/ flaps, though they do produce a distinction. If the underlying voicing contrast is not perceptible, then it is of no help to the hearer—a speaker's effort to maintain this contrast would be wasted. I argue that in spite of this finding, the origins of incomplete neutralization—perceptible or otherwise—are indeed in the speaker-hearer relationship. Linguistic mechanisms that are motivated in one context may apply in other contexts as well, even where they are not motivated (for further discussion, see Section 2.8). Perceptible incomplete neutralization, then, is motivated by the speaker-hearer relationship, and any machinery needed to handle it may also be used for the purposes of imperceptible incomplete neutralization.

1.3.4 Predicting the Direction of Incomplete Neutralization

A final issue to be discussed is the *direction* of incomplete neutralization. When a contrast for feature F, on which segments A and segment B differ, is incompletely neutralized, it will be realized somewhere on a continuum between the canonical realizations of A and B.[9] To take a more concrete example, in the case of final devoicing, where a [voice] contrast is being reduced, a devoiced /d/ will be either more [d]-like or more [t]-like. Which pole an incompletely neutralized segment ends up at is not random. Recall the Directionality Observation in (7), repeated from (3):

(7) **Directionality Observation**

Summary: Incompletely neutralized segments fall somewhere between the canonical realizations of the two neutralized segments.

Definition: For an underlying segment /α/ in an incompletely neutralized contrast with /β/:

- Given a continuum $\mathbb{C}$ along the acoustic cue(s) separating the canonical realization of [α] from the canonical realization of [β]:
  - Incompletely neutralized [α] will fall closer to [β] than the canonical realization of [α] is, and
  - Incompletely neutralized [α] will not surpass the canonical realization of [β].

In our devoicing example, the Directionality Observation predicts that devoiced /d/ will surface more [d]-like than underlying /t/ surfaces (that is, devoiced /d/ will surface somewhere between the canonical realizations of [d] and [t]). This, in fact, is the case. In their study, Port and O'Dell (1985) found that devoiced stops — as compared with underlyingly voiceless stops—had longer preceding preceding vowels (approximately 15 ms), less aspiration (approximately 15 ms), more voicing into closure (by about 5 ms), and marginally shorter closure duration. These match with the canonical realization of voiced segments (Chen 1970, Kluender et al. 1988, Kingston and Diehl 1994, Hombert et al. 1979,

[9] It is worth noting that this need not have been the case. One could imagine a system of contrast preservation that maintained a voicing contrast by some completely unrelated means, such as fronting or palatalization. The successful theory of incomplete neutralization will rule out such possibilities.

see, e.g.,), albeit to a lesser degree.

Not all theories of incomplete neutralization straightforwardly account for the Directionality Observation. As described in Section 1.2, The Turbidity Model of incomplete Neutralization (van Oostendorp 2008) argues that the phonetics can differentiate between, for example, underlyingly voiceless and devoiced stops. Short of an *ad hoc* requirement on realizing devoiced stops in a particular manner, though, there is no guarantee that such segments will follow the Directionality Observation.

Under the Candidate Chain analysis of incomplete neutralization (Gouskova and Hall 2009) theory, in which Candidate Chains begin at the input and progress, one markedness or faithfulness violation at a time, to the output. This approach fares better at generating the Directionality Observation. Under the (fairly generous) assumption that any step along the chain will be on the continuum between the canonical realization of the contrasted segments, then regardless of which step the phonetics chooses to realize, it will be on the continuum.

In Chapter 4, I argue that the directionality of incomplete neutralization is best modeled by transderivational identity (Benua 1997). Building upon Steriade (2000) and Yu (2007a), I suggest that paradigm uniformity (specifically identity to a typewise-frequent base) is the force that pressures subphonemic distinctions to resist complete neutralization. Under this model, an incompletely neutralized form faces pressure to neutralize, but at the same time to remain faithful to a morphologically-related form at the phonetic level. These competing pressures, then, result in a compromise: incomplete neutralization.

1.3.5 Summary

To summarize, the dissertation poses the following following questions and proposes their accompanying responses:

Q: Is incomplete neutralization due entirely to experimental artifacts such as hyperarticulation and influence from orthography?

A: In Chapter 2, for American English Flapping, I show that two experimental tasks—one designed to exaggerate such experimental artifacts (the ‘minimal pair reading task’), and one designed to reduce such artifacts (the ‘wug task’)—both result in a similar degree of neutralization. That is, incomplete neutralization was found in both task types. I argue that incomplete neutralization is therefore not due solely to these types of effects.

Q: Do cases of ‘truly phonological’ incomplete neutralization exist, or can all proposed instances be relegated to phonetic implementation?

A: In Chapter 3, I show that monomoraic vowel lengthening in Japanese results in the incomplete neutralization of the short/long vowel length contrast. The bimoraicity requirement that drives this lengthening is ‘deep’ in the phonology of Japanese (Itô 1990), and cannot be considered a byproduct of phonetic implementation. For further discussion, see Section 3.3.3.1.

Q: What sorts of contrasts can be incompletely neutralized?

A: While the most widely reported cases of incomplete neutralization focus on feature- and segment-level phenomena (e.g., final devoicing), I argue that processes operating over suprasegmental structure can also lead to incomplete neutralization. The Japanese case study exemplifies this point: monomoraic vowel lengthening operates over moras, and results in incomplete neutralization. For further discussion, see Section 3.4.3.2.

Q: What motivates incomplete neutralization?

A: I argue that incomplete neutralization runs along a continuum from highly perceptible to imperceptible. Perceptible incomplete neutralization is straightforwardly motivated by speakers' desires to maintain a contrast (see, e.g., Lindblom 1986, 1990, Diehl and Kluender 1989, Kingston and Diehl 1994, Flemming 2001, 2010, Scarborough 2003, 2010, Syrett and Kawahara under revision). Further, imperceptible incomplete neutralization is an extension of this desire: speakers are motivated to preserve a contrast, even if they end up failing. For further discussion, see Section 4.4.1.

Q: How can the 'direction' of incomplete neutralization be predicted?

A: The Directionality Observation captures the fact that incompletely neutralized segments fall somewhere between the canonical realizations of the segments that represent the underlying contrast. This notion, codified in (3) above, is captured in the model described in Chapter 4. For further discussion, see Section 4.4.

Q: How should we model incomplete neutralization?

A: I argue that incomplete neutralization is best modeled as a compromise between faithfulness to a morphologically-related base form and pressure to hit phonetic targets. Specifically, in Chapter 4, I model this compromise in a phonetics with weighted constraints. Paradigm Uniformity constraints, which penalize deviation from a morphologically-related base form, compromise with constraints that enforce phonetic targets in order to yield forms that are only partially neutralized.

Chapter 2

Flapping in American English

In American English, /d/ and /t/ surface as [ɾ] in certain prosodic configurations (Kahn 1980), a process known as 'flapping' or 'tapping'. A number of studies have examined the proposition that the underlying voicing status of a flap might be discernible on the surface—that is, that flapping incompletely neutralizes the underlying voicing contrast (see, e.g., Fisher and Hirsh 1976, Fox and Terbeek 1977, Zue and Laferriere 1979, Joos 1942, Port 1976, Huff 1980, Herd et al. 2010). This chapter presents a suite of production and perception experiments on flapping in American English, with several goals in mind.

First, as is discussed in Section 2.1, the experimental evidence regarding the (in)complete neutralization of the voicing contrast in flapping is mixed. The production experiments (Experiments 1 and 2, in Sections 2.3 and 2.4, respectively), aim to add to—and clarify—this body of evidence. Second, and also addressed by these production experiments, are the possible effects of hyperarticulation and orthography on the findings of incomplete neutralization in flapping. Specifically, the production experiments show that the underlying voicing distinction between /d/ and /t/ is, indeed, maintained on the surface in flaps—though it is quite reduced—and that this finding holds even when the effects

Portions of this chapter stem from Braver (2010), Braver (2011), Braver (2013), and Braver (revision invited).

of hyperarticulation and orthography are mitigated. Finally, the perception experiments (Experiments 3, 4, and 5, in Sections 2.5, 2.6, and 2.7) show that listeners' poor performance in identifying /d/ flaps from /t/ flaps (e.g., Herd et al. 2010, Experiment 3) extends to discrimination of these sounds as well. The inability of listeners to properly categorize—let alone discriminate—this reduced contrast implies that speakers do not maintain it in a manner that is useful for the hearer.

Chapter Contents

2.1 Previous Work 29
2.2 Remaining Issues and Motivations for the Current Studies 32
2.3 Experiment 1: Production of Flaps in American English (I) 35
2.4 Experiment 2: Production of Flaps in American English (II) 61
2.5 Experiment 3: Identification Perception of Flaps in American English 69
2.6 Experiment 4: ABX Perception of Flaps in American English 74
2.7 Experiment 5: 2AFC Perception of Flaps in American English 80
2.8 Chapter Discussion and Conclusions 83

2.1 Previous Work

While final devoicing has received much attention in experimental studies (as detailed in Section 1.2), flapping in American English has also been the subject of some interest to those studying incomplete neutralization. In particular, a number of studies report that for some speakers, /d/ flaps and /t/ flaps are, in fact, distinguishable from one another—in other words, they maintain some trace of the voicing status of their input correspondent. These reported distinctions between /d/ flaps and /t/ flaps include differences in the duration of the preceding vowel, degree of intensity dip during closure, and duration of closure (Fisher and Hirsh 1976, Fox and Terbeek 1977, Zue and Laferriere 1979). Other studies, however, report that flapping is completely neutralizing, at least for some speakers (Joos 1942, Port 1976, the latter of which only examined flapping contexts including [ɹ]).

Further complicating matters, some studies have found mixed results. Huff (1980), in his examination of New York City speakers, found that the F1 and F2 of /æ/ and /ay/ preceded by /d/ flaps and /t/ flaps pattered with those preceded by non-flapped /d/ and /t/, though this was not the case for /aw/.[1] Similarly, while this study found that all vowels were significantly longer before /d/ than before /t/ in monosyllabic words (where flapping occurred across a word boundary), only /ay/ matched this pattern in a disyllabic flapping context (where flapping occurred within a word).

Studies of the perceptibility of /d/ flaps and /t/ flaps has shown equally mixed results. Sharf (1960) presents an identification task with English minimal pairs recorded by two speakers—one male, and one female. When identifying tokens from the female speaker, listeners were accurate 86% of the time; tokens from the male speaker, however, were only accurately identified 61% of the time. Sharf notes, though, that the female speaker may have produced a [t] rather than a flap in the /t/ words, perhaps accounting for this difference.[2] In a similar experiment, Malécot and Lloyd (1968) found that their 50 listeners performed at only 56.6% accuracy. Fisher and Hirsh (1976), in addition to a production study, found that five out of six phonetically-trained judges were able to correctly categorize /d/ and /t/ flaps to some degree, though some judges were better than others (χ^2 values ranged from 4.24 to 24.52).[3]

A more recent examination of both the production and perception of flapping in

[1] As in Chapter 1, I use Huff's (1980) vowel transcriptions in describing his results.

[2] Sharf (1960, p. 105) uses the term 'voiced t' to refer to flaps, describing this sound as 'the post-stress, intervocalic sound which occurs in American speech to represent orthographic *t* and *d* in words such as *ladder* and *latter*'.

[3] Fisher and Hirsh (1976, p. 187) note that one judge could not confidently determine whether a flap was voiced or voiceless, and so labeled all flaps as 'flap, but I can't say whether voiceless or voiced.'

American English was undertaken by Herd et al. (2010). This study found evidence of incomplete neutralization in actual English words: vowels preceding /d/ flaps were, on average, 6 ms longer than vowels preceding /t/ flaps—a statistically significant difference. This result trended stronger in bimorphemic words (e.g., 'wetting' vs. 'wedding'; 8 ms average difference) than in monomorphemic words (e.g., 'petal' vs. 'pedal'; 3 ms average difference). The distinction between monomorphemic and bimorphemic words is based on a small sample (three monomorphemic pairs and five bimorphemic pairs), and did not reach statistical significance. Additionally, the duration of /d/ flaps themselves were found to be on average 0.9 ms longer than /t/ flaps—a difference in duration opposite of the canonical realization of unflapped /t/ and /d/ (see, e.g., Chen 1970). While this difference was found to be statistically significant, Herd et al. (2010) themselves suggest that this is likely due to the large sample size, and dismiss the notion that flap duration actually serves as a cue to the underlying voicing status of a flap.

Herd et al. (2010) also present a forced-choice identification task using four pairs of actual English words. By carefully selecting recordings from their production experiment, three conditions were created, listed in (1):

(1) a. 'mean' condition: pre-/d/ vowels *longer* than pre-/t/ vowels by 7–9 ms
    b. 'enhanced' condition: pre-/d/ vowels *longer* than pre-/t/ vowels by 22–34 ms
    c. 'opposite' condition: pre-/d/ vowels *shorter* than pre-/t/ vowels by 5–21 ms

None of the tokens' vowel durations were artificially manipulated—the three conditions were created simply by selecting tokens that fit the criteria of each condition. Listeners heard a token from one of the four word-pairs, and were asked to select which member of the pair they had heard (e.g., upon hearing [ˈliɾɚ], they chose between 'leader' and 'liter'). On average, listeners fell near chance, scoring 52% correct in both the mean and the enhanced conditions, and 48% correct in the opposite condition. Tokens that contained an underlying /d/ (e.g., 'leader') were more often identified correctly (57%) than those containing an underlying /t/ (e.g., 'liter'; 44% correct).

2.2 Remaining Issues and Motivations for the Current Studies

In spite of the prior work on incomplete neutralization generally, and the case of flapping in particular, a number of issues remain. Chief among these issues is the potential effect of extragrammatical factors in producing subphonemic distinctions. As discussed in Section 1.2, some studies have suggested that hyperarticulation in the laboratory setting, or the influence of orthography may play a greater role in incomplete neutralization than more purely grammatical notions (Fourakis and Port 1986, Warner et al. 2006). The production experiments described in this chapter (Experiments 1 and 2, in Sections 2.3 and 2.4) were designed to tease apart such effects. Each of these experiments was comprised of two tasks: one designed to increase the effects of hyperarticulation and orthographic influence (the 'minimal pair' task), while the other (the 'wug' task, Berko 1958) was designed to reduce these effects. (This paradigm was largely inspired by Fourakis and Iverson's 1984 study of incomplete neutralization in German. That study reports two tasks—an 'elicitation' task, in which speakers conjugated verbs based on oral cues, and a 'reading' task, in which speakers read target words from index cards.) By showing that speakers incompletely neutralize under both conditions, I argue that incomplete neutralization is not 'just' an effect of the task at hand.

In the work on the perception of the /d/ flap vs. /t/ flap distinction (and incomplete neutralization more generally), the vast majority of experiments have probed the ability of listeners to identify and categorize incompletely neutralized segments, as opposed to discriminating among them (though see Matsui 2011 for a counterexample, and discussion of this trend).[4] One can ask whether listeners' poor performance on identification tasks like the one presented by Herd et al. (2010) are due to the imperceptibility of the contrast being tested, or rather, to the difficulty of the task itself. In other words, could an unidentifiable incompletely neutralized contrast be perceived in a different type of task? Experiments 3, 4, and 5 (Sections 2.5, 2.6, and 2.7) test this possibility. Experiment 3, an identification task, similar to the one conducted by Herd et al. (2010), shows that listeners do indeed find identification nearly impossible. Experiments 4 and 5, an ABX task and a 2-Alternative Forced Choice (2AFC) task, respectively, were designed to test listeners ability to *discriminate* among /d/ flaps and /t/ flaps. I argue that the results of these experiments, when taken together, suggest that listeners cannot discriminate between the two types of flaps.[5]

Given the relatively small distinction between /d/ flaps and /t/ flaps (e.g., pre-/d/ flap vowels were only 6 ms longer than pre-/t/ flap vowels in Herd et al.'s 2010 study), the question of perceptibility has serious implications for a theory of incomplete neutralization. If a hearer can perceive a contrast, however reduced, a speaker might reasonably expend effort to preserve that contrast. If, on the other hand, a very small surface distinction is indistinguishable, the speaker has little incentive to maintain it if their only motivations are cooperative communication and increasing intelligibility (see, e.g., Lindblom 1990, Smiljanic and Bradlow 2004). I argue that while the speaker-hearer relationship may well be at the heart of contrast preservation when incomplete neutralization is perceptible, it does not directly motivate cases like flapping, where contrasts are not recoverable in an idealized laboratory setting — let alone a normal

[4] For the sake of precision, the terms 'distinguish(ability)', 'perceive', and 'perceptibility' should be taken as cover terms for identification/categorization and discrimination. The more specific terms are used as appropriate.

[5] To further reduce any friction from the particular experimental task, listeners were aided by a practice phase at the start of each task (including a section with both real English words and a section with nonce words), and feedback was provided on each trial. (Herd et al. 2010 do not specify whether feedback was used in their perception task).

conversational environment. This idea is discussed further in Sections 2.8 and 4.4.1.

Two further considerations remain, both methodological in nature. First, lexical frequency is known to affect identification tasks on incomplete neutralization. In their perception study, Herd et al. (2010) found lexical frequency effects of two sorts. Words with high lexical frequency were more often correctly identified than words with low lexical frequency (59% accurate vs. 42% accurate). Additionally, there was an interaction between lexical frequency and underlying /d/ vs. /t/: words with underlying /d/ had a relatively small difference when divided between frequent and infrequent words (62% correct for words with high lexical frequency, 51% accurate for words with low lexical frequency); lexical frequency, however, had a much greater impact on words with underlying /t/ (55% correct for words with high lexical frequency, 33% correct for words with low lexical frequency). In order to mitigate possible effects of lexical frequency, all experiments presented in this chapter make use of nonce words only. As an added benefit, using nonce words also allows for a greater number of stimuli—English has only a limited number of actual minimal pairs that differ just by /d/ or /t/ in the flapping environment under investigation here.

The second methodological consideration concerns bias. In their identification task, Herd et al. (2010) found that /d/-words were accurately identified 57% of the time, while /t/-words were accurately identified only 44% of the time. The results of the perception tasks presented here are given in d' which, unlike the percent-correct measure reported by Herd et al. (2010), teases apart sensitivity from bias (Macmillan and Creelman 2005). Bias of this sort can lead to misinterpretation of experimental results. For example, if a listener says that they heard a /d/ word on all trials—regardless of what they had actually heard—they would still be accurate on 100% of all /d/ trials. (They would, of course, also score 0% correct on /t/ trials.) The percent-correct measure could lead to an interpretation that listeners are good at *finding* /d/ words (and bad at finding /t/ words), when in reality the results are due only to listeners' bias toward *responding* /d/.

2.3 Experiment 1: Production of Flaps in American English (I)

Experiment 1 is a production study of flaps in American English, designed to address a number of the issues laid out in Section 2.2. The experiment is divided into two nonce-word production tasks: (a) a 'minimal pair' task, in which speakers read from a screen that presents both members of a minimal pair simultaneously, and (b) a 'wug' task, in which speakers are asked to change nonce verbs, one at a time, from their plain form (e.g., *unteed*) to their *-ing* form (e.g., *unteeding*). While the minimal pair task was designed to emphasize the potential effects of hyperarticulation and orthography, the wug task was designed to mitigate these effects. Tokens from both tasks show evidence of incomplete neutralization of /d/ and /t/ in the flapping environment, suggesting that the incomplete neutralization of the voicing contrast in this context is not due solely to these extragrammatical effects.

2.3.1 Methods

The experiment was divided between the two tasks described above—the minimal pair task, and the wug task. All speakers participated in both tasks, with the order balanced across speakers. The tasks were separated by a short set of arithmetic problems, lasting approximately 5 minutes. Each task was preceded by instructions and a short practice phase.

2.3.1.1 Speakers and Recording

14 native English speakers, all undergraduates raised in New Jersey, participated in this experiment. One speaker failed to accurately read a majority of the stimuli as presented, so this speaker's data was not included in any analyses. Of the remaining 13 speakers, 10 were female, 3 were male.

All recordings were performed in a sound-attenuated recording booth in the Rutgers Phonology and Field Research Laboratory, using an AT44040 cardioid capacitor microphone with a pop filter, amplified through an ART TubeMP microphone pre-amplifier and JVC RX-554V receiver. The speech was digitized as WAV files at a sampling rate of 44.1kHz using Audacity. Acoustic analysis was performed using Praat (Boersma and Weenink 2009).

In both tasks, each token was recorded twice. Only the second repetition of each token was analyzed, since in many cases the participants stumbled slightly in their pronunciation of the first repetition. When the second token was deviant, or boundaries could not be determined, the first token was used instead. If both repetitions were deviant, the item was excluded.

2.3.1.2 The Minimal Pair Task

Stimuli Thirty disyllabic nonce verbs were created—half with final /d/, and half with final /t/—forming 15 minimal pairs. The first syllable of each nonce verb was either *re-* or *un-*, actual verbal prefixes in English, in order to encourage participants to accept the nonce words as possible English verbs. The onset of the second syllable was always voiceless, and was varied so as not to clash with existing English words. The vowels in the second syllable were divided evenly among high, mid, and low vowels. To these disyllabic nonce verbs, the English progressive *-ing* suffix was then added, putting the formerly word-final /d/ or /t/ into a flapping environment. The full list of stimuli is provided in Table 2.1.

Procedure On each trial, speakers were presented with both members of the target minimal pair, in frame sentences, as in (2):

(2) John was <u>unketting</u> this whole week.

John was <u>unkedding</u> this whole week.

rekadding	~	rekatting
resteeding	~	resteeting
unkadding	~	unkatting
reskeeding	~	reskeeting
respedding	~	respetting
unpadding	~	unpatting
rekeeding	~	rekeeting
unskadding	~	unskatting
unskeeding	~	unskeeting
unskedding	~	unsketting
unspedding	~	unspetting
restadding	~	restatting
unstedding	~	unstetting
restedding	~	restetting
unsteeding	~	unsteeting

Table 2.1: Experiment 1: Stimuli from the Minimal Pair Task

This procedure was repeated for all 15 minimal pairs, randomized, with 30 distractor items. Within each minimal pair, the order of /d/ and /t/ forms was randomized.

This task was designed to increase speakers' attention to pronunciation and orthography (see the 'reading task' in Fourakis and Iverson 1984). Since speakers saw both members of a minimal pair on the screen at the same time, the task highlighted the distinction between voiced and voiceless stimuli.

2.3.1.3 The Wug (Paradigm Completion) Task

The wug (paradigm completion) task follows a modified version of Berko's (1958) 'wug' task in order to direct participants focus away from pronunciation. In other words, participants' focus should be directed at filling in a morphological paradigm, rather than worrying about correct pronunciation—which might lead to hyperarticulation. This type of distraction to avoid hyperarticulation is motivated by a similar study by Fourakis and Iverson (1984), who used a morphological paradigm completion task to elicit forms in their study of the incomplete neutralization of German devoicing.

Stimuli 30 disyllabic nonce verbs, distinct from those in the minimal pair reading task were created—half ending in /d/, and half ending in /t/. These verbs were not organized as minimal pairs. The tokens were of similar form to those in the minimal pair task, with the first syllable consisting of either *re-* or *un-*, and the second syllable consisting of voiceless onsets, an equal number of high, mid, and low vowels, and a /d/ or /t/ coda. As with the minimal pair task, when these tokens are put into the progressive form, the formerly word-final /d/ or /t/ is placed into a flapping environment. The full list of stimuli is given in Table 2.2

unteed, retedd, unkeed, rekedd, unpede, repedd, resheed, reskedd, unspeed, rekad, respad, unspad, unstad, untedd, retad
reteet, retet, unkeet, unket, resket, respeet, reket, unteet, untet, unspat, retat, unpete, untat, unstat, repat

Table 2.2: Experiment 1: Stimuli from the Wug Task

Procedure Each nonce verb was put into a frame sentence that unambiguously identified it as a verb, as in (3). Note that these sentences do not contain the progressive *-ing* form of the stimuli.

(3) John learned how to <u>unteed</u> this week.

Speakers were first shown this sentence on a screen. This was followed by a frame sentence with a blank, designed to elicit to the progressive *-ing* form of that nonce verb, as in (4):

(4) In fact, he was __________ this whole week.

After an initial practice phase, speakers were reliably able to fill in the desired form when reading aloud (e.g., they would read (4) as "In fact, he was unteeding this whole week.").

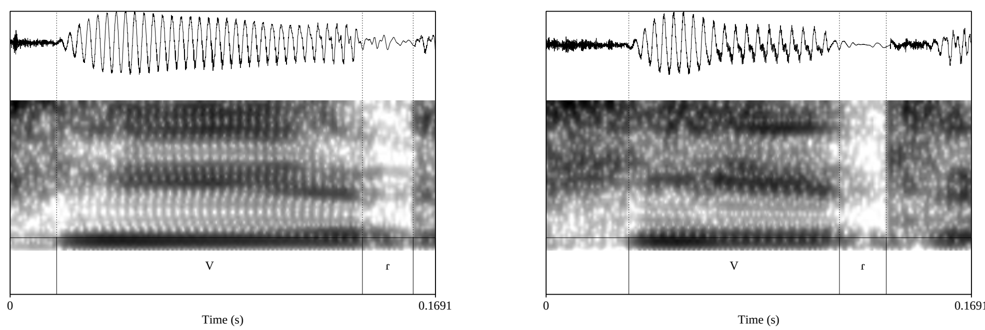
This procedure was designed to reduce effects of both hyperarticulation and of orthography (again, see Fourakis and Iverson 1984). Since speakers were engaged in a morphological task, their attention was directed away from pronunciation. Similarly, since speakers were not seeing minimal pairs, their attention was not drawn to a potential voicing contrast. Additionally, since the screen displayed a blank, rather than the actual target, speakers were less likely to be influenced by the orthographic form of the target. There is still, of course, the potential that the orthographic form of the uninflected nonce stimulus that speakers saw could have affected their production. This procedure was chosen over one involving auditory presentation of stimuli, though, to avoid effects of participants mimicking phonetic details of the stimuli (see Babel 2009, especially chapter 2, for an overview of phonetic accommodation in both the laboratory setting and in other contexts).

2.3.1.4 Measurements

In considering whether a particular phenomenon is completely or incompletely neutralizing, there is a question of which acoustic effects should 'count'. Since phonological features like voicing have a variety of acoustic correlates (Lisker 1986, Kingston and Diehl 1994, Stevens and Blumstein 1981), one can reasonably ask how many of these correlates must be divergent between two forms in order to consider neutralization incomplete. In studies of incomplete neutralization involving voicing contrasts (like both final devoicing and flapping), duration of the preceding vowel has been a commonly examined phonetic reflex of the phonological voicing contrast. Voiced consonants are known to be preceded by longer vowels than voiceless consonants, even in putatively neutralizing contexts

(Chen 1970),[6] so a common hypothesis is that in cases of incomplete neutralization, vowels preceding underlyingly, phonologically voiced consonants should be longer than those preceding underlyingly, phonologically voiceless consonants.

In this study, preceding vowel duration was one of the measurements considered. The duration of vowels was measured from the onset of voicing to the onset of flap closure. The onset of flap closure was marked as the location on a spectrogram with a marked reduction in formant structure accompanied by a drop in intensity and periodic energy, as shown in the waveform. See Figure 2.1 for representative spectrograms.



'unteeting' *'unteeding'*

Figure 2.1: Experiment 1: Representative spectrograms (from Speaker 13, minimal pair task)

In addition to the duration of the preceding vowel, the duration of the flap closure itself was measured, since voiced segments tend to have shorter closures than voiceless ones (Kluender et al. 1988). The onset of flap closure was measured as above; the end of flap closure was determined by a rise in amplitude, the resumption of formants, and, if present, burst. F0 and F1 were also measured at the onset and offset of flap closure since they tend to be lower next to voiced consonants than voiceless consonants (Kingston and Diehl 1994, Hombert et al. 1979). Finally, the amount of intensity dip during flap closure was

[6] Chen's study included words from Russian with word-final underlying voiced and voiceless stops—for example, /glub'/ and /glup/. In footnote 6 on page 135, Chen notes that there is a devoicing process word-finally in Russian, but argues that this is actually evidence for the underlying voicing distinction being maintained by length of the preceding vowel, 'even though voicing itself is absent phonetically'.

measured as the difference (in dB) between the maximum intensity during the preceding vowel and the minimum intensity during flap closure, as determined by Praat. Voiced segments, since they have voicing during the closure, have a smaller intensity dip than voiceless consonants. Additionally, Warner et al. (2009) found that the size of the dip in intensity serves as a cue in the perception of /t/ and /d/ as flaps—a larger intensity dip leads to increased likelihood of perception of a flap.

In order to remove tokens that may not have had a flap articulation, the distribution of the /d/ tokens and the /t/ tokens for the various measures listed here, as well as for the percent of closure voicing in the /d/ or /t/ (as determined by Praat's voicing report function) was examined. The distribution, on visual inspection, was not obviously bimodal for these measures, which would have indicated a categorical distinction between flaps and unreduced alveolar stops. This may be due to the fact that flap articulation is highly varied (Warner et al. 2009). As a cutoff point, tokens that measured larger than 2 standard deviations above the mean for these measurements were considered outliers and discarded.

2.3.1.5 Statistical Analyses

For each of the measurements discussed in §2.3.1.4, a linear mixed model (Baayen 2008) was run using the `lme4` package (Bates and Maechler 2009) in R (R Development Core Team 2009). The measurements were regressed against a model in which underlying voicing status (/t/ or /d/), task, and vowel quality were fixed factors, and speaker and item were random factors. Vowel quality, which was contrast coded, was included to soak up variability—vowel quality is known to affect the measurements used here, including vowel duration (Peterson and Lehiste 1960, Lehiste 1970) and F0 (Whalen and Levitt 1995, Lehiste and Peterson 1961). An interaction term between underlying voicing status and task was also included to investigate an effect of task of degree of neutralization—that is, the interaction term reflects effects of underlying voicing status that may be present in one task, but not the other. The procedure for calculating degrees of freedom in this type of model has not yet been discovered, so significance of the coefficients were checked by the Markov Chain Monte Carlo method, using the `pvals.fnc` function of the `languageR` package for R (Baayen 2009).

Post-hoc ANOVAs were conducted for each speaker to locate differences in the measured properties based on underlying voicing status and task, with vowel quality as an additional factor. The underlying voicing status/task interaction was again included. An additional goal of the post-hoc ANOVAs was to locate which, if any, speakers contributed heavily to the significance of effects of underlying voicing status.

The question of multiple comparisons arises in both the linear mixed model and the post-hoc ANOVAs. Of the measurements taken, F0 at onset and F0 at offset are related comparisons, as are F1 at onset and F1 at offset. As such, the significance level for these measurements in the linear mixed model was Bonferronized and set to $\alpha = 0.05/2 = 0.025$, since each F0 measurement and F1 measurement were related to one other measurement. The post-hoc ANOVAs, which were designed in part to indicate which speakers showed the strongest effects, compare 13 speakers, so $\alpha = 0.05/13 = 0.0038$.

2.3.2 Results

2.3.2.1 The Effect of Underlying Voicing Status (/t/ vs. /d/)

To support the hypothesis that flapping in American English incompletely neutralizes /t/ and /d/, a significant distinction would need to be found between /t/-flaps and /d/-flaps on some correlate of voicing. The overall means for each of the correlates of voicing, by underlying voicing status, are shown in Figure 2.2. The linear mixed models showed a significant effect of underlying voicing status on pre-flap vowel duration, but not on any of the other measured correlates of voicing, as can be seen in Table 2.3.

In terms of vowel duration, for example, the mean /d/ flap duration was 8.76 ms longer than the mean /t/ flap duration (see Figure 2.3, enlarged from the first panel of Figure

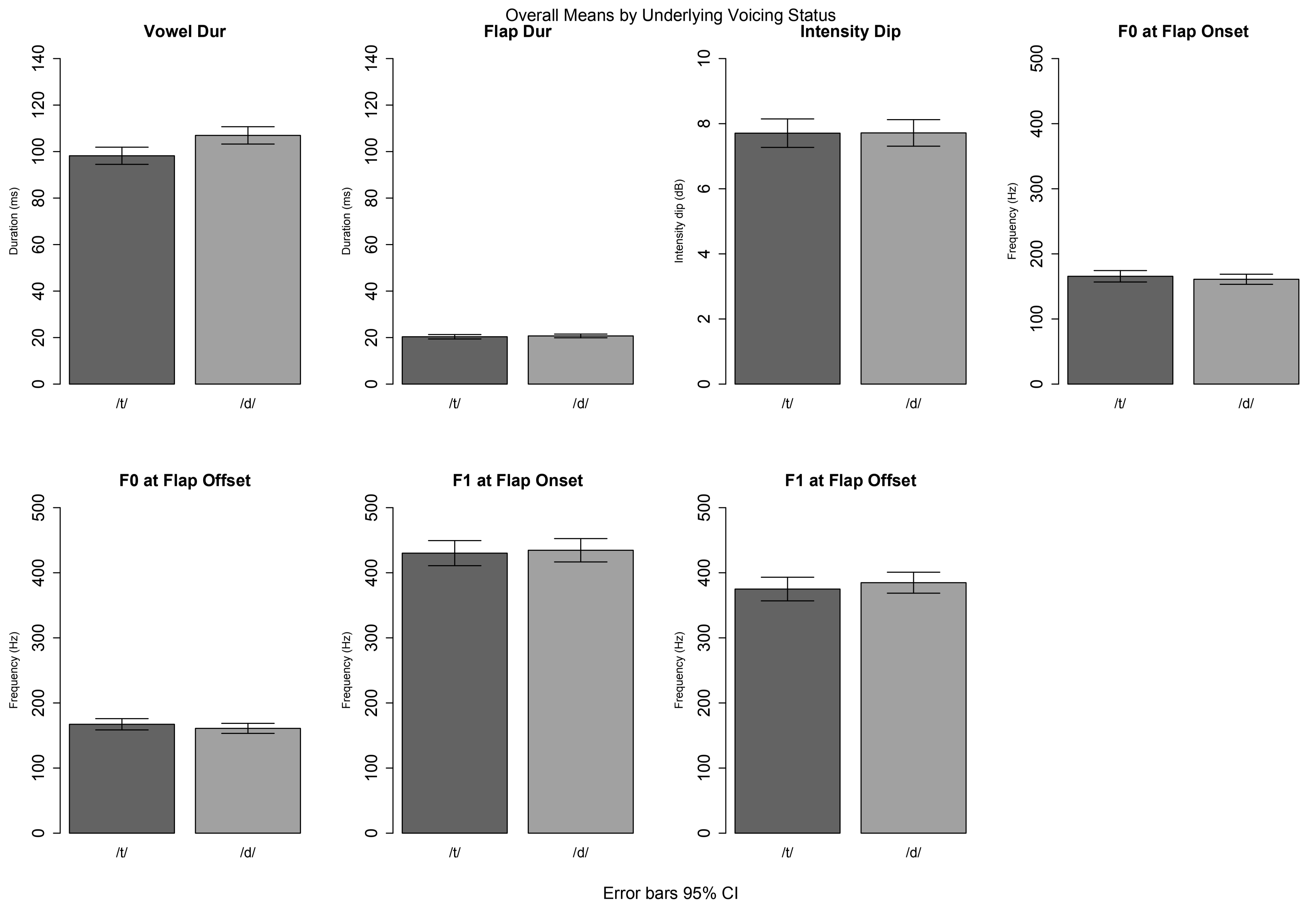


Figure 2.2: Experiment 1: Overall means by underlying voicing status across minimal pair and wug tasks combined

Effect of Underlying Voicing Status		
Dependent Variable	t	p
Vowel Duration	-2.105	< 0.05
Flap Duration	-1.076	$(n.s.)$
Intensity Dip	-0.883	$(n.s.)$
F0 at Onset	0.490	$(n.s.)$
F0 at Offset	1.229	$(n.s.)$
F1 at Onset	0.080	$(n.s.)$
F1 at Offset	0.892	$(n.s.)$

Table 2.3: Experiment 1: Results of linear mixed models (Section 2.3.1.5): effect of underlying voicing status (/d/ vs. /t/). (F0 and F1 measures: $\alpha = 0.025$ due to Bonferronization; other measures: $\alpha = 0.05$)

2.2, and Table 2.4). This distinction was significant ($t = -2.105, p < 0.05$). Mean flap closure duration did not vary significantly between /t/ flaps and /d/ flaps, even before Bonferronization (Figure 2.4, repeated in part from Figure 2.2, $t = -1.076, n.s.$). The mean difference in flap closure duration between /d/ flaps and /t/ flaps was 0.36 ms. An effect of underlying voicing status also failed to reach significance for intensity dip during flap closure, with a mean difference between /d/ flaps and /t/ flaps of 0.09 dB (Figure 2.5, $t = -0.883, n.s.$). Plots for the effect of underlying voicing status on F0 and F1 measures are included as part of Figure 2.2.

	Mean	SD
Pre-/d/ flap	106.96	25.49
Pre-/t/ flap	98.20	24.94

Table 2.4: Experiment 1: Mean and standard deviation of vowel durations (in ms) for each category

2.3.2.2 The Effect of Task

The main effect of task represents the effect that the two different tasks—the wug task on the one hand, and the minimal pair task on the other—have directly on the various correlates of voicing that were measured. This effect does not take into account

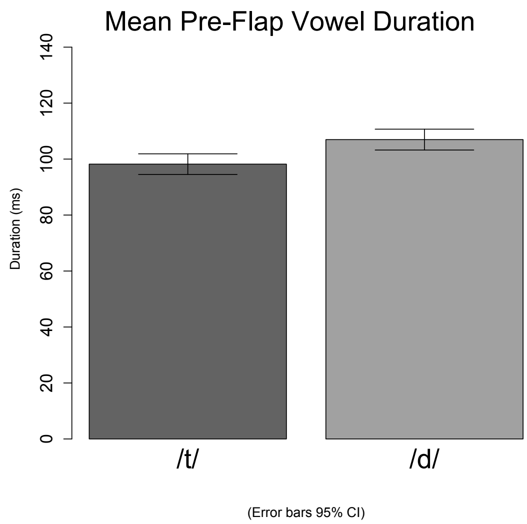


Figure 2.3: Experiment 1: Mean duration of pre-flap vowels by underlying voicing status

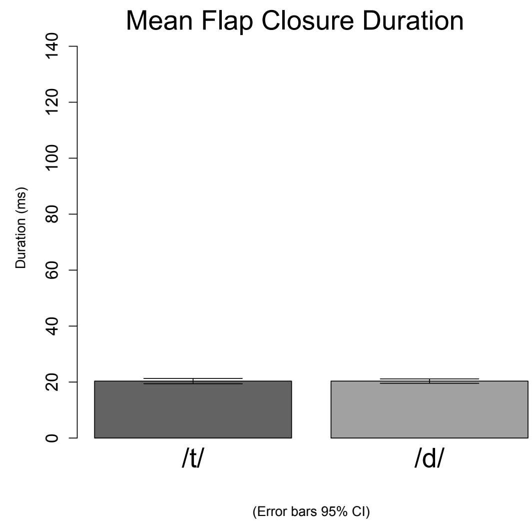


Figure 2.4: Experiment 1: Mean flap closure duration by underlying voicing status

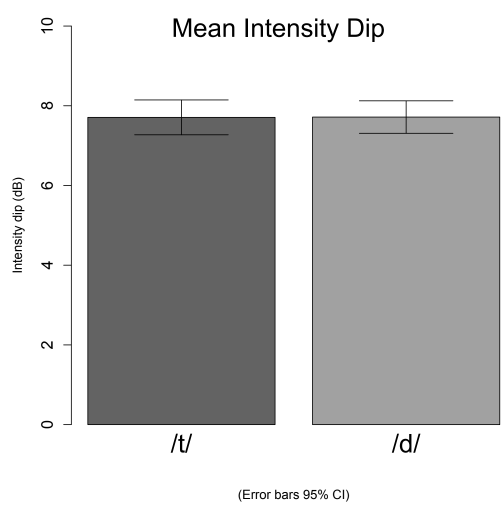


Figure 2.5: Experiment 1: Mean intensity dip by underlying voicing status

underlying voicing status—and thus does not show the effect of task on the degree of neutralization. Rather, it shows the effect of task on the various measures. For example, this main effect of task shows whether vowel duration in the wug task was, overall, longer or shorter than vowel duration in the minimal pair task.

On the other hand, the interaction of task and voicing status shows whether task has an effect on how much the underlying voicing status of a segment affects the individual measures of voicing. In other words, this interaction effect shows the effect of task on the degree of incomplete neutralization. For example, the interaction effect shows whether the change in vowel duration between /d/ flaps and /t/ flaps varies based on whether that token was from the wug task or the minimal pair task.

I begin by discussing the main effect of task below, followed by the effect of the interaction between task and voicing status.

Main effect of task There was a significant main effect of task on several of the measured correlates of voicing. Pre-flap vowels were longer in the wug task than in the minimal pair task, with a mean difference of 12.21 ms (Figure 2.6, $t = 4.378, p < 0.001$). Similarly, F0 at the offset of flap closure was significantly higher in the wug task than in the minimal pair task, with a mean difference of 1.1967 Hz (see Figure 2.7, $t = 3.124, p < 0.005$).

A main effect of task F1 at flap onset (Figure 2.8) showed significance ($t = 2.770, p < 0.01$), and F1 at flap offset trended in the same direction but did not reach the Bonferronized significance level of $\alpha = 0.025$ (Figure 2.9, $t = 2.136, n.s.$). An overview of the main effect of task in the linear mixed models is shown in Table 2.5.

Main Effect of Task		
Dependent Variable	t	p
Vowel Dur	4.378	< 0.001
Flap Dur	-0.238	($n.s.$)
Intensity Dip	1.843	($n.s.$)
F0 at Onset	1.365	($n.s.$)
F0 at Offset	3.124	< 0.005
F1 at Onset	2.770	< 0.001
F1 at Offset	2.136	($n.s.$)

Table 2.5: Experiment 1: Results of linear mixed models (Section 2.3.1.5): effect of task (wug vs. minimal pair). (F0 and F1 measures: $\alpha = 0.025$ due to Bonferronization; other measures: $\alpha = 0.05$)

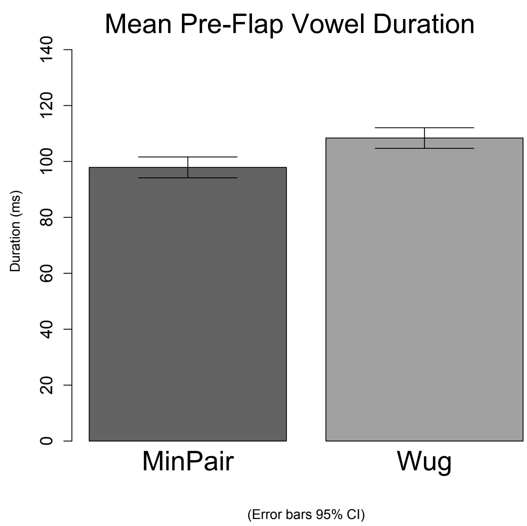


Figure 2.6: Experiment 1: Mean vowel duration by task

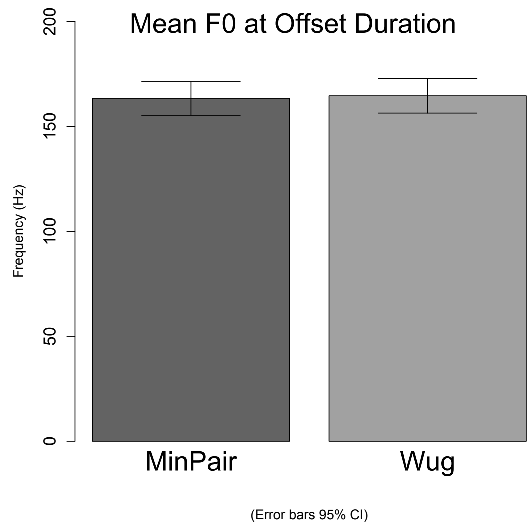


Figure 2.7: Experiment 1: Mean F0 at flap offset by task

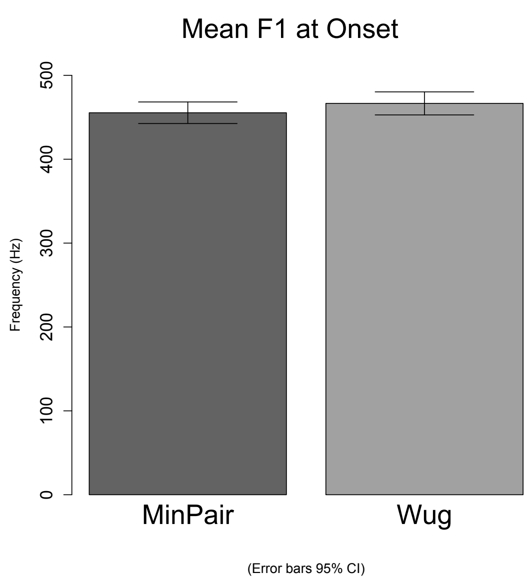


Figure 2.8: Experiment 1: Mean F1 at flap onset by task

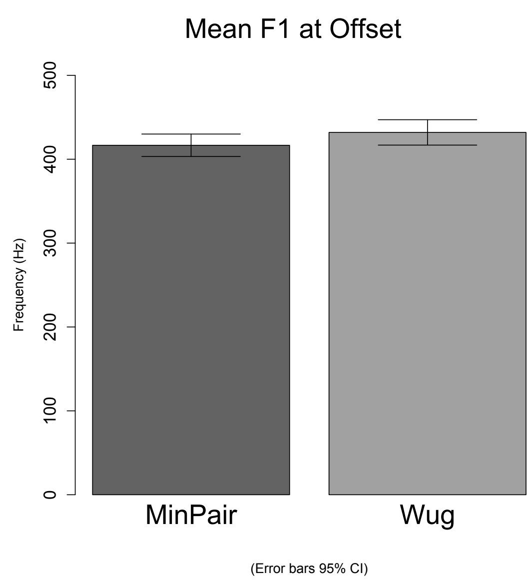


Figure 2.9: Experiment 1: Mean F1 at flap offset by task

Interaction of task with underlying voicing status (/d/ vs. /t/) The interaction of task with voicing status serves as an indicator of a task effect on the degree of neutralization on each of the measures of voicing. The results of the linear mixed models, summarized in Table 2.6, showed no significant effects of the interaction of task with voicing status on any of the measured correlates of voicing.

Effect of Interaction Between Task and Underlying Voicing		
Dependent Variable	t	p
Vowel Dur	−0.245	(*n.s.*)
Flap Dur	0.970	(*n.s.*)
Intensity Dip	1.117	(*n.s.*)
F0 at Onset	0.499	(*n.s.*)
F0 at Offset	−0.015	(*n.s.*)
F1 at Onset	−0.510	(*n.s.*)
F1 at Offset	−0.584	(*n.s.*)

Table 2.6: Experiment 1: Results of linear mixed models (Section 2.3.1.5): effect of interaction between task and voicing. (F0 and F1 measures: $\alpha = 0.025$ due to Bonferronization; other measures: $\alpha = 0.05$)

2.3.2.3 Summary of Overall Effects

The overall linear mixed models showed a significant effect of underlying voicing status on pre-flap vowel duration—pre-/d/-flaps were longer than pre-/t/-flaps. There was a main effect of task on vowel duration, F0 at flap offset, and F1 at flap onset. The interaction of task and underlying voicing status, which reflects the effects of underlying voicing status present in one task but not in another, was not significant on any of the measures.

2.3.2.4 Variation Among Speakers

Several speakers, individually, showed effects that were not identical to those seen in aggregate. In this section, I highlight some such findings in individual speakers.

Underlying voicing status 10 of the 13 speakers showed effects of underlying voicing status on the various correlates of voicing, suggesting that these speakers do not have entirely identical /t/-flaps and /d/-flaps. 3 of the speakers showed significant effects of underlying voicing status on pre-flap vowel duration—the one measure that showed overall significance—even to the Bonferronized $\alpha = 0.0038$ (an additional 7 of the 13 speakers showed similar trends that did not reach significance). I highlight here two such speakers who showed significance on various measures of underlying voicing status—speakers 6 and 12. The mean pre-flap vowel duration for each speaker, by underlying voicing status, are shown in Figure 2.10.

Speaker 6 showed significant effects of underlying voicing status on both the duration of the preceding vowel (mean difference 13.88 ms, $F(1,52) = 12.5996, p < 0.001$), and the dip in intensity during flap closure (mean difference 1.38 dB, $F(1,53) = 11.406, p < 0.005$), as can be seen in Figures 2.11 and 2.12, respectively. The full ANOVA results for speaker 6 are shown in Table 2.7.

Like speaker 6, speaker 12 showed a significant effect of underlying voicing status on pre-flap vowel duration (mean difference: 11.15 ms, $F(1,50) = 10.0784, p < 0.005$), as can be seen in Figure 2.13. Unlike speaker 6, however, speaker 12 did not show a significant effect of underlying voicing status on intensity dip during flap closure (see Figure 2.14, mean difference: 1.38 dB, $F(1,51) = 4.1662, n.s.$), but did show a significant effect of underlying voicing status on F0 at flap onset (see Figure 2.15, mean difference: 5.45 Hz, $F(1,51) = 10.3332, p < 0.005$). While the dip in intensity was not significantly different between /t/-flaps and /d/-flaps after Bonferronization, it did trend strongly towards a larger dip in intensity for /d/-flaps than /t/-flaps—the opposite direction as compared to speaker 6. The full ANOVA results for speaker 12 are shown in Table 2.8.

Interaction of task and underlying voicing status The post-hoc ANOVAs did not indicate that effects of the interaction of task and underlying voicing status rose to

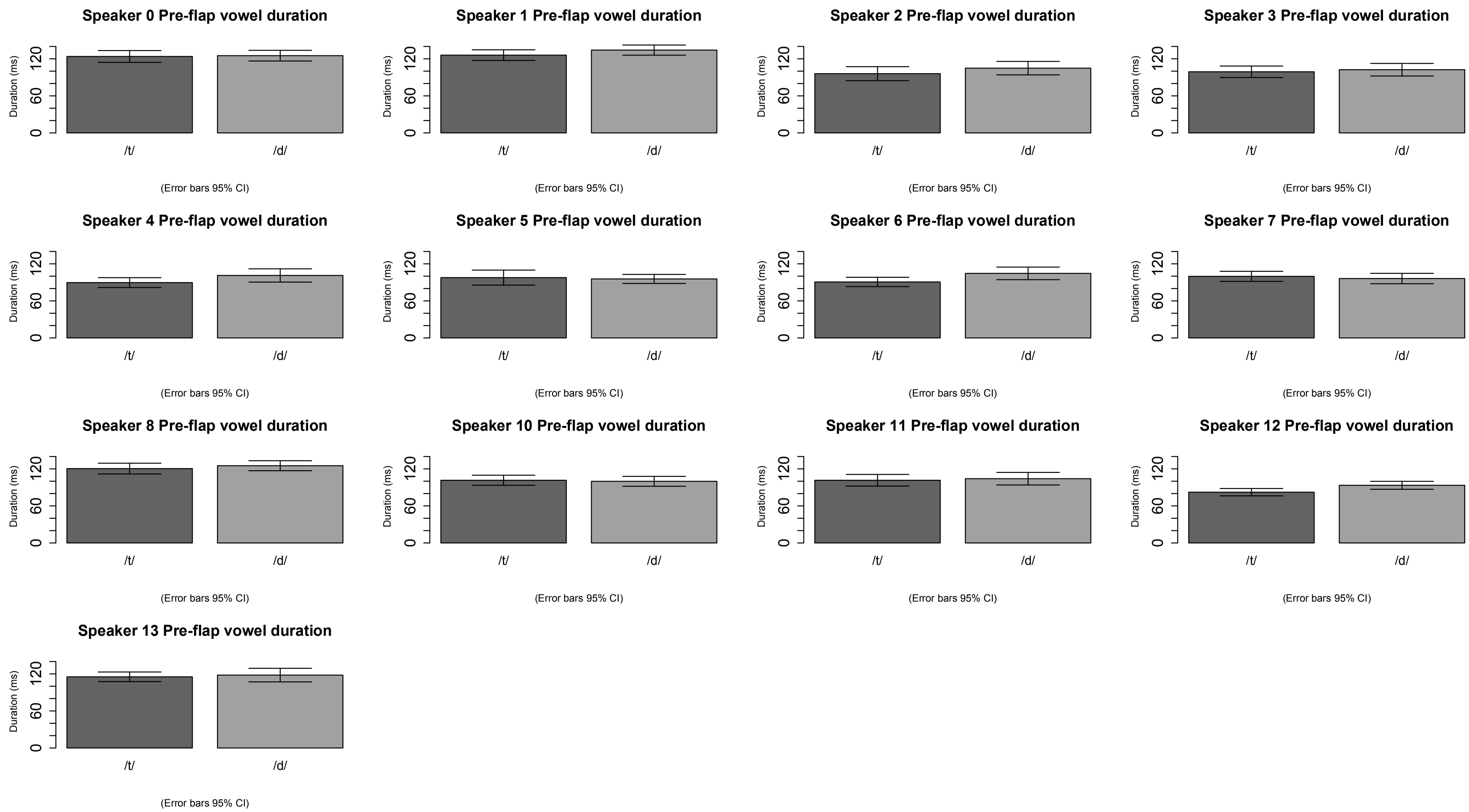


Figure 2.10: Experiment 1: Pre-flap vowel duration for each speaker, by underlying voicing status.

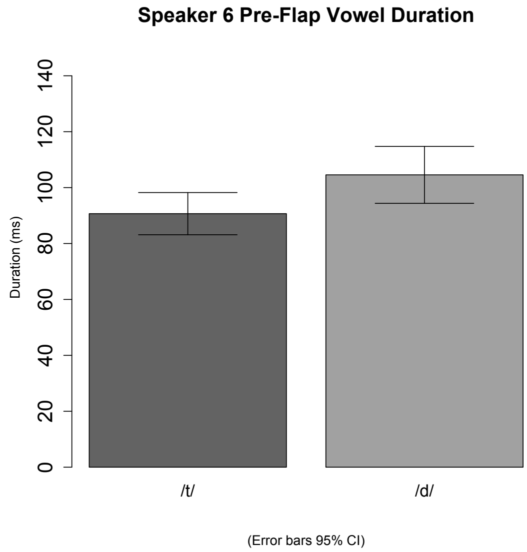


Figure 2.11: Experiment 1: Mean pre-flap vowel duration for speaker 6, by underlying voicing status

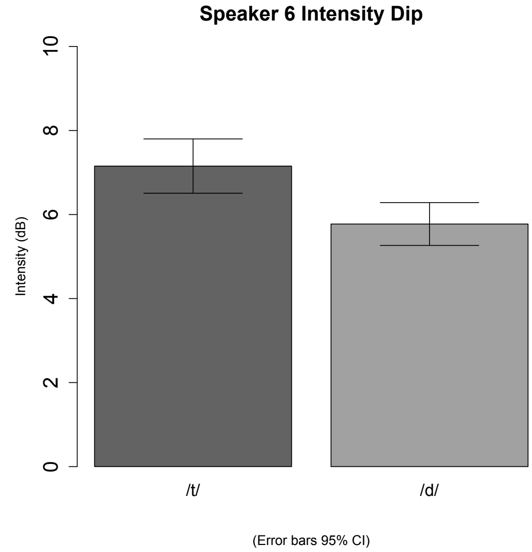


Figure 2.12: Experiment 1: Mean intensity for speaker 6, by underlying voicing status

Effect of Underlying Voicing Status: Speaker 6			
Dependent Variable	F	dfs	p
Vowel Dur	12.5996	1,52	< 0.001
Flap Dur	0.6188	1,52	(*n.s.*)
Intensity Dip	11.4060	1,53	< 0.0025
F0 at Onset	6.3148	1,52	(*n.s.*)
F0 at Offset	3.7853	1,50	(*n.s.*)
F1 at Onset	0.3012	1,53	(*n.s.*)
F1 at Offset	2.0358	1,53	(*n.s.*)

Table 2.7: Experiment 1: The effect of underlying voicing status on each of the measurements for speaker 6, as determined by the post-hoc ANOVAs described in Section 2.3.1.5. ($\alpha = 0.0038$ due to Bonferronization)

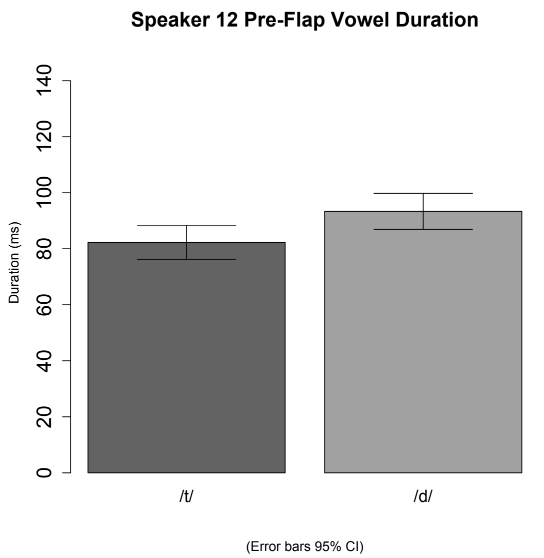


Figure 2.13: Experiment 1: Mean pre-flap vowel duration for speaker 12, by underlying voicing status

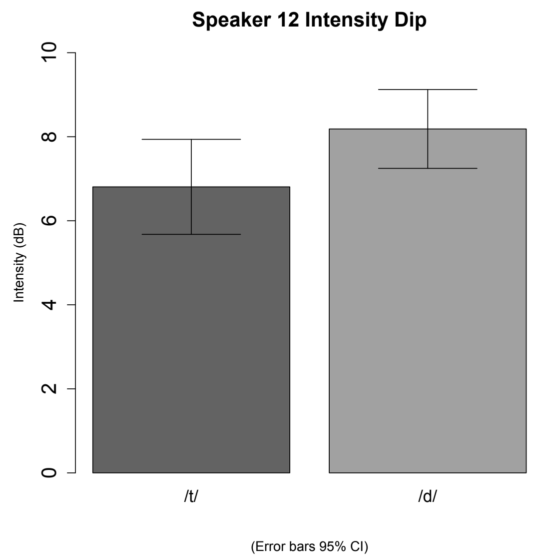


Figure 2.14: Experiment 1: Mean intensity dip for speaker 12, by underlying voicing status, n.s.

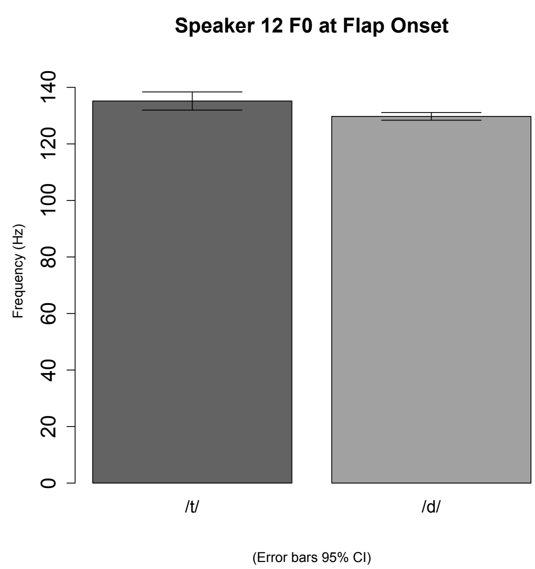


Figure 2.15: Experiment 1: Mean pre-flap vowel duration for speaker 12, by underlying voicing status

significance (after Bonferronization) for any speakers. I review here two speakers who had relatively strong trends in the direction of such effects. This interaction is represented for all speakers in the line graphs in Figure 2.16: similar slopes for each task represent similar degrees of neutralization; differing slopes represent different degrees of

Effect of Underlying Voicing Status: Speaker 12			
Dependent Variable	F	dfs	p
Vowel Dur	10.0784	1,50	< 0.0038
Flap Dur	1.2948	1,51	(*n.s.*)
Intensity Dip	4.1662	1,51	(*n.s.*)
F0 at Onset	10.3332	1,51	< 0.0025
F0 at Offset	6.0232	1,51	(*n.s.*)
F1 at Onset	6.8134	1,50	(*n.s.*)
F1 at Offset	0.1236	1,44	(*n.s.*)

Table 2.8: Experiment 1: The effect of underlying voicing status on each of the measurements for speaker 12, as determined by the post-hoc ANOVA described in Section 2.3.1.5. ($\alpha = 0.0038$ due to Bonferronization)

neutralization.

Speaker 1 trended strongly towards /d/ flap and /t/ flap having a larger difference in F0 at flap offset in the minimal pair task than in the wug task. In other words, the mean difference between the F0 at flap offset for /d/-flaps and for /t/-flaps was greater in the minimal pair task (16.57 Hz) than in the wug task (-12.98 Hz), as can be seen in Figure 2.17 ($F(1,44) = 5.15, n.s.$). This distinction does not reach significance after Bonferronization.

Similarly, speaker 10 showed a non-significant, but trending effect of the interaction of task and underlying voicing status on the dip in intensity during flap closure ($F(1,47) = 5.10, n.s.$). The mean difference between the intensity dip in /d/-flaps and /t/-flaps in the minimal pair task was 1.07 dB, and the mean difference in the wug task was -1.85dB, as can be seen in Figure 2.18.

2.3.2.5 Summary of Results

The linear mixed model showed a significant overall effect of underlying voicing status on the duration of pre-flap vowels, but not on any of the other various correlates of voicing. There was, additionally, a main effect of task on duration of the preceding vowel, on F0 at flap offset, and on F1 at flap onset, suggesting that the task at hand had

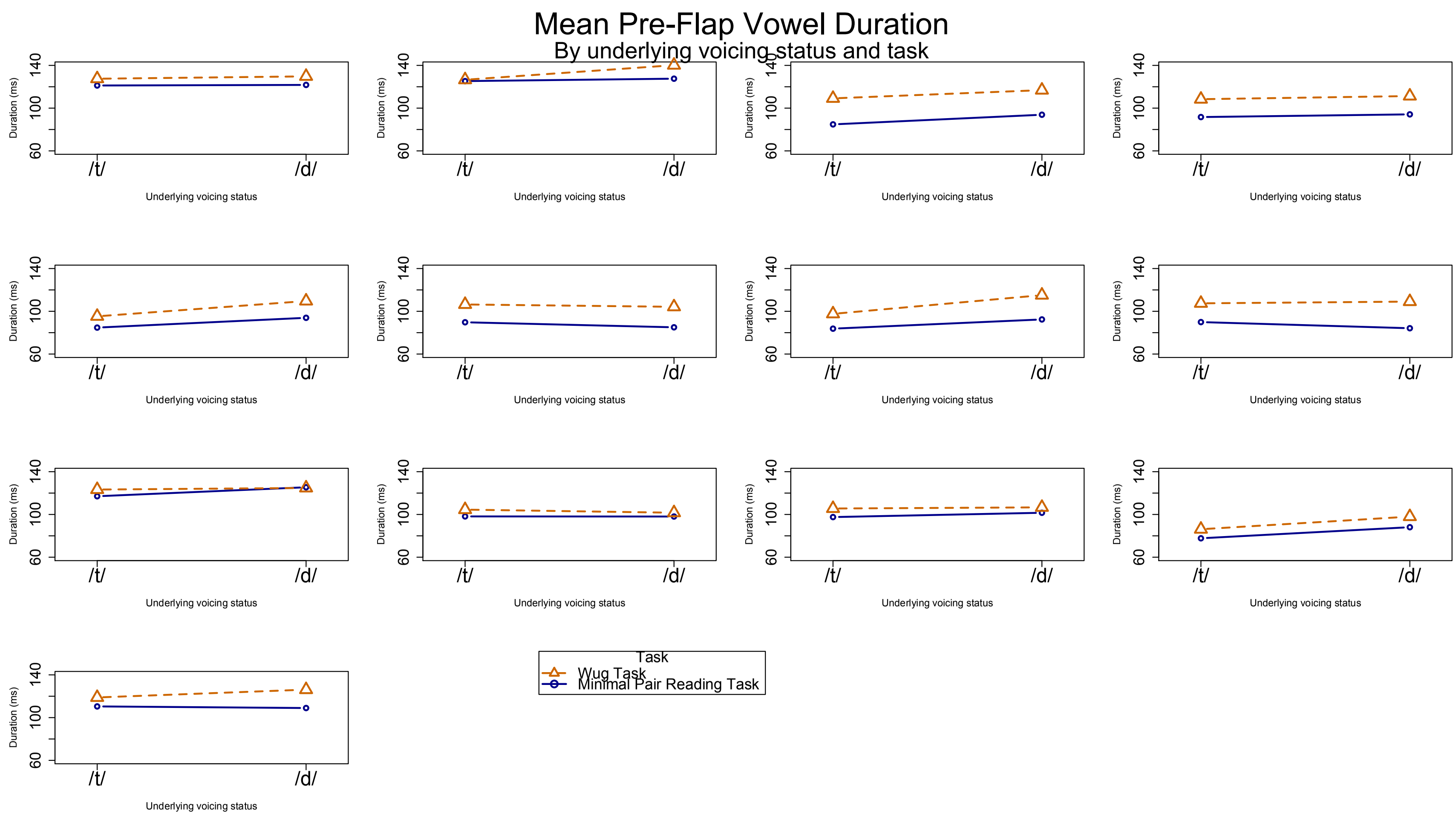


Figure 2.16: Experiment 1: Pre-flap vowel duration for each speaker, by underlying voicing status and task.

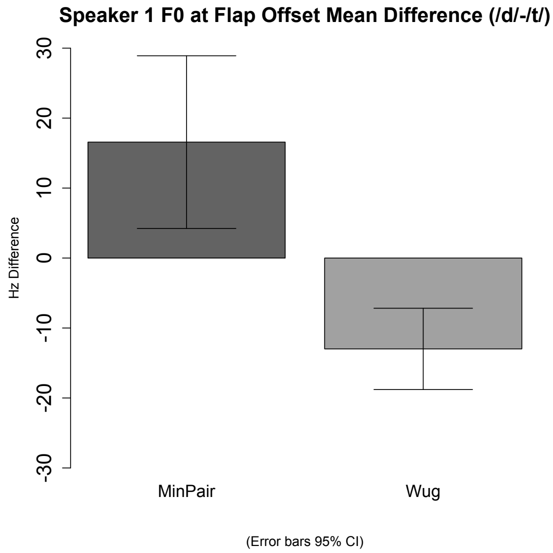


Figure 2.17: Experiment 1: Mean /d/-flap F0 at flap offset – mean /t/-flap F0 at flap offset for speaker 1, by task

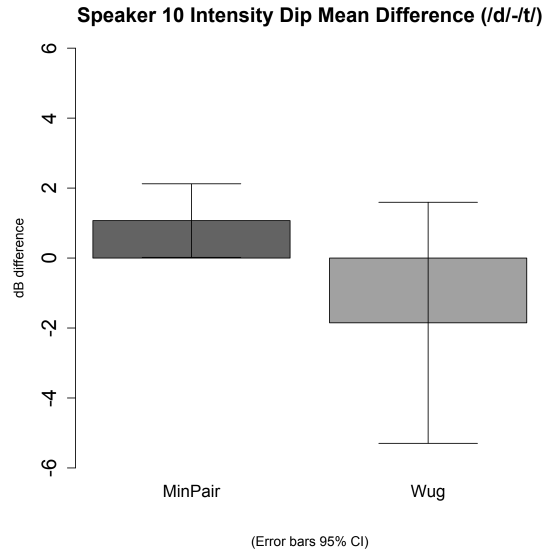


Figure 2.18: Experiment 1: Mean /d/-flap intensity dip – mean /t/-flap intensity dip for speaker 10, by task

a direct influence on these measurements. While the main effect of task was found to be significant for these measures, there was no effect of the interaction of task with voicing status on any measures of voicing, suggesting that task does not strongly affect the degree of neutralization on these measures.

Some individual speakers showed effects that were not significant in the aggregated data. Speaker 6 showed significant effects of underlying voicing status on pre-flap vowel duration (which was significant overall) and also intensity dip (which was not significant overall), suggesting that speaker 6 incompletely neutralizes /d/-flaps and /t/-flaps on both of these correlates of voicing. Similarly, speaker 12 showed effects of underlying voicing status on pre-flap vowel duration and on F0 at flap onset. The mean intensity dip for speaker 12 differed between /t/-flaps and /d/-flaps, but did not rise to the level of significance after Bonferronization. Additionally, the direction of this distinction—with /t/-flaps having a smaller dip than /d/-flaps—is opposite of that of speaker 6. No individual speakers showed significant effects of an interaction between task and underlying voicing status, which would have indicated a task effect on degree of neutralization.

2.3.3 Discussion

2.3.3.1 Incomplete Neutralization of Voicing in Flapping

As summarized in Section 2.1, there has been disagreement in the literature as to whether flapping in American English is incompletely neutralizing. This study shows an effect of underlying voicing status on pre-flap vowel duration (though not on any of the measures of voicing examined). In other words, whether a given flap token originated as a /t/ or a /d/ had an impact on the duration of the preceding vowel (/d/-flaps had longer preceding vowels than /t/-flaps, a mean difference of 8.76 ms). This suggests that, overall, the /t/ flaps and /d/ flaps were *incompletely* neutralized. While not all of the speakers individually showed significantly longer pre-/d/ flap vowel durations than pre-/t/ flap vowel durations, 10 of the 13 speakers showed at least a trend in this direction.

One difference between this study and previous production tasks on the incomplete neutralization of flapping (e.g., Herd et al. 2010, Fisher and Hirsh 1976, Fox and Terbeek 1977, Zue and Laferriere 1979, Huff 1980), is that this study used nonce-words instead of actual English words. The studies cited in Section 2.1 as having found incomplete neutralization (Fisher and Hirsh 1976, Fox and Terbeek 1977, Zue and Laferriere 1979, Huff 1980) all used actual English words.[7] Experiments with actual words have the disadvantage of being prone to effects of lexical frequency (though many, such as Herd et al. 2010 take pains to mitigate these effects to some degree). If the words with /t/ were more frequent than the words with /d/, for example, we might expect that the /t/ words were reduced more than the /d/ words and thus had a shorter vowel duration and flap

[7] It should be noted that the study in Port (1976), which did not find any effects of underlying voicing status on preceding vowel duration or flap closure duration, used ‘possible English words’, rather than actual English words, and was thus not susceptible to effects of lexical frequency. His study concerned the effect of speech rate on medial stops in English. The experiment to which I refer here, described in chapter 2 of Port’s dissertation, examines the effects of tempo on vowel duration and stop closure duration. He notes, however, that ‘/t/ and /d/ are not statistically different at any tempo’ (p. 19). Relevant to the current discussion, he used six tokens in this experiment, two of which (‘ditter’ and ‘didder’) are nonce words with alveolar stops in flap-inducing position. The other tokens had velar and bilabial stops.

closure duration. There is, though, also a potential pitfall to the use of nonce words in this type of study. Since words with higher lexical frequencies are more likely to reduce (Bybee 2000, 2001), and since nonce-words are by definition of frequency zero, we might expect nonce words to resist reduction—perhaps to the point of maintaining a strong /t/~/d/ contrast even in a position that allows flapping. I take the significant findings of incomplete neutralization in *both* types of studies, however, as evidence for the existence of incomplete neutralization in American English flapping.

2.3.3.2 Task

The main effect of task, which shows the impact of task directly on the measured correlates of voicing, was significant for vowel duration, F0 at flap offset, and F1 at flap onset. Pre-flap vowel durations were longer (by 12.21 ms) in the wug task, which was designed to decrease hyperarticulation, than in the minimal pair task, which was designed to increase hyperarticulation. This is unexpected, as we might predict vowel duration to increase as speakers hyperarticulate. Similarly, F0 at flap offset was significantly higher in the wug task (by 1.1967 Hz), and F0 at flap onset trended in this direction as well. F1 at flap onset was significantly higher in the wug task than in the minimal pair task, with a similar trend in F1 at flap offset (mean differences: at onset 11.1558 Hz, at offset 15.3480 Hz).

Taken together, the longer vowel duration and the higher F0 and F1 could indicate that, contrary to the intent of the tasks, speakers focused or hyperarticulated more in the wug task than in the minimal pair task. One possible explanation is that the speakers may have treated the target words in the wug task as an answer to be reported—effectively highlighting the target word—whereas in the minimal pair task the ‘answer’ was given to them on the screen. If we take the wug task, then, to have encouraged more hyperarticulation than the minimal pair reading task, it is perhaps striking that the speakers did not produce incomplete neutralization, even in this task.

The hypothesis that a hyperarticulation-inducing task and a hyperarticulation-reducing task might alter the degree of neutralization was not supported by this experiment— the interaction of task and underlying voicing status did not have a significant effect on any of the measures of voicing. As such, this factor cannot be said to be responsible for the difference between the studies which have found incomplete neutralization and the studies that have not.

2.3.3.3 Variation

While not all speakers showed significant effects of underlying voicing status on pre-flap vowel duration, 10 of the 13 did trend in the direction shown in the aggregate data—longer vowels before /d/-flaps than before /t/-flaps. Individual speakers (such as speakers 6 and 12, as outlined in Section 2.3.2.4) showed significant effects indicating incomplete neutralization on other correlates of voicing as well.

As noted in Section 2.3.2.4, no significant effects of an interaction between task and underlying voicing status on any of the measures of voicing were found for any of the speakers. Two speakers had relatively strong trends in the direction of such effects, but they did not reach significance after Bonferronization. On balance, the evidence does not support a strong task effect on the degree of neutralization, even for these speakers.

The speakers in this study were largely homogenous—all were in New Jersey, all were between 18–25 years old, and 10 out of the 13 speakers were female. No other sociological data was collected; as such, it is impossible to draw any conclusions about the impact of such factors on the results presented here.

2.3.3.4 Effect Size and Just Noticeable Differences

The overall mean pre-flap vowel duration difference between /d/-flaps and /t/-flaps was quite small—just 8.76 ms. These figures are near or above the level of just noticeable difference (JND) for vowel duration — in a study of Dutch vowels, Nooteboom and

Doodeman (1980) found that the JND for vowel duration was about 5ms in synthetic vowels of approximately 90ms. This suggests that the distinctions in vowel duration found in this study is right on the border of perceptibility—a notion to be examined in Experiments 3, 4, and 5 (Sections 2.5, 2.6, and 2.7).

2.3.3.5 Preceding Context

In order to generate a maximal number of stimuli, the onset preceding the target vowel was varied between simple (single voiceless segment) and complex ([s] followed by voiceless stop). The linear mixed model with pre-flap vowel duration from the main experiment was rerun with onset type and the interaction of onset type and underlying voicing status as fixed factors. The type of onset had a significant effect on the duration of pre-flap vowels ($t = 7.493, p < .001$ with complex onset vowels longer by an average of 11.936ms.), but the interaction of underlying voicing status and preceding context did not have a significant effect ($t = 0.148, n.s.$, mean difference between (complex /d/ - complex /t/) and (simple /d/ - simple /t/)=2.089ms).

Additionally, place of articulation of preceding consonants was varied among {p, t, k}. The linear mixed model from the main experiment was modified to include a term for consonant place of articulation (tokens with complex onsets were excluded from this analysis). Neither /p/ nor /t/ showed a significantly different vowel duration than /k/ (the reference level) (p: $t = -0.580, n.s.$; t: $t = -0.113, n.s.$), nor did the interaction terms with place of articulation and underlying voicing status raise to significance (p: $t : -0.312$, t: $t = 0.209$).

Given these results, Experiment 2, the follow-up to this experiment, uses only simple onsets, and retains the use of {p, t, k}.

2.3.3.6 Actual English Roots

The stimuli in this experiment were composed of a prefix (either *un-* or *re-*), followed by a root. While the combinations were not actual English words, some of the roots were. In order to examine the possibility that this affected the results, the linear mixed model with pre-flap vowel duration from the main experiment was rerun with root type and the interaction of root type and underlying voicing status as fixed factors.

Whether a root was an actual English root had a significant main effect overall—stimuli with actual English roots had longer pre-flap vowels than those without by an average of 1.970ms ($t = -2.429, p < 0.025$). There was not a significant interaction of underlying voicing status and root type ($t = 0.478, n.s.$).

Given these results, a modification was made for Experiment 2. In that experiment, the tokens are presented without actual prefixes (e.g., *un-keet* in Experiment 1 might become *buhkeet* in Experiment 2). Since speakers have no cue to break down this nonce word morphologically, any substring within the word that happens to overlap with an actual English word should not be analyzed as a morpheme within the monomorphemic nonce word.

2.3.4 Conclusion

This study found evidence incomplete neutralization in American English flapping—most notably, vowels preceding /d/ flaps were found to be longer than those preceding /t/ flaps. This effect was found in two task types—a minimal pair task, designed to increase the effects of hyperarticulation and orthography, and a wug task, designed to decrease these effects. This, while effects of incomplete neutralization in other languages have been argued to be due to effects of hyperarticulation or orthography, the results presented here do not suggest that this is the case—at least for flapping.

While there was an overall significant effect of incomplete neutralization in pre-flap vowel duration, and 10 of the 13 speakers trended in this direction, there was variation among the speakers. Some speakers showed significant effects of underlying voicing status on other measures, such as F0 at flap onset and intensity dip, while others did not show any such effects.

The follow-up study in Experiment 2 replicates these results, and addresses some of the issues raised in Section 2.3.3.

2.4 Experiment 2: Production of Flaps in American English (II)

Experiment 2 is the follow-up to the study presented in Experiment 1; As such, it follows a format quite similar to Experiment 1. The experiment is comprised of two tasks, with the goals of showing that (a) speakers produce words with /d/ flaps differently from those with /t/ flaps, and (b) this result is robust across task types. As in Experiment 1, the minimal pair task was designed to increase the effects of hyperarticulation and orthographic influence, while the 'wug task' was designed to reduce these effects (see Fourakis and Iverson 1984 for a similar design in an experiment on German final devoicing). By showing that speakers incompletely neutralize the voicing distinction in flapping under both conditions, it is argued that extragrammatical factors are not solely responsible for these and previous results.

2.4.1 Stimuli

The nonce-word stimuli in this experiment were used in both the wug task and the minimal pair task, with minor differences described below. Minimal pairs of disyllabic nonce words were created, ending in either /d/ or /t/, such that adding the '-ing' suffix to these stimuli would place the alveolar stop in a post-tonic intervocalic context where flapping occurs.

The initial (non-target) syllable in each token was composed of a simple onset (one of {p,b,t,d}) with a schwa nucleus and no coda. The second (target) syllable in each token was composed of a simple onset (one of {p,t,k}), a vocalic nucleus (one of {æ, ɛ, i}), and a final /d/ or /t/. A Python script was used to generate all possible combinations of these parameters, yielding 72 tokens, comprised of 36 minimal pairs. The stimuli were written in English orthography, with the second syllable capitalized to indicate stress, and with schwas indicated by 'uh' (for example, puh-PAT~puh-PAD, and tuh-KEET~tuh-KEED). The stimuli are listed in Table 2.9.[8]

/σ-pi{t,d}/		/σ-ti{t,d}/		/σ-ki{t,d}/	
puhPEET	puhPEED	puhTEET	puhTEED	puhKEET	puhKEED
buhPEET	buhPEED	buhTEET	buhTEED	buhKEET	buhKEED
tuhPEET	tuhPEED	tuhTEET	tuhTEED	tuhKEET	tuhKEED
duhPEET	duhPEED	duhTEET	duhTEED	duhKEET	duhKEED
/σ-pɛ{t,d}/		/σ-tɛ{t,d}/		/σ-kɛ{t,d}/	
puhPEHT	puhPEHD	puhTEHT	puhTEHD	puhKEHT	puhKEHD
buhPEHT	buhPEHD	buhTEHT	buhTEHD	buhKEHT	buhKEHD
tuhPEHT	tuhPEHD	tuhTEHT	tuhTEHD	tuhKEHT	tuhKEHD
duhPEHT	duhPEHD	duhTEHT	duhTEHD	duhKEHT	duhKEHD
/σ-pæ{t,d}/		/σ-tæ{t,d}/		/σ-kæ{t,d}/	
puhPAT	puhPAD	puhTAT	puhTAD	puhKAT	puhKAD
buhPAT	buhPAD	buhTAT	buhTAD	buhKAT	buhKAD
tuhPAT	tuhPAD	tuhTAT	tuhTAD	tuhKAT	tuhKAD
duhPAT	duhPAD	duhTAT	duhTAD	duhKAT	duhKAD

Table 2.9: Experiment 2: Stimuli

[8]One stimulus item, puh-TEET, is similar to the actual English adjective 'petite'. 'Petite', however, is ungrammatical in the context in which speakers were asked to produce these items. Neither the flap closure duration nor the pre-flap vowel duration of puh-TEET differed significantly from other stimuli ending in TEET (closure duration: $t(29.82) = -1.78, n.s.$; vowel duration: $t(15.47) = -0.79, n.s.$).

2.4.2 Participants and Equipment

12 speakers participated in both tasks in this experiment. All speakers reported being raised primarily in New Jersey (though one also reported significant time spent in Brooklyn). Each speaker made two visits to the Rutgers Phonology and Field Research Laboratory, exactly seven days apart, with task order balanced across speakers. Recordings were made in a sound-attenuated booth, using an AT44040 Cardioid Capacitor microphone with a pop filter, amplified through an ART TubeMP microphone pre-amplifier and a JVC RX-554V receiver. The speech was digitized as WAV files at a sampling rate of 44.1kHz using Audacity (Audacity Team 2008). Acoustic analysis was performed using Praat (Boersma and Weenink 2009).

2.4.3 Methods

2.4.3.1 The Minimal Pair Task

The minimal pairs constructed as described in Section 2.4.1 were put into the progressive '-ing' form (e.g., the tokens 'puh-PAT' and 'puh-PAD' became 'puh-PAD-ing' and 'puh-PAT-ing', respectively). The stimuli were then presented visually to speakers on a computer screen. On each trial, speakers saw two sets of sentences, consisting of two sentences each, on the screen. In the first set of sentences in each trial, speakers saw one sentence with a nonce word in its 'plain' form (i.e., without the '-ing' suffix), and a second sentence with the same nonce word in its '-ing' form. The second set of sentences in this trial was identical, except that the nonce word from the first set was replaced with the other member from its minimal pair. Whether the /d/- or /t/-member of the minimal pair was displayed in the first or second set was randomized. On each trial, then, speakers saw something like (5):

(5) John learned how to puh-PAT this week.
He was puh-PAT-ing this whole week.

John learned how to puh-PAD this week.
He was puh-PAD-ing this whole week.

Speakers were asked to read each sentence aloud naturally. This procedure was repeated for all 36 minimal pairs (=72 target items), randomized, with 36 filler pairs (=72 total filler items).

As with the minimal pair task in Experiment 1, this task was designed to increase speakers' attention to both pronunciation and orthography (see the 'reading task' in Fourakis and Iverson 1984). Since speakers saw both members of a minimal pair on the screen at the same time, the task highlighted the distinction between the voiced and voiceless stimuli.

2.4.3.2 The Wug Task

In the wug task, speakers were shown the 'plain' form of the 72 stimuli, and were asked to fill the '-ing' form of these words into a blank in a frame sentence. Unlike the minimal pair task, only one stimulus was shown on the screen at a time—the stimuli were not presented as minimal pairs. The order of stimuli was randomized, with 72 filler items.

On a given trial, speakers saw a screen like (6):

(6) John learned how to puh-PAT this week.
He was ________ this whole week.

After an initial training phase, speakers reliably filled in the '-ing' form of the nonce word, producing, for example, (7):

(7) John learned how to puh-PAT this week.
He was puh-PAT-ing this whole week.

As with the wug task in Experiment 1, this task was designed with Fourakis and Iverson's (1984) 'elicitation task' in mind. While the morphology of the English progressive is simpler than the German strong verb system employed in their task, the task presented here also seeks to reduce the influence of orthography, and at the same time ‘diminish the obviousness of focus on pronunciation’ (Fourakis and Iverson 1984, p. 142). By instructing speakers to focus on filling in the blank, rather than on their pronunciation, the task aimed to elicit more natural productions of the target words. Similarly, while speakers did see the plain form of the nonce word on the screen, they did not see the target ‘-ing’ form, mitigating the effects of orthography, especially as compared to the minimal pair reading task. Finally, unlike in the minimal pair task, speakers did not see both members of a minimal pair on a given trial, reducing the immediacy of the distinction between the voiced and voiceless stimuli.

2.4.3.3 Acoustic Measurements and Statistical Analysis

As discussed in Section 2.3.1.4, the [voice] feature has a number of known acoustic correlates, including preceding vowel duration (Chen 1970), closure duration (Kluender et al. 1988), and effects on the F0 and F1 of surrounding vowels (Hombert et al. 1979, Kingston and Diehl 1994). Pre-flap vowel duration, flap closure duration, percent closure during voicing, and F0 and F1 slope at flap onset and offset were measured. The duration of vowels preceding the flapped segments were measured from the onset of voicing to the onset of the flap closure. The onset of flap closure was marked as the location on the spectrogram with a marked reduction in formant structure, accompanied by a drop in intensity and periodic energy as seen in the waveform. The offset of flap closure was marked as the location on the spectrogram where formant structure resumed, and intensity and periodic energy increased as seen in the waveform. See Figure 2.19 for representative examples.

Tokens more than 2 standard deviations from the mean on any of the recorded measurements were considered outliers and discarded. A linear mixed model (Baayen 2008) was run using the `lme4` package (Bates and Maechler 2009) in R (R Development Core Team 2009). Pre-flap vowel duration was regressed against a model in which

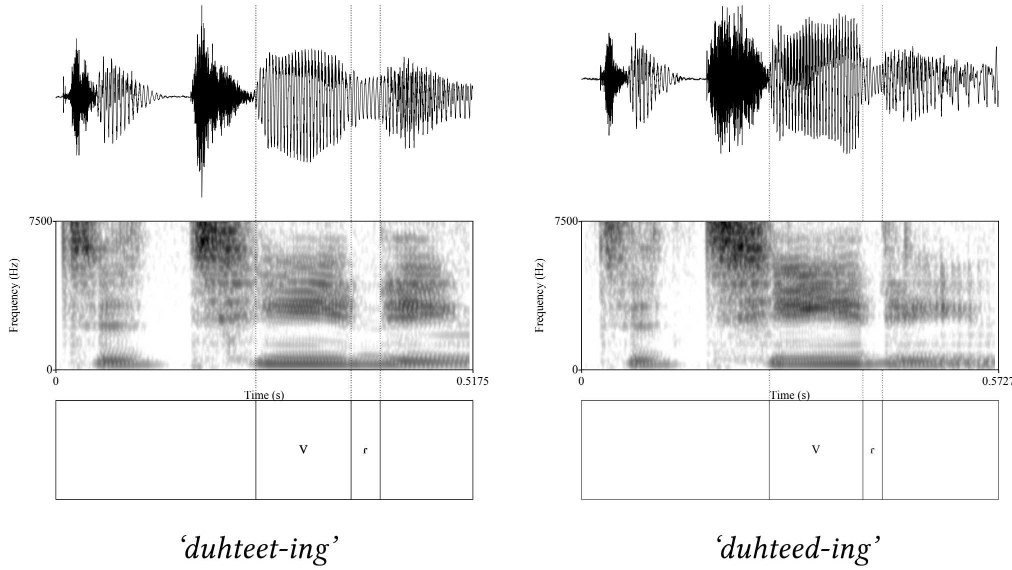


Figure 2.19: Experiment 2: Representative spectrograms, speaker 6, minimal pair task

underlying voicing status and task were fixed factors. Vowel height (contrast coded) was also included as a fixed factor to soak up variability, since it is known to affect vowel duration (Peterson and Lehiste 1960, Lehiste 1970). An interaction term between underlying voicing status and task was also included in order to examine the effects of underlying voicing status on neutralization that might be present in one task but not the other. Speaker and item were included as random factors.[9]

While preceding vowel duration tends to be the focus of studies of incomplete neutralization, it is far from the only known correlate of voicing (Kluender et al. 1988, Kingston and Diehl 1994, Hombert et al. 1979). As such, identical models were run, substituting the following measurements as the fixed factor: closure duration, percent voicing during closure (as determined by Praat's 'Voice Report'), and the slope of F0 and F1 over 10ms preceding and 10ms following flap closure.

[9]The procedure for calculating degrees of freedom in this type of model is unknown, so the significance of the coefficients was checked by the `pvals.fnc` function of R's `languageR` package (Baayen 2009), which uses the Markov Chain Monte Carlo method.

2.4.4 Results

I focus in this section on the results pertaining to preceding vowel duration, as the other measures investigated were found not to be significantly impacted by underlying voicing status. The results of these other measures are summarized in Appendix A.

2.4.4.1 The Effect of Task

Task did not have a significant main effect on target vowel duration—pre-flap vowels were similar in duration in the wug task and the minimal pair task (mean difference: 1.74 ms, $t = 0.54, n.s.$). Similarly, the effect of the interaction between task and underlying voicing on pre-flap vowel duration was not significant ($t = 0.82, n.s.$), suggesting that the difference in vowel duration before /d/ and /t/ flaps was consistent across tasks. This result is reflected in Figure 2.20: the difference between the bars for pre-/t/ and pre-/d/ vowel durations in the minimal pair task is not significantly different from the difference between pre-/t/ and pre-/d/ vowel durations in the wug task. Since task was not a significant factor in vowel duration, the remaining results are reported with stimuli pooled together from both tasks.

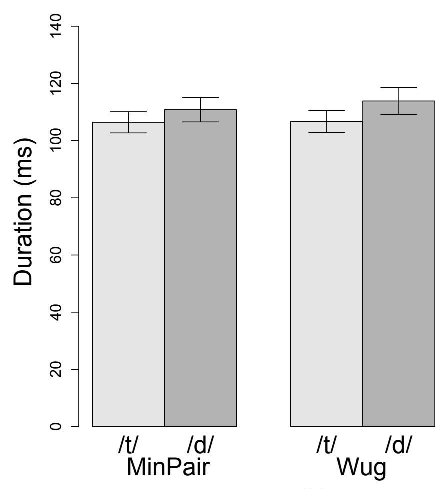


Figure 2.20: Experiment 2: Mean pre-flap vowel duration by underlying voicing status and task

2.4.4.2 The Effect of Underlying Voicing Status

Among all stimuli (pooled from both tasks), underlying voicing status had a significant effect on vowel duration. Vowels preceding /d/ flaps were longer than those preceding /t/ flaps (mean difference: 5.69 ms, $t = 3.77, p < 0.001$). The mean duration of vowels preceding /d/ flaps and /t/ flaps is shown in Figure 2.21 and Table 2.10, with these results broken down by speaker in Table 2.11.

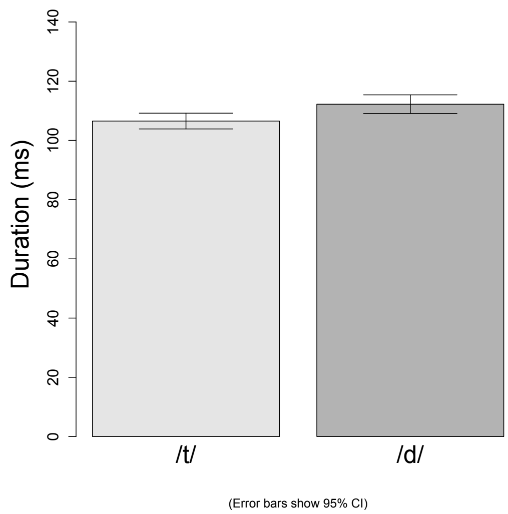


Figure 2.21: Experiment 2: Mean pre-flap vowel duration by underlying voicing status (averaged across both tasks)

	Mean	SD
Pre-/d/ flap	112.23	30.06
Pre-/t/ flap	106.54	26.75

Table 2.10: Experiment 2: Mean and standard deviation of vowel durations (in ms) for each category

2.4.5 Discussion

The results of Experiment 2 suggest that flapping in American English is incompletely neutralizing, at least for some speakers. The underlying voicing status of a flap is reflected on the surface in the duration of the preceding vowel—vowels before /d/

Speaker	1	2	3	4	5	6
Before /d/-flap	122.65	86.79	86.02	90.76	71.31	131.11
Before /t/-flap	114.61	87.61	88.43	102.11	73.98	115.83
Difference	8.04	-0.83	-2.41	-11.36	-2.67	15.28
Speaker	7	8	9	10	11	12
Before /d/-flap	119.50	120.24	103.2	94.47	105.58	138.46
Before /t/-flap	114.94	118.88	88.6	93.44	98.71	132.92
Difference	4.55	1.36	14.60	1.04	6.87	5.54

Table 2.11: Experiment 2: Mean pre-flap vowel duration for all speakers in the acoustic experiment, across both tasks. All values in ms.

flaps are longer than vowels before /t/ flaps. This pattern mirrors the distinction in preceding vowel duration seen in voiced segments generally (Chen 1970). The fact that the vowel duration distinction between /d/ flaps and /t/ flaps was not significantly different between the two tasks suggests that the results of this experiment, and those of previous studies on incomplete neutralization (including Experiment 1) are not due primarily to extragrammatical factors such as hyperarticulation and orthography.

2.5 Experiment 3: Identification Perception of Flaps in American English

The speaker-hearer relationship plays a crucial role in incomplete neutralization. Indeed, Port and Crawford (1989) argue from experimental evidence that speakers vary the completeness of incompletely neutralized contrasts depending on the communicative situation. (This result is also found in other phenomena in which speakers manipulate their productions for the benefit of hearers. See, e.g., Lindblom 1990, Scarborough 2003, 2010, Flemming 2010, Syrett and Kawahara under revision.) Along these lines, it is worth exploring whether hearers can actually perceive the differences that speakers produce in cases of incomplete neutralization. Do speakers' attempts to maintain a contrast actually succeed? The vowel duration distinctions found in Experiments 1 and 2 are quite small;

as such, the perceptibility of these distinctions bears investigating.

For flapping in American English, Herd et al. (2010) showed that on a basic identification task, listeners perform poorly. Experiment 3 was designed to replicate this result with nonce words, taking bias into consideration through the use of d' as a measure of sensitivity. It follows the basic format of similar studies into the perception of /t/ and /d/ flaps (Sharf 1960, Malécot and Lloyd 1968, Fisher and Hirsh 1976, Herd et al. 2010), with the important distinction that the stimuli in this task—which were taken from speaker's productions in Experiment 2—are nonce words, rather than actual words of English.

2.5.1 Participants and equipment

21 undergraduates participated in this experiment. All participants were native speakers of English. A plurality of participants were born in New Jersey (42.9%), and 76.2% of the participants report having been raised mostly in New Jersey. The experiment took place at the Rutgers Phonetics Laboratory, with stimuli displayed and responses recorded by SuperLab 4.5 (Cedrus Corporation 2010) through Sennheiser HD 280 Professional headphones.

2.5.2 Tokens

Tokens were selected from tokens recorded in Experiment 2 (see Section 2.4.1). From among the tokens of all 12 speakers in the that experiment, tokens were chosen from the three speakers who had the biggest difference between pre-/d/ and pre-/t/ vowel duration, and who accurately produced a sufficient number of tokens. Tokens were chosen from each speaker to maximize the pre-flap vowel duration difference between members of a pair, while at the same time balancing onset and vowel of the target syllable, as well as for the voicing of the target segment (/d/ or /t/). The experiment was blocked such that a given block had only tokens from a single speaker. The full set of tokens, used in all of the perception experiments reported in this chapter, is listed in Table 2.12.

Speaker 6			Speaker 11			Speaker 12		
/d/	/t/	Task	/d/	/t/	Task	/d/	/t/	Task
buhKADing	buhKATing	wug	buhKEEDing	buhKEETing	min	puhTEHDing	puhTEHTing	min
buhKEHDing	buhKEHTing	min	buhKEEDing	buhKEETing	wug	buhKADing	buhKATing	min
buhPADing	buhPATing	min	buhKEHDing	buhKEHTing	min	buhKEEDing	buhKEETing	wug
buhPEEDing	buhPEETing	wug	buhTADing	buhTATing	min	buhPADing	buhPATing	min
buhTADing	buhTATing	min	buhTEHDing	buhTEHTing	min	buhTADing	buhTATing	min
buhTEEDing	buhTEETing	min	buhTEHDing	buhTEHTing	wug	buhTEEDing	buhTEETing	wug
buhTEHDing	buhTEHTing	wug	duhPADing	duhPATing	wug	duhKEEDing	duhKEETing	wug
duhKEEDing	duhKEETing	min	duhPEEDing	duhPEETing	wug	duhKEHDing	duhKEHTing	min
duhKEHDing	duhKEHTing	min	puhKADing	puhKATing	min	duhPEHDing	duhPEHTing	min
duhPADing	duhPATing	min	puhKEHDing	puhKEHTing	min	duhTADing	duhTATing	min
puhPEEDing	puhPEETing	wug	puhTADing	puhTATing	min	puhKEHDing	puhKEHTing	wug
puhPEHDing	puhPEHTing	min	puhTEEDing	puhTEETing	min	puhPEEDing	puhPEETing	min
puhTADing	puhTATing	wug	puhTEHDing	puhTEHTing	wug	puhTEEDing	puhTEETing	min
puhTEEDing	puhTEETing	min	tuhKADing	tuhKATing	min	puhTEHDing	puhTEHTing	wug
tuhKADing	tuhKATing	min	tuhPADing	tuhPATing	wug	tuhKADing	tuhKATing	min
tuhKEEDing	tuhKEETing	min	tuhPEEDing	tuhPEETing	min	tuhPADing	tuhPATing	wug
tuhKEHDing	tuhKEHTing	wug	tuhPEHDing	tuhPEHTing	min	tuhPEEDing	tuhPEETing	min
tuhPEHDing	tuhPEHTing	min	tuhTEEDing	tuhTEETing	min	tuhPEHDing	tuhPEHTing	min

Table 2.12: Tokens used in Experiments 3, 4, and 5. ‘Speaker’ indicates which speaker from Experiment 2 produced the token. ‘Task’ refers to the task from Experiment 2 during which the token was produced. Experiment 3 used every token here as a stimulus. Experiments 4 and 5 used stimuli created from each minimal pair in the table.

2.5.3 Procedure

Prior to the actual experimental task, participants read instructions for the task, practiced with both English and nonce words that had been recorded by the author, and were given an opportunity to ask the experimenter any questions about the procedure.

On each experimental trial, listeners heard a single token (as described above), and were directed to press one of two buttons indicating whether the sound immediately preceding the '-ing' was a /d/ or a /t/. For example, if listeners heard 'buhKEED-ing', they would press the button indicating that the sound immediately preceding the '-ing' was a /d/.

As an additional measure to reduce the difficulty of this task, visual feedback was given on each trial. After a participant's response (or failure to respond after 1500 ms) the correct response was shown for 500 ms (colored green if they had been correct, and red if they had been incorrect).

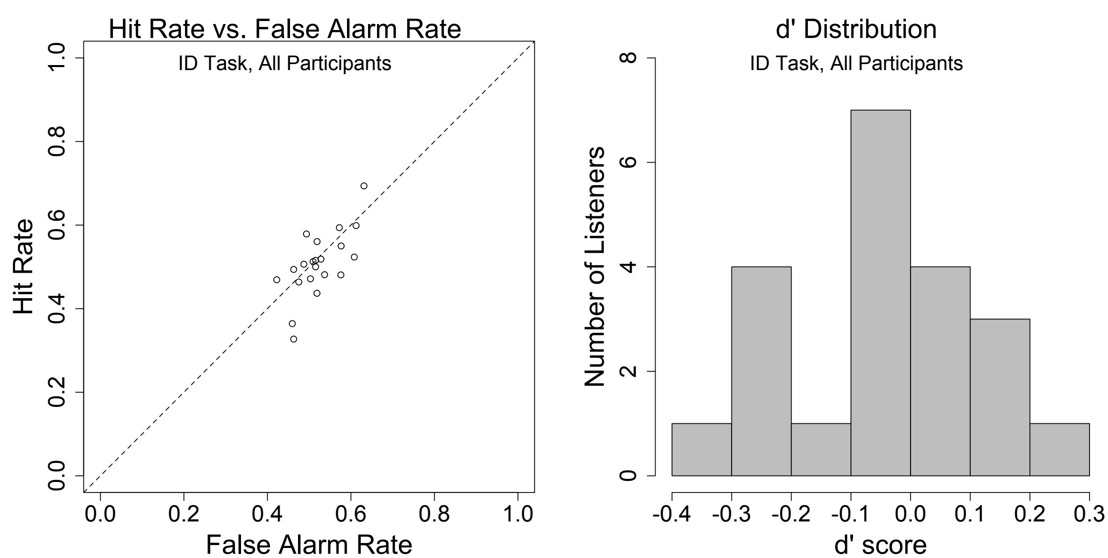
The task consisted of three blocks (one for each of the three speakers from whom tokens were taken). Each block consisted of 36 trials (half /d/ and half /t/), randomized, with three repetitions (=108 total trials). Listeners were allowed to take a short break between each block. Block order was balanced (Latin Square) across all listeners.

2.5.4 Analysis

The results of this task are reported using d' as a measure of sensitivity (Macmillan and Creelman 2005). Unlike the percent-correct measure, d' teases apart listeners' actual sensitivity from bias, because it takes both hit rate (how often a listener is correct) and false alarm rate (how often a listener mistakenly identifies a non-target as a target) into account. In this task, d' is calculated as $d' = z(H) - z(F)$, where H is the hit rate and F is the false alarm rate (Macmillan and Creelman 2005). A d' score of zero indicates an inability to discriminate; as d' scores rise above zero, they indicate improving ability to discriminate. Given the strong bias in favor of /d/ found in Herd et al. (2010), d' is a more appropriate measure of sensitivity than percent-correct for this study.

2.5.5 Results

The mean d' score across all listeners in Experiment 3 was $d' = -0.04$, which is not significantly different from zero (Wilcoxon test: $V = 76, n.s.$). In other words, listeners responded that they had heard a /d/ just as often when they had heard a /d/ as when they had heard a /t/. Figure 2.22(a) shows a plot of the hit rate vs. false alarm rate for participants in the identification task. Listeners cluster around the hit rate = false alarm rate line—they had as many hits (saying '/d/' when they heard a /d/) as they had false alarms (saying '/d/' when they had actually heard a /t/). In other words, saying that they had heard a /d/ had little relation to whether they had actually heard a /d/. Given that d' was calculated as the difference between the z-transformed hit and false alarm rates, similar hit and false alarm rates will result in d' scores near zero. This is reflected in the frequency distribution of d' scores shown in Figure 2.22(b), which center around zero.



(a) Hits vs. False Alarms

(b) Histogram of d' scores

Figure 2.22: Experimeng 3: Results of the identification experiment

2.5.6 Discussion

Listeners in this study were unable to correctly categorize /d/-flaps and /t/-flaps.[10] Even though this study uses nonce word stimuli, the results comport with those seen in previous identification tasks of incompletely neutralized flaps using actual English words. One question that remains, however, is whether listeners will show better performance on easier tasks.

Given the relative difficulty of identification tasks, two additional experiments are reported in the sections that follow—an ABX task (Experiment 4) and a 2-alternative forced choice (2AFC) task (Experiment 5). These tasks serve to address two questions: (a) whether listeners' poor performance in identification studies of flapping is due in part to the difficulty of such tasks, and (b) whether listeners' inability to discriminate the relevant sounds, rather than lack of categorical knowledge, is the key to understanding such results.

2.6 Experiment 4: ABX Perception of Flaps in American English

In an ABX task, listeners are given three stimuli per trial, and so are able to make comparisons to aid in their decision-making. The results of this relatively easier task can show whether listeners' poor performance in identification tasks reflects an inability to perceive a given distinction, or whether task type plays a role in listeners' performance.

[10] Since nonce words were used in this identification task, unlike the one presented in Herd et al. (2010), participants saw only 'd' and 't' on the screen, rather than whole words. In order to assuage the possible concern that the 'd or t' type task is easier than a task in which listeners see whole-word minimal pairs on the screen, an identical version of the task reported in this section was conducted with 12 new listeners, in which the nonce-word minimal pairs were presented on the screen instead of 'd' and 't'. The results are nearly identical to those of the task presented in the main body of this section: participants were unable to correctly categorize /d/-flaps and /t/-flaps ($d' = -0.02$), which is not significantly different from 0 (Wilcoxon $V = 26$, *n.s.*).

The ABX task reported in this section was designed to address this possibility. Listeners in the identification task (Experiment 3), as well as in the similar task reported in Herd et al. (2010), show a poor ability to distinguish between /t/ and /d/ flaps. The ABX task reported here was designed to determine whether these results are truly due to an inability to perceive the distinction at hand, or whether listeners perform better in a different type of task.

Additionally, it is possible that listeners perform poorly in identification tasks not because they lack the relevant phonological categories, but rather because they cannot discriminate between the relevant classes of sounds. While the ABX task reported here was designed with the intention of tapping into listeners' categorical knowledge, listeners seemed to actually apply a more auditory mode of perception (see, Sections 2.6.3 and 2.6.6, and Pisoni 1973, Gerrits and Schouten 2004). As such, this task speaks to listeners' abilities to discriminate /d/ and /t/ flaps.

2.6.1 Participants and equipment

21 undergraduates, none of whom had participated in Experiments 1 or 2, participated in this experiment. All participants were native speakers of English. The vast majority of participants were born in New Jersey (90.5%), and 95.2% of the participants report having been raised mostly in New Jersey. The experiment took place at the Rutgers Phonetics Laboratory, with stimuli displayed and responses recorded by SuperLab 4.5 (Cedrus Corporation 2010) through Sennheiser HD 280 Professional headphones.

2.6.2 Stimuli

On each trial, listeners heard a total of three nonce words. On each trial, two of the tokens were physically identical, and the remaining token was the other member of a minimal pair. These minimal pairs were drawn from the set of tokens used in Experiment 3 (described above in Section 2.5.2).

2.6.3 Procedure

Prior to the actual experimental task, participants read instructions for the task, practiced with both English and nonce words, and were given an opportunity to ask the experimenter any questions about the procedure.

On each trial, listeners heard three stimuli (A, then B, then X), and were asked to press a button indicating whether the third word (X) was the same as the first (A) or the second (B). For example, if listeners heard 'buhKEED-ing - buhKEET-ing - buhKEED-ing', they would press the button indicating that the first sound, A, was the same as X.

An ISI of 250 ms was used between the first two stimuli (A and B), and an ISI of 500 ms was used between the second stimulus and the final, target stimulus (B and X). The 500 ms B-X ISI was used in order to allow listeners time to make an attempt at category labeling of the first two stimuli before hearing the third. Pisoni (1973) found that within-category vowel discrimination (where category labeling is not expected) peaked with a 250 ms ISI and degraded as the interval between the vowels was lengthened, but that ISI has a less profound effect on between-category discrimination (where category labeling *is* expected). Pisoni (1975) further suggests that the differences between categorical and auditory modes of perception can be attributed to the ability (or inability) to retrieve auditory information from short-term memory. Presumably, allowing the auditory traces of the A and B stimuli to decay before hearing the X stimulus would invite listeners to employ a more categorical, rather than auditory, mode of perception (see Gerrits and Schouten 2004 for a discussion of these modes of perception, as well as of the effect of ISI).

As with Experiment 3, visual feedback was given on each trial. After a participant's response (or failure to respond after 1500 ms) the correct response was shown for 500 ms (colored green if they had been correct, and red if they had been incorrect).

This experiment consisted of three blocks (one for each of the three speakers from whom tokens were selected). Each block was randomized, and consisted of 18 trials of each of the following 4 configurations (72 trials total per block), such that /d/-words and /t/-words each appeared equally as stimulus A and stimulus B, and such that the target stimulus (X) matched A and matched B an equal number of times: (i) A=/d/, B=/t/, X=/t/, (ii) A=/d/, B=/t/, X=/d/, (iii) A=/t/, B=/d/, X=/d/, (iv) A=/t/, B=/d/, X=/t/. Block order was balanced (Latin Square) across all listeners.

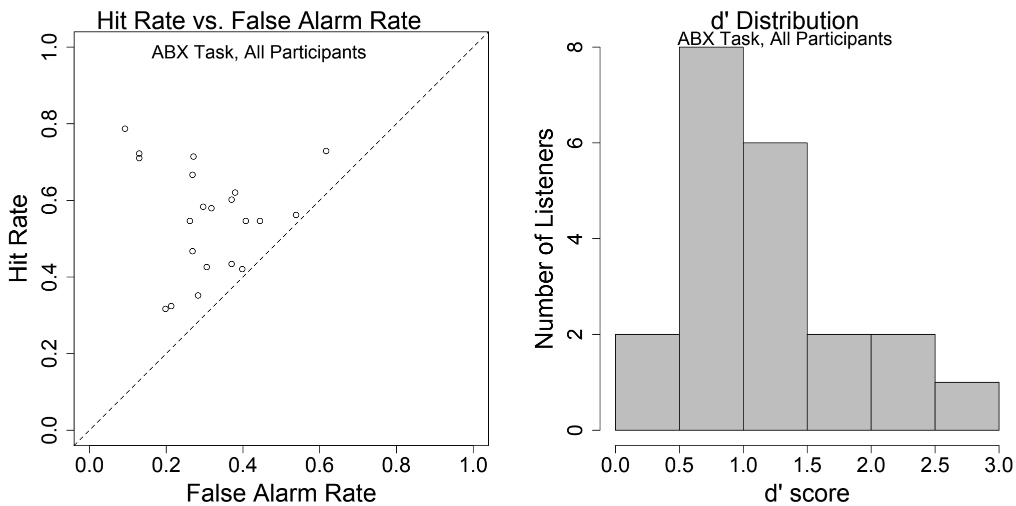
2.6.4 Analysis

As with Experiment 3, listeners' sensitivity was calculated using d' as a measure of sensitivity. Given that this task employed a roving ABX design, I assume a differencing strategy (Macmillan and Creelman 2005, pp. 221–225, 233). d' values were calculated using the `dprime.ABX` function of the `psyphy` package (Knoblauch 2011) with the `method="diff"` option in R (R Development Core Team 2009).

2.6.5 Results

The mean d' score across all listeners in Experiment 4 was $d' = 1.24$, which is significantly different from zero (Wilcoxon test: $V = 231, p < 0.001$). In other words, listeners responded that A was the same as X more often when that was actually the case than when it was not the case. Figure 2.23(a) shows a plot of the hit rate vs. false alarm rate for participants in the ABX task. All listeners had higher hit rates than false alarm rates—they had more hits (saying 'A is the same as X' when that was the case) than false alarms (saying 'A is the same as X' when that was not the case). This fact is reflected in listeners' d' scores, which are all above zero, as can be seen in the frequency distribution in Figure 2.23(b).

After completing this task, participants were informally asked what strategy they had used. Many listeners indicated that they had used cues unrelated to the /t/~/d/ distinction, such as the pitch contour of individual tokens. Similarly, some listeners related having



(a) Hits vs. False Alarms (b) Histogram of d' scores

Figure 2.23: Experiment 4: Results of the ABX experiment

listened for physical similarities between the two identical tokens, rather than focusing on the /t/∼/d/ distinction as they had been instructed.

2.6.6 Discussion

The 500 ms B-X ISI employed in this experiment, which was intended to induce listeners to employ a categorical mode of perception, did not appear to have the desired effect. If listeners had employed a categorical mode of perception, the focus on pitch contour and other physical differences between tokens would not have been as prominent. It is possible that, in spite of the longer ISI, listeners employed a more auditory mode of perception because they were unable to categorize the stimuli that they heard.

In order to verify that the relatively good performance on this task was due to listeners' use of cues unrelated to the underlying voicing distinction, the five best and five worst performing minimal pairs were analyzed. If the five best performing minimal pairs all showed differences unrelated to the underlying voicing distinction, perhaps listeners used these cues rather than, for example, pre-flap vowel duration, to make their decisions on these pairs. Similarly, if the five worst-performing minimal pairs had relatively strong canonical indicators of the underlying voicing distinction, but listeners still performed poorly, it stands to reason that these cues did not help listeners correctly distinguish /d/ and /t/ flaps.

Among the best-performing pairs, a difference unrelated to the voicing distinction between the two tokens was always evident. For example, in the pair 'tuhkad-ing'~'tuhkat-ing' (from speaker 12 in the minimal pair task), the '-ing' syllable of 'tuhkad-ing' had creaky voice while the matching syllable in 'tuhkat-ing' did not. Similarly, in the pair 'duhkeed-ing'~'duhkeet-ing' (from speaker 12 in the wug task), the mean pitch of the target syllable differed by 16.13 Hz (243.93 Hz for 'duhkeed-ing', 260.06 Hz for 'duhkeet-ing').[11] The vowel duration difference in this pair was only 1.63 ms (130.45 ms for 'duhkeed-ing', 128.82 ms for 'duhkeet-ing'), which is not likely long enough for listeners to have used as a cue to underlying voicing status.

At the same time, the worst-performing pairs were not necessarily those with the smallest vowel duration differences. In fact, of the five worst-performing pairs analyzed, every pair had a larger vowel duration difference than the 'duhkeeding'~'dukeeting' pair (which was among the five best-performing pairs). For example, the pair 'puhteed-ing'~'puhteet-ing' (from speaker 12 in the minimal pair task) had a vowel duration difference of 10.21 ms (114.54 ms for 'duhteed-ing', 104.33 ms for 'duhteet-ing').

These results, taken together, suggest that listeners used cues unrelated to underlying voicing (such as pitch differences and creaky voice) to distinguish among members of at least some minimal pairs. In other words, it is possible that listeners' relatively strong performance on this task does not reflect their general ability to distinguish /d/ flaps from /t/ flaps, but rather to distinguish the particular recorded tokens of /d/ and /t/ flaps that they happen to have heard.

[11] Among the five worst-performing pairs, some pitch differences between pair members above 10 Hz were observed. This does not, however, negate the possibility that listeners used the pitch differences in the best-performing pairs.

2.7 Experiment 5: 2AFC Perception of Flaps in American English

In order to test the hypothesis that listeners' good performance in the ABX task in Experiment 4 was due to the use of distinctions unrelated to voicing between stimuli on a given trial, this Experiment makes use of a 2-Alternative Forced Choice task. This 2AFC task differs from the ABX task in that on a given trial, listeners hear only two stimuli, and they are always different from one another. This experimental design is still easier than an identification task—listeners are able to compare two stimuli (Macmillan and Creelman 2005, pp. 167-170). Since there are not two physically identical stimuli on a given trial, however, listeners cannot employ the 'physical similarities' strategy that was possible in the ABX task.

2.7.1 Participants and Equipment

24 undergraduates participated in this experiment, none of whom participated in the previous experiments. All participants were native speakers of English. 75% of the participants were born in New Jersey, and 95.8% reported being raised primarily in New Jersey. This experiment took place at the Rutgers Phonetics Laboratory, with stimuli displayed and responses recorded by SuperLab 4.5 (Cedrus Corporation 2010) through Sennheiser HD 280 Professional headphones.

2.7.2 Stimuli

All tokens were taken from the same set of stimuli used in Experiments 3 and 4.

2.7.3 Procedure

On each trial, listeners heard two tokens—members of a minimal pair—separated by 250 ms of silence. Listeners were instructed to pay attention to the sound immediately preceding the '-ing' in each word. Half of the listeners were asked whether the /d/-member of the pair came first or second. The other half of the listeners were asked whether the /t/-member of the pair came first or second. Participants in the 'find /d/' variation were instructed to push one of two buttons corresponding to the two words they saw on the screen, indicating whether the first word or the second word had the /d/. Participants in the 'find /t/' variation indicated whether the first word or the second word had the /t/. For example, if a listener who was told to 'find /d/' heard 'buhKEED-ing - buhKEET-ing', they would press the button corresponding to the first word, since it is the one that contains a /d/ immediately preceding '-ing'.

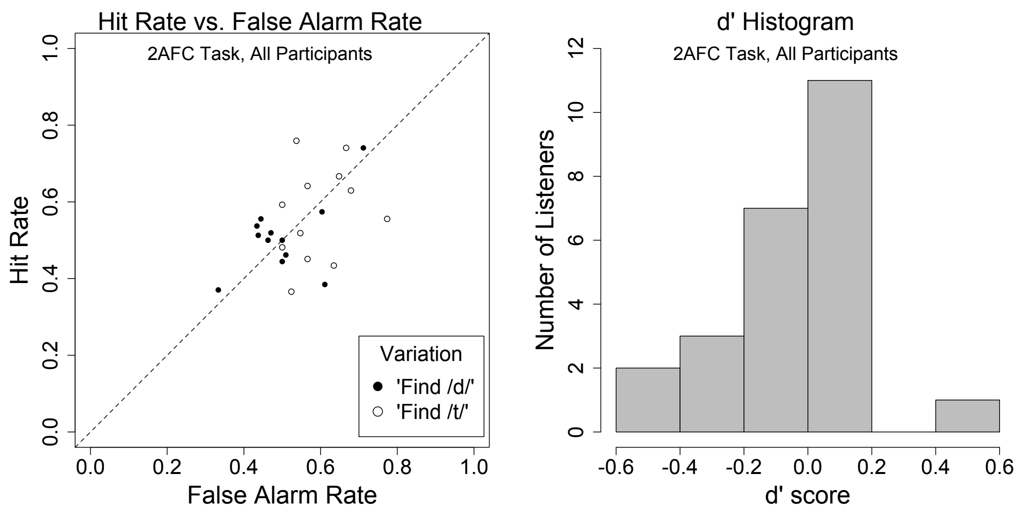
Each of the three blocks consisted of 36 randomized trials (half /d/ and half /t/), all from the same speaker. Block order was balanced (Latin Square) across all listeners. As with the previous perception studies, feedback was provided on each trial, and the next trial was presented if listeners did not respond within 1500 ms.

2.7.4 Analysis

As with the previous perceptual tasks, d' was used as a measure of sensitivity. Because 2-Alternative Forced Choice tasks such as this one are easier than identification tasks, d' is calculated here as $d' = \frac{z(H) - z(F)}{\sqrt{2}}$ (Macmillan and Creelman 2005, p. 167-170).

2.7.5 Results

The mean d' score across all listeners for the 2AFC task was $d' = -0.016$, which is not significantly different from zero (Wilcoxon test: $V = 138, n.s.$). In other words, listeners indicated that the target sound (/d/ or /t/, depending on the listener) had been in the first word of a given trial just as often when the target sound had indeed been in the first word as when the target sound had actually been in the second word. Figure 2.24(a) shows a plot of the hit rate vs. false alarm rate for participants in the 2AFC task. Figure 2.24(b) shows the frequency distribution of d' scores.



(a) Hits vs. False Alarms

(b) Histogram of d' scores

Figure 2.24: Experiment 5: Results of the 2AFC experiment

2.7.6 Discussion

Listeners in Experiment 5 were unable to accurately distinguish between /d/ flaps and /t/ flaps. This result, taken together with the post-hoc analysis of the best- and worst-performing tokens in Experiment 4 suggests that listeners in the ABX task (Experiment 4) used a 'physical similarities' strategy. In the 2AFC task (Experiment 5), where this strategy was not available to them, listeners' performance was significantly worse (i.e., at chance level).

These results, in combination with the identification task, suggest that listeners are neither able to accurately categorize nor distinguish /d/ flaps from /t/ flaps.

2.8 Chapter Discussion and Conclusions

The results of the production task in Experiments 1 and 2 suggest that flapping in American English is, indeed, incompletely neutralizing. Vowels preceding /d/ flaps were, on average, longer than vowels preceding /t/ flaps (by 8.76 ms in Experiment 1, and by 5.69 ms in Experiment 2)—a small, but measurable trace of the underlying voicing distinction. This result holds in both a minimal pair reading task and a paradigm completion (wug) task, suggesting that this and previous findings of incomplete neutralization in American English are not due solely to extragrammatical or task-related factors such as hyperarticulation and orthographic effects.

While there is a significant difference between vowels preceding /d/ flaps and /t/ flaps on the surface, listeners can neither categorize nor distinguish between the two. This result holds for nonce words, and with the d' measure of sensitivity which takes listener bias into account. In the identification task (Experiment 3), listeners showed an inability to distinguish between /d/ and /t/ flaps—a result in line with the recent study from Herd et al. (2010). While listeners in the ABX task (Experiment 4) showed d' scores higher than chance level, this result appears to be due to listeners' use of cues unrelated to the voicing distinction. Confirming this hypothesis, listeners were unable to distinguish between /d/ and /t/ flaps in the 2AFC task (Experiment 5), in which this strategy was not available to them.

Taken together, these results suggest that listeners' poor performance on identification tasks of /d/ and /t/ flaps is due to their inability to discriminate between the sounds at play, rather than listeners' categorical knowledge. Not all cases of incomplete neutralization, however, are imperceptible. Port and O'Dell (1985), for example, report a 15ms vowel duration difference in German final devoicing, and show that listeners can identify these tokens better than chance. Previewing the following chapter, incomplete neutralization in monomoraic noun lengthening in Japanese also results in a relatively large distinction. As such, I argue that incomplete neutralization is actually a continuum of completeness: at one pole, identification and discrimination are not possible (as in flapping), and at the other, perception of the contrast large enough to be perceivable (as in German or Japanese).

Given the existence of imperceptible incomplete neutralization, one might conclude that the driving force behind incomplete neutralization (at least of the non-distinguishable type) cannot be contrast preservation or conversational cooperation (c.f. Lindblom 1990, Syrett and Kawahara under revision). If a hearer cannot perceive the preserved contrasts, why should the speaker preserve them? I argue that the linguistic machinery required to account for the distinguishable type of incomplete neutralization—which is motivated by the speaker-hearer relationship—is 'overapplied'. This idea, as well as the specifics of the linguistic machinery required, are further developed in Chapter 4, especially Section 4.4.1.

Chapter 3

Bimoraicity-Related Vowel Lengthening in Japanese

Chapter Contents

3.1 Introduction 85
3.2 Background 86
3.3 Experiment 6: Vowel Lengthening in Monomoraic Nouns (I) 90
3.4 Experiment 7: Vowel Lengthening in Monomoraic Nouns (II) 100
3.5 General Discussion and Conclusion 111

3.1 Introduction

Since Port and O'Dell's classic finding on German coda devoicing, incomplete neutralization has been found in a number of other cases, including epenthesis in Levantine Arabic (Gouskova and Hall 2009), flapping in American English (Braver, revision invited; Herd et al. 2010), insertion of intrusive stops in English (Fourakis and Port 1986), tonal

Portions of this chapter stem from Braver and Kawahara (2012) and Braver and Kawahara (to appear).

neutralization in Cantonese (Yu 2007b), voicing assimilation in Russian (Burton and Robblee 1997), [ə]-insertion in English speakers' pronunciation of non-native clusters (Davidson 2006), and coda aspiration in Eastern Andalusian Spanish (Gerfen 2002).

While the vast majority of previously described cases of incomplete neutralization center on feature- and segment-level contrasts, the aim of this chapter—through a case study on Japanese—is to provide evidence of a novel case of incomplete neutralization in the domain of duration-based length contrasts.

This study centers on a prosodic constraint in Japanese which requires every Prosodic Word to be minimally bimoraic. When monomoraic nouns appear in isolation, they must lengthen to meet this prosodic minimality requirement (Mori 2002). Experiments 6 and 7 show that these lengthened nouns are not as long as underlyingly long nouns, thus constituting a case of incomplete neutralization. These studies expand the typology of incomplete neutralization by showing that duration-based length contrasts can be incompletely neutralized.[1]

3.2 Background

There is a large body of evidence showing that Japanese has a bimoraic minimality requirement for Prosodic Words (Itô 1990, Poser 1990, Mester 1990, Itô and Mester 1992, Mori 2002). This bimoraicity requirement is observed in many word formation patterns, all of which are based on a bimoraic template, including nickname formation, geisha client name formation, loanword abbreviation, verbal root reduplication, scheduling compounds, and telephone number recitation.

[1]Myers (2005) shows a case where a short/long vowel length contrast surfaces as three different durations (short vowels, lengthened short vowels before NC sequences, and long vowels). As Myers himself argues, however, the distinction between lengthened and long vowels is best described as coarticulatory shortening of vowels in closed syllables (Fowler 1983, Maddieson 1985). Since this case is explained by factors of phonetic implementation, it does not constitute evidence of true incomplete neutralization of a duration-based length contrast.

For instance, in the nickname formation pattern, a full name must be truncated to two moras before the suffix *-chan*[2] can be applied. For example, the five-mora name *Wasaburoo* can be truncated to two moras as in (1b), but not one, as in (1c). Similarly, the three-mora name *Kotomi* can be truncated to either two monomoraic syllables, as in (2b), or a single bimoraic syllable, as in (2c). *Kotomi* cannot, however, be shortened to a single mora, as in (2d).

(1)	a.	wasaburoo	(full name)		(2)	a.		kotomi	(full name)
	b.	wasa(-chan)	(2 moras)			b.		koto(-chan)	(2 moras)
	c.	* wa(-chan)	(1 mora)			c.		koc(-chan)	(geminate; 2 moras)
						d.	*	ko(-chan)	(1 mora)

The bimoraicity requirement is evident, too, in telephone number recitation patterns (Itô 1990). In the recitation of telephone numbers, monomoraic digits (e.g. *ni* 'two') are lengthened, as in (3a). Additionally, those digits which have both a monomoraic and a bimoraic allomorph (e.g., *shi*~*yon* 'four') always surface as the bimoraic allomorph, as in (3b).

(3) a.

6	5	1	-	3	2	8	6
roku	{ **goo** }	ichi	(no)	san	{ **nii** }	hachi	roku
	*go				*ni		

b.

4	6	1	-	3	8	9	6
{ **yon** }	roku	ichi	(no)	san	hachi	{ **kyuu** }	roku
*shi						*ku	

What nickname formation and telephone number recitation—as well as numerous

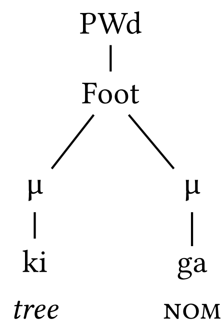
[2]Here and throughout, Japanese morphemes are given in the standard Romaji romanization, except when enclosed in [square brackets], in which case, they are given in IPA.

other morphophonological processes in Japanese (Itô 1990, Poser 1990, Mester 1990, Itô and Mester 1992, Mori 2002)—have in common is that they are all based on the requirement that a Prosodic Word must be binary at the moraic level. A Prosodic Word must contain at least one foot, and the foot must be binary (at the moraic level in Japanese; McCarthy and Prince 1986, 1993), as in (4).

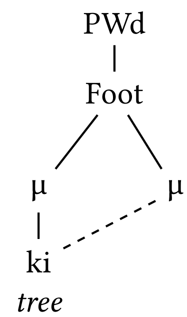
(4)

In spite of this bimoraicity requirement, there are monomoraic nouns in the Japanese lexicon; e.g., [ki] ‘tree’, [i] ‘stomach’, and [e] ‘picture’. Itô (1990) argues that the bimoraic minimality requirement holds only for morphologically derived words. However, Mori (2002) shows that when these monomoraic nouns appear in isolation within a prosodic word (e.g., without case particles), lengthening occurs. She found that monomoraic nouns lengthen in this context by 40–50%, while underlyingly bimoraic nouns do not show such lengthening in the same environment. Therefore, Mori concludes that this lengthening is caused by a phonological bimoraic minimality requirement: monomoraic nouns with a case particle in the same Prosodic Word satisfy the bimoraicity requirement (by virtue of the particle’s mora), as in (5a), while monomoraic nouns must gain an additional mora to satisfy this requirement, as in (5b).

(5) a. No lengthening with a particle

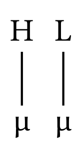


b. Lengthening without a particle

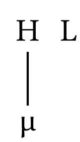


It should be noted that the lengthening must have a phonological component to it, as the lengthened nouns are demonstrably bimoraic on the surface. Pitch accent in Japanese is comprised of a HL tonal contour, with an H tone on one vocalic mora, and an L tone falls on the next, as in (6a) (see McCawley 1968 and Kawahara (to appear) for an overview of the phonology of Japanese accent). In finally-accented words, only the H of the HL tonal complex is realized on the final vowel (unless there is following phonological material; Kawahara (to appear)). The prediction, then, is that only bimoraic (or longer) words will show both the H and the L of the accent's HL tonal complex, and monomoraic words will show only an H tone, as in (6b). Japanese does not allow tonal crowding (as suggested by the behavior of finally accented words), so a word with an HL tonal complex on the surface cannot simply be a single mora with two tones, as in (6c).

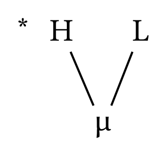
(6) a.



b.



c.



Although Mori does not include underlyingly long vowels in her stimulus set, she does refer to previous studies (Beckman 1982, Hoequist 1983) which have shown that Japanese heavy syllables are are generally 66–80% longer than light syllables. A more recent phonetic study by Hirata (2004) shows that long vowels in Japanese can be up to 150% longer than short vowels. This difference between Mori's results (40-50% longer) and other studies on Japanese length distinctions implies, as Mori herself suggests, that we may be observing a case of incomplete neutralization. Experiments 6 and 7 set out to directly test this hypothesis by comparing the vowel duration of lengthened nouns to that of underlyingly long nouns.

3.3 Experiment 6: Vowel Lengthening in Monomoraic Nouns (I)

In this experiment (reported in Braver and Kawahara (to appear)), native speakers of Japanese were asked to read sets of sentences aloud. Each set contained three sentences, forming a minimal triplet differing only in their noun length and presence/absence of a particle: (i) monomoraic noun with a particle ('short/prt'); (ii) monomoraic noun without a particle ('short/Ø'); (ii) underlyingly long noun ('long'). The incomplete neutralization hypothesis would suggest that monomoraic nouns without a particle undergo lengthening, but that they do not become as long as underlyingly long nouns.

From the previous studies discussed above, we expect that (i) monomoraic nouns are lengthened without case particles, as Mori (2002) found, but that (ii) the lengthened nouns are not as long as underlyingly long vowels.

3.3.1 Method

3.3.1.1 Stimuli

11 sets of 3 sentences were constructed, as described above. Nouns within each set had the same segmental content, modulo vowel length. A sample set is provided in Table 3.1; the full list of stimuli is given in Appendix B.

A few remarks are in order about the stimuli, which I come back to in more detail in Section 3.3.3.2. Accent, which is lexically contrastive in Japanese, was controlled in 9 of the 11 sets. In the long condition, long vowels were orthographically indicated by

	Condition	Japanese orthography	Transcription	Gloss
(a)	short/prt	麩が美味しい。	*fu* ga oishii	gluten NOM delicious
(b)	short/Ø	麩美味しい。	*fu* oishii	gluten delicious
(c)	long	封がとれた。	*fuu* ga toreta	seal NOM came.off

Table 3.1: Experiment 6: Sample stimulus set

either (a) kanji alone, or (b) kana with a length mark (ー). Some morphemes written with kanji, had they been written in hiragana, would have been written as diphthongs. Such morphemes, however, are generally pronounced with long monophthongs (see Vance 2008, pp.63-68, for discussion).

3.3.1.2 Participants

Participants were seven female native speakers of Japanese. They were undergraduate and graduate students at Japanese universities. One speaker was excluded from analyses since she explicitly stated that she noticed the lengthening phenomenon during the recording. The remaining six speakers were all from the Kantō area of Japan (the region around Tokyo).

3.3.1.3 Recording

Recording took place in a sound-attenuated room at International Christian University (Tokyo, Japan), using a TASCAM DR-40 recorder, with a 44.1kHz sampling rate. SuperLab (ver. 4, Cedrus Corporation 2010) was used to present the stimuli. Speakers practiced all items once to ensure that they read kanji as intended. Each speaker read all 33 sentences (11 sets × 3 sentences) in random order. The sentences were re-randomized, and the speaker re-read the sentences. Each speaker read each sentence a total of ten times. Speakers were instructed not to pause mid-sentence, in order to prevent them from inserting a pause or glottal stop rather than lengthening for the short/Ø condition.

3.3.1.4 Acoustic Measurements

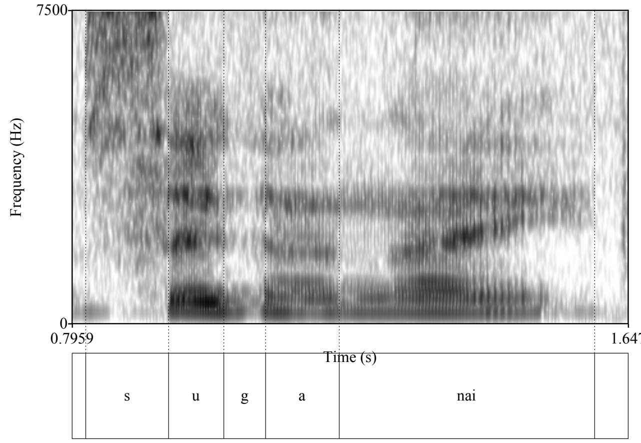


Figure 3.1: Experiment 6: A representative segmented spectrogram. Speaker 14, *su ga nai* (short/prt), repetition 9.

For each sentence, vowel durations of the target noun was measured. Segmental boundaries were placed using the start/end of visible F2 and/or F3. A sample spectrogram in Figure 3.1 illustrates the segmentation procedure.

3.3.1.5 Statistical Analysis

A linear mixed model was run using the `lme4` package in R. Vowel duration was regressed against condition (short/prt, short/Ø, long) as a fixed factor, and speaker and item as random factors. Condition was treatment coded, to produce comparisons between short/prt vs. lengthened (short/Ø) nouns, and lengthened vs. underlyingly long nouns. Since the way to calculate degrees of freedom for these analyses are not yet known (Baayen 2008), the significance values are calculated by the Markov Chain Monte Carlo method using the `pvals.fnc` function of the `languageR` package (Baayen 2009).

3.3.2 Results

Figure 3.2 and Table 3.2 illustrates the general results. On average, short nouns with a particle were 73.54 ms, short nouns without a particle were 119.19 ms, and underlyingly long nouns were 145.74 ms. These results show that short nouns lengthen without case particles ($t = -8.018, p < 0.001$), but they do not become as long as underlyingly long nouns ($t = 1.369, p < 0.05$).[3] This three-way distinction holds across all of the stimulus sets and all of the speakers in this experiment, as can be seen in Figures 3.3 and 3.4, respectively.

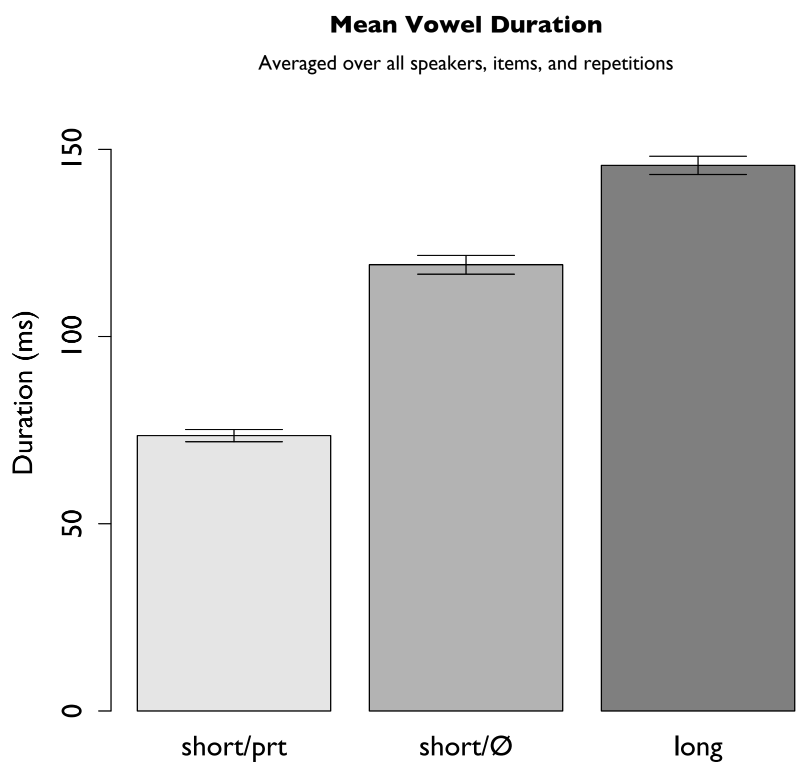


Figure 3.2: Experiment 6: Mean vowel durations, averaged over all speakers and all items.

[3]The effect size may seem small, despite the fact that the mean difference in raw duration is 26.55 ms. A post-hoc t-test confirms the significance of the lengthened vs. long vowel length distinction ($t(1278.99) = -14.90, p < 0.001$).

	Mean	SD
short/prt	73.54	20.58
short/Ø	119.19	32.56
long	145.74	31.21

Table 3.2: Experiment 6: Mean and standard deviation of vowel durations (in ms) for each category

Before concluding that the Japanese monomoraic lengthening pattern instantiates a novel type of incomplete neutralization, there is one alternative possibility. That is, if the current participants did not lengthen some short/Ø nouns at all, and also *completely* neutralized some other short/Ø nouns, we might have obtained the above results, even if they did not incompletely neutralize: intermediate values can arise from the averaging of short and long values. If that were the case, the vowel durations in the short/Ø condition would show a bimodal distribution, with one peak overlapping with the short/prt condition and the other peak overlapping with the long condition. Figure 3.5 presents histograms to addresses this possibility. We observe that lengthened nouns are unimodal, distributing between the short condition and the long condition. I thus conclude that the Japanese lengthening is a genuine case of incomplete neutralization.

3.3.3 Discussion

3.3.3.1 A New Case for Phonological Incomplete Neutralization

Monomoraic nouns in Japanese become longer when they appear in isolation, in order to satisfy the bimoraicity requirement. However, the lengthened nouns do not become as long as underlying long nouns. These results hold for all eleven triplets tested, across all six speakers. The lengthening of monomoraic nouns in Japanese is incompletely neutralizing. While we might have expected lengthened nouns to become identical to underlyingly long nouns, a trace of their underlying phonological length is apparent on the surface; i.e., lengthened nouns have a trace of their underlying phonological

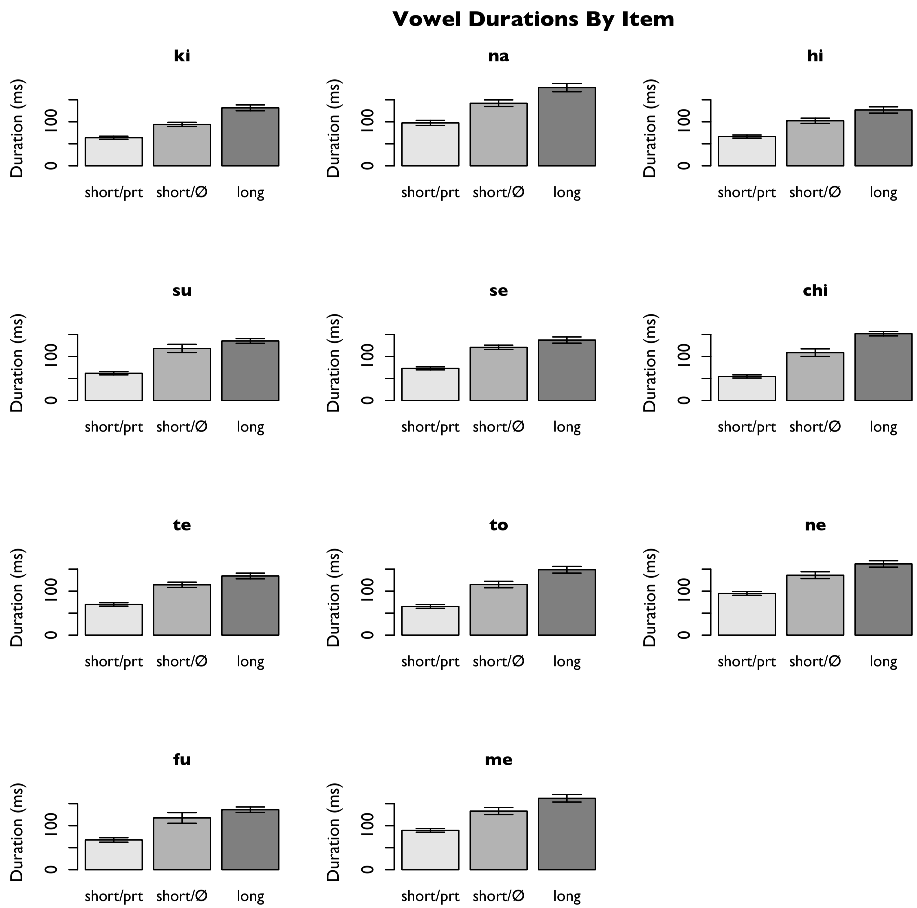


Figure 3.3: Experiment 6: Mean vowel durations, by item.

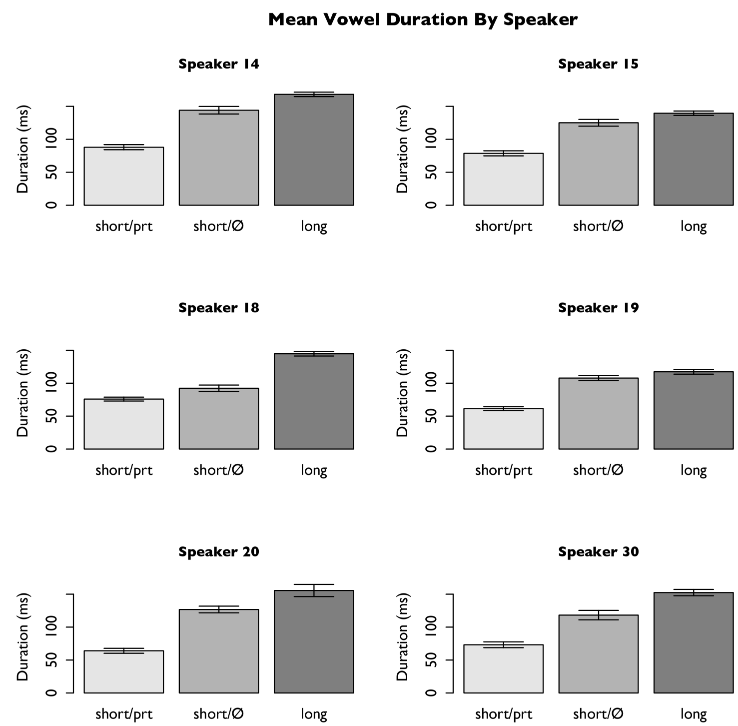


Figure 3.4: Experiment 6: Mean vowel durations, by speaker.

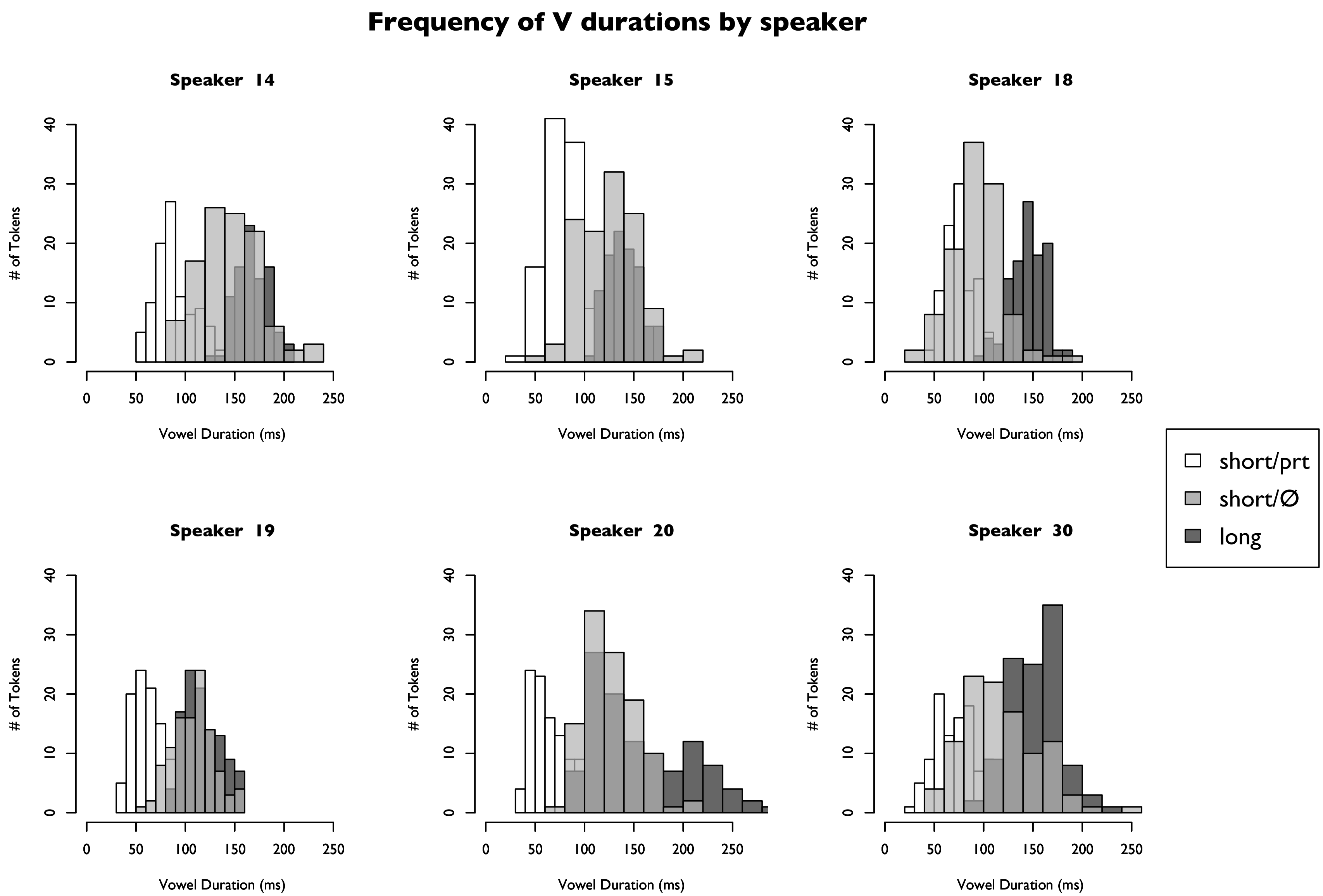


Figure 3.5: Experiment 6: Distribution of vowel duration by condition, for each speaker.

shortness—they are shorter than underlyingly long nouns.

The lengthening pattern cannot be relegated to a matter of phonetic implementation. The lengthening is motivated by a prosodic bimoraic word minimality requirement, which interacts with many morphophonological patterns (Poser 1990, Itô 1990) that yield new morphologically derived words. In fact, Ito (1990) argues that the bimoraicity holds only in morphologically derived environments, suggesting that the prosodic bimoraicity requirement exists "deep" in Japanese phonology. I therefore maintain that this case of incomplete neutralization cannot be relegated to a phonetic implementation rule.[4]

3.3.3.2 Discussion of the Current Stimuli

Although I believe that the current experiment demonstrates that the Japanese monomoraic lengthening pattern is a solid case of incomplete neutralization, there are some details of the stimuli that may merit discussion. The full list of stimuli from Experiment 6 is given in Appendix B, and is annotated to indicate the stimuli to which each of these discussion points pertain.

Pitch accent As discussed in Section 3.3.1.1, 2 of the 11 stimulus sets did not match for accent. In other words, the short noun (used in the short/prt and and short/Ø conditions) was accented while the long noun was not (or vice versa). Hoequist (1983) finds a small effect of pitch on syllable duration in Japanese (high pitch:low pitch = 1.08:1). For comparison, the results of the current experiment show that the overall long:short/Ø ratio = 1.22:1, and in the two accent mismatched sets this same ratio was 1.35:1. Given the small effect size found by Hoequist, the small difference in long vowel duration between accent

[4] I make the more or less standard assumption that phonetics does not affect morphological word formation patterns (i.e. there are no morphological patterns that are sensitive to raw phonetic properties). Given this assumption, since the bimoraicity constraint governs many morphological processes, it must be phonological rather than phonetic. See also Cohn (1998) for arguments that the English minimal word requirement is based on abstract, phonological lengths rather than on raw phonetic duration.

matched and accent mismatched sets, and the consistent pattern across all lexical sets, I argue that the accent-mismatched sets are legitimate evidence.

Orthographic diphthongs Also discussed in Section 3.3.1.1 was the fact that some stimuli in the long condition, had they been written in hiragana, would have been written as diphthongs. (Note that no stimuli were written in this way—all such stimuli were represented by kanji.) For example, *sei* 'positive', was written with the kanji「背」, however, had it been written in hiragana, it would have been rendered as「せい」(se+i). In any event, *sei* is pronounced [se:]. For ease of exposition, I refer to 'diphthong sets' and 'non-diphthong sets' to distinguish those sets containing long vowels that could have been written with orthographic diphthongs—even though no actual diphthongs (orthographic or pronounced) were involved in any of the sets.

Diphthong sets had long vowels, on average, 15.25 ms shorter than long vowels in non-diphthong sets (non-diphthong sets: 148.21 ms, diphthong sets: 132.87 ms, $t(154.93) = 4.42$, $p < 0.001$). Given that long vowels in diphthong sets were shorter, we might expect that in these sets, the short/Ø vowels (which do not have even the hypothetical potential for diphthongization, and therefore should not have depressed duration) would be as long as the depressed long vowels. In other words, in the diphthong sets, the depressed vowel duration in the long condition might lead to the appearance of complete neutralization, since the lengthening short/Ø vowels have less duration to make up to meet that of the long vowels. In spite of these facts, there is still a distinction between short/Ø vowels and long vowels in these diphthong sets (short/Ø: 122.31 ms, long: 132.87 ms, $t(198.95) =$, $p < 0.005$).

Quoted/expressive long nouns 5 of the 11 long nouns were quoted expressives or interjections (see Appendix B). Quoted long vowels were on average 18.42 ms longer than non-quoted ones (quoted: 155.48 ms, non-quoted: 137.06 ms, $t(516.365) = -7.55$, $p <$

0.001). It is conceivable that the difference between short/Ø and long vowel durations in the quoted sets is partly attributable to the apparent lengthening effect of quotation.

However, I argue that this is not the case. All vowels in the quoted sets—not just long vowels—are longer than vowels in the non-quoted sets. (Quoted short/prt: 83.17 ms, non-quoted short/prt: 65.26 ms, difference: 17.91 ms, $t(492.702) = -11.56, p < 0.001$). This suggests that something about the quoted sets—which pertains to both the short/prt and the long vowels—is increasing their duration. Under the hypothesis that quotation itself was causing this durational difference, we would expect only long vowels to be impacted, since only long vowels were quoted in these sets.

One possibility is that the onsets in the quoted sets allowed for greater expansion of the vowels than the non-quoted sets. This hypothesis is supported by the rather large difference in consonant duration across the two sets (quoted: 57.23 ms, not-quoted: 92.97 ms).

It should also be noted that the main results from this experiment still hold if the quoted sets are removed from the data. Lengthening occurs (short/Ø vowels are longer than short/prt vowels, mean difference: 48.84 ms, $t = -9.68, p < 0.001$) and neutralization is incomplete (short/Ø vowels are not as long as underlyingly long vowels, mean difference: 22.98 ms, $t = 2.50, p < 0.05$).

While I believe based on the discussion here that none of these three factors are an actual confound, I nevertheless present a follow-up study, reported here as Experiment 7.

3.3.3.3 Conclusion

In Japanese, monomoraic nouns become longer when they appear in isolation in order to satisfy the bimoraic minimality requirement. The lengthened vowels, however, do not become as long as underlying long vowels. This three-way distinction does not arise from averaging over the results of an optional process (Section 3.3.2), or from some possible factors that affect duration independently (Section 3.3.3.2). Experiment 7, the follow-up to this experiment, further bolsters this result. I thus argue that the vowel lengthening pattern in Japanese instantiates a case of incomplete neutralization.

3.4 Experiment 7: Vowel Lengthening in Monomoraic Nouns (II)

In this experiment—a follow-up to Experiment 6, a similar procedure was employed. Given post-hoc analysis of the stimuli from Experiment 6, Experiment 7 used a new set of stimuli which contained no quoted/expressive forms, and which included better controls for sentence-level mora count through identical frame sentences within each set.

3.4.1 Method

3.4.1.1 Stimuli

15 sets of minimal triplet sentences were constructed, each containing: (a) a monomoraic noun followed by the particle *mo* ('short/prt' condition), (b) a monomoraic noun without a particle ('short/Ø' condition), and (c) an underlyingly long noun without a particle ('long' condition). A sample set is given in Table 3.3.

	Condition	Japanese orthography	Transcription	Gloss
(a)	short/prt	木もなくしたよ。	ki mo nakushita yo	tree ALSO lost DISC
(b)	short/Ø	木なくしたよ。	ki nakushita yo	tree lost DISC
(c)	long	キーなくしたよ。	kii nakushita yo	key lost DISC

Table 3.3: Experiment 7: Sample stimulus set

Within each set, the nouns' segmental content was identical, with the exception of vowel length in the long condition. Non-approximant consonants were used as onsets (if present) in the target nouns to facilitate clear segmentation. Experiment 6 used the nominative particle *ga*, since it is arguably the default case marker in Japanese subjects (Fukui 1986, Inoue 1997). In that study, however, [g] sometimes spirantized, which made the segmentation more difficult. Therefore, in this study, the commitative particle *mo* was used in the short/prt condition in order to facilitate segmentation. No particle was included in the long vowel condition, because the main target comparison was between the short/Ø condition and the long condition.[5] Since Mori (2002) had already shown that long nouns are barely affected in duration by the presence/absence of case particles, this dimension was not varied in the stimuli. Unlike in Experiment 6, all three items within a given set had the same predicate to control for any sentence-level duration compensation effects. The predicate always started with a non-approximant consonant to make the segmentation more straightforward. A sentence-final discourse particle [yo] was attached at the end of each sentence to make the stimulus sentences more colloquial, and to further make the absence of case particles more natural. The list of all the stimuli used in this experiment is provided in Appendix C.

3.4.1.2 Participants

Twelve native speakers of Japanese participated in the experiment. They were all undergraduate students at International Christian University (Tokyo, Japan). They were paid ¥500 for their time, and each signed a consent form before participating in the experiment.

3.4.1.3 Procedure

The recording session took place in a sound-attenuated room at International Christian University. Superlab version 4.0 (Cedrus Corporation 2010) was used to present the

[5] Due to an error, one stimulus set contained the particle *mo* in the long condition. A post-hoc analysis shows no substantial difference between this set and other sets.

stimuli. The stimuli were written in the standard Japanese orthography, with a mixture of kanji, katakana, and hiragana (see Appendix C).

In each block, every stimulus was presented once, and speakers were asked to read the stimuli as they were presented on the screen. The speakers were allowed to take a short break after each block. The order of the stimuli within each block was randomized by Superlab. Each speaker read each sentence a total of 7 times. Each speaker was assigned 30 minutes for the experiment.

Before the main session, as practice, each speaker read all the stimuli once to familiarize themselves with the stimuli and the task. After the practice phase, speakers were able to ask clarifying questions about the procedure. Their speech was directly recorded onto a portable recorder (TASCAM DR-40) with a 44.1kHz sampling rate and a 16 bit quantization level.

The duration of each vowel was measured, starting at the offset of the preceding consonant and ending at the end of visible F2/F3, using Praat (Boersma and Weenink 2009). A representative spectrogram is given in Figure 3.6.

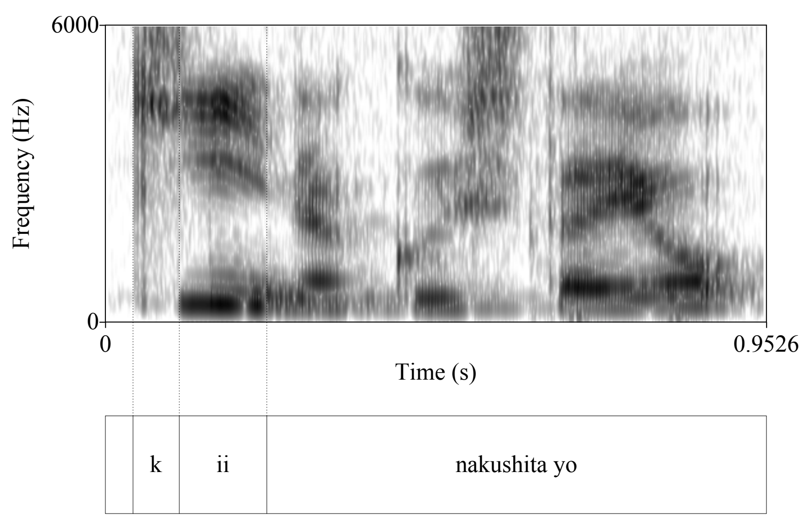


Figure 3.6: Experiment 7: A representative segmented spectrogram. Speaker 43, *kii nakushita yo* (long), repetition 7.

3.4.1.4 Statistical Analysis

Statistical significance was assessed with a linear mixed model (Baayen 2008) in which vowel duration was regressed against condition (short/prt, short/Ø, long) as a fixed factor and with speaker and item as random factors. Condition was treatment coded to produce comparisons between short/prt vs. short/Ø (to assess whether lengthening occurs) and short/Ø vs. long (to assess whether lengthened nouns are as long as underlyingly long nouns). Since the way to calculate degrees of freedom for these analyses are not yet known (Baayen 2008), the significance values are calculated by the Markov Chain Monte Carlo method using the `pvals.fnc` function of the `languageR` package (Baayen 2009).

3.4.2 Results

Figure 3.7 and Table 3.4 shows the overall results, averaging over all speakers and all items. Short nouns are lengthened when they appear without case particles, and hence are longer than short nouns that appear with particles (mean difference: 69.98 ms, $t = 15.692, p < 0.001$). However, the lengthened nouns are not as long as underlyingly long vowels (mean difference: 32.47 ms, $t = 7.047, p < 0.001$). Therefore, the Japanese lengthening pattern instantiates a case of incomplete neutralization.

	Mean	SD
short/prt	54.99	21.89
short/Ø	124.98	34.91
long	157.45	39.21

Table 3.4: Experiment 7: Mean and standard deviation of vowel durations (in ms) for each category

To investigate whether this three-way distinction holds across speakers, Figure 3.8 shows the patterns of all 12 speakers analyzed. As can be seen in the figure, all speakers show incomplete neutralization: lengthened nouns are not as long as underlyingly long

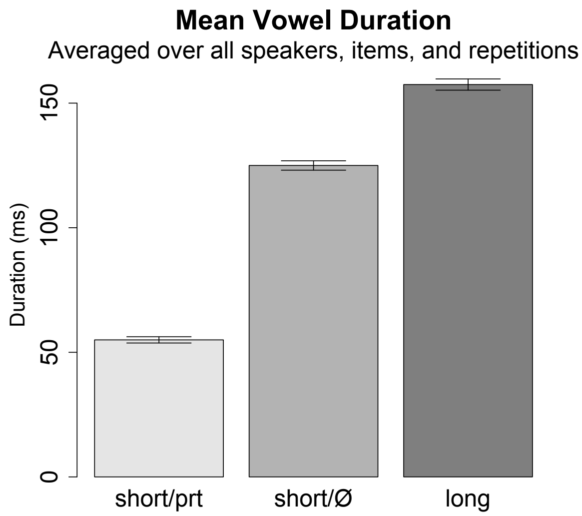


Figure 3.7: Experiment 7: Vowel duration over all speakers and all items.

nouns for any speaker.[6]

Finally, to investigate the item effect, Figure 3.9 shows a by-item analysis, with results for each of the 15 lexical sets. It is again observed that within each set, all short nouns are lengthened without particles, but they are not as long as underlyingly long nouns.

3.4.3 Discussion

3.4.3.1 Looking Deeper into the Data

I first take a deeper look at the data, discussing some aspects of the stimulus sets and the results.

[6] The two speakers with the smallest mean differences between short/Ø and long vowels were speakers 44 and 46. The difference for speaker 46 is significant (short/Ø mean: 139.12, long mean: 147.94, mean difference: 8.82, $t = 19.43, p < 0.001$). The difference for speaker 44 trends in the same direction as the other speakers, but does not reach significance (short/Ø mean: 125.79, long mean: 131.45, mean difference: 5.66, $t = .928, n.s.$).

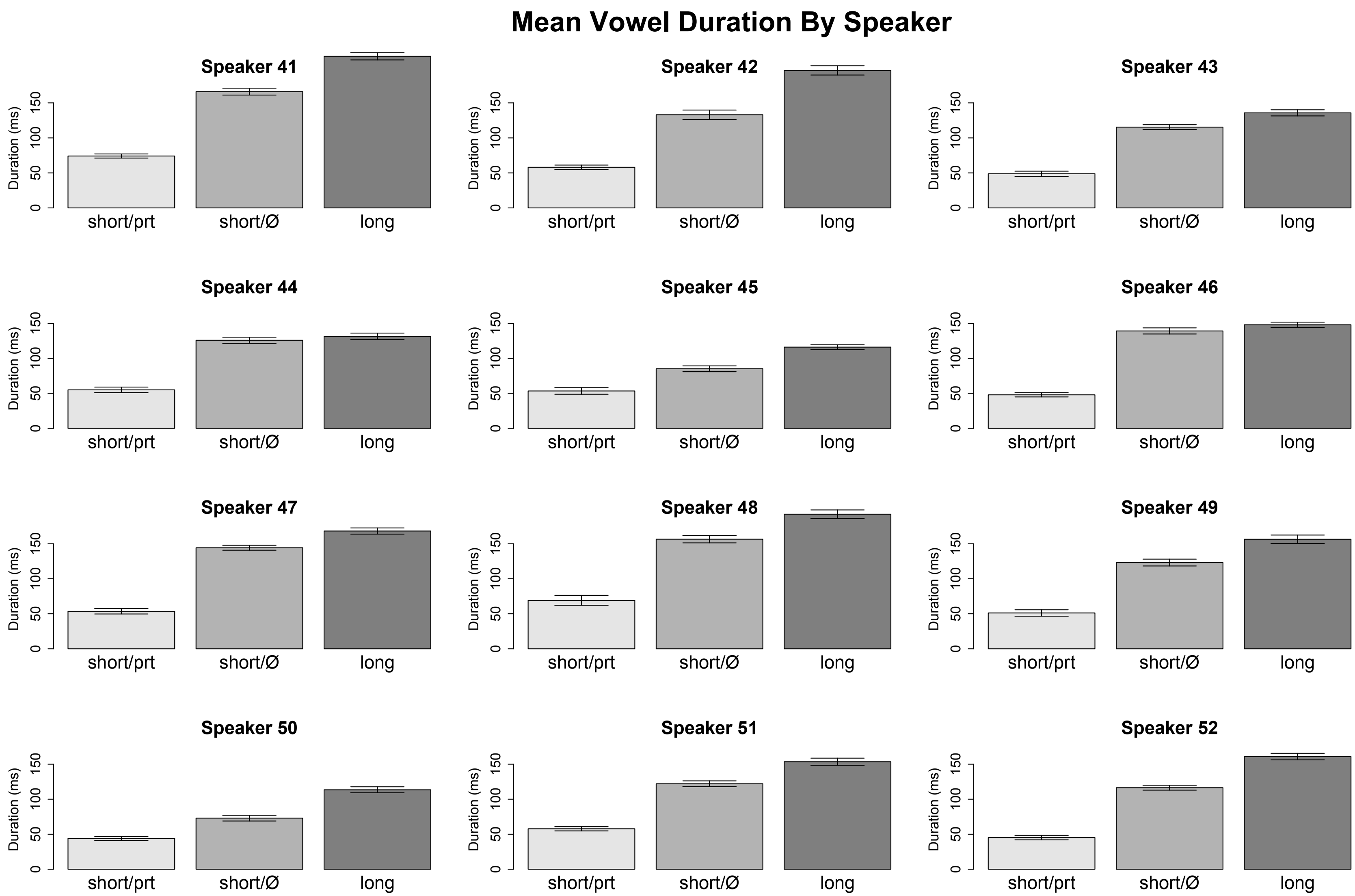


Figure 3.8: Experiment 7: Vowel duration by speaker, averaged across items.

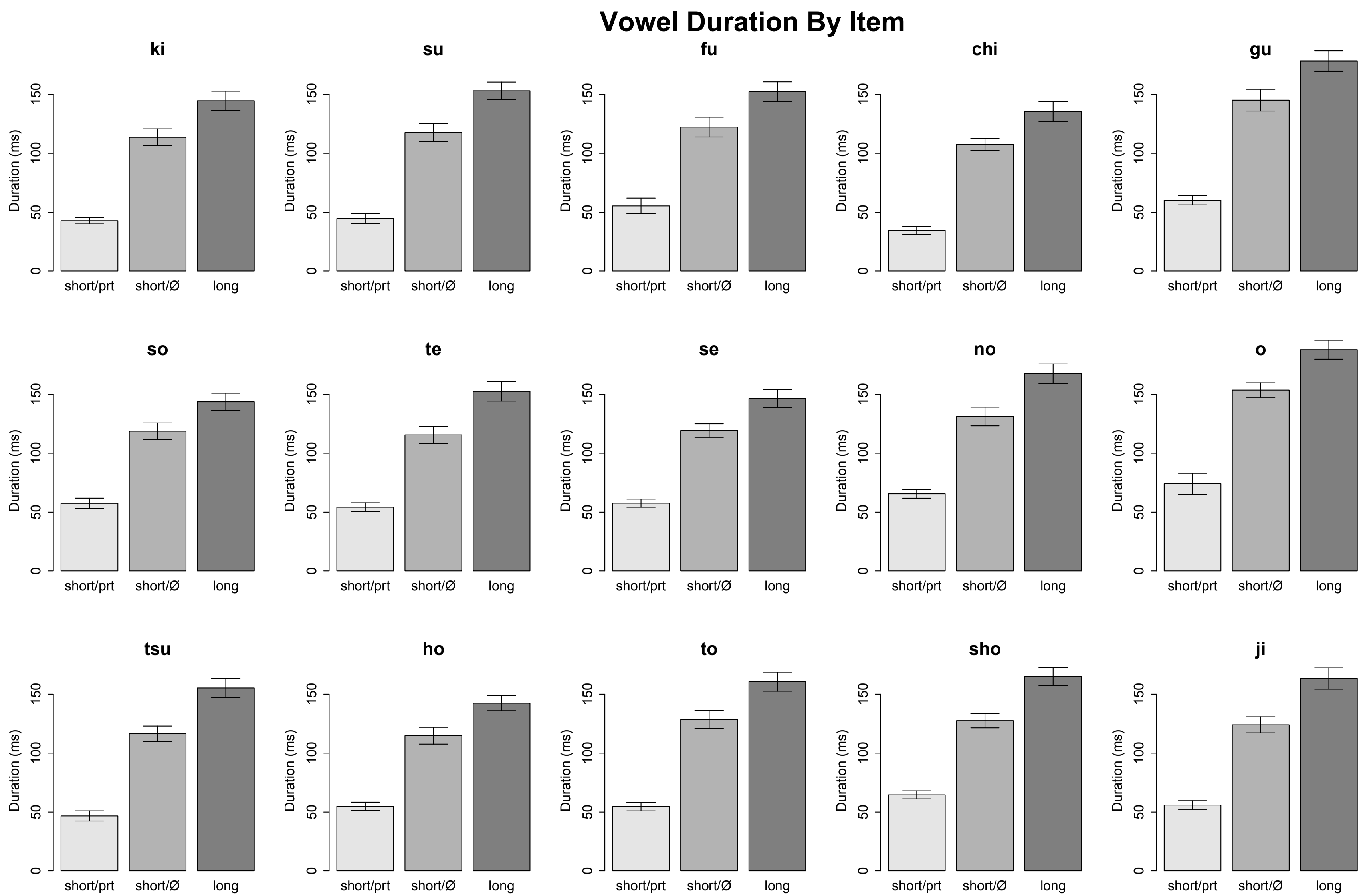


Figure 3.9: Experiment 7: Vowel duration by item, averaged across speakers.

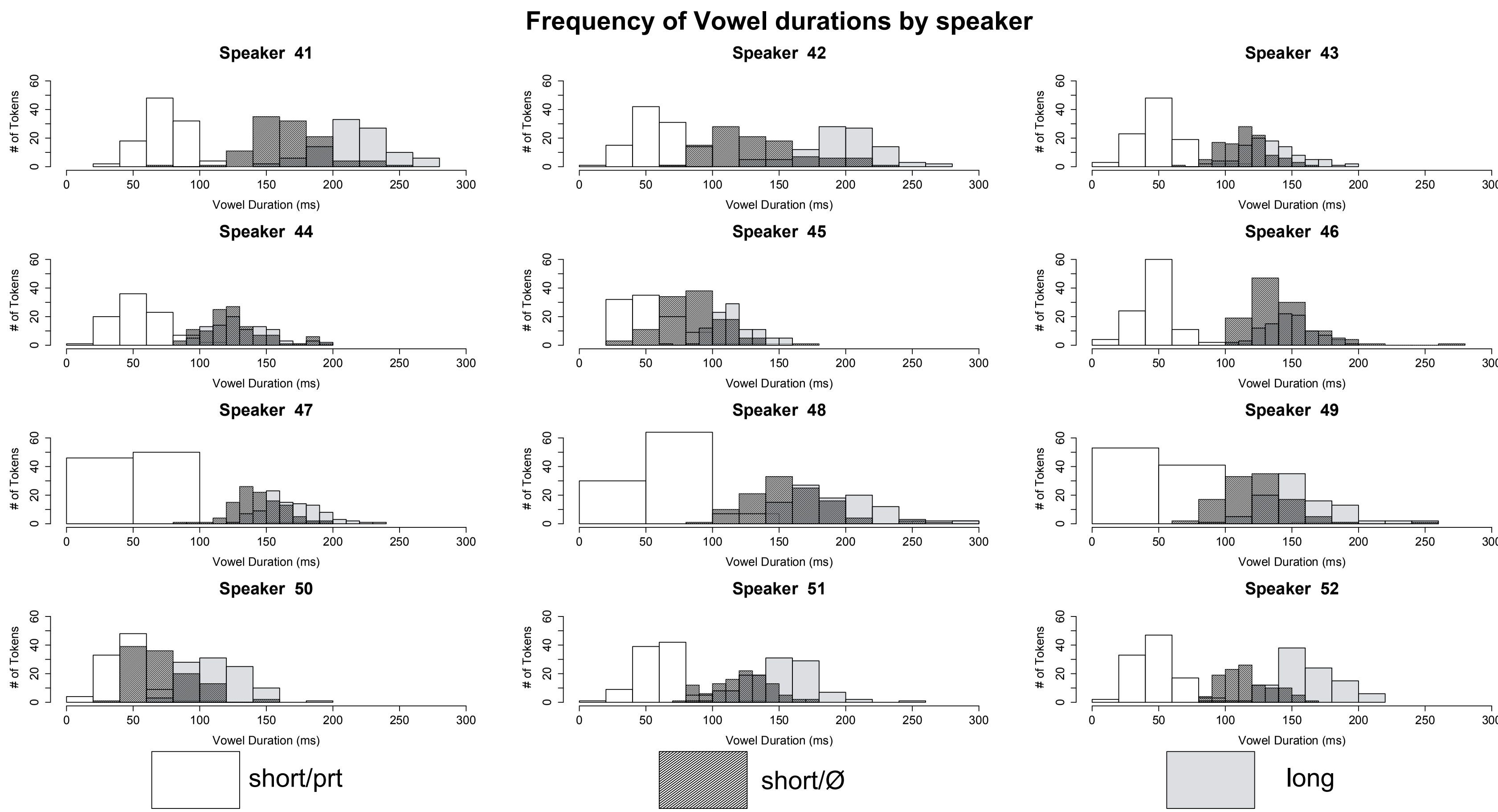


Figure 3.10: Experiment 7: Distribution of vowel duration by condition, for each speaker.

Distribution of conditions within each speaker One might argue that this case of incomplete neutralization derives from optional application of vowel lengthening. If speakers apply lengthening of short/Ø nouns optionally, they would produce both short and long nouns in the short/Ø condition—averaging over these tokens would result in an intermediate duration between the short/prt and long conditions. To address this possibility, Figure 3.10 provides histograms of each condition for each speaker. This alternative hypothesis predicts that lengthened nouns should show a bimodal distribution—one portion of the short/Ø tokens overlapping with the short/prt condition and the other portion overlapping with the long condition. It is observed that, contrary to the hypothesis entertained above, lengthened nouns have a unimodal distribution which is intermediate between the short condition distribution and the long condition distribution.

Orthographic diphthongs According to Japanese writing convention, some long vowels are represented as 'orthographic diphthongs' when spelled out in the hiragana syllabary. For example, *nou* [noo] 'brain' would be written in hiragana as 'のう' (no + u). As with Experiment 6, some long vowels in Experiment 7, had they been written in hiragana, would have been rendered this way (e.g., *nou* as above, and *tei* [tee] 'base' as 'てい' (te + i)), however, all long target stimuli were rendered in logographic kanji or the katakana syllabary which renders long vowels with a length mark (ー). These orthographic diphthongs are generally pronounced as long vowels (Labrune 2012; see Vance 2008, pp. 63–68 for discussion), and thus were not expected to be a confound. Recall from Figure 3.9 that the three-way incomplete neutralization holds for all lexical sets—only some of which would have had orthographic diphthongs had they been written in hiragana. Incomplete neutralization holds in both the 7 sets where orthographic diphthongs would have been present had hiragana been used for the long nouns, as well as in the 8 remaining sets where the writing system did not call for orthographic diphthongs (see Table 3.5).

		Duration		Difference		
Orthographic diphthong sets ($n = 7$)	short/prt:	60.53ms	}	66.60ms	$t = 3.87$,	$p < 0.001$
	short/Ø:	127.13ms	}	32.65ms	$t = 8.01$,	$p < 0.001$
	long:	159.78ms				
Non-orthographic diphthong sets ($n = 8$)	short/prt:	51.12ms	}	72.91ms	$t = 5.78$,	$p < 0.001$
	short/Ø:	124.03ms	}	31.67ms	$t = 13.89$,	$p < 0.001$
	long:	155.70ms				

Table 3.5: Experiment 7: Results from sets containing orthographic diphthongs in long vowel conditions, and those without orthographic diphthongs.

Accent mismatches Finally in some sets, short nouns and long nouns differ in accent (e.g., *fu* is unaccented while *fu'u* is accented) (see Appendix C). However, since the effect of accent on Japanese vowel duration is minute (8% increase in Hoequist 1983) and 12 out 15 sets are controlled in terms of their accentuation, the finding of a durational difference between lengthened nouns and long nouns cannot be attributed to accentual differences. Recall again that the tripartite incomplete neutralization holds in all sets, regardless of whether the short nouns and long nouns agree in accent. Incomplete neutralization holds in the 3 sets with accent mismatches, as well as in the 12 sets with no such mismatch (see Table 3.6).

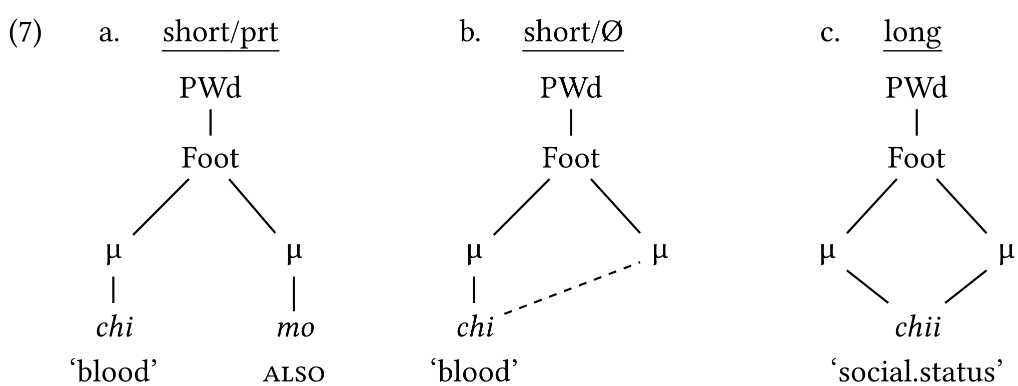
3.4.3.2 General Implications

The current results suggest that the short/long vowel length distinction in Japanese is incompletely neutralized when monomoraic nouns without case particles are lengthened: these lengthened nouns must have two moras on the surface to meet the Japanese bimoraicity requirement (Itô 1990, Poser 1990, Mester 1990, Itô and Mester 1992, Mori 2002), yet their vowel durations are intermediate between those of underlyingly short

		Duration		Difference		
Accent mismatch sets ($n = 3$)	short/prt:	50.19ms	}	74.79ms	$t = 2.03$,	$p < 0.05$
	short/Ø:	124.98ms	}	30.39ms	$t = 5.08$,	$p < 0.001$
	long:	155.37ms				
Non-accent mismatch sets ($n = 12$)	short/prt:	56.12ms	}	68.86ms	$t = 6.90$,	$p < 0.001$
	short/Ø:	124.98ms	}	33.05ms	$t = 14.91$,	$p < 0.001$
	long:	158.03ms				

Table 3.6: Experiment 7: Results from sets with accent mismatches, and those without accent mismatches.

and underlyingly long vowels. As an example, take the set given in (7). Since *chi mo* (short/prt), in (7a), and *chii* (long), in (7c), both have two underlying moras within their Prosodic Word, no lengthening occurs in these conditions. In order to meet the bimoraicity requirement, *chi* (short/Ø), in (7b) must link to a second additional mora, since there is no other available underlying segmental content. This study shows, however, that lengthened vowels like those in (7b), are not as long as underlyingly long vowels, like those in (7c).



Having established that the Japanese case is indeed a case of incomplete neutralization, some remarks are in order. First the current results expand the typology of processes that can lead to incompletely neutralized contrasts to include not just processes at the segment- and feature-level, but also processes motivated by suprasegmental structure.

Second, since the lengthening is triggered by a clearly phonological constraint, it cannot be treated as a matter of phonetic implementation—unlike a number of proposed cases of incomplete neutralization. For example, Ohala (1974) and Fourakis and Port (1986) treat the case of intrusive stops in English as a matter of phonetic implementation. Similarly, Davidson (2006) treats [ə]-insertion in English speakers' pronunciation of non-native clusters, which results in an apparent case of incomplete neutralization, as resulting from gestural mis-coordination. If the phenomenon in question is a matter of phonetic implementation, it is not strictly speaking a case of incomplete neutralization, as two segments are not neutralized phonologically.

In the current case, however, lengthening is motivated by a clearly phonological, rather than phonetic, bimoraic minimality constraint in Japanese. The constraint is deeply tied into the morphophonology of Japanese, as it governs many Japanese morphophonological patterns (Itô 1990, Poser 1990, Mester 1990, Itô and Mester 1992, Mori 2002). I thus conclude that lengthening is phonological, as it is triggered by a phonological constraint, and cannot be relegated to a matter of phonetic implementation.

3.5 General Discussion and Conclusion

The experiments in this chapter show that the short/long vowel length distinction in Japanese is incompletely neutralized in the context of monomoraic noun lengthening (Experiments 6 and 7). These results suggest the novel finding that duration-based length contrasts can be incompletely neutralized, as shown in both Experiments 6 and 7). Given this fact, the typology of processes susceptible to incomplete neutralization must be expanded to include processes—like monomoraic noun lengthening—that affect a contrast of length or prosodic structure.

It should be noted that Japanese speakers can distinguish vowel duration to an accuracy of less than 10 ms (Fujisaki et al. 1975). Given that the mean difference between short/Ø vowels and long vowels was much larger—26.55 ms (=145.74-119.19) in

Experiment 6 and 32.47 ms (=157.45-124.98) in Experiment 7, it is likely that this difference is perceptible by speakers of Japanese. Considering again the continuum from relatively imperceptible incomplete neutralization to very perceptible incomplete neutralization, Japanese falls on the more perceptible end than either flapping (2) or German (Port and O'Dell 1985).

Chapter 4

Modeling Incomplete Neutralization with Frequency-Based Transderivational Identity in a Weighted Constraint Grammar

In the preceding chapters, I have shown that incomplete neutralization can take multiple forms. On the one hand, in Chapter 2, I argued that incomplete neutralization in American English flapping can result in distinctions so small as to be imperceptible. On the other hand, in Chapter 3, I showed that incomplete neutralization in Japanese monomoraic noun lengthening results in a relatively large—though still subphonemic—distinction.

In this chapter, I argue for a model that derives the continuum of degrees of completeness in incomplete neutralization (for the sake of exposition I will call the two poles of this continuum 'imperceptible' and 'perceptible' incomplete neutralization, respectively).

The model combines two major existing ideas. The first is transderivational identity, or Output-Output identity (Benua 1997). While Benua's (1997) output-output correspondence constraints were originally intended to apply to categorical phonological phenomena, I argue (following Steriade 2000[1]) that this family of constraints can contribute to incomplete neutralization. Further, I argue that Output-Output identity, at least in the case of incomplete neutralization, should require faithfulness to a *frequent* base, rather than a morphologically simple base.

The second main mechanism behind the proposed model is a weighted constraint grammar in the phonetics (see, e.g., Legendre et al. 1990, Pater 2009 on weighted constraint models generally, and Zsiga 2000, Flemming 2001 for such models in phonetics). This type of grammar allows for competing phonetic pressures to be modeled such that conflict is resolved through compromise (as opposed to a ranked constraint grammar with strict dominance, like Optimality Theory (Prince and Smolensky 1993)). I argue that such compromise is key to incomplete neutralization—a process that results in outputs that are in many ways a middle path between two extremes.

Chapter Contents

4.1 Overview of the Model 115

4.2 Japanese 127

4.3 Flapping 143

4.4 Discussion and Conclusion 156

[1]It should be noted that Steriade (2000) addresses flapping in a model of phonetic analogy based on paradigm uniformity. Two remarks are in order. First, the data behind this claim was not supported in a follow up study by Riehl (2003). Second, the model does not take frequency into account in determining basehood, contrary to the model I propose in this chapter.

4.1 Overview of the Model

As discussed above, the model I argue for here combines a system of weighted constraints in the phonetics with transderivational identity. These ideas are briefly summarized in the following subsections, then the model is applied to the case studies of incomplete neutralization discussed in the previous chapters of the dissertation.

4.1.1 Weighted Constraints and Phonetics: Zsiga (2000) and Flemming (2001)

At the heart of the model being proposed is the idea that competing phonetic demands must compromise. This idea is familiar from the phonetic literature—for example, both Zsiga (2000) and Flemming (2001) argue for weighted constraint grammars (Legendre et al. 1990, Pater 2009) that can operate over phonetic information. They differ, however, on the relationship between these constraints and the phonological module—Zsiga (2000) argues for a set of 'phonetic alignment constraints' that apply separately from the phonology, Flemming (2001) articulates a unified model of phonetics and phonology, deriving both categorical and gradient phenomena with the same sort of weighted constraints.

More specifically, Zsiga (2000) envisions a model with two constraint-based grammars. First, a phonological grammar which operates along the lines of classic Optimality Theory, manipulates phonological representations (in terms of, e.g., features, segments, and prosodic units). The winning candidate from this grammar then serves as the input to the next grammar: a phonetic component. In this phonetic component, the output from phonology (now the input to phonetics) is associated with a set of candidates which consist of phonetic realizations (in terms of gestural targets). Constraints in the phonetic component can require alignment of gestural targets, or specify goals such as "be distinct" or "conserve energy".

Flemming's (2001) proposal is conceptually similar to this model, with the major distinction that the phonetic and phonological components are merged. Constraints in this joint phonetics/phonology module can refer to both phonological structures and raw phonetic values (e.g., milliseconds, Hertz, etc.). For example, in order to model cross-linguistic differences in coarticulatory effects, Flemming's constraint *F2(C)=L* imposes a cost for every Hertz different between the F2 of a consonant and its F2 target, *L*. When paired with a constraint *F2(V)=T*, which penalizes vowels that fail to reach their F2 targets, and a constraint *F2(C)=F2(V)*, which prefers abutting CV segments to have an identical F2 (in order to minimize effort), a tension is set up between the consonant and vowel in a CV string hitting their respective F2 targets vs. the segments assimilating in F2. Varying the weight of these constraints determines how the conflict will be resolved.

As Flemming himself points out (p. 16), this constraint-based unified model of phonetics and phonology is conceptually similar to optimality-theoretic phonology (Prince and Smolensky 1993) in that the output is the candidate which best satisfies conflicting violable constraints. The model differs from OT, however, in two crucial ways. First, the constraints refer to representations of both phonological *and* phonetic detail. Second, constraint conflict is adjudicated by constraint *weighting* rather than a strict-dominance ranking system (e.g., Harmonic Grammar: Legendre et al. 1990, Pater 2009).

The major benefit of weighted constraints over strictly-dominating ranked constraints is that they can generate *compromise*. With ranked constraints, conflict is resolved by acceding to the demands of the highest-ranked constraint. Under weighted constraints, neither conflicting constraint must be satisfied absolutely to yield an optimal output. The candidate which represents a compromise between two highly-weighted constraints may yield a lower overall cost than a candidate which follows completely the demands of any one constraint. This attribute of weighted constraint systems, as argued by Zsiga (2000, see especially pp. 96-97), makes them optimal in the analysis of gradient phenomena. (For arguments in favor of weighted constraints specifically with respect to the constraints proposed in this chapter, see Section 4.2.1.)

I take a middle path between Zsiga (2000) and Flemming (2001). Specifically, I assume a two-grammar model along the lines of Zsiga (2000), in which an OT-like grammar applies phonological processes first, followed by a weighted-constraint phonetic grammar which takes the output of the phonology as its input. I borrow from Flemming (2001), however, the idea of reference to raw phonetic details—in particular raw phonetic duration—in computing cost. As such, many of the constraints defined below are similar in spirit to the constraints like *F2(C)=L* and *F2(V)=T* described briefly above, in which some phonetic property of a segment in a candidate is penalized for its distance from some target.

4.1.2 Transderivational Identity and Paradigm Uniformity

The second major component of this model comes from Benua's (1997) Transderivational Identity (also known as Output-Output-, or OO-correspondence), which argues that morphologically-related words share a special relationship. Specifically, pairs of surface forms enter into OO-correspondence relationships, and are compelled by OO-faithfulness constraints to be similar to one another. This requirement to be faithful to morphologically related forms, then, can force deviations from language-wide phonological patterns. As Benua (1997, p. vi) puts it, "When OO-correspondence constraints take precedence, phonology misapplies," favoring a paradigm-specific pattern over a language-general one.

Note that Benua's (1997) theory of transderivational identity was originally intended to apply to categorical, phonological, patterns. That is, OO-correspondence constraints were defined over phonological representations — not phonetic ones, and constraint conflict was assessed by strict domination—not constraint weighting. I argue that these original intentions should be expanded.

First, as Steriade (2000) has argued, words are sometimes pressured to be faithful to morphologically related forms—even in terms of phonetic values such as duration.

Steriade (2000) models this pressure in terms of Paradigm Uniformity (PU) constraints (in a classic, ranked constraint, OT model). These PU constraints require a given segment to be, e.g., 'durationally equivalent' to their corresponding segment in a morphologically related form, as in (1), slightly modified from Steriade (2000):

(1) PU (DURATION)

If two consonants, C and C′, stand in correspondence and C is morpheme initial in the careful pronunciation of the relevant morpheme, C′ is durationally equivalent to C.

In the same way that classic, categorical OO-correspondence requires morphologically related forms to deviate from language-wide phonological patterns, gradient OO-correspondence can nudge morphologically related forms closer to one another phonetically.

This brings us to the second divergence from Benua's (1997) classical conception of OO-correspondence: weighted constraints. Note, that under a ranked constraint model such as classic OT, constraints such as the one in (1) cannot compromise with other pressures in the system. As argued above in Section 4.1.1, and below in Section 4.2.1, such compromise is crucial: the interplay between hitting a target and transderivational identity/paradigm uniformity is precisely what allows for incompletely neutralized contrasts.

4.1.3 Basehood and Canon

In transderivational identity, the two outputs in correspondence are in an asymmetric relationship with one another. The first of these two outputs is the candidate under evaluation; the second is the *base*—the output form to which the candidate corresponds. In other words, the base serves as a sort of target which the candidate strives to meet.

Different authors have given different definitions of basehood, but in the main, they come back to the idea that a base is in some sense a canonical word, as can be seen in the definitions in (2)—(4):

(2) Kager (1999, p. 282): "Definition of Base: (a) the base is a free-standing output form —a *word.* (b) The base contains a *subset of the grammatical features* of the derived form."

(3) Benua (1997, p. 29): "The base of an OO correspondence relation is *a licit output word*, which is both morphologically and phonologically well formed."

(4) Kenstowicz (1996, p. 8): "Base-Identity: Given an input structure [X Y] output candidates are evaluated for how well they match [X] and [Y] if the latter *occur as independent words*."

A second commonality among most conceptions of basehood is the notion of Base Priority, as defined by Benua (1997) in (5):

(5) Base Priority: "Transderivational identity relations are asymmetrical, in that the derived word can mimic the base, but the base cannot mimic the phonology of the derived word… The base never deviates from canonical patterns in order to mimic its derived counterpart…" (Benua 1997, p. 240)

When Base Priority is combined with stipulations that a base be a subset or substring of a candidate (such as in the defintions of basehood provided by Kager 1999 in (2) and Kenstowicz 1996 in (4)), morphologically simple forms are predicted to serve as bases for more complex ones. Steriade (2013) points out that phonology generally abides by the classic conception of Base Priority; however, she argues that this is not the whole picture.

More specifically, Steriade (2013) argues that basehood may be determined, at least in part, by frequency (typewise or tokenwise). Further, she claims that more frequent forms are better known, therefore more resistant to modification, and less likely to require on-line computation. As such, this idea can be thought of as “a grammatical reflex of strategies for generating forms that are not well known, with assistance from related but better known ones” (Steriade 2013, March 5, p. 2).[2]

This idea actually has its roots nearly half a century earlier. Mańczak (1958) argued that less frequent forms are in some cases analogized to more frequent ones. While Steriade (2013) allows that basehood may be determined by either type or token frequency, Mańczak (1980, p. 284–285) argues that of more frequent *forms* serve as the target of analogical leveling, implying token frequency only. The model proposed here makes use only of type frequency; however this is due largely to the data being considered—I see no reason the model could not accommodate basehood based on token frequency instead.

4.1.3.1 Defining and Evaluating Bases

To summarize the spirit of this idea, then, a base is in some way a more *canonical* version of the candidate. For the purposes of this model, the type of canon to which bases belong is that of type frequency. Specifically, as laid out in (6), the base of a given candidate will belong to the most frequent type within that candidate’s inflectional paradigm. (It must also, of course, have the same root as the candidate.)

(6) **Basehood**

- For a candidate α
  - $\alpha \in$ inflectional paradigm $\mathbb{P}$
  - MostFreq($\mathbb{P}$) is the most frequent type within $\mathbb{P}$
  - $\sqrt{\alpha}$ is the root of α
  - Base(α) is of type MostFreq($\mathbb{P}$), and $\sqrt{\text{Base}(\alpha)} = \sqrt{\alpha}$

Put more simply, a candidate’s base is the form that results from applying the most

[2] As Steriade (2013) herself points out, Albright (2002, *et seq.*) shows that in some cases, frequency alone is not sufficient to identify the base.

frequent (typewise) member of the candidate's inflectional paradigm to the candidate's root. For example, given the English noun *penguins*, and assuming that the most frequent type of inflection in the nominal paradigm is the phonologically null nominative singular, then the base of *penguins* is *penguin*.

This frequency-based model of basehood presents a distinct, empirical advantage over those based in (morphological) simplicity: frequent, but morphologically complex forms sometimes serve as bases. As is discussed in more detail in Section 4.2, this is the case in Japanese monomoraic lengthening. Monomoraic nouns with no case particle (e.g., *ki* 'tree') lengthen, but do not reach the same length as underlyingly long nouns (e.g., *kii* 'key'), due to faithfulness to a base which contains a monomoraic noun followed by a case particle (e.g., *ki ga* 'tree NOM'). Given that this base contains two morphemes, as compared to the candidate monomoraic noun's single morpheme, morphological simplicity cannot be the whole story.

Further, I adopt, with a few modifications, **?**'s (?)pp. 33–39]benua-diss Recursive Evaluation analysis of paradigms. Specifically, I assume that at the speech time, the speaker first determines which underlying form should serve as the base, then applies the language's canonical phonology and phonetics to that form. At this point, the candidate is evaluated with respect to this freshly-minted base. This view has the benefit of neatly accounting for paradigm uniformity effects in nonce words: given a nonce word for which the speaker can work out an underlying representation, the speaker need only apply the language's canonical phonology and phonetics to this form, and evaluate the nonce candidate with respect to the generated base. No knowledge of word-specific frequency is needed to generate the base, as type frequency is computed over the entire inflectional paradigm of the candidate (e.g., all nouns, all verbs, etc.). Under a theory of basehood in which the phonetic detail of a base is pre-generated, nonce words are predicted to fail, since the base will not be available at the time of evaluation. This is counter to the evidence—speakers are able to produce incompletely neutralized forms, which implies faithfulness to a base.

4.1.3.2 Canon-hood

Along similar lines, the model sets targets for various phonetic properties of a segment, e.g., vowel duration, based on the canonical realization of that segment. The canonical realization of a segment will be taken to be the realization of that segment in the phonological environment in which it appears most frequently. This idea is formalized in (7):

(7) **Canon-hood**

- For a segment φ
  - MostFreq(φ) is the most frequent type of conditioning environment in which φ appears
  - $\mathbb{R}$(MostFreq(φ), φ) is the realization of φ in MostFreq(φ)
  - The canonical realization of φ is $\mathbb{R}$(MostFreq(φ), φ)

4.1.3.3 Correspondence within the Paradigm

As discussed above, Output-Output correspondence models require that output forms within a paradigm share relations with one another; various constraints on these relations can then be enforced via constraints of a type familiar from Input-Output correspondence relationships. I assume that for a given triplet <input, candidate, base>, there are three types of relations. First, there are Input-Output (IO) correspondence relations between the input and the candidate ($\mathfrak{R}_{IO(cand)}$). The second type of relation is also one between an input and an output: $\mathfrak{R}_{IO(base)}$ is the relation between an input and its base — a fully computed form. Finally, the candidate and the base share Output-Output (OO) correspondence with one another in the $\mathfrak{R}_{OO}$ relationship.

I will assume here that $\mathfrak{R}_{IO(cand)}$, $\mathfrak{R}_{IO(base)}$, and $\mathfrak{R}_{OO}$, are defined over segments, and that a segment has at most one relation of each type. Further, I assume that these correspondence relations are order-preserving (though not necessarily bijective: insertion/deletion is permitted). Given these assumptions, I do not consider here the implications of metathesis or coalescence.

These relations are shown schematically in Figure 4.1. In this hypothetical example, the segments of input /pata/, candidate [pada], and base form [past] correspond via the three relations under discussion. Beginning with $\mathfrak{R}_{IO(cand)}$, note that every segment in the input has exactly one correspondent in the candidate, and vice versa. This is not the only possible situation, however, as is evident from the $\mathfrak{R}_{IO(base)}$ relation: the final segment of the input has no correspondent in the base, and the penultimate segment in the base has no correspondent in the input. As is familiar from canonical IO correspondence (e.g. McCarthy and Prince 1995), the lack of correspondence in these cases represents deletion of the final vowel and insertion of an [s] into the base.

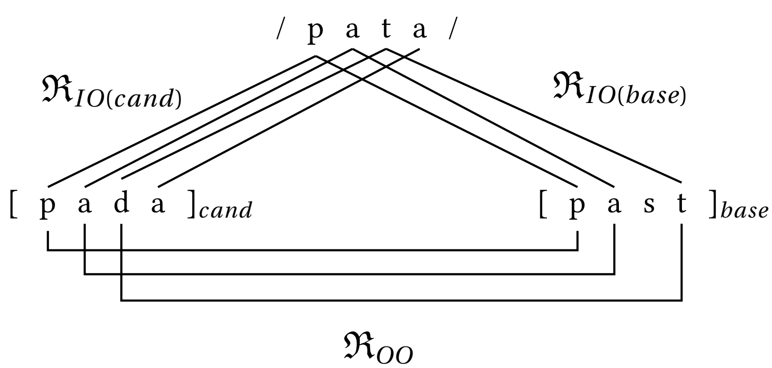


Figure 4.1: Schematic showing $\mathfrak{R}_{IO(cand)}$, $\mathfrak{R}_{IO(base)}$, and $\mathfrak{R}_{OO}$ for hypothetical input /pata/, candidate [pada], and base [past].

The $\mathfrak{R}_{OO}$ relation functions in a similar fashion. Note that neither the final segment of the candidate nor the penultimate segment of the base are members of a $\mathfrak{R}_{OO}$ relation. Intuitively, we see that this is because the candidate-final [a] has been deleted from the base, and the base-penultimate [s] was never inserted into the candidate. To cast this idea in a slightly more formal notion, a segment in the candidate will have a correspondent in the base if and only if it has a correspondent in the input, *and* that input correspondent has a correspondent in the base. Similarly, a segment in the base will have a correspondent in the candidate if and only if it has a correspondent in the input, *and* that input correspondent has a correspondent in the candidate. This notion, the Correspondent Transitivity Rule, will become crucial for constraints that enforce similarity between corresponding segments.

It should be noted at this point that this view represents a mirror image of that described by Tesar (in press, ch. 2) for output-drivenness. Under such a view, the crucial relationship is not Output-Output correspondence, but rather Input-Input (II) correspondence. II-correspondence is not a part of the grammar per se, but rather allows the analyst to examine the relative similarity of two different candidates which have different inputs but identical outputs. This is the inverse of OO-correspondence, in which the grammar employs a relation between two different outputs (i.e., the candidate and the base), which share a single input.

4.1.3.4 Relation to Prototype and Exemplar Models

The notions of a canonical realization and a base determined by frequency are, in many ways, reminiscent of prototype models (Posner and Keele 1968, Reed 1972, Rosch 1973, Smith and Minda 2000) and exemplar models (Nosofsky 1986, 1990, Lacerda 1995, Goldinger 1996, Johnson 1997, Pierrehumbert 2001, Wedel 2004, 2006). Such models are useful when thinking about categorization in terms of typicality or canonhood.

To take an example from Kirby (2010), imagine that we encounter a previously unknown creature, and we wish to categorize it as a bird or a non-bird. One approach would be to list features inherent to birds, such that the category BIRD would be defined by the set of features [+WINGS, +FEATHERS, +LAYS-EGGS, +BEAK, +CAN-FLY ...]. Problems with this approach are apparent as soon as we observe an animal that has many, but not all, of these features. Penguins, for instance, lay eggs, have wings, feathers, and beaks, but cannot fly. If we decide that penguins should, nonetheless be classified as birds, this would seem to indicate that penguins are somehow 'less bird' than other 'real' birds. Prototype and exemplar models seek to define precisely what it means to be a typical, more canonical, instance of a category.

In prototype models (Posner and Keele 1968, Reed 1972, Rosch 1973, Smith and Minda 2000), a single category-member is selected as the prototype—usually the most 'typical' member of the category, or sometimes an average of all category members. New stimuli are categorized based on their distance from the various category prototypes: in some models this distance is computed as the number of shared features, while in others the distance between a stimulus and each prototype is calculated along the dimensions of continuous data.

Exemplar models (Nosofsky 1986, 1990, Lacerda 1995, Goldinger 1996, Johnson 1997, Pierrehumbert 2001, Wedel 2004, 2006) do not pick out a single, prototypical member of each category. Instead, each category is represented by a set of labeled exemplars (a 'cloud'). Categorization is based on comparison with previously categorized exemplars from all categories (with a possible category bias). To return to the bird/non-bird categorization problem, all previously categorized birds would be available for comparison —even those that may not have had the full set of bird-like features. Should we have previously encountered an ostrich, and categorized it as a bird, this exemplar would be available for comparison during the categorization of penguin. Since ostriches and penguins share similar features (e.g., the bird-like features, minus the ability to fly), they are likely to be categorized together. In other words, even if an existing exemplar is not the most typical example of the category, a new, non-typical observation that shares traits with the existing exemplar is likely to be correctly classified.

Such models (and their close kin) have been applied to the study of incomplete neutralization and related phenomena. Notably, Yu (2007b) provides an analysis of a case of near merger in Cantonese tone in terms of an exemplar-based model. Under this view, near merger occurs when two or more exemplar clouds begin to overlap, but category membership still remains distinct.

Using a computational model with a number of similarities to exemplar models, Kirby (2010) analyzed the case of incomplete neutralization in Dutch word-final obstruents described by Warner et al. (2004). This work showed that when only a single acoustic cue is considered, complete neutralization may be predicted, but that when more cues are added to the model, this prediction is often reversed. In other words, a given cue that predicts complete neutralization on its own does not necessarily do so when other cues are involved as well.

Further evidence that typicality might play a role in incomplete neutralization comes from an experiment on final devoicing in German conducted by Röttger et al. (2012). Participants were presented auditorily with nonce stimuli in the plural form—which preserves the underlying voicing distinction (e.g. Gope/Gobe), and were then asked to produce the stimuli in the singular form, where devoicing occurs (e.g. Gop/Gob). The auditory stimuli, however, were manipulated: the duration of the vowel preceding the stop was either enhanced, kept as normal, completely neutralized or reversed. All conditions resulted in a positive incomplete neutralization effect, however, there was a significant interaction between underlying voicing and manipulation condition. One interpretation of this data is a lexical co-activation effect, though it is also conceivable that the true cause of the interaction is imitation of the stimuli or phonetic accommodation (e.g., Babel 2009).

While the model developed in this chapter is not formally related to either prototype models or exemplar models, it does share the idea of 'typicality', as enshrined in the notions of 'canon-hood' and 'basehood' as defined in (7) and (6) above, respectively. An incompletely neutralized form is pressured to become more like a canonical instance of the category (and in that way is more like a prototype model than an exemplar model), but also receives pressure to neutralize.

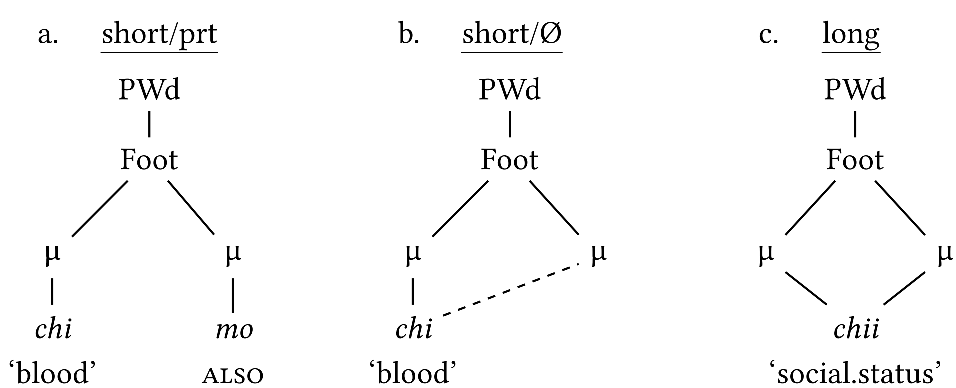
4.1.4 Exordium

That the word ‘model(ing)’ appears so frequently within the first few paragraphs of this chapter is not an accident. The goal of the theory proposed here is to model the actual phonetic output of incomplete neutralization—in other words, to capture real-world and experimental speech data in the terms of this framework. Such data being central to the model, the sections that follow make use of the experiments on incomplete neutralization presented in Chapters 2 and 3 to fill in the details of this proposal.

4.2 Japanese

In Chapter 3, I presented experiments on the case of incomplete neutralization in Japanese monomoraic vowel lengthening. Japanese has a requirement that feet (and hence, PWds) be bimoraic (Itô 1990, Poser 1990). In spite of this requirement, Japanese has monomoraic nouns. When a monomoraic noun is followed by a particle (which also contributes a mora), as in (8a), the bimoraicity requirement is satisfied. Similarly, underlyingly bimoraic nouns, as in (8c) satisfy the requirement by virtue of having two moras. Monomoraic nouns, when not followed by a case particle , as in (8b), lengthen (Mori 2002).

(8)



In both Experiment 6 (Section 3.3) and Experiment 7 (Section 3.4), vowels in monomoraic nouns without case particles (short/Ø condition) were found to lengthen (i.e., they were longer than those in identical monomoraic nouns followed by case particles—the short/prt condition). Further, in both experiments, short/Ø nouns' vowels were not as long as underlyingly bimoraic nouns' (long condition). These results are summarized in Table 4.1.

	Experiment 6		Experiment 7	
	Mean	SD	Mean	SD
short/prt	73.54	20.58	54.99	21.89
short/Ø	119.19	32.56	124.98	34.91
long	145.74	31.21	157.45	39.21

Table 4.1: Mean and standard deviation of vowel durations (in ms) for each category in Experiments 6 and 7

As discussed in Section 4.1.3, the goal of this chapter is to model actual speech data. For the moment, let us adopt the data from Experiment 7 (with some rather heavy-handed rounding for ease of explication), summarized in Table 4.2.[3] Note that a key feature of the unrounded data—that the lengthened short/Ø vowel durations are closer to long vowel durations than short/prt vowel durations—is preserved in this rounded data.

	Target Data
short/prt	50
short/Ø	125
long	150

Table 4.2: Data to be modeled: Average vowel durations (in ms), based on an idealized version of the figures from Experiment 7.

Let us further assume that there is some language-specific target duration for any given underlying segment. Let us also assume that all segments which bear one mora in Japanese have a target duration of *TargetDur(μ)*, and all segments which bear two moras have a target duration of *TargetDur(μμ)*. From the data under analysis in Table 4.2, we know that bimoraic nouns have an average vowel duration of 150 ms, and that short nouns

[3] The analysis presented here applied to the data directly from Experiment 7, as opposed to the rounded data in Table 4.2, is provided in Appendix D.

with particles have an average vowel duration of 50 ms. Given that we have no evidence of an alternation in duration with long vowels, let us assume that the long vowel target is 150 ms, and that it has, on average, been met. Further, short/prt vowels face no pressure to lengthen, since their particle provides the second mora necessary to fulfill the Japanese bimoraicity requirement. Given this fact, and that we have no evidence suggesting that short/prt vowels shorten, let us assume that the target for short vowels in Japanese is 50 ms, and that, on average, short/prt vowels meet this target. Therefore, for Japanese, *TargetDur(μ)*=50 ms and *TargetDur(μμ)*=150 ms.

While Japanese is a mora-timing language (see Warner and Arai 2001, and references therein), and therefore mora duration is relatively stable across segment type, it should be acknowledged that segments may have different inherent durations and may be influenced by their environment: for example, vowel quality impacts duration (House 1961), and high-pitch vowels are slightly longer than low pitch vowels in Japanese (Hoequist 1983), among other contextual factors (e.g. Klatt 1973). (See also Lehiste 1970 for discussion of intrinsic factors affecting vowel duration, and Hirata 2004 on the effect of speech rate.) I abstract away from these details, though see no reason they should disrupt the analysis presented here.

Candidates are pressured to conform to the language-specific duration targets by two constraints, both of the family I term Dur(*x*)=TargetDur(*x*) (see Flemming's 2001 C-duration and σ-duration constraints). Dur(μ)=TargetDur(μ), in (9), prefers monomoraic segments to match their target duration; Dur(μμ)=TargetDur(μμ), in (12) prefers bimoraic segments to match their target duration. First, consider the definition of Dur(μ)=TargetDur(μ):

(9) **Dur(μ)=TargetDur(μ)**

In brief: The duration of a mora-bearing unit in the candidate, which bears a single mora in the output, should match the target (canonical) output duration of that mora-bearing unit (when it bears one mora) in the language at large.

Definition:[4]

- For a mora-bearing unit α which bears one mora in the output, and is spoken at speech rate $\mathfrak{R}$, let:
  - *TargetDur(μ)* be the canonical output duration of α when bearing one mora in the output, and spoken at speech rate $\mathfrak{R}$
  - *Dur(μ)* be the actual duration of α in the candidate under evaluation
  - w_μ be the weight applied to this constraint
- $cost = w_\mu(\mathit{TargetDur}(\mu) - \mathit{Dur}(\mu))^2$

To see DUR(μ)=TARGETDUR(μ) in action, consider the tableau in (10) (which assumes a constraint weight w_μ=1). Under the DUR(μ)=TARGETDUR(μ) column, the large, black numbers show the cost of each candidate for that constraint; the small gray numbers show the calculation used to reach that cost. For example, in candidate (10)a, the cost associated with DUR(μ)=TARGETDUR(μ) is calculated as in (11), where the candidate's duration is *Dur(μ)*=30, and the target duration for a single mora (as per our idealized data) is *TargetDur(μ)*=50.

[4]The cost function in (9) (and for the following constraints) squares the difference between the target and the actual duration in order to create a quadratic equation. Doing so makes it possible to mathematically compute the minimum value for a cost function. (To visualize this, recall that a quadratic equation results in a parabola. Finding the minimum possible cost is equivalent to finding the bottom of the parabola. Without squaring this term, the graph would be a straight line—which does not have a bottom.)

(10) Cost of Dur(μ)=TargetDur(μ) for various candidates (for w_{μ}=1)

	ki mo nakushita yo (short/prt)	Dur(μ)=TargetDur(μ)
a.	V dur = 30 ms	400 $1(50-30)^2$
b.	V dur = 40 ms	100 $1(50-40)^2$
c. ☞	V dur = 50 ms	0 $1(50-50)^2$
d.	V dur = 60 ms	100 $1(50-60)^2$
e.	V dur = 70 ms	400 $1(50-70)^2$

(11) $$cost = w_{\mu}(TargetDur(\mu) - Dur(\mu))^2$$
$$= 1(50-30)^2$$
$$= 400$$

As can be seen from the tableau in (10), Dur(μ)=TargetDur(μ) prefers short/prt candidates whose vowel duration is as near to the short vowel target (50 ms) as possible; as candidates' vowel durations diverge from this target, cost increases.

Dur(μμ)=TargetDur(μμ) works in a nearly identical fashion, as defined in (12), and exemplified in (13).

(12) **Dur(μμ)=TargetDur(μμ)**

In brief: The duration of a mora-bearing unit in the candidate, which bears two moras in the output, should match the target (canonical) output duration of that mora-bearing unit (when it bears two moras) in the language at large.

Definition:

- For a mora-bearing unit ℵ which bears two moras in the output, and is spoken at speech rate 𝔑, let:
  - *TargetDur(μμ)* be the canonical output duration of ℵ when bearing two moras in the output, and spoken at speech rate 𝔑

- $Dur(\mu\mu)$ be the actual duration of ℵ in the candidate under evaluation
- $w_{\mu\mu}$ be the weight applied to this constraint

- $cost = w_{\mu\mu}(TargetDur(\mu\mu) - Dur(\mu\mu))^2$

(13) Cost of Dur(μμ)=TargetDur(μμ) for various candidates (for $w_{\mu\mu}$=1)

			Dur(μμ)=TargetDur(μμ)
kii mo nakushita yo (long)	a.	V dur = 130 ms	400 $1(150-130)^2$
	b.	V dur = 140 ms	100 $1(150-140)^2$
	c. ☞	V dur = 150 ms	0 $1(150-150)^2$
	d.	V dur = 160 ms	100 $1(150-160)^2$
	e.	V dur = 170 ms	400 $1(150-170)^2$
ki Ø nakushita yo (short/Ø)	f.	V dur = 130 ms	400 $1(150-130)^2$
	g.	V dur = 140 ms	100 $1(150-140)^2$
	h. ☞	V dur = 150 ms	0 $1(150-150)^2$
	i.	V dur = 160 ms	100 $1(150-160)^2$
	j.	V dur = 170 ms	400 $1(150-170)^2$

The tableau in (13) shows that as surface-bimoraic vowels (both short/Ø and long categories) approach a duration of 150 ms, their cost is minimized.

It is clear, then, how short/prt and long vowels are pressured to meet their targets. We are left, though, to deal with the short/Ø nouns' vowels (mean: 125 ms), which meet neither the target for short nouns (50 ms) nor the target for long nouns (150 ms). Phonologically speaking, short/Ø nouns have two moras on the surface (as argued in Chapter 3); however the vowels in these nouns do not comply entirely with the preference of Dur(μμ)=TargetDur(μμ). I argue that output-output correspondence pressures short/Ø vowels to maintain a degree of similarity to canonical short vowels. This accounts for the difference between short/Ø vowels and long vowels, which are both phonologically bimoraic on the surface, but which differ in their realized phonetic duration.

To see how this works, let us first work with a single example extrapolated from the idealized data, and then proceed to formalism. Let us assume the three utterances in (14), from the short/prt, short/Ø, and long conditions, respectively, with vowel durations of the underlined words in parentheses:

(14)	a.	ki	mo	nakushita	yo		(50 ms)
		tree	ALSO	lost	DISC		
	b.	ki	Ø	nakushita	yo		(125 ms)
		tree		lost	DISC		
	c.	kii	Ø	nakushita	yo		(150 ms)
		key	ALSO	lost	DISC		

The target duration for the vowel in *ki(+mo)* (14a), which has one mora, is 50 ms; the target for *kii* in (14c), which has two moras, is 150 ms. *ki(+Ø)* in (14b) has two moras in the output, and therefore is motivated by Dur(μμ)=TargetDur(μμ) to hit a target duration of 150 ms. Why, then, does *ki(+Ø)* in (14b) not reach this target? Recall from Chapter 3 that, generally speaking, nouns in Japanese are followed by case particles. That is to say that sentences of the type in (14a) are more common than sentence of the type in (14b), and thus when speakers have heard *ki*, it has usually been in a context which includes a second mora in its prosodic word. Most of the time, then, speakers hear a short/prt *ki(+mo)* (or *ki*(+ga), or *ki* with some other particle) with a target duration *and* an actual duration of 50 ms. I argue that speakers are therefore faithful to this canonical *ki* vowel duration of 50 ms, even in cases like the *ki(+Ø)* in (14b), which has two moras on the surface (this idea should be familiar from Section 4.1.3 and references therein).

This faithfulness to the canonical duration of a segment is enforced by an output-output constraint, OO-ID-Dur. More specifically, OO-ID-Dur enforces faithfulness between a given segment in a candidate and the corresponding segment in the candidate's base, where the base is the form that has the same root as the candidate, and is a member of the most frequent type within the candidate's inflectional paradigm (as defined in (6)).

This constraint is spelled out in (15):

(15) **OO-ID-Dur**

In brief: The duration of a segment in the candidate should be faithful to the duration of the same segment in the base—the most frequent type in the candidate's inflectional paradigm as applied to the candidate's root.

Definition:

- For a segment α in the candidate, let:
  - *Dur(Cand)* be the duration of segment α in the output
  - *Dur(Base)* be the duration of the segment β in the base such that $(\alpha, \beta) \in \mathfrak{R}_{OO}$, where:
    - the base is the form that has the same root as α, and is a member of the most frequent type within the candidate's inflectional paradigm (as defined in (6))
    - if $(\alpha, \beta) \notin \mathfrak{R}_{OO}$ (determined via the Correspondence Transitivity Rule, §4.1.3.3), then this constraint is vacuously satisfied with a total cost of zero
  - w_{ID} be the weight applied to this constraint
- $cost = w_{\mathrm{ID}}(Dur(Cand) - Dur(Base))^2$

The base form to which this constraint requires allegiance is, as per (6), the form which has the same root as the candidate and is of the same type as the most frequent type in the candidate's inflectional paradigm. Continuing with our example from (14), if our candidate is *ki(+Ø)*, as in the tableau in (16), the most frequent type within the nominal/case paradigm in Japanese is nouns with nominative case (*-ga*). The form which fits this type and which has the same root as the candidate is *ki(+ga)*. Since this form meets the bimoraicity requirement in Japanese without lengthening, the duration of the target vowel (*i*) is 50 ms.

(16) Cost of OO-ID-Dur for various candidates (for w_{ID}=1)

	ki Ø nakushita yo (short/Ø)	OO-ID-Dur
a.	V dur = 25 ms	625 $1(25-50)^2$
b. ☞	V dur = 50 ms	0 $1(50-50)^2$
c.	V dur = 75 ms	625 $1(75-50)^2$
d.	V dur = 100 ms	2,500 $1(100-50)^2$
e.	V dur = 125 ms	5,625 $1(125-50)^2$
f.	V dur = 150 ms	10,000 $1(150-50)^2$
g.	V dur = 175 ms	15,625 $1(175-50)^2$

As can be seen in the tableau in (16), OO-ID-Dur prefers short/Ø candidates with a vowel duration approaching 50 ms, and increasingly penalizes candidates as they diverge from that target.

One possible concern regarding OO-ID-Dur involves its role in the recursive evaluation of paradigms. Specifically, if bases are generated on the fly, and OO-ID-Dur is active during the computation of a base β, might β be influenced by *its* base β′? Given the definition of basehood in (6), this is not possible: the base of β will always be β itself. Recall that basehood is defined as the form that results from applying the most frequent member of the candidate's inflectional paradigm to the candidate's root. As such, any candidates that share a root, and are in the same inflectional paradigm, will share a unique base υ. When OO-ID-Dur is applied to our base β (which is now acting as a candidate), its base β′ will, by the definition of basehood, share a root, inflectional paradigm, and unique base υ with β. Since we know that β is also a base for this root and inflectional paradigm, it must be equivalent to the unique base υ. For the sake of concreteness, I assume that during the recursive evaluation of a paradigm, if a base's base is that base itself, the recursion halts, and that base is the one to which the candidate must be faithful.

Note also that a definition of basehood that relies on morphological similarity and/or simplicity (e.g., Benua 1997, Kager 1999, p. 282) fails to account for this data. Consider a short/Ø candidate *ki(+Ø)* in such a system. It is crucial that the base to which *ki(+Ø)* strive for identity be *ki(+ga)*, since this form surfaces with the canonical short vowel duration of 50 ms. If basehood is determined solely by morphological simplicity, or is not explicitly determined by the model, there are arguably three possible outcomes. First, *ki(+Ø)* could simply not have a base, since there is no simpler form in the paradigm. Second, and perhaps more implausibly, *ki(+Ø)* could serve as its own base (though it is unclear what sort of predictions this type of recursive basehood would make). Third, if making the assumption that *ki* is morphologically more simple than *ki(+Ø)*, then the base of *ki(+Ø)* could indeed by the noun *ki* by itself. This option is problematic, though, since *ki* never surfaces without a second mora's worth of phonological content, as in *ki(+ga)*, so the target duration to which the candidate would strive to be faithful is unclear.

With the constraint definitions now settled, let us now consider their interaction. First, it should be noted that Dur(μ)=TargetDur(μ) is irrelevant for assessing short/Ø inputs. This is because Dur(μ)=TargetDur(μ) only penalizes mora-bearing units with a single mora. Since the only short/Ø candidates that are able to survive the phonological module have two moras, Dur(μ)=TargetDur(μ) does not apply. As such, it will not be shown in the tableaux and calculations that follow.

In essence, OO-ID-Dur is least costly for short/Ø candidates that are closer to the target for short vowels, while Dur(μμ)=TargetDur(μμ) is least costly for short/Ø candidates that are closer to the target for long vowels. Since the base for short/Ø candidates, as discussed above, is in essence their short/prt counterpart, OO-ID-Dur pressures short/Ø candidates towards 50 ms; since *TargetDur(μμ)*=150 ms, Dur(μμ)=TargetDur(μμ) pressures short/Ø candidates towards 150 ms. The relative importance of these constraints—as reflected in their weighting—determines how the conflict between these two pressures is resolved. As the weight of Dur(μμ)=TargetDur(μμ) increases relative to that of OO-ID-Dur, the lowest cost candidates move towards 150 ms. Conversely, as the weight of OO-ID-Dur increases relative to that of Dur(μμ)=TargetDur(μμ), the lowest cost candidates move towards 50 ms.

The total cost of a given candidate is the sum of the cost of all constraints; the winning candidate is the one which is the least costly overall. The constraints at play here are OO-ID-Dur and Dur(μμ)=TargetDur(μμ), so the total cost for a candidate is given in (17):

(17) $\textit{Total Cost} = \textit{cost}(\text{OO-ID-Dur}) + \textit{cost}(\text{Dur}(\mu\mu)=\text{TargetDur}(\mu\mu))$

Using the constraint definitions in (12) and (15), we can expand the cost terms of the equation:

(18) $\textit{Total Cost} = w_{\text{ID}}(\textit{Dur(Cand)}-\textit{Dur(Base)})^2 + w_{\mu\mu}(\textit{TargetDur}(\mu\mu)-\textit{Dur}(\mu\mu))^2$

As an example, consider (19), with weights (presciently) set at w_{ID}=1, $w_{\mu\mu}$=3, and a short/Ø candidate with a vowel duration of 125 ms. Working first with the OO-ID-Dur term, *Dur(Cand)* in this example is 125 ms (the duration of the short/Ø vowel), and *Dur(Base)* is 50 ms (the duration of the more frequent, short version of the noun). In the Dur(μμ)=TargetDur(μμ) term, *TargetDur(μμ)* is 150 ms (since this is the actual average long vowel duration), and *Dur(μμ)* is 125 ms (the duration of our candidate's lengthened short/Ø vowel, which has two moras):

(19)
$$\begin{aligned} \textit{Total Cost} &= w_{\text{ID}}(\textit{Dur(Cand)}-\textit{Dur(Base)})^2 &&+ w_{\mu\mu}(\textit{TargetDur}(\mu\mu)-\textit{Dur}(\mu\mu))^2 \\ &= 1(125-50)^2 &&+ 3(150-125)^2 \\ &= 1(5625) &&+ 3(625) \\ &= 7500 \end{aligned}$$

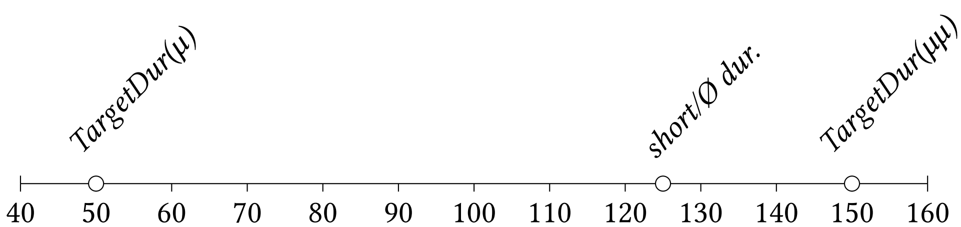
In this way, given a set of weights <w_{ID}, w_{μ}>, and a set of candidates, the cost for each candidate can be computed, and the least costly candidate can be selected as the output of the phonetic module. This is shown in Table 4.3, for weights <w_{ID}=1, w_{μ}=3>, and maintaining the assumption that *TargetDur(μ)*=50 ms, *TargetDur(μμ)*=150 ms, and that our desired short/Ø vowel duration is 125 ms.

Short/Ø dur.(ms)	*cost*(OO-ID-DUR) $w_{\text{ID}}(\textit{Dur(Cand)}-\textit{Dur(Base)})^2$	*cost*(DUR(μμ)=TARGETDUR(μμ)) $w_{\mu\mu}(\textit{TargetDur}(\mu\mu)-\textit{Dur}(\mu\mu))^2$	Total Cost
75.00	$1(75-50)^2$	$3(150-75)^2$	17,500.00
83.33	$1(83.33-50)^2$	$3(150-83.33)^2$	14,445.56
100.00	$1(100-50)^2$	$3(150-100)^2$	10,000.00
116.17	$1(116.17-50)^2$	$3(150-116.17)^2$	7,811.88
125.00	$1(125-50)^2$	$3(150-125)^2$	7,500.00
133.83	$1(133.83-50)^2$	$3(150-133.83)^2$	7,811.88
150.00	$1(150-50)^2$	$3(150-150)^2$	10,000.00

Table 4.3: Costs for given short/Ø vowel durations, where w_{ID}=1, $w_{\mu\mu}$=3, *TargetDur(μ)*=50ms, and *TargetDur(μμ)*=150ms

As is evident from Table 4.3, with the weighting <w_{ID}=1, $w_{\mu\mu}$=3>, as short/Ø vowel duration approaches 125 ms (our desired outcome), the total cost decreases. As candidate short/Ø vowel duration diverges from 125 ms—either longer or shorter—cost increases. As can be verified mathematically, a short/Ø duration of 125 ms is the minimum of the cost function with this weighting (see Appendix E for a generalized implementation in R code). The model has accurately produced the desired result: the lengthened short/Ø vowels are closer in duration to the target for long vowels than they are to the target for short vowels, as can be seen in the numberline in (20).[5]

(20)



[5]Note that this is not a violation of the Directionality Observation. In the contrast between short vowels and long vowels, the lengthened (short/Ø) vowels are intermediate in duration between the canonical realization of short/prt vowels and underlyingly long vowels.

Crucially, varying these constraint weights varies the minimum of the cost function. Consider Table 4.4, which maintains the same target durations as above. The weights of w_{ID} and w_{μ} are given in the first two columns, varying from 1–4. Given in the third column is the predicted short/Ø duration for that set of weights—that is, the short/Ø vowel duration which results in the minimum cost for that set of weights (see Appendix E for an implementation in R code).

w_{ID}	$w_{\mu\mu}$	Short/Ø duration (ms)
4	1	70.00
3	1	75.00
2	1	83.33
1	1	100.00
1	2	116.17
1	3	125.00
1	4	130.00

Table 4.4: Predicted short/Ø vowel durations for sample weightings of OO-ID-Dur (w_{ID}) and Dur(μμ)=TargetDur(μμ) ($w_{\mu\mu}$), where *TargetDur(μ)*=50 ms and *TargetDur(μμ)*=150 ms.

The patterns in Table 4.4 follow intuitive expectations: as the value of w_{ID} increases relative to $w_{\mu\mu}$, short/Ø duration decreases. Conversely, as the value of $w_{\mu\mu}$ increases relative to w_{ID}, short/Ø duration increases. Since OO-ID-Dur, in essence, provides pressure for short/Ø vowels to remain similar to their base (short/prt) counterpart, any increase in the relative value of w_{ID} should indeed lead to shorter vowels. Similarly, since Dur(μμ)=TargetDur(μμ) prefers that bimoraic short/Ø vowels meet the target of 150 ms, increased values of $w_{\mu\mu}$ should correspondingly increase the duration of short/Ø vowels.

4.2.1 Weighted vs. Ranked Constraints

We can now see the utility of weighted, rather than ranked constraints: their ability to make compromises. If a compromise can be reached between OO-ID-Dur, which prefers short/Ø *ki* to remain near 50 ms, and Dur(μμ)=TargetDur(μμ), which prefers short/Ø *ki* to reach a target of 150 ms, the model will generate an output with a vowel duration somewhere between 50–150 ms. The location along this duration continuum is determined by the relative importance (as determined by the relative weights) of the two constraints. This idea—that compromise is modeled as the relative weighting of constraints—can be seen in Figure 4.2. In this figure, the x- and y-axes represent the weights of the constraints OO-ID-Dur and Dur(μμ)=TargetDur(μμ), and the z-axis represents the predicted duration of the lengthened short/Ø vowels (computed as the length of short/Ø vowels which, for the set of $<w_{\text{ID}}=x,\ w_{\mu\mu}=y>$, results in a minimum cost). *TargetDur(μ)* was set to 50 ms, and *TargetDur(μμ)* was set to 150 ms, as per the data being modeled. A point $<x, y, z>$ on the graph, then, represents the duration z of short/Ø vowels where $w_{\text{ID}}=x$ and $w_{\mu\mu}=y$.

As can be surmised from Figure 4.2, for $w_{\text{ID}}=x$ and $w_{\mu\mu}=y$, where $0 < \{x, y\} \leq 5$, the model can generate a wide range of possible short/Ø vowel durations, depending on the constraint weights. It is also clear that for any given pair of weights, the model predicts a single short/Ø vowel duration.

Let us consider, briefly, how this model diverges from a strict-dominance weighted constraint model. If, under a model of ranked constraints like Optimality Theory (Prince and Smolensky 1993), OO-ID-Dur were most highly ranked, short/Ø *ki(+Ø)* would show no lengthening: while Dur(μμ)=TargetDur(μμ) pressures surface-bimoraic vowels to reach the target duration for long vowels of 150 ms, the canonical realization of short/Ø *ki(+Ø)* has a duration of 50 ms. Given the high ranking of OO-ID-Dur and the strict-dominance nature of the model in this exercise, this conflict would be resolved in favor of OO-ID-Dur: short/Ø *ki(+Ø)* would be faithful to its base, with a duration of 50 ms, at any expense.

This idea—that using these constraints in a strictly-dominating system leads to an "all or nothing" type of conflict resolution—is shown in the tableaux in (21) and (22), which places the constraints defined above in a ranked system. In (21), where OO-ID-Dur ≫

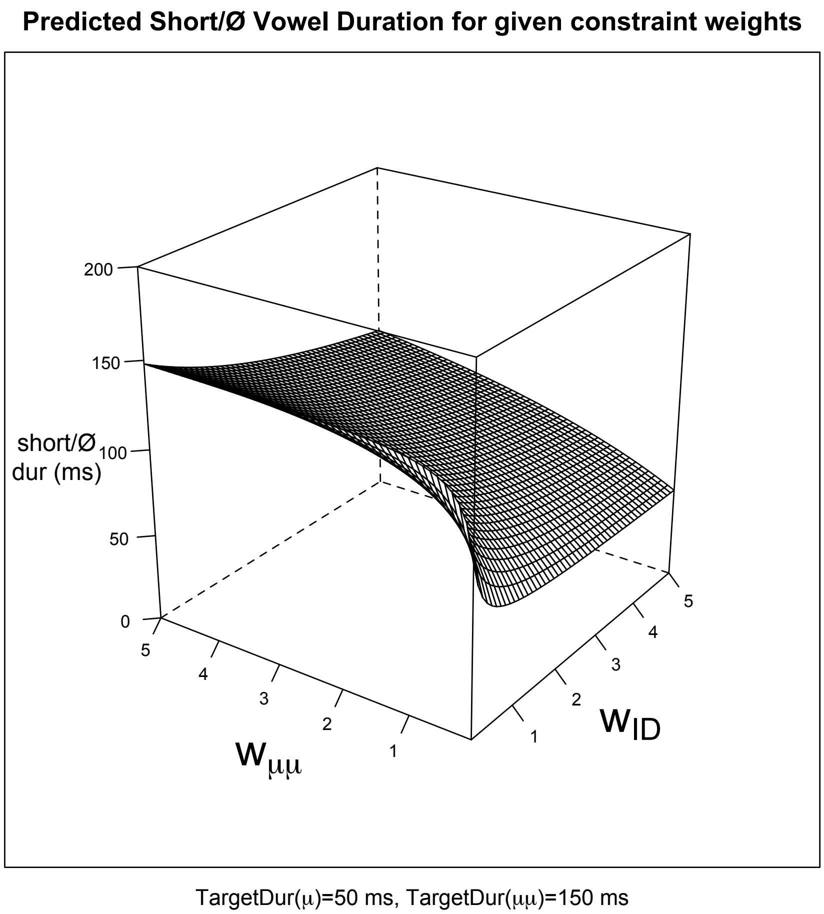


Figure 4.2: Experiment 7: Predicted short/Ø vowel duration for the idealized data, given weights for Dur(μμ)=TargetDur(μμ) and OO-ID-Dur. *TargetDur(μ)*=50 ms, and *TargetDur*(μμ)=150 ms.

Dur(μμ)=TargetDur(μμ), the desired candidate (marked with a frown glyph ☹), with a vowel duration of 125 ms, cannot win, because the winning (but undesired) candidate with a vowel duration of 50 ms (marked with a bomb glyph 💣) has fewer violations on the higher ranked constraint. (The number of constraint violations here are calculated in the same manner as *cost* in each constraint's definition, with the exceptions that (a) no weighting factor is applied, and (b) the result is not given as the square of the difference between a candidate and its target/base, but rather the absolute value of this difference.[6])

[6]The use of absolute value here, rather than squares, is purely for ease of explication and to keep violation counts manageable. In the weighted constraint system, squares are used to make the cost function

(21) OO-ID-Dur ≫ Dur(μμ)=TargetDur(μμ)

	ki Ø nakushita yo (short/Ø)	OO-ID-Dur	Dur(μμ)= TargetDur(μμ)
a.	V dur = 25 ms	25 (|25 – 50|)	125 (|150 – 25|)
b. 💣	V dur = 50 ms	(|50 – 50|)	100 (|150 – 50|)
c.	V dur = 75 ms	25 (|75 – 50|)	75 (|150 – 75|)
d.	V dur = 100 ms	50 (|100 – 50|)	50 (|150 – 100|)
e. ☹	V dur = 125 ms	75 (|125 – 50|)	25 (|150 – 125|)
f.	V dur = 150 ms	100 (|150 – 50|)	(|150 – 150|)
g.	V dur = 175 ms	125 (|175 – 50|)	25 (|150 – 175|)

The situation in fact, gets even more grim for the ranked approach to these constraints. The desired candidate (V dur = 125 ms) turns out to be collectively harmonically bounded by candidates (b) and (f) (Samek-Lodovici and Prince 1999)—it cannot win under any ranking. This can be seen in tableau (22), which has the reverse ranking of (21), and still does not select our desired optimum.

(22) Dur(μμ)=TargetDur(μμ) ≫ OO-ID-Dur

	ki Ø nakushita yo (short/Ø)	Dur(μμ)= TargetDur(μμ)	OO-ID-Dur
a.	V dur = 25 ms	125 (|150 – 25|)	25 (|25 – 50|)
b.	V dur = 50 ms	100 (|150 – 50|)	(|50 – 50|)
c.	V dur = 75 ms	75 (|150 – 75|)	25 (|75 – 50|)
d.	V dur = 100 ms	50 (|150 – 100|)	50 (|100 – 50|)
e. ☹	V dur = 125 ms	25 (|150 – 125|)	75 (|125 – 50|)
f. 💣	V dur = 150 ms	(|150 – 150|)	100 (|150 – 50|)
g.	V dur = 175 ms	25 (|150 – 175|)	125 (|175 – 50|)

quadratic, so that therefore a cost minimum can be found.

To summarize, then, the constraints outlined here do not produce the attested results when implemented in a ranked constraint system with strict-dominance.

4.3 Flapping

We now turn our attention to the second case of incomplete neutralization at the heart of this dissertation: flapping in American English. To briefly re-familiarize the reader with the basic facts, Table 4.5 shows the mean duration of vowels preceding /d/ flaps and those preceding /t/ flaps, from Experiments 1 and 2 (Sections 2.3 and 2.4, respectively). In both of these experiments, speakers produced nonce words containing flaps, which were based on nonce-roots ending in either /d/ or /t/, plus a vowel-initial suffix (e.g., *puhKEED~puhKEET* $\xrightarrow{+ing}$ puhKEE[ɾ]ing). As can be seen in Table 4.5, vowels preceding /d/ flaps were, on average, longer than those preceding /t/ flaps.

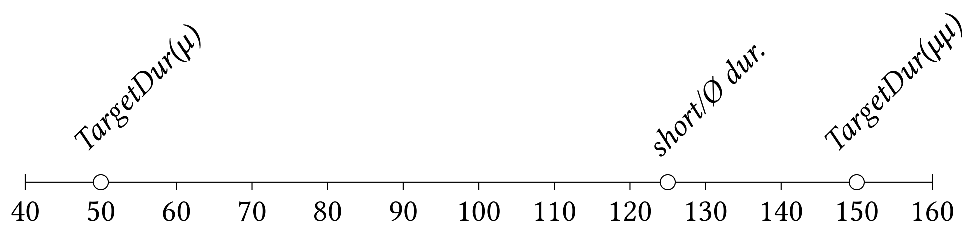
	Experiment 1		Experiment 2	
	Mean	SD	Mean	SD
Pre-/d/ flap	107.60	25.49	112.23	30.06
Pre-/t/ flap	102.51	24.94	106.54	26.75

Table 4.5: Mean and standard deviation of vowel durations (in ms) for each category in Experiments 1 and 2

One difference between this case of incomplete neutralization and the case of Japanese described above is the relatively small difference at hand. In Experiment 1, vowel duration differed, on average, by only 5.08 ms; in Experiment 2, the difference was only 5.69 ms. Given the results of the series of perception experiments in Chapter 2, listeners are unable to perceive the difference between /d/ flaps and /t/ flaps, even in ideal laboratory settings. While a perception study on the Japanese case remains to be pursued, the large difference between conditions (32.56 ms between long and short/Ø in Experiment 7) suggests that listeners may well be able to perceive the difference in Japanese. Regardless of listeners' performance on such a task, it is worth noting that incomplete neutralization can be 'more' incomplete (as in Japanese), or 'less' incomplete (as in American English flapping).

The keen reader will, at this point, recall that in the analysis of Japanese, it was shown that the model under discussion operates on a principle of compromise. In the case of Japanese, the compromise favored short/Ø candidates closer to the duration of canonical long vowels than canonical short vowels, but this preference was not overwhelming. This idea can be seen in the numberline in (23) (repeated from (20)). The actual duration of short/Ø vowels is closer to *TargetDur(μμ)* than it is to *TargetDur(μ)*, but it is not identical.

(23)



Before comparing this picture with flapping, let us first consider the parallel targets. Recall from Chapter 2 that the stimuli used in Experiments 1 and 2 consisted of disyllabic nonce words (e.g. *puhKEED*), which were then given a verbal suffix (e.g., *puhKEED-ing*). Two factors likely influence a difference between the stressed vowels in the base form of the nonce words from those in the inflected forms. First, vowels which precede flaps are generally shorter than vowels preceding other segments (Fox and Terbeek 1977). Second, the addition of a syllable to a word decreases the duration of syllable nuclei (Klatt 1973).[7] Let us assume that these phenomena together result in a 30% duration reduction of the nucleus of the stressed syllable in the (trisyllabic) flapped stimuli (e.g., puhKEED-ing) as

[7] Klatt (1973), following Lehiste (1972) and Barnwell (1971), shows that disyllabic nonce words have shorter stressed syllables than monosyllabic words. He further argues that there may be some interaction with the voicing of the following consonant. However for the sake of clarity, I make the simplified assumption above.

compared to their non-flapped (disyllabic) root forms (e.g., puhKEED).

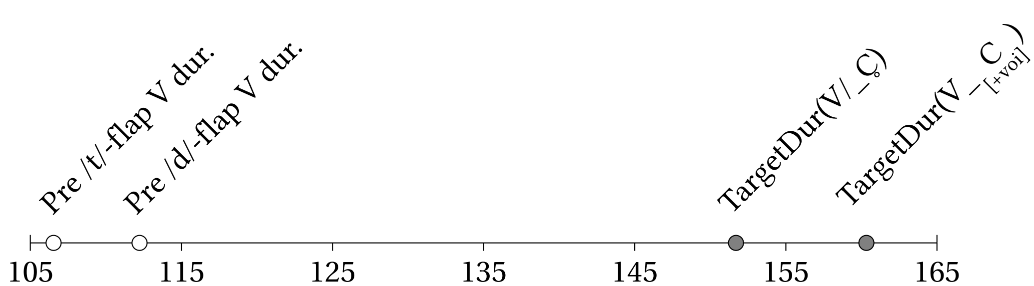
As non-flapped stimuli were not measured in Experiments 1 and 2, the target durations for vowels which precede voiced and voiceless segments in non-flapped contexts is computed based on the flapped tokens, and the assumption that flapped tokens are 30% shorter than non-flapped tokens. The mean duration of pre-/d/ flap tokens and pre-/t/ flap tokens from Experiment 2 were divided by 0.7 (to find the value of which they are reduced by 30%), resulting in non-flap context vowel duration targets of 160.33 ms for vowels preceding voiced segments, and 152.20 ms for vowels preceding voiceless segments. Given these targets, and the results from Experiment 2, the data to be modeled below is summarized in Table 4.6.

	Target Data
Pre-/d/ flap	112.23
Pre-/t/ flap	106.54
TargetDur(V/_$C_{[+voi]}$)	160.33
TargetDur(V/_C̥)	152.20

Table 4.6: Data to be modeled: Average vowel durations (in ms), based on Experiment 2, and extrapolated target durations for vowels in non-flapping contexts.

As can be seen from Table 4.6, and from the numberline in (24), the distinction between vowels preceding /d/ flaps and those preceding /t/ flaps is much smaller than the distinction modeled in the Japanese case (c.f., (20)).

(24)



To begin modeling this data, we begin by defining a set of constraints very similar in character to those seen previously. The first two constraints, much like DUR(μ)=TARGETDUR(μ) and DUR(μμ)=TARGETDUR(μμ) in the Japanese case, set target durations for certain kinds of segments. In particular, DUR($V/\_\underset{[+voi]}{C}$)=TARGETDUR($V/\_\underset{[+voi]}{C}$) and DUR($V/\_\mathring{C}$)=TARGETDUR($V/\_\mathring{C}$) set target durations for vowels preceding voiced and voiceless consonants, respectively. These constraints are defined in (25) and (27), and examples of their application to sample candidates (with weights set to 1) are provided in (26) and (28)

(25) **DUR($V/\_\underset{[+voi]}{C}$)=TARGETDUR($V/\_\underset{[+voi]}{C}$)**

In brief: The duration of a vowel in the candidate, which precedes a voiced consonant, should match the canonical (as defined in (7)) output duration of that vowel (when it precedes a voiced consonant) in the language at large.

Definition:

- For a vowel α which precedes a voiced consonant in the output, and is spoken at speech rate $\mathfrak{R}$, let:
  - $TargetDur(V/\_\underset{[+voi]}{C})$ be the canonical output duration of α when preceding a voiced consonant in the output, and spoken at speech rate $\mathfrak{R}$
  - $Dur(V/\_\underset{[+voi]}{C})$ be the actual duration of α in the candidate under evaluation
  - $w_{\underset{[+voi]}{C}}$ be the weight applied to this constraint
- $cost = w_{\underset{[+voi]}{C}}(TargetDur(V/\_\underset{[+voi]}{C}) - Dur(V/\_\underset{[+voi]}{C}))^2$

(26) Cost of DUR(V/_$C_{[+voi]}$)=TARGETDUR(V/_$C_{[+voi]}$) for various candidates (for $w_{C_{[+voi]}}$=1)

			DUR(V/_$C_{[+voi]}$)=TargetDur(V/_$C_{[+voi]}$)
/puhKEED/	a.	V dur = 150.00 ms	106.71 $1(160.33 - 150.00)^2$
	b. ☞	V dur = 160.33 ms	0.00 $1(160.33 - 160.33)^2$
	c.	V dur = 170.00 ms	93.51 $1(160.33 - 170.00)^2$
/puhKEED+ing/	d.	V dur = 150.00 ms	106.71 $1(160.33 - 150.00)^2$
	e. ☞	V dur = 160.33 ms	0.00 $1(160.33 - 160.33)^2$
	f.	V dur = 170.00 ms	93.51 $1(160.33 - 170.00)^2$
/puhKEET/	g.	V dur = 150.00 ms	0.00 *n.a.*
	h.	V dur = 160.33 ms	0.00 *n.a.*
	i.	V dur = 170.00 ms	0.00 *n.a.*
/puhKEET+ing/[8]	j.	V dur = 150.00 ms	106.71 $1(160.33 - 150.00)^2$
	k. ☞	V dur = 160.33 ms	0.00 $1(160.33 - 160.33)^2$
	l.	V dur = 170.00 ms	93.51 $1(160.33 - 170.00)^2$

(27) **DUR(V/_C̥)=TARGETDUR(V/_C̥)**

In brief: The duration of a vowel in the candidate, which precedes a voiced consonant, should match the canonical (as defined in (7)) output duration of that vowel (when it precedes a voiced consonant) in the language at large.

Definition:

- For a vowel α which precedes a voiceless consonant in the output, and is spoken at speech rate $\mathfrak{R}$, let:

[8] I assume that all candidates for this input have undergone flapping in the phonology, and as such, the target vowels precede a flap. These vowels, therefore, precede voiced segments, and are subject to the restrictions imposed by DUR(V/_$C_{[+voi]}$)=TargetDur(V/_$C_{[+voi]}$).

- *TargetDur(V/_C̥)* be the canonical output duration of α when preceding a voiceless consonant in the output, and spoken at speech rate $\mathfrak{R}$
- *Dur(V/_C̥)* be the actual duration of α in the candidate under evaluation
- $w_{(V/\_\mathring{C})}$ be the weight applied to this constraint

• $cost = w_{(V/\_\mathring{C})}(TargetDur(V/\_\mathring{C}) - Dur(V/\_\mathring{C}))^2$

(28) Cost of DUR(V/_C̥)=TARGETDUR(V/_C̥) for various candidates (for $w_{C_{[+voi]}}$ =1)

			DUR(V/_C̥)=TARGETDUR(V/_C̥)
/puhKEED/	a.	V dur = 130.00 ms	0.00 *n.a.*
	b.	V dur = 152.20 ms	0.00 *n.a.*
	c.	V dur = 170.00 ms	0.00 *n.a.*
/puhKEED+ing/	d.	V dur = 130.00 ms	0.00 *n.a.*
	e.	V dur = 152.20 ms	0.00 *n.a.*
	f.	V dur = 170.00 ms	0.00 *n.a.*
/puhKEET/	g.	V dur = 130.00 ms	492.84 $(152.20 - 130.00)^2$
	h. ☞	V dur = 152.20 ms	0.00 $(152.20 - 152.20)^2$
	i.	V dur = 170.00 ms	316.84 $(152.20 - 170.00)^2$
/puhKEET+ing/[9]	j.	V dur = 130.00 ms	0.00 *n.a.*
	k.	V dur = 152.20 ms	0.00 *n.a.*
	l.	V dur = 170.00 ms	0.00 *n.a.*

As can be seen in the tableau in (28), DUR(V/_C̥)=TARGETDUR(V/_C̥) prefers candidates with pre-flap vowels that approach the target of 160.33 ms. Note that this constraint does not assess any cost for forms like /puhKEET/, as there are no voiced segments. The tableau in (26) shows that DUR(V/_$C_{[+voi]}$)=TARGETDUR(V/_$C_{[+voi]}$) behaves in the inverse manner: since only forms like /puhKEET/ have vowels which precede surface-voiceless

[9] As in (26), I assume that all candidates for this input have undergone flapping in the phonology, and as such, the target vowels precede a flap. These vowels, therefore, precede voiced segments, and are not subject to the restrictions imposed by DUR(V/_C̥)=TARGETDUR(V/_C̥).

segments, the constraint can only assess a cost in these cases.

Also necessary is a constraint that motivates pre-flap vowel shortening. This is accomplished by setting a target duration of zero for such vowels, and penalizing candidates for distance from this target. Due to the nature of the weighted constraint grammar, a low target such as this does not doom candidates to infinitely short durations. As long as any competing constraints provide sufficient pressure for vowel duration to remain at reasonable levels, the relative weighting of these constraints will produce the desired result. The constraint SHORTEN(V/_ɾ), which accomplishes this, is provided in (29), and is exemplified in the tableau in (30).

(29) **SHORTEN(V/_ɾ)**

In brief: The duration of a vowel in the candidate, which precedes a flap, should be as near to zero as possible.

Definition:

- For a vowel α which precedes a flap in the output, and is spoken at speech rate $\mathfrak{R}$, let:
  - *TargetDur(V/_ɾ)* = 0
  - *Dur(V/_ɾ)* be the actual duration of α in the candidate under evaluation
  - $w_{(\text{V/\_ɾ})}$ be the weight applied to this constraint
- $cost = w_{(\text{V/\_ɾ})}(\textit{TargetDur(V/\_ɾ)} - \textit{Dur(V/\_ɾ)})^2$

(30) Cost of SHORTEN(V/_ɾ) for various candidates (for $w_{(V/\_ɾ)}$=1)

			SHORTEN(V/_ɾ)
/puhKEED/	a.	V dur = 100.00 ms	0.00 *n.a.*
	b.	V dur = 106.54 ms	0.00 *n.a.*
	c.	V dur = 112.23 ms	0.00 *n.a.*
	d.	V dur = 120.00 ms	0.00 *n.a.*
/puhKEED+ing/	e. ☞	V dur = 100.00 ms	10,000.00 $(0.00 - 100.00)^2$
	f.	V dur = 106.54 ms	11,350.77 $(0.00 - 106.54)^2$
	g.	V dur = 112.23 ms	12,595.57 $(0.00 - 112.23)^2$
	h.	V dur = 120.00 ms	14,400.00 $(0.00 - 120.00)^2$
/puhKEET/	i.	V dur = 100.00 ms	0.00 *n.a.*
	j.	V dur = 106.54 ms	0.00 *n.a.*
	k.	V dur = 112.23 ms	0.00 *n.a.*
	l.	V dur = 120.00 ms	0.00 *n.a.*
/puhKEET+ing/	m. ☞	V dur = 100.00 ms	10,000.00 $(0.00 - 100.00)^2$
	n.	V dur = 106.54 ms	11,350.77 $(0.00 - 106.54)^2$
	o.	V dur = 112.23 ms	12,595.57 $(0.00 - 112.23)^2$
	p.	V dur = 120.00 ms	14,400.00 $(0.00 - 120.00)^2$

In the tableau in (30), we see that SHORTEN(V/_ɾ) is least costly for flapped candidates whose pre-flap vowels approach the target of 100 ms among the candidates depicted.

Also relevant for the analysis of flapping is OO-ID-DUR, repeated here in (31) from (15).

(31) **OO-ID-DUR**

In brief: The duration of a segment in the candidate should be faithful to

[9] As in (26), I assume that all candidates for this input have undergone flapping in the phonology, and as such, the target vowels precede a flap. These vowels, therefore, precede voiced segments, and are not subject to the restrictions imposed by DUR(V/_Ç̥)=TARGETDUR(V/_Ç̥).

the duration of the same segment in the base—the most frequent type in the candidate's inflectional paradigm as applied to the candidate's root.

Definition:

- For a segment α in the candidate, let:
  - *Dur(Cand)* be the duration of segment α in the output
  - *Dur(Base)* be the duration of the segment β in the base such that $(\alpha, \beta) \in \Re_{OO}$, where:
    - the base is the form that has the same root as α, and is a member of the most frequent type within the candidate's inflectional paradigm (as defined in (6))
    - if $(\alpha, \beta) \notin \Re_{OO}$ (determined via the Correspondence Transitivity Rule, §4.1.3.3), then this constraint is vacuously satisfied with a total cost of zero
  - w_{ID} be the weight applied to this constraint
- $cost = w_{\mathrm{ID}}(Dur(Cand) - Dur(Base))^2$

For the flapped nonce word stimuli currently under analysis, the base form targeted by OO-ID-Dur is crucially a non-flapped form. This is because in the English verbal paradigm, such forms are more frequent than flapped forms. Table 4.7 shows that in the spoken data subset of the *Corpus of Contemporary American English* (*COCA*, Davies 2008-), both the third person singular (3sg, *-s*) and infinitive forms are more frequent than the past tense *-ed* and present progressive *-ing* forms which would have the potential to condition flapping. The data in Table 4.7 was compiled by finding the 100 (first column) and 1,000 (second column) most frequent words tagged with each verb inflection. Within each inflection type, the token frequency of these 100 or 1,000 verbs was summed to give a total frequency score for that inflection type, as indicated in the table.[10]

[10] It should be noted that one weakness of this method is that more frequent forms are more likely to be irregular (see, e.g., Francis and Kučera 1982, Pinker 1999)—a contingency which was not controlled for here. Given that past tense forms are more likely to be irregular in English verbs than present tense forms, however, this issue seems unlikely to have impacted the outcome.

	Top 100	Top 1,000
3sg (-s)	3,150,504	3,332,235
infinitive	2,967,269	3,754,817
-ed	2,560,261	2,923,198
plain (finite)	2,019,278	2,400,311
-ing	1,374,797	1,791,630

Table 4.7: Combined frequency of all words in the top 100 (first column) and top 1,000 (second column) most frequent words for each verb inflection type, from COCA's spoken data.

Given the corpus findings, the base to which OO-ID-Dur compels identity in flapped tokens should be either the 3sg form or the infinitive form of the candidate, depending on whether we assume that sampling the 100 most frequent or the 1,000 most frequent verbs of each inflection type provides the truest measure of type frequency. In any event, neither of these forms would induce flapping in the inputs relevant to the discussion at hand. As such, I will assume that the base for flapped forms, for the purposes of OO-ID-Dur, consists of the verb's infinitive form, which is identical to its root.

With the question of base forms settled, it is now possible to view the effects of OO-ID-Dur on candidates from the flapping data in the tableau in (32).

(32) Cost of OO-ID-Dur for various candidates (for w_{ID}=1)

			OO-ID-Dur
/puhKEED/	a.	V dur = 150.00 ms	106.71 $(160.33 - 150.00)^2$
	b. ☞	V dur = 160.33 ms	0.00 $(160.33 - 160.33)^2$
	c.	V dur = 170.00 ms	93.51 $(160.33 - 170.00)^2$
/puhKEED+ing/	d.	V dur = 106.54 ms	2,893.36 $(160.33 - 106.54)^2$
	e.	V dur = 112.23 ms	2,313.61 $(160.33 - 112.23)^2$
	f. ☞	V dur = 160.33 ms	0.00 $(160.33 - 160.33)^2$
	g.	V dur = 170.00 ms	93.51 $(160.33 - 170.00)^2$
/puhKEET/	h.	V dur = 130.00 ms	492.84 $(152.20 - 130.00)^2$
	i. ☞	V dur = 152.20 ms	0.00 $(152.20 - 152.0)^2$
	j.	V dur = 170.00 ms	316.84 $(152.20 - 170.00)^2$
/puhKEET+ing/	k.	V dur = 106.54 ms	2,084.84 $(152.20 - 106.54)^2$
	l.	V dur = 112.23 ms	1,597.60 $(152.20 - 112.23)^2$
	m. ☞	V dur = 152.20 ms	0.00 $(152.20 - 152.20)^2$
	n.	V dur = 160.00 ms	60.84 $(152.20 - 160.00)^2$

OO-ID-Dur, as is evident from the tableau in (32), prefers candidates whose base form has a voiced segment preceding the target vowel (e.g., puhKEED, puhKEED-ing) to reach the target duration for vowels which precede voiced segments (160.33 ms); candidates whose base form has a voiceless segment preceding the target vowel (e.g., puhKEET, puhKEET-ing), on the other hand, are pressured by this constraint to reach the target for vowels that precede voiceless segments (152.20 ms).

Let us now consider the interaction of the constraints at hand. First, as can be seen in the tableau in (28), Dur(V/_C̥)=TargetDur(V/_C̥) is not applicable to any candidates that result in flapping. This is because Dur(V/_C̥)=TargetDur(V/_C̥) only exerts pressure on vowels preceding voiceless segments. The target vowels in our data, however, precede flaps (which are voiced), and are therefore not subject to the penalties of Dur(V/_C̥)=TargetDur(V/_C̥), which will not be shown in the tableaux below. Second, as is evident from the fact that, in the tableau in (26), Dur(V/_$\underset{[+voi]}{C}$)=TargetDur(V/_$\underset{[+voi]}{C}$) prefers candidates with the same vowel duration for both /d/ inputs and /t/ inputs, this constraint does not drive the incomplete neutralization, and so will also be absent from future tableaux.

The two remaining constraints, OO-ID-Dur and Shorten(V/_ɾ), are in conflict with one another. OO-ID-Dur prefers flapped candidates to be faithful to their base (which as discussed above, is in essence, the verb's infinitive form). In other words, OO-ID-Dur results in a higher cost for candidates whose target vowels are higher or lower than that of the base. Shorten(V/_ɾ), however, pressures pre-flap vowels to shorten as much as possible. Assuming that the base form's target vowel has a duration greater than zero ms, these constraints cannot both be completely satisfied at the same time. As the weight of OO-ID-Dur increases relative to Shorten(V/_ɾ), the winning candidate's target vowel will increase in duration until it meets the duration of its base. Conversely, as the weight of Shorten(V/_ɾ) increases relative to OO-ID-Dur, the winning candidate's target vowel will decrease in duration. (As in the Japanese case above, the candidate whose total sum of all constraint costs is lowest is the winning candidate.)

(33) Tableau modeling flapping data from Experiment 2, with $<w_{\mathrm{ID}}=2.3\overline{3},\ w_{(\mathrm{V}/\_\text{ɾ})}=1>$, $\mathit{TargetDur}(V/\_\underset{[+voi]}{C})=160.33$, $\mathit{TargetDur}(V/\_\underset{\circ}{C})=152.20$

		OO-ID-DUR	SHORTEN(V/_ɾ)	Total Cost
/puhKEED+ing/	a. V dur = 100.00 ms	8,492.66 $2.3\overline{3}(100.00-160.33)^2$	10,000.00 $1(0.00-100.00)^2$	18,492.66
	b. V dur = 106.54 ms	6,751.17 $2.3\overline{3}(106.54-160.33)^2$	11,350.77 $1(0.00-106.54)^2$	18,101.94
	c. ☞ V dur = 112.23 ms	5,398.42 $2.3\overline{3}(112.23-160.33)^2$	12,595.57 $1(0.00-112.23)^2$	17,993.99
	d. V dur = 120.00 ms	3,795.19 $2.3\overline{3}(120.00-160.33)^2$	14,400.00 $1(0.00-120.00)^2$	18,195.19
/puhKEET+ing/	e. V dur = 100.00 ms	6,357.96 $2.3\overline{3}(100.00-152.20)^2$	10,000.00 $1(0.00-100.00)^2$	16,357.96
	f. ☞ V dur = 106.54 ms	4,864.63 $2.3\overline{3}(106.54-152.20)^2$	11,350.77 $1(0.00-106.54)^2$	16,215.40
	g. V dur = 112.23 ms	3,727.73 $2.3\overline{3}(112.23-152.20)^2$	12,595.57 $1(0.00-112.23)^2$	16,323.30
	h. V dur = 120.00 ms	2,419.29 $2.3\overline{3}(120.00-152.20)^2$	14,400.00 $1(0.00-120.00)^2$	16,819.29

As can be seen in the tableau in (33), with the weighting $<w_{\mathrm{ID}}=2.3\overline{3},\ w_{(\mathrm{V}/\_\text{ɾ})}=1>$, as vowel duration approaches 112.23 ms for /d/-flap candidates and 106.54 ms for /t/-flap candidates (the desired outcome in each case, respectively) the total cost decreases. (See Appendix E for an implementation of the cost function and its minimization in R code.)

As in the Japanese case, varying the weights of these constraints—changing the terms of the compromise between them—generates different results. This principle is visualized in the graph in Figure 4.3, in which the x- and y- axes represent the weights of OO-ID-DUR and SHORTEN(V/_ɾ) ($w_{\mathrm{ID}}=x$, $w_{(\mathrm{V}/\_\text{ɾ})}=y$, $0< \{x,\ y\} \leq 5$), and the z-axis represents the predicted duration of pre-flap vowels. The top layer in the graph shows pre-/d/ flap vowels; the lower layer shows pre-/t/ flap vowels. Note that the the model never allows a language in which pre-/t/ flap vowels are longer than pre-/d/ flap vowels (i.e., the upper and lower layers in the graph never cross, such that the lower, pre-/t/ vowel layer appears above the upper pre-/d/ vowel layer).

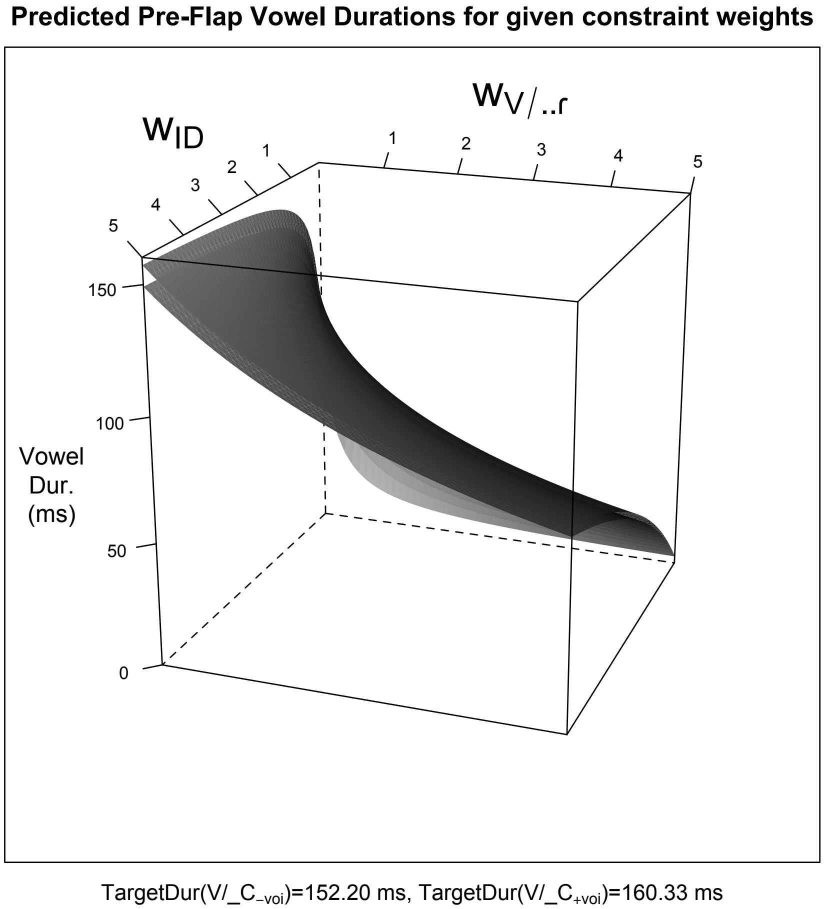


Figure 4.3: Experiment 2: Predicted pre-flap vowel duration for /d/-flaps (top sheet) and /t/-flaps (bottom sheet), given weights for w_{ID} and $w_{(\text{V}/\_ɾ)}$. *TargetDur(V/_C̥)*=138.50 ms, and *TargetDur(V/_C$_{[+voi]}$)*=145.90.

4.4 Discussion and Conclusion

As we have seen, the model described in this chapter successfully captures the experimental data from the earlier chapters of the dissertation. Crucially, it provides an intuitive mechanism through which the *degree* of incomplete neutralization can be tweaked. That is to say, the continuum from 'imperceptible' incomplete neutralization (like flapping) to 'perceptible' incomplete neutralization (like Japanese) is captured in terms of the relative weighting of constraints that impose targets on phonetic properties (e.g., Dur(μμ)=TargetDur(μμ), Shorten(V/_ɾ)) and constraints that require identity to a relatively frequent base (OO-ID-Dur). Varying these weights can yield systems not only at the poles of completely of neutralization, but can also generate systems in between (of which German (Port and O'Dell 1985) is perhaps a case).[11]

This ability to model the degree of incomplete neutralization in a controlled fashion is a distinct advantage to this system. Recall, for example, the proposal for incomplete neutralization in final devoicing based in Turbidity Theory (van Oostendorp 2008, Goldrick 2001) (summarized in Section 1.2.3). At its core, the Turbidity Model argues that devoiced segments have a different output structure than underlyingly voiceless ones. This means that the phonetics is able to maintain a(n incomplete) difference between underlyingly voiceless and devoiced segments, since in essence they are distinct inputs to the phonetics. It is not obvious, though, how distinct the phonetics must make these different inputs. While the phonetics, presumably, aims to make devoiced segments more 'voiced-like' than underlyingly voiced segments, the degree to which this occurs is not predicted by any inherent properties of the Turbidity Model.

Underlying the assumption that the phonetics in the Turbidity Model aims to make devoiced segments more like their voiced underlying representations than underlyingly voiceless segments is the idea of the projection relationship. Under this theory, segments stand in two possible relations with a given feature: a projection relationship, and a pronunciation relationship. Under this account, voiced segments are related to the [+voice] feature by both the abstract, structural projection relationship, and the pronunciation relationship, which describes the actual phonological output. Devoiced segments, however, have only the projection relationship with the [+voice] feature. (Underlyingly voiceless segments have no relationships with the [+voice feature]).

One possibility under the Turbidity Model is that the phonetics should interpret a

[11] I note here that the sort of frequency described in the definition of basehood in (6) seems unsuitable for the case of German final devoicing. Under a model of OO-faithfulness, the ideal base for underlyingly voiced segments in the German case should be realized as voiced. Under the assumption that the (generally null) nominative singular is the most frequent type in the German nominal paradigm, this is not the case. I assume that in this case, either a more traditional form of basehood based on morphological simplicity is active, or that a variety of frequency other than the type active in Japanese and English flapping is at play.

projected feature in the manner in which that feature is canonically produced—but in a reduced manner. "For instance, if the length of the preceding vowel is a cue to voicing, this lengthening may also be triggered slightly by projected voicing features" (van Oostendorp 2008, p. 1371). Note, however, that this is not a restrictive system — there is nothing stopping a projected voicing feature being interpreted as, e.g., a drop in pitch. Similarly, without further clarity with respect to the requirement that the triggering must be 'slight', it is conceivable that the phonetics might lengthen vowels before devoiced segments to a fair degree. The model argued for in this chapter provides a more explicit method of determining the manner in which an incompletely neutralized unit will be produced.

A slightly more nuanced issue is that of the Directionality Observation (Section 1.3.4), which claims that the realization of two incompletely neutralized categories is predictable: on a continuum of the acoustic cue(s) that differentiate the categories, an incompletely neutralized segment must fall somewhere between the canonical realizations of the two categories. Under the Turbidity Model (van Oostendorp 2008), this generalization can be enforced only by requirement stipulated above.

Given the constraints described in this chapter, the Directionality Observation is secure, regardless of constraint weights. This concept can be seen in Figure 4.3, in which the layer representing pre-/d/ flap vowels is always higher than that representing pre-/t/ flap vowels. Taking the example of Japanese, even if the constraint weightings were extremely unbalanced in favor of Dur(μμ)=TargetDur(μμ) (the constraint which favors surface-bimoraic candidates to approach the canonical duration of long vowels), that constraint would pressure both lengthening short/Ø vowels and underlyingly long vowels to reach the same target duration, while OO-ID-Dur would exert (a very small amount of) pressure on short/Ø vowels to decrease in duration from this target. In other words, lengthened short/Ø vowels never become longer than underlyingly long vowels.

The Turbidity Model (van Oostendorp 2008) might be thought of as the phonetic module accessing the underlying phonological representation via a mediating layer of structure. An alternative approach would be to posit that the phonetics could access underlying representations directly. Under such a model, before producing a segment, the phonetics would 'look up' its underlying status, which could impact its production (via, perhaps, a form of Input-(Phonetic)Output correspondence constraints).

Such a model faces several problems. First, it is susceptible to some of the same critiques as the Turbidity Model—for instance, it is not clear when the model would predict 'perceptible' or 'imperceptible' incomplete neutralization. Second, for features like voicing that have multiple cues, which one(s) would the phonetics be faithful to? Further, assuming that the phonetics would strive for identity with the underlying form, then underlying forms would likely need to be enriched with a high level phonetic detail (e.g., duration in ms., pitch in Hz., etc.). Along these lines, the Directionality Observation presents a problem for this type of theory. Without definite targets to meet, expressed in raw phonetic terms, it is unclear how to constrain the phonetic module's interpretation of faithfulness to an input.

4.4.1 Speakers, Hearers, and Constraints Out of Their Motivating Context

One explanation for incomplete neutralization is that speakers maintain the underlying contrast for the benefit of their interlocutors. Indeed, Port and Crawford (1989) argue from experimental evidence that speakers vary the completeness of incompletely neutralized contrasts depending on the communicative situation. (This result mirrors those from other phenomena in which speakers manipulate their productions for the benefit of hearers. See, e.g., Lindblom 1990, Scarborough 2003, 2010, Flemming 2010, Syrett and Kawahara under revision.) This view provides a clear motivation for incomplete neutralization at the perceptible end of the continuum; why, though, should languages have incomplete neutralization nearer the imperceptible end? Is a speaker motivated to preserve a contrast in a manner that will not necessarily be perceived by their audience?

OO-ID-Dur is grounded in a speaker's desire to speak clearly. Whether or not any manipulations made to the speech signal actually aid a hearer is irrelevant—the speaker's desire is still active. In other words, speakers will try to maintain a contrast, even though they know that it might be in vain. Like a child intent on flying a kite on a windless day, OO-ID-Dur encodes speakers' intent to maintain a contrast regardless of the perceptual outcome.

In languages without 'perceptible' incomplete neutralization, such constraints must still receive some weighting—albeit a relatively low one—since speakers' desire to speak clearly is always present. In other words, the constraints that drive 'perceptible' incomplete neutralization exist, and we should expect that in languages without 'perceptible' incomplete neutralization, constraints like OO-ID-Dur might still cause a more limited form of the phenomenon (in spite of their low weighting).

4.4.2 Conclusion

Phonetic constraints of the type described in this chapter differ from phonological constraints, as in classic OT, in two crucial ways. First, phonetic constraints make reference to both phonological structure *and* raw, quantitative phonetic information (Zsiga 2000, p. 96; Flemming 2001). Second, phonetic constraints must be evaluated in a weighted constraint grammar, rather than a grammar that operates on ranking and strict dominance (Zsiga 2000, p. 96–97).[12]

I have shown that a weighted constraint grammar in the phonetic module can serve to model incomplete neutralization. The model imposes a conflict between meeting a target (through constraints of the form Dur(x)=TargetDur(x)) and faithfulness to a base

[12]Flemming (2001, p. 33) points out that it may be conceptually possible to generate compromise between two constraints in a strict-dominance model by decomposing each constraint in a set of ranked sub-constraints and interleaving them. As he argues, however, this would essentially entail quantizing the phonetic value being targeted.

(through the constraint OO-ID-Dur). Assuming a notion of basehood that centers on type frequency ensures that the proper base is selected, even if the base is not morphologically more simple than the candidate.

This model has two major virtues, especially as compared to competing analyses. First, it clearly and explicitly produces languages along a continuum of completeness of neutralization—from languages where the neutralization is nearly complete and imperceptible, as in flapping in American English, to languages where the partially-neutralized distinction is likely to be perceptible, as in Japanese. Second, this model accurately predicts the Directionality Observation—candidates that violate this generalization can never win, regardless of constraint weightings.

Chapter 5

Conclusion

In this dissertation, I have presented a series of experiments on two case studies in incomplete neutralization — flapping in American English and monomoraic noun lengthening in Japanese. These two cases differ crucially in their *degree* of neutralization. While the distinction being made in the flapping case is so small as to be imperceptible, the distinction made in the Japanese case is large enough that listeners can likely distinguish tokens of each category. I argued for a model of incomplete neutralization in which the phonetic module, operating as a weighted-constraint grammar, resolves the conflict between constraints that pressure candidates to meet various targets, and constraints that reward similarity to a (typewise) frequent base.

In the case of flapping, I showed in Experiments 1 and 2 that vowels preceding /d/ flaps were longer than vowels preceding /t/ flaps. This result held across two task types, suggesting that incomplete neutralization in flapping is not simply an effect of hyperarticulation or orthography. Further, the use of nonce word stimuli in these tasks showed that these results are not impacted by lexical frequency. I also showed, through perception tasks in Experiments 3, 4, and 5, that listeners cannot distinguish between tokens with /d/ flaps and tokens with /t/ flaps.

Further, I showed in Experiments 6 and 7 that when Japanese monomoraic nouns lengthen, the resulting distinction between lengthened vowels and underlyingly long vowels is incompletely neutralized. While almost all prior research into incomplete neutralization has focused on voicing or other feature- and segment-level contrasts, this result shows that incomplete neutralization can apply to suprasegmental contrasts as well.

Crucially, the distinction found in the flapping case was much smaller than the one found in the Japanese case. I argued that incomplete neutralization is not a homogenous process, but rather, consists of a continuum from almost completely neutralized (and imperceptible) to relatively less neutralized (and perceptible). In Chapter 4, I proposed a model of the phonetic module with weighted constraints that—unlike many previous theories of incomplete neutralization—predicts this continuum.

A second key benefit of the model is that it accurately and explicitly accounts for the Directionality Observation—candidates that violate this generalization are never able to surface, regardless of constraint weightings.

While incomplete neutralization on the perceptible end of the neutralization continuum is motivated on communicative/cooperative grounds, the direct benefit to the speaker or hearer is less clear at the other end of the spectrum. I argue that since constraints required for 'perceptible' incomplete neutralization are motivated, we should not be surprised to find them at various weightings in different languages — when weighted sufficiently low, the result is 'imperceptible' incomplete neutralization.

In conclusion, this dissertation provides experimental evidence for at least two points along the continuum of degree of neutralization, and provides a phonetic grammar capable of generating these phenomena and others along a continuum from nearly complete to nearly non-existent neutralization.

Appendix A

Summary of All Measures from Experiment 2

Voicing Correlate	Mean /d/	Mean /t/	Mean difference	t	p
Pre-flap vowel duration	112.23 ms	106.54 ms	5.69 ms	3.77	p<0.001
Flap closure duration	28.54 ms	28.16 ms	0.38 ms	0.04	n.s.
Percent closure voicing	98.14%	97.40%	0.74%	-0.68	n.s.
F0 slope at flap onset	0.40 Hz/ms	0.43 Hz/ms	-0.03 Hz/ms	-0.45	n.s.
F1 slope at flap onset	4.64 Hz/ms	4.91 Hz/ms	-0.27 Hz/ms	-0.80	n.s.
F0 slope at flap offset	-0.06 Hz/ms	0.07 Hz/ms	-0.13 Hz/ms	-1.29	n.s.
F1 slope at flap offset	-0.75 Hz/ms	0.21 Hz/ms	-0.96 Hz/ms	-0.79	n.s.

Table A.1: Experiment 2: Summary of results from all measured correlates of voicing. *t*-values are from the linear mixed models described in §2.4.3.3. *p*-values are calculated using the `pvals.fnc` function of R's `languageR` package (Baayen 2009), which uses the Markov Chain Monte Carlo method.

Appendix B

Stimuli from Experiment 6

(Table begins on next page)

Japanese orthography	Transcription	Gloss	Accent Mismatch	Diphthong in hiragana	Long noun is quoted
木が倒れた。	*ki'* ga taoreta	tree NOM fell			
木倒れた。	*ki'* taoreta	tree fell			
キー見つかった？	*ki'i* mitsukaata	key found			
菜が煮えた。	*na'* ga nieta	vegetable NOM cooked			
菜煮えた。	*na'* nieta	vegetable cooked			✓
「なー」と言われた。	*na'a* to iwareta	DISC COMP was.said			
火が消えた。	*hi'* ga kieta	fire NOM went.out			
火消えた。	*hi'* kieta	fire went.out			✓
「ひー」と叫んだ。	*hi'i* to sakenda	INTERJ COMP shouted			
酢がない。	*su'* ga nai	vinegar NOM NEG			
酢ない。	*su'* nai	vinegar NEG			
スーが見つからない。	*su'u* ga mitsukara nai	Sue NOM NEG			
背がのびた。	*se'* ga nobita	spine NOM stretched			
背のびた。	*se'* nobita	spine stretched		✓	
正の整数。	*se'i* no seisuu	positive MOD integer			
血がでた。	*chi* ga deta	blood NOM went.out			
血でた。	*chi* deta	blood went.out	✓		
地位がある。	*chi'i* ga aru	social.status NOM have			

(continued...)

Japanese orthography	Transcription	Gloss	Accent Mismatch	Diphthong in hiragana	Long noun is quoted
手がしびれた。	*te'* ga shibireta	hand NOM became.numb			
手しびれた。	*te'* shibireta	hand became.numb		✓	
低の長さ。	*te'i* no nagasa	base.of.shape MOD length			
戸が壊れた。	*to* ga kowareta	door NOM broke			
戸壊れた。	*to* kowareta	door broke	✓		✓
「とー」と叫んだ。	*to'o* to sakenda	INTERJ COMP shouted			
根がぬけた。	*ne'* ga nuketa	root NOM pulled.out			
根ぬけた。	*ne'* nuketa	root pulled.out			✓
「ねー」と言われた。	*ne'e* to iwareta	DISC COMP was.said			
麩が美味しい。	*fu'* ga oishii	gluten NOM delicious			
麩美味しい。	*fu'* oishii	gluten delicious			
封がとれた。	*fu'u* ga toreta	seal NOM came.off			
目が腫れた。	*me'* ga hareta	eye NOM swelled			
目腫れた。	*me'* hareta	eye swelled			✓
「メー」と鳴いた	*me'e* to naita	“baa (sheep sound)” COMP made.animal.sound			

All stimulus sets from Experiment 6. Target nouns are in boldface. Accents are shown for target nouns only.

Appendix C

Stimuli from Experiment 7

Japanese orthography	Transcription	Gloss
木もなくしたよ。	**ki'** mo nakushita yo	tree ALSO lost DISC
木なくしたよ。	**ki'** nakushita yo	tree lost DISC
キーなくしたよ。	**ki'i** nakushita yo	key lost DISC
酢も見つけたよ。	**su'** mo mitsuketa yo	vinegar ALSO found DISC
酢見つけたよ。	**su'** mitsuketa yo	vinegar found DISC
スー見つけたよ。	**su'u** mitsuketa yo	Sue found DISC
麸も残したよ。	**fu** mo nokoshita yo	gluten ALSO left DISC
麸残したよ。	**fu** nokoshita yo	gluten left DISC
封残したよ。	**fu'u** nokoshita yo	seal left DISC
血も捧げたよ。	**chi** mo sasageta yo	blood ALSO dedicated DISC
血捧げたよ。	**chi** sasageta yo	blood dedicated DISC
地位捧げたよ。	**chi'i** sasageta yo	social.status dedicated DISC
具も出したよ。	**gu** mo dashita yo	ingredients ALSO served DISC
具出したよ。	**gu** dashita yo	ingredients served DISC
グー出したよ。	**gu'u** dashita yo	fist served DISC
ソも確かめたよ。	**so'** mo tashikameta yo	so ALSO confirmed DISC
ソ確かめたよ。	**so'** tashikameta yo	so confirmed DISC
層確かめたよ。	**so'u** tashikameta yo	layer confirmed DISC
手も測ったよ。	**te'** mo hakatta yo	hand ALSO measured DISC
手測ったよ。	**te'** hakatta yo	hand measured DISC
低測ったよ。	**te'i** hakatta yo	base measured DISC

(continued...)

Japanese orthography	Transcription	Gloss
背も違うよ。	**se'** mo chigau yo	height ALSO is-different DISC
背違うよ。	**se'** chigau yo	height is-different DISC
性違うよ。	**se'i** chigau yo	gender is-different DISC
野も持ってるよ。	**no'** mo motteru yo	field ALSO have DISC
野持ってるよ。	**no'** motteru yo	field have DISC
脳持ってるよ。	**no'u** motteru yo	brain have DISC
尾も出てきたよ。	**o'** mo detekita yo	tail ALSO appeared DISC
尾出てきたよ。	**o'** detekita yo	tail appeared DISC
王出てきたよ。	**o'u** detekita yo	king appeared DISC
津も買収したよ。	**tsu'** mo baishuushita yo	Tsu ALSO bought/bought.off DISC
津買収したよ。	**tsu'** baishuushita yo	Tsu bought/bought.off DISC
通買収したよ。	**tsu'u** baishuushita yo	expert bought/bought.off DISC
帆も叩いたよ。	**ho'** mo tataita yo	sail ALSO hit DISC
帆叩いたよ。	**ho'** tataita yo	sail hit DISC
ほおも叩いたよ。	**ho'o** tataita yo	cheek hit DISC
都も独占したよ。	**to'** mo dokusenshita yo	city ALSO monopolized DISC
都独占したよ。	**to'** dokusenshita yo	city monopolized DISC
塔独占したよ。	**to'u** dokusenshita yo	tower monopolized DISC
書も独占したよ。	**sho'** mo dokusenshita yo	book ALSO monopolized DISC
書独占したよ。	**sho'** dokusenshita yo	book monopolized DISC
章独占したよ。	**sho'u** dokusenshita yo	chapter monopolized DISC
字も公開したよ。	**ji'** mo koukaishita yo	letter ALSO publicized DISC
字公開したよ。	**ji'** koukaishita yo	letter publicized DISC
爺公開したよ。	**ji'i** koukaishita yo	grandpa publicized DISC

All stimulus sets from Experiment 7. Target nouns are in boldface. Accents are shown for target nouns only.

Appendix D

Application of the Model to Unrounded Experiment 7 Results

	Mean	SD
short/prt	54.99	21.89
short/Ø	124.98	34.91
long	157.45	39.21

Table D.1: Mean and standard deviation of vowel durations (in ms) for each category in Experiments 7

Short/Ø dur.(ms)	*cost(*OO-ID-DUR*)* w_{ID}*(Dur(Cand)–Dur(Base))*^{2}	*cost(*DUR(μμ)=TARGETDUR(μμ)*)* $w_{\mu\mu}$*(TargetDur(μμ)–Dur(μμ))*^{2}	Total Cost
75.00	$1(75-54.99)^2$	$2.1\overline{5}(157.45-75)^2$	15,053.87
83.33	$1(83.33-54.99)^2$	$2.1\overline{5}(157.45-83.33)^2$	12,645.29
100.00	$1(100-54.99)^2$	$2.1\overline{5}(157.45-100)^2$	9,140.32
116.17	$1(116.17-54.99)^2$	$2.1\overline{5}(157.45-116.17)^2$	7,811.88
124.98	$1(125-54.99)^2$	$2.1\overline{5}(157.45-125)^2$	7,171.20
133.83	$1(133.83-54.99)^2$	$2.1\overline{5}(157.45-133.83)^2$	7,418.34
150.00	$1(150-54.99)^2$	$2.1\overline{5}(157.45-150)^2$	9,146.54

Table D.2: Costs for given short/Ø vowel durations, where w_{ID}=1, $w_{\mu\mu}$=$2.1\overline{5}$, *TargetDur(μ)*=54.99 ms, and *TargetDur(μμ)*=157.45 ms

As is evident from Table D.2, with the weighting <w_{ID}=1, $w_{\mu\mu}$=$2.1\overline{5}$>, as short/Ø vowel duration approaches 124.98 ms (our desired outcome), the total cost decreases. As candidate short/Ø vowel duration diverges from 124.98 ms—either longer or shorter—cost increases.

Appendix E

R code

R code used to model monomoraic noun lengthening in Japanese

```
library(lattice)  #for 3D graphs

shortTarget<<-50    #Target duration for short vowels
longTarget<<-150    #Target duration for long vowels

#Set default weights for the constraints
w1<<-1  #OO-ID-Dur
w2<<-1  #Dur(uu)=TargetDur(uu)


##################################################
# cost(x):
# For a given vowel duration (x)
# return the cost of that candidate
# for weights <OOIDDUR=w1, Shorten(V/_)=w2>
##################################################
cost<-function(x) {
  lengthened<-x     #Actual duration of lengthened vowel
            #OO-ID             #Dur(uu)=TargetDur(uu)
  cost<-w1*(lengthened-shortTarget)^2 + w2*(lengthened-longTarget)^2#
  return(cost)
}


##################################################
```

```
# duration(x,y):
# For a given set of constraint weights (x,y)
# where x is the weight of OOIDDur and y is the
# weight of Shorten(V/_ ), return the vowel duration
# that is least costly
##################################################
duration<-function(x,y) {
  w1<<-x    # NB: these weights, which are set to the arguments of this function
  w2<<-y    #   are set as global variables, and so will apply in the cost3D function

  #Find theLength, which is the duration that minimizes the cost function
  theLength<-optimize(cost, interval=c(0,500))
  return(theLength)
}



##################################################
# Create 3D plot:
# For a given set of constraint weights ranging
# from 0.1 to 5, plot the predicted vowel duration
##################################################

#Populate a grid with x/y ranging from 0.1 to 5, by 0.1
x<-seq(0.1,5,0.1)
y<-seq(0.1,5,0.1)
g<-expand.grid(x=x,y=y)

#Populate grid with result of length(x,y) (the preducted vowel
# durations for each weight)
lengthVec<-c()
for (i in 1:nrow(g)){
  lengthVec<-append(lengthVec,duration(g[i,]$x,g[i,]$y)$minimum)
}
g$length<-lengthVec

newcolors<-colorRampPalette(c("grey100", "grey100"))  #grayscale for graph

```

```
#Create the graph
wireframe(length ~ x*y, data=g,
    drape=T, scales=list(arrows=F), colorkey=F,
    screen=list(z = 55, x = -65, y = 0),
    col.regions=newcolors(1),
    ylab=list(label=expression(paste(w[mu][mu], sep="")),cex=2),
    xlab=list(label=expression(paste(w[ID],sep="")),cex=2),
    zlab=list(label="short/∅\ndur (ms)",cex=1.1),
    zlim=c(0,200),
    main="Predicted Short/∅ Vowel Duration for given constraint weights",
    sub=expression(paste("TargetDur(",mu,")=50 ms, TargetDur(",mu,mu,")=150 ms",
         sep=""))
)
```

R code used to model Flapping in American English

```
library(lattice) #for 3D graphs

nmlVTdur<<-152.2016     #Target duration for vowels preceding voiceless segments
nmlVDdur<<-160.33     #Target duration for vowels preceding voiced segments

#Set default weights for the constraints
w1<<-1  #OO-ID-Dur
w2<<-1  #Shorten(V/_ )

#By default, cost(x) will operate over /d/ tokens
td<<-"d"
#td<<-"t"   #uncomment to operate over /t/ tokens


##################################################
# cost(x):
# For a given vowel duration (x)
# return the cost of that candidate
# for weights <OOIDDUR=w1, Shorten(V/_ )=w2>
##################################################
cost<-function(x) {
  actualV<-x          #Actual duration of the targetvowel

  if (td=="d") targetDur<-nmlVDdur else targetDur<-nmlVTdur
```

```

  flapVTarget<-0        #Set target for pre-flap vowel duration to 0

          #OO-ID-DUR           #Shorten(V/_ )
  cost<-w1*(actualV - targetDur)^2 + w2*(actualV - flapVTarget)^2

  return(cost)
}


##################################################
# duration(x,y):
# For a given set of constraint weights (x,y)
# where x is the weight of OOIDDur and y is the
# weight of Shorten(V/_ ), return the vowel duration
# that is least costly
##################################################

duration<-function(x,y) {
  w1<<-x    # NB: these weights, which are set to the arguments of this function
  w2<<-y    #   are set as global variables, and so will apply in the cost3D
      function

  #For both "t" and "d", find theLength, which is the duration that minimizes
      the cost function
  td<<-"t"
    theLength<-optimize(cost, interval=c(0,500))
    str<-paste("t: ", theLength$minimum, " (", theLength$objective, ")        ",
         sep="")
  td<<-"d"
    theLength<-optimize(cost, interval=c(0,500))
    str<-paste(str, "d: ", theLength$minimum, " (", theLength$objective, ")",
        sep="")
  return(str)
}



##################################################
```

```
# Create 3D plot:
# For a given set of constraint weights ranging
# from 0.1 to 5, plot the predicted vowel duration
##################################################

# Function to get a vector of predicted vowel durations
length3DWireFramer<-function(x,y,thetd) {
  w1<<-x
  w2<<-y
  td<<-thetd
  theLength<-optimize(cost, interval=c(0,500))
  return(theLength)
}

#Populate a grid with x/y ranging from 0.1 to 5, by 0.1
x<-seq(0.1,5,0.1)
y<-seq(0.1,5,0.1)
g<-expand.grid(x=x,y=y)

#Populate grid with the predicted vowel durations
lengthVec<-c()
for (i in 1:nrow(g)){
  lengthVec<-append(lengthVec,length3DWireFramer(g[i,]$x,g[i,]$y, "d")$minimum)
}
  g$d<-lengthVec

lengthVec<-c()
for (i in 1:nrow(g)){
  lengthVec<-append(lengthVec,length3DWireFramer(g[i,]$x,g[i,]$y, "t")$minimum)
}
  g$t<-lengthVec


#Create the graph
wireframe(d + t ~ x*y, data=g,
    drape=F, scales=list(arrows=F), colorkey=F,
    screen=list(z=-90-20, x=-70, y =0),
    shade=T,
    shade.colors = function(irr,ref,height,w=0.9)
```

```
            grey(w * irr + (1 - w) * (1 - (1 - ref)^0.4)),
    ylab=list(label=expression(paste(w[V/ ..], sep="")),cex=2),
    xlab=list(label=expression(paste(w[ID],sep="")),cex=2),
    zlab=list(label="Vowel\nDur.\n(ms)",cex=1.1),
    zlim=c(00,160),
    main="Predicted Pre-Flap Vowel Durations for given constraint weights",
    sub=expression(paste("TargetDur(V/_",C[-voi],")=152.20 ms, TargetDur(V/_",C
        [+voi],")=160.33 ms", sep=""))
)
```

References

Albright, Adam (2002). *The Identification of Bases in Morphological Paradigms.* Doctoral dissertation, University of California, Los Angeles.

Anderson, Stephen R. (1975). *On the Interaction of Phonological Rules of Various Types.* Journal of Linguistics 11(1):39–62.

Audacity Team (2008). *Audacity v. 1.2.6.* Computer program. `http://audacity.sourceforge.net`.

Baayen, Harald (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R.* Cambridge University Press, Cambridge.

Baayen, Harald (2009). *languageR: Data Sets and Functions with "Analyzing Linguistic Data: A Practical Introduction to Statistics".* R package. `http://CRAN.R-project.org/package=languageR`.

Babel, Molly Elizabeth (2009). *Phonetic and Social Selectivity in Speech Accomodation.* Doctoral Dissertation, University of California, Berkeley.

Barnwell, Thomas P. (1971). *An Algorithm for Segment Durations in a Reading Machine Context.* Technical Report 479, MIT, Cambridge, Mass.

Bates, Douglas and Maechler, Martin (2009). *lme4: Linear Mixed-Effects Models Using S4 Classes.* R package. `http://CRAN.R-project.org/package=lme4`.

Beckman, Mary (1982). *Segmental Duration and the 'Mora' in Japanese.* Phonetica 39:113–135.

Benua, Laura (1997). *Transderivational Identity: Phonological Relations between Words.* Doctoral dissertation, University of Massachusetts, Amherst.

Berko, Jean (1958). *The Child's Learning of English Morphology.* Word 14:150–177.

Bloomfield, Leonard (1933/1984). *Language.* University of Chicago Press, Chicago.

Boersma, Paul and Weenink, David (2009). *Praat: Doing Phonetics by Computer.* Computer program. `http://www.praat.org`.

Braver, Aaron (2010). *Incomplete Neutralization in American English Flapping.* Ms. Rutgers, The State University of New Jersey.

Braver, Aaron (2011). *Incomplete Neutralization in American English Flapping: A Production Study*. In *Proceedings of the 34th Annual Penn Linguistics Colloquium*, volume 17 of *University of Pennsylvania Working Papers in Linguistics*. Penn Linguistics Club. http://repository.upenn.edu/pwpl/vol17/iss1/5/.

Braver, Aaron (2013). *Perception of Incompletely Neutralized /d/ and /t/ Flaps in American English*. In Stefan Keine and Shayne Slogget (eds.) *Proceedings of the 42nd Annual Meeting of the North Eastern Linguistic Society*, volume 1, pp. 93–104. GLSA, Amherst.

Braver, Aaron (revision invited). *Imperceptible Incomplete Neutralization: Production, Identification, and Discrimination of /d/ and /t/ Flaps in American English*.

Braver, Aaron and Kawahara, Shigeto (2012). *Complete and Incomplete Neutralization in Japanese Monomoraic Lengthening*. ms, Rutgers University.

Braver, Aaron and Kawahara, Shigeto (to appear). *Incomplete Vowel Lengthening: Japanese Monomoraic Lengthening as Incomplete Neutralization*. In *Proceedings of WCCFL 31*.

Burton, Martha W. and Robblee, Karen E. (1997). *A Phonetic Analysis of Voicing Assimilation in Russian*. Journal of Phonetics 25(2):97–114.

Bybee, Joan (2000). *The Phonology of the Lexicon: Evidence from Lexical Diffusion*. In Michael Barlow and Suzanne Kemmer (eds.) *Usage-based models of language*, pp. 65–85. CSLI, Stanford.

Bybee, Joan (2001). *Phonology and Language Use*. Cambridge University Press, Cambridge.

Cedrus Corporation (2010). *Superlab v. 4.5*. Computer program.

Chen, Matthew (1970). *Vowel Length Variation as a Function of the Voicing of the Consonant Environment*. Phonetica 22(129-159).

Cohn, Abigail (1998). *The Phonetics-Phonology Interface Revisited: Where's Phonetics?* Texas Linguistic Forum 41:25–40.

Davidson, Lisa (2006). *Phonology, Phonetics, or Frequency: Influences on the Production of Non-Native Sequences*. Journal of Phonetics 34:104–137.

Davies, Mark (2008-). *The Corpus of Contemporary American English: 450 million words, 1990-present*. Available online at http://corpus.byu.edu/coca/.

Diehl, Randy and Kluender, Keith (1989). *On the Objects of Speech Perception*. Ecological Psychology 1:121–144.

Dinnsen, Daniel (1985). *A Re-Examination of Phonological Neutralization*. Journal of Linguistics 21(2):265–279.

Dinnsen, Daniel and Charles-Luce, Jan (1984). *Phonological Neutralization, Phonetic Implementation and Individual Differences*. Journal of Phonetics 12:49–60.

Dinnsen, Daniel A. and Garcia-Zamor, Marie (1971). *The Three Degrees of Vowel Length in German.* Papers in Linguistics 4:111–126.

Dmitrieva, Olga (2005). *Incomplete Neutralization in Russian Final Devoicing: Acoustic Evidence from Native Speakers and Second Language Learners.* Master's Thesis, University of Kansas, Lawrence, Kansas.

Fisher, William M. and Hirsh, Ira J. (1976). *Intervocalic Flapping in English.* In *Papers from the Twelfth Regional Meeting of the Chicago Linguistic Society*, pp. 183–198. Chicago Linguistic Society.

Flemming, Edward (2001). *Scalar and Categorical Phenomena in a Unified Model of Phonetics and Phonology.* Phonology 18:7–44.

Flemming, Edward (2010). *Modeling Listeners: Comments on Pluymaekers et al. and Scarborough.* In Cécile Fougeron, Barbara Küknert, Mariapaola D'Imperio, and Nathalie Vallée (eds.) *Laboratory Phonology 10*, pp. 587–606.

Fougeron, Cécile and Steriade, Donca (1997). *Does Deletion of French Schwa Lead to Neutralization of Lexical Distinctions?* In *Proceedings of the 5th European Conference on Speech Communication and Technology*, volume 7, pp. 943–946.

Fourakis, Marios and Iverson, Gregory (1984). *On the 'Incomplete Neutralization' of German Final Obstruents.* Phonetica 41:140–149.

Fourakis, Marios and Port, Robert (1986). *Stop Epenthesis in English.* Journal of Phonetics 14(2):197–221.

Fowler, Carol A. (1983). *Converging Sources of Evidence on Spoken and Perceived Rhythms of Speech: Cyclic Production of Vowels in Sequences of Monosyllabic Stress Feet.* Journal of Experimental Psychology: General 112:386–412.

Fox, Robert A. and Terbeek, Dale (1977). *Dental Flaps, Vowel Duration, and Rule Ordering in American English.* Journal of Phonetics 5:27–34.

Francis, Nelson and Kučera, Henry (1982). *Frequency Analysis of English Usage: Lexicon and Grammar.* Houghton Mifflin, Boston.

Fujisaki, H.; Nakamura, K.; and Imoto, T. (1975). *Auditory Perception of Duration of Speech and Non-Speech Stimuli.* In G. Fant and M. A. A. Tatham (eds.) *Auditory Analysis and Perception of Speech*, volume 197-219. Academic, London.

Fukui, Naoki (1986). *Theory of Projection in Syntax.* CSLI, Stanford.

Gahl, Susanne (2008). Time *and* Thyme *are not Homophones: The Effect of Lemma Frequency on Word Durations in Spontaneous Speech.* Language 84(3):474–496.

Gerfen, Chip (2002). *Andalusian Codas.* Probus 14:247–277.

Gerfen, Chip and Hall, Kathleen Currie (2001). *Coda Aspiration and Incomplete Neutralization in Eastern Andalusian Spanish.* ms. University of North Carolina at Chapel Hill.

Gerrits, Ellen and Schouten, M.E.H. (2004). *Categorical Perception Depends on the Discrimination Task.* Perception and Psychophysics 66(3):363–376.

Goldinger, Stephen D. (1996). *Words and Voices: Episodic Traces in Spoken Word Identification and Recognition Memory.* Journal of Experimental Psychology: Learning, Memory, and Cognition 22(5):1166–1183.

Goldrick, Matthew (2001). *Turbid Output Representations and the Unity of Opacity.* In Masako Hirotani, Andries Coetzee, Nancy Hall, and Ji-Yung Kim (eds.) *Proceedings of the North East Linguistics Society 30*, pp. 231–245. GLSA Publications, Amherst, Mass.

Gouskova, Maria and Hall, Nancy (2009). *Acoustics of Unstressable Vowels in Lebanese Arabic.* In Steve Parker (ed.) *Phonological Argumentation: Essays on Evidence and Motivation.* Equinox Books.

Hayes, Bruce (1992/1995). *On What to Teach the Undergraduates: Some Changing Orthodoxies in Phonological Theory.* In *Linguistics in the Morning Calm*, volume 3, pp. 59–77. Hanshin, Seoul.

Herd, Wendy; Jongman, Allard; and Sereno, Joan (2010). *An Acoustic and Perceptual Analysis of /t/ and /d/ Flaps in American English.* Journal of Phonetics 38:504–516.

Hirata, Yukari (2004). *Effects of Speaking Rate on the Vowel Length Distinction in Japanese.* Journal of Phonetics 32(4):565–589.

Hoequist, Charles E. (1983). *Durational Correlates of Linguistic Rhythm Categories.* Phonetica 40:19–31.

Hombert, Jean-Marie; Ohala, John J.; and Ewan, Willam G. (1979). *Phonetic Explanations for the Development of Tones.* Language 55(1):37–58.

House, Arthur S (1961). *On Vowel Duration in English.* Journal of the Acoustical Society of America 33:1174–1178.

Huff, Charles T. (1980). *Voicing and Flap Neutralization in New York City English.* Research in Phonetics 1:233–256.

Hyman, Larry (1975). *Phonology: Theory and Analysis.* Holt, Rinehart and Winston, New York.

Inoue, Kazuko (1997). *Case Marking vs. Case Checking in Japanese Generative Grammar: An Alternative Proposal.* In *Proceedings of the electronic conference "The 40th Anniversary of Generativism"*, Web Journal of Formal, Computational, & Cognitive Linguistics. FCCL. `http://fccl.ksu.ru/papers/inoue.htm`.

Itô, Junko (1990). *Prosodic Minimality in Japanese.* In Michael Ziolkowski, Manual Noske, and Karen Deaton (eds.) *Proceedings of Chicago Linguistic Society 26: Parasession on the Syllable in Phonetics and Phonology*, pp. 213–239. Chicago Linguistic Society, Chicago.

Itô, Junko and Mester, Armin (1992). *Weak Layering and Word Binarity.* Ms. University of California, Santa Cruz.

Jakobson, Roman; Fant, Gunnar; and Halle, Morris (1952/1975). *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates.* MIT Press, Cambridge.

Johnson, Keith (1997). *Speech Perception without Speaker Normalization.* In Keith Johnson and J. Mullennix (eds.) *Talker variability in speech processing*. Academic Press, San Diego.

Joos, Martin (1942). *A Phonological Dilemma in Canadian English.* Language 18(2):141–144.

Kager, René (1999). *Optimality Theory.* Cambridge University Press, Cambridge.

Kahn, Daniel (1980). *Syllable-based Generalizations in English Phonology.* Garland, New York.

Kawahara, Shigeto (to appear). *The Phonology of Japanese Accents.* In Haruo Kubozono (ed.) *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology.* Mouton.

Keating, Pat (1988). *The Phonology-Phonetics Interface.* In F.J. Newymeyer (ed.) *Linguistics: The Cambridge Survey*, volume 1, pp. 281–302. Cambridge University Press, Cambridge.

Kenstowicz, Michael (1996). *Base Identity and Uniform Exponence: Alternatives to Cyclicity.* In Jacques Durand and Bernard Laks (eds.) *Current Trends in Phonology: Models and Methods*, pp. 363–393. European Studies Research Institute, Salford, Manchester.

Kingston, John and Diehl, Randy L (1994). *Phonetic Knowledge.* Language 3:419–454.

Kirby, James P. (2010). *Cue Selection and Category Restructuring In Sound Change.* Doctoral Dissertation, University of Chicago.

Klatt, Dennis H. (1973). *Interaction between Two Factors that Influence Vowel Duration.* Journal of the Acoustical Society of America 54:1102–1104.

Kluender, Keith R; Diehl, Randy L; and Wright, Beverly A (1988). *Vowel-Length Differences Before Voiced and Voiceless Consonants: An Auditory Explanation.* Journal of Phonetics 16:153–169.

Knoblauch, Kenneth (2011). *Psyphy: Functions for Analyzing Psychophysical Data in R.*

Labov, William; Jaeger, Malcah; and Steiner, Richard (1972). *A Quantitative Study of Language Change in Progress.* Technical Report NSF-GS-3287, University of Pennsylvania, US Regional Survey.

Labrune, laurence (2012). *The Phonology of Japanese.* Oxford University Press, Oxford.

Lacerda, Francisco P. (1995). *The Perceptual-Magnet Effect: An Emergent Consequence of Exemplar-Based Phonetic Memory.* In *Proceedings of the Thirteenth International Congress of the Phonetic Sciences*, pp. 140–147. Stockholm.

Legendre, Géraldine; Miyata, Yoshiro; and Smolensky, Paul (1990). *Harmonic Grammar—a Formal Multi-Level Connectionist Theory of Linguistic Well-Formedness: An Application.* In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, pp. 884–891. Lawrence Erlbaum Associates, Mahwah, NJ.

Lehiste, Ilse (1970). *Suprasegmentals.* MIT Press, Cambridge.

Lehiste, Ilse (1972). *Timing of Utterances and Linguistic Boundaries.* Journal of the Acoustical Society of America 51:2018–2024.

Lehiste, Ilse and Peterson, Gordon (1961). *Some Basic Considerations in the Analysis of Intonation.* Journal of the Acoustical Society of America 33(4):419–425.

Lindblom, Björn (1986). *Phonetic Universals in Vowel Systems.* In John Ohala and Jeri Jaeger (eds.) *Experimental Phonology*, pp. 13–44. Academic Press, Orlando.

Lindblom, Björn (1990). *Explaining Phonetic Variation: A Sketch of the H&H Theory.* In W. J. Hardcastle and A. Marchal (eds.) *Speech Production and Speech Modeling*, pp. 403–439. Kluwer, Dordrecht.

Lisker, Leigh (1986). *"Voicing" in English: A Catalog of Acoustic Features Signaling /b/ Versus /p/ in Trochees.* Language and Speech 29:3–11.

Macmillan, Neil A. and Creelman, C. Douglas (2005). *Detection Theory: A User's Guide.* Lawrence Erlbaum Associates Inc., Mahwah, NJ, 2nd edition.

Maddieson, Ian (1985). *Phonetic Cues to Syllabification.* In Victoria Fromkin (ed.) *Phonetic Linguistics*, pp. 203–221. Academic Press, London.

Malécot, André and Lloyd, Paul M. (1968). *The /t/:/d/ distinction in American alveolar flaps.* Lingua 19:264–272.

Mańczak, Witold (1958). *Tendences Générales Des Changements Analogiques.* Lingua 7:298–325, 387–420.

Mańczak, Witold (1980). *Laws of Analogy.* In Jacek Fisiak (ed.) *Historical Morphology*, pp. 283–288. Mouton, The Hague.

Matsui, Mayuki (2011). *The Identifiability and Discriminability Between Incompletely Neutralized Sounds: Evidence from Russian.* In *Proceedings of the International Congress of Phonetic Sciences XVII*, pp. 1342–1345.

McCarthy, John J. (2007). *Hidden Generalizations: Phonological Opacity in Optimality Theory.* Equinox, London.

McCarthy, John J. and Prince, Alan (1986). *Prosodic Morphology.* Ms. University of Massachusetts and Rutgers University.

McCarthy, John J. and Prince, Alan (1993). *Prosodic Morphology: Constraint Interaction and Satisfaction.* ROA482. RuCCS-TR-3.

McCarthy, John J. and Prince, Alan (1995). *Faithfulness and Reduplicative Identity.* In Jill Beckman, Laura Walsh, and Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, number 18 in University of Massachusetts Occasional Papers. Graduate Linguistic Student Association, Amhest, MA. ROA60.

McCawley, James D. (1968). *The Phonological Component of a Grammar of Japanese.* Mouton, The Hague.

Mester, Armin (1990). *Patterns of Truncation.* Linguistic Inquiry 21:475–485.

Mitleb, Fares M. (1981a). *Segmental and Non-segmental Structure in Phonetics: Evidence from Foreign Accent.* Doctoral Dissertation, Indiana University, Bloomington.

Mitleb, Fares M. (1981b). *Temporal Correlates of 'Voicing' and its Neutralization in German.* Research in Phonetics 2:173–192.

Mori, Yoko (2002). *Lengthening of Japanese Monomoraic Nouns.* Journal of Phonetics 30(4):689–708.

Myers, Scott (2005). *Vowel Duration and Neutralization of Vowel Length Contrasts in Kinyarwanda.* Journal of Phonetics 33(4):427–446.

Nooteboom, Sieg G. and Doodeman, Gert J. N. (1980). *Production and Perception of Vowel Length in Spoken Sentences.* Journal of the Acoustical Society of America 67(1):276–287.

Nosofsky, Robert M. (1986). *Attention, Similarity, and the Identification-Categorization Relationship.* Journal of Experimental Psychology: General 115(1):39–57.

Nosofsky, Robert M. (1990). *Relations Between Exemplar-Similarity and Likelihood Models of Classification.* Journal of Mathematical Psychology 34:393–418.

Ohala, John J. (1974). *Experimental Historical Phonology.* In J. M. Naderson and Charles Jones (eds.) *Historical Linguistics II: Theory and Description in Phonology. Proceedings of the First International Linguistic Conference on Historical Linguistics*, pp. 353–389. Elsevier, New York.

Pater, Joe (2009). *Weighted Constraints in Generative Linguistics.* Cognitive Science pp. 1–37.

Peterson, Gordon E. and Lehiste, Ilse (1960). *Duration of Syllable Nuclei in English.* The Journal of the Acoustical Society of America 32(6):693–703.

Pierrehumbert, Janet B. (2001). *Exemplar Dynamics: Word Frequency, Lenition, and Contrast.* In Joan Bybee and Paul Hopper (eds.) *Frequency Effects and the Emergence of Linguistic Structure*, pp. 137–157. John Benjamins.

Pinker, Steven (1999). *Words and Rules.* Basic Books, New York.

Pisoni, David (1973). *Auditory and Phonetic Memory Codes in the Discrimination of Consonants and Vowels.* Perception & Psychophysics 13:253–260.

Pisoni, David (1975). *Auditory Short-Term Memory and Vowel Perception.* Memory and Cognition 3(1):7–8.

Port, Robert (1976). *The Influence of Speaking Tempo on the Duration of Stressed Vowel and Medial Stop in English Trochee Words.* Doctoral Dissertation, University of Connecticut.

Port, Robert and Crawford, Penny (1989). *Incomplete Neutralization and Pragmatics in German.* Journal of Phonetics 17:257–282.

Port, Robert and O'Dell, Michael (1985). *Neutralization and Syllable-Final Voicing in German.* Journal of Phonetics 13:455–471.

Poser, William (1990). *Evidence for Foot Structure in Japanese.* Language 66:78–105.

Posner, Michael I. and Keele, Steven W. (1968). *On the Genesis of Abstract Ideas.* Journal of Experimental Psychology 77:353–363.

Prince, Alan and Smolensky, Paul (1993). *Optimality Theory: Constraint Interaction in Generative Grammar.* ROA537. RuCCS-TR-2.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Reed, Stephen K. (1972). *Pattern Recognition and Categorization.* Cognitive Psychology 3:383–407.

Rialland, Annie (1986). *Schwa et Syllabes en Français.* In W. Leo Wetzels and Engin Sezer (eds.) *Studies in Compensatory Lengthening*, pp. 187–226. Foris, Dordrecht.

Riehl, Anastasia K. (2003). *American English Flapping: Perceptual and Acoustic Evidence Against Paradigm Uniformity with Phonetic Features.* Working Papers of the Cornell Phonetics Laboratory 15:271–331.

Rosch, Eleanor (1973). *Natural Categories.* Cognitive Psychology 4:328–350.

Röttger, Timo B.; Winter, Bodo; Kirby, James P.; and Grice, Martine (2012). *Incomplete Neutralization—A Cognitive By-product of Lexical Co-activation?* Poster presented at the 13th Conference on Laboratory Phonology.

Rudin, Catherine (1980). *Phonetic Evidence for a Phonological Rule: g-Deletion in Turkish.* Research in Phonetics 1:217–232.

Samek-Lodovici, Vieri and Prince, Alan (1999). *Optima.* Ms. University of London and Rutgers University.

Scarborough, Rebecca Anne (2003). *Lexical Confusability and Degree of Coarticulation.* In *Proceedings of the Twenty-Ninth Annual Meeting of the Berkeley Linguistics Society*, pp. 367–378.

Scarborough, Rebecca Anne (2010). *Lexical and Contextual Predictability: Confluent Effects on the Production of Vowels.* In Cécile Fougeron, Barbara Küknert, Mariapaola D'Imperio, and Nathalie Vallée (eds.) *Laboratory Phonology 10*, pp. 557–586.

Sharf, Donald J. (1960). *Distinctiveness of 'Voiced T' words.* American Speech 35(2):105–109.

Slowiaczek, Louisa M. and Dinnsen, Daniel (1985). *On the Neutralizing Status of Polish Word-Final Devoicing.* Journal of Phonetics 13:325–341.

Slowiaczek, Louisa M. and Szymanska, Helena (1989). *Perception of Word-Final Devoicing in Polish.* Journal of Phonetics 17:205–212.

Smiljanic, R. and Bradlow, A. (2004). *Contrast Enhancement in Croatian and English Clear Speech.* In *Ninth Conference on Laboratory Phonology*. Champaign-Urbana, IL.

Smith, J. David and Minda, John Paul (2000). *Thirty Categorization Results in Search of a Model.* Journal of Experimental Psychology: Learning, Memory, and Cognition 25:3–27.

Steriade, Donca (2000). *Paradigm Uniformity and the Phonetics-Phonology Boundary.* In Janet Pierrehumbert and Michael Broe (eds.) *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, chapter 22, pp. 313–334. Cambridge University Press.

Steriade, Donca (2013). *The Analysis of Cyclic and Pseudo-Cyclic Phenomena.* Lecture notes (24.964, co-taught with David Pesetsky, Spring 2013). `https://stellar.mit.edu/S/course/24/sp13/24.964/materials.html`.

Stevens, Kenneth and Blumstein, Sheila (1981). *The Search for Invariant Acoustic Correlates of Phonetic Features.* In Peter Eimas and Joanne Miller (eds.) *Perspectives on the study of speech*, pp. 1–38. Erlbaum Associates, Hillsdale, NJ.

Syrett, Kristen and Kawahara, Shigeto (under revision). *Production and Perception of Listener-Oriented Clear Speech in Child Language.*

Taylor, Dennis Q. (1975). *The Inadequacy of Bipolarity and Distinctive Features: The German "Voiced/Voiceless" Consonants.* In Peter A. Reich (ed.) *The Second LACUS Forum*, pp. 107–119. Hornbeam Press, Inc., Columbia, SC.

Tesar, Bruce (in press). *Output Driven Phonology.* Cambrdige University Press.

Trubetzkoy, Nikolai S. (1939/1969). *Grundzüge der Phonologie [Principles of Phonology].* Vandenhoeck and Ruprecht [Translated by Christiane A. M. Baltaxe 1969, University of California Press], Güttingen.

van Oostendorp, Marc (2008). *Incomplete Devoicing in Formal Phonology.* Lingua 118:1362–1374.

Vance, Timothy J. (2008). *The Sounds of Japanese.* Cambridge University Press, Cambridge.

Warner, Natasha and Arai, Takayuki (2001). *Japanese Mora-Timing: A Review.* Phonetica 58:1–25.

Warner, Natasha; Fountain, Amy; and Tucker, Benjamin V. (2009). *Cues to the Perception of Reduced Flaps.* Journal of the Acoustical Society of America 125:3317–3327.

Warner, Natasha; Good, Erin; Jongman, Allard; and Sereno, Joan (2006). *Orthographic vs. Morphological Incomplete Neutralization Effects.* Journal of Phonetics 34(2):285–293.

Warner, Natasha; Jongman, Allard; Sereno, Joan; and Kemps, Rachèl (2004). *Incomplete Neutralization and other Sub-Phonemic Durational Differences in Production and Perception: Evidence from Dutch.* Journal of Phonetics 32:251–276.

Wedel, Andrew B. (2004). *Category Competition Drives Contrast Maintenance Within an Exemplar-Based Production/Perception Loop.* In John Goldsmith and Richard Wicentowski (eds.) *Proceedings of the Seventh Meeting of the ACL Special Interest Group on Computational Phonology*, pp. 1–10.

Wedel, Andrew B. (2006). *Exemplar Models, Evolution and Language Change.* The Linguistic Review 23:247–274.

Whalen, D. H. and Levitt, A. G. (1995). *The Universality of Intrinsic F0 of Vowels.* Journal of Phonetics 23:349–366.

Yu, Alan C. L. (2007a). *Tonal Phonetic Analogy.* In Jürgen Trouvain and William J. Barry (eds.) *Proceedings of the International Congress of Phonetic Sciences XVI*, pp. 1749–1752. Saarland University, Saarbrücken.

Yu, Alan C. L. (2007b). *Understanding Near Mergers: The Case of Morphological Tone in Cantonese.* Phonology 24:187–214.

Yu, Alan C. L. (2011). *Contrast Reduction.* In John Goldsmith, Jason Riggle, and Alan C. L. Yu (eds.) *The Handbook of Phonological Theory, Second Edition*, chapter 9, pp. 291–318. Blackwell.

Zsiga, Elizabeth (2000). *Phonetic Alignment Constraints: Consonant Overlap and Palatalization in English and Russian.* Journal of Phonetics 28:69–102.

Zue, Victor W. and Laferriere, Martha (1979). *Acoustic Study of Medial /t, d/ in American English.* Journal of the Acoustical Society of America 66:1039–1050.

Curriculum Vitae

Aaron Braver

Education

2007 – 2013	Ph.D. in Linguistics Rutgers, The State University of New Jersey. New Brunswick, NJ
2003 – 2007	B.A. in Linguistics/Cognitive Science (with Honors) and Anthropology Brandeis University. Waltham, MA

Positions

2013 –	Assistant Professor of Linguistics, Department of English, Texas Tech University
2011 – 2013	Teaching Assistant, Writing Program, Rutgers University
2009 – 2011	Teaching Assistant, Linguistics Department, Rutgers University
2009 – 2010, 2012	Research Assistant to Shigeto Kawahara, Rutgers Phonetics Laboratory
2007 – 2009	Graduate Fellow, Rutgers University

Publications

to appear	Incomplete Vowel Lengthening: Japanese Monomoraic Lengthening as Incomplete Neutralization. (With Shigeto Kawahara) *Proceedings of WCCFL 31.*
(ms)	Complete and Incomplete Neutralization in Japanese Monomoraic Lengthening. (With Shigeto Kawahara)
2013	The Phonetics of Emphatic Vowel Lengthening in Japanese. *Open Journal of Modern Linguistics* 3(2), pp. 141-148. (With Shigeto Kawahara)
revision invited	Imperceptible Incomplete Neutralization: Production, Non-Identifiability, and Non-Discriminability in American English Flapping.
2013	Perception of Incompletely Neutralized /d/ and /t/ Flaps in American English. In Stefan Keine and Shayne Slogget (eds.) *Proceedings of the 42nd Annual Meeting of the North Eastern Linguistic Society*, volume 1, pp. 93–104. GLSA, Amherst.
2011	Incomplete Neutralization in American English Flapping: A Production Study. In *University of Pennsylvania Working Papers in Linguistics* 17(1): Proceedings of the 34th Annual Penn Linguistics Colloquium.